# A SURVEY OF THE SCHRÖDINGER PROBLEM AND SOME OF ITS CONNECTIONS WITH OPTIMAL TRANSPORT

Christian Léonard

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

Abstract. This article is aimed at presenting the Schrödinger problem and some of its connections with optimal transport. We hope that it can be used as a basic user's guide to Schrödinger problem. We also give a survey of the related literature. In addition, some new results are proved.

1. **Introduction.** This article is aimed at presenting the Schrödinger problem and some of its connections with optimal transport. We hope that it can be used as a basic user's guide to Schrödinger problem. We also give a survey of the related literature. In addition, some new results are proved.

We denote by $P(Y)$ and $M_+(Y)$ the sets of all probability and positive measures on a space $Y$.

In 1931, Schrödinger [65, 66] addressed a problem which is translated in modern terms[1] as follows. Let $\mathcal{X} = \mathbb{R}^n$ or more generally a complete connected Riemannian manifold without boundary, $\Omega = C([0,1], \mathcal{X})$ be the space of all continuous $\mathcal{X}$-valued paths on the unit time interval $[0,1]$ and denote $R \in M_+(\Omega)$ the law of the reversible Brownian motion on $\mathcal{X}$, i.e. the Brownian motion with the volume measure as its initial distribution. Remark that $R$ is an unbounded measure on $\Omega$ whenever the manifold $\mathcal{X}$ is not compact. Define the relative entropy of any probability measure $P$ with respect to $R$ by

$$H(P|R) = \int_\Omega \log\left(\frac{dP}{dR}\right) dP \in (-\infty, \infty], \qquad P \in P(\Omega)$$

if $P$ is absolutely continuous with respect to $R$ and the above integral is meaningful, and $H(P|R) = \infty$ otherwise. A precise definition of the relative entropy with respect to an unbounded measure $R$ is presented at the Appendix. The *dynamical Schrödinger problem* is

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

where $\mu_0, \mu_1 \in P(\mathcal{X})$ are prescribed values of the initial and final time marginals $P_0 := P(X_0 \in \cdot)$ and $P_1 := P(X_1 \in \cdot)$ of $P$. Here $(X_t)_{0 \le t \le 1}$ is the canonical process on $\Omega$.

---

[1]Recall that in the early 30's the axioms of probability theory were not settled and that the notion of random process still relied entirely on physical intuition.

This is a convex minimization problem since $H(\cdot|R)$ is a convex function and the constraint set $\{P \in \mathrm{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\} = \{P \in \mathrm{M}(\Omega) : P \geq 0, P_0 = \mu_0, P_1 = \mu_1\}$ is a convex subset of the vector space $\mathrm{M}(\Omega)$ of all bounded measures on $\Omega$. Furthermore, as $H(\cdot|R)$ is *strictly* convex, if $(\mathrm{S_{dyn}})$ admits a solution, it must be unique. Let $\widehat{P} \in \mathrm{P}(\Omega)$ be this solution (if it exists). We shall see at Proposition 1 that it disintegrates as

$$\widehat{P}(\cdot) = \int_{\mathcal{X}^2} R^{xy}(\cdot) \, \widehat{\pi}(dxdy) \tag{1}$$

where for all $x, y \in \mathcal{X}$, $R^{xy} := R(\cdot \mid X_0 = x, X_1 = y)$ is the Brownian bridge from $x$ to $y$ and $\widehat{\pi} \in \mathrm{P}(\mathcal{X} \times \mathcal{X})$ is the unique solution to the following *static Schrödinger problem*

$$H(\pi|R_{01}) \to \min; \qquad \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1. \tag{S}$$

Here,

$$R_{01}(dxdy) := R((X_0, X_1) \in dxdy) \propto \exp(-d(x, y)^2/2) \, \mathrm{vol}(dx)\mathrm{vol}(dy) \tag{2}$$

is the joint law of the initial and final positions of the reversible Brownian motion $R$, $d$ is the Riemannian distance and $\pi_0 := \pi(\cdot \times \mathcal{X})$ and $\pi_1 := \pi(\mathcal{X} \times \cdot)$ are the first and second marginals of $\pi \in \mathrm{P}(\mathcal{X}^2)$.

The disintegration formula (1) means that $\widehat{P}$ shares its bridges with $R$, that is: $\widehat{P}^{xy} = R^{xy}$ for almost all $x, y$, and that this mixture of bridges is governed by the unique solution $\widehat{\pi}$ to the static Schrödinger problem (S). It also follows from (1) that the values of the dynamical and static problems are equal: $\inf (\mathrm{S_{dyn}}) = \inf (\mathrm{S})$.

The structure of problem (S) is similar to Monge-Kantorovich problem's one:

$$\int_{\mathcal{X}^2} c(x, y) \, \pi(dxdy) \to \min; \qquad \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1 \tag{MK}$$

where $c : \mathcal{X}^2 \to [0, \infty)$ represents the cost for transporting a unit mass from the initial location $x$ to the final location $y$. Both are convex optimization problems, but unlike (S), the linear program (MK) might admit infinitely many solutions. Since (2) writes as $R_{01}(dxdy) \propto \exp \big( -c(x, y) \big) \, \mathrm{vol}(dx)\mathrm{vol}(dy)$ with

$$c(x, y) = d^2(x, y)/2, \quad x, y \in \mathcal{X},$$

it happens that the Schrödinger problem (S) is connected to the quadratic Monge-Kantorovich optimal transport problem (MK) which is specified by this quadratic cost function. The natural dynamical version of (MK) is

$$\int_\Omega C \, dP \to \min; \qquad P \in \mathrm{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1 \tag{$\mathrm{MK_{dyn}}$}$$

with

$$C(\omega) = \int_{[0,1]} |\dot{\omega}_t|^2_{\omega_t}/2 \, dt \in [0, \infty], \quad \omega \in \Omega, \tag{3}$$

where we put $C(\omega) = \infty$ when $\omega$ is not absolutely continuous.

Let us comment on the choice of this dynamical version of (MK). For all $x, y \in \mathcal{X}$, we have

$$c(x, y) = \inf\{C(\omega); \omega \in \Omega : \omega_0 = x, \omega_1 = y\} \tag{4}$$

and this infimum is attained at the constant speed geodesic path $\gamma^{xy}$ between $x$ and $y$, which is assumed to be unique for any $(x, y)$, for simplicity. Therefore, the solutions of (MK) and $(\mathrm{MK_{dyn}})$ are in one-one correspondence:

- If $\widehat{P}$ solves $(\mathrm{MK_{dyn}})$, then $\widehat{P}_{01} := \widehat{P}((X_0, X_1) \in \cdot)$ solves (MK);

- If $\widehat{\pi}$ solves (MK), then the solution of $(\mathrm{MK_{dyn}})$ is

$$\widehat{P}(\cdot) = \int_{\mathcal{X}^2} \delta_{\gamma^{xy}}(\cdot)\, \widehat{\pi}(dxdy) \tag{5}$$

where $\delta_a$ denotes the Dirac measure at $a$.

Furthermore, we have equality of the values of the problems: $\inf(\mathrm{MK}) = \inf(\mathrm{MK_{dyn}}) \in [0,\infty]$.

Again, the dynamical Schrödinger problem $(\mathrm{S_{dyn}})$ and the dynamical Monge-Kantorovich problem $(\mathrm{MK_{dyn}})$ are similar. Comparing their respective solutions (1) and (5), we see that the $\widehat{\pi}$'s solve their respective static problem (S) and (MK), while for each $(x,y)$ the bridge $R^{xy} \in \mathrm{P}(\Omega)$ in (1) plays the role of $\delta_{\gamma^{xy}} \in \mathrm{P}(\Omega)$ in (5).

All the notions pertaining to the Monge-Kantorovich optimal transport problems (MK) and $(\mathrm{MK_{dyn}})$ which are going to be invoked below are discussed in great detail in C. Villani's textbook [72].

1.1. **Displacement interpolations in** $\mathrm{P}_2(\mathcal{X})$. Let $\mathrm{P}(\mathcal{X})$ denote the set of all probability measures on $\mathcal{X}$ and $\mathrm{P}_2(\mathcal{X}) := \left\{p \in \mathrm{P}(\mathcal{X}) : \int_{\mathcal{X}} d^2(x_0,y)\, p(dy) < \infty\right\}$. If $\mu_0, \mu_1$ are in $\mathrm{P}_2(\mathcal{X})$, then (MK) and $(\mathrm{MK_{dyn}})$ admit a solution. Let $\widehat{P}$ be a solution of $(\mathrm{MK_{dyn}})$. We consider

$$\mu_t := \widehat{P}(X_t \in \cdot) \in \mathrm{P}_2(\mathcal{X}), \quad t \in [0,1] \tag{6}$$

the time-marginal flow of $\widehat{P}$. The $\mathrm{P}_2(\mathcal{X})$-valued path $[\mu_0, \mu_1] = (\mu_t)_{0 \le t \le 1}$ is called a *displacement interpolation* between $\mu_0$ and $\mu_1$. These interpolations were discovered by R. McCann in his PhD thesis [44]. They encode *geometric* properties of the manifold $\mathcal{X}$: although $\mathrm{P}_2(\mathcal{X})$ is not endowed with a Riemannian metric, $[\mu_0, \mu_1]$ is a minimizing constant speed geodesic path on $\mathrm{P}_2(\mathcal{X})$, in the length space sense. In particular, (5) shows that for each $x, y$, $[\delta_x, \delta_y] = (\delta_{\gamma_t^{xy}})_{0 \le t \le 1}$. Therefore, displacement interpolations lift the notion of minimizing constant speed geodesic paths from the state space $\mathcal{X}$ up to $\mathrm{P}_2(\mathcal{X})$. Based on this property, Otto [30, 60] discovered that the heat equation can be seen as the gradient flow of the Boltzmann entropy with respect to the Riemannian-like distance $W_2$ on $\mathrm{P}_2(\mathcal{X})$ which is defined as follows: the common value of the Monge-Kantorovich problems

$$W_2^2(\mu_0, \mu_1)/2 := \inf(\mathrm{MK}) = \inf(\mathrm{MK_{dyn}})$$

allows to define the Wasserstein distance $W_2(\mu_0, \mu_1)$ between $\mu_0$ and $\mu_1$. Saying that $\gamma^{xy}$ has a constant speed means that for all $0 \le s \le t \le 1$, $d(\gamma_s^{xy}, \gamma_t^{xy}) = (t-s)d(x,y)$. With (5), we see that $[\mu_0, \mu_1]$ inherits this constant speed property: for all $0 \le s \le t \le 1$, $W_2(\mu_s, \mu_t) = (t-s)W_2(\mu_0, \mu_1)$. It is a remarkable fact that $W_2(\mu_0, \mu_1)$ also admits the following Benamou-Brenier representation [1]:

$$W_2^2(\mu_0, \mu_1) = \inf_{(\nu,v)} \left\{ \int_{[0,1]\times\mathcal{X}} |v_t(x)|_x^2\, \nu_t(dx)dt \right\} = \int_{[0,1]\times\mathcal{X}} |\nabla\psi_t(x)|_x^2\, \mu_t(dx)dt \tag{7}$$

where the infimum is taken over all $(\nu,v)$ such that $\nu = (\nu_t)_{0 \le t \le 1} \in C([0,1], \mathrm{P}_2(\mathcal{X}))$, $v$ is a smooth vector field and these quantities are linked by the following current equation (in a weak sense) with boundary values:

$$\begin{cases} \partial_t \nu + \nabla\cdot(\nu\, v) = 0, & t \in (0,1) \\ \nu_0 = \mu_0,\ \nu_1 = \mu_1. \end{cases}$$

The last equality in (7) states that the infimum is attained at $\nu = \mu$: the displacement interpolation (6), and some gradient vector field $v = \nabla\psi$ which might not be smooth. The optimal couple $(\mu, \nabla\psi)$ solves the forward-backward coupled system

$$(a) \left\{ \begin{array}{ll} \partial_t\mu + \nabla\cdot(\mu\,\nabla\psi) = 0, & t \in (0,1] \\ \mu_0, & t = 0 \end{array} \right. \qquad (b) \left\{ \begin{array}{ll} \partial_t\psi + \frac{1}{2}|\nabla\psi|^2 = 0, & t \in [0,1) \\ \psi_1, & t = 1 \end{array} \right. \tag{8}$$

for some measurable function $\psi_1$ which is designed for obtaining $\mu_1$ at time 1. The potential $\psi$ is the unique viscosity solution of the Hamilton-Jacobi equation (8)-(b); it admits the Hopf-Lax representation

$$\psi_t(z) = \inf_y \left\{ \frac{d^2(z,y)}{2(1-t)} + \psi_1(y) \right\}, \quad 0 \leq t < 1, y \in \mathcal{X}. \tag{9}$$

In particular, if $\psi_1$ is bounded, $\psi$ is locally Lipschitz continuous and almost everywhere differentiable.

Based on these properties of the displacement interpolations, F. Otto [30, 60] developed an informal theory aimed at considering the metric space $(P_2(\mathcal{X}), W_2)$ as a Riemannian manifold. This informal approach relies on the idea that, in view of the current equation (8)-(a), $\partial_t\mu_{|t=0} = -\nabla\cdot(\mu\,\nabla\psi_0)$ is a candidate to be a tangent vector at $\mu_0$. Second order calculus necessitates to take also (8)-(b) into account.

The analogue of displacement interpolation exists with $(\text{MK}_{\text{dyn}})$ replaced by $(\text{S}_{\text{dyn}})$. This *entropic interpolation* also enjoys properties which are similar to (7) and (8). They are discussed below.

## 1.2. The Monge-Kantorovich problem is a limit of Schrödinger problems.
It is well-known that taking $R^k$ to be the reversible Brownian motion with variance $1/k$, i.e. the Markov measure associated to the Markov generator

$$L^k = \Delta/(2k)$$

with the volume measure as its initial distribution, the bridges of $R^k$ converge: for each $(x,y)$, we have

$$\lim_{k\to\infty} R^{k,xy} = \delta_{\gamma^{xy}} \in P(\Omega) \tag{10}$$

with respect to the usual narrow topology $\sigma(P(\Omega), C_b(\Omega))$. This result is an easy consequence of Schilder's theorem which is a large deviation result (as $k$ tends to infinity) whose rate function is precisely the dynamical cost function $C$ given at (3), see [16].

In fact, the dynamical and static Monge-Kantorovich problems are respectively the $\Gamma$-limits of sequences of dynamical and static Schrödinger problems associated to the sequence $(R^k)_{k\geq 1}$ in $M_+(\Omega)$ of reference path measures [49, 40]. More precisely (but still informally), considering the sequence of re-normalized Schrödinger problems

$$H(P|R^k)/k \to \min; \qquad P \in P(\Omega) : P_0 = \mu_0, P_1 = \mu_1, \tag{$\text{S}_{\text{dyn}}^k$}$$

we have

$$\Gamma\text{-}\lim_{k\to\infty} (\text{S}_{\text{dyn}}^k) = (\text{MK}_{\text{dyn}}) \tag{11}$$

and similarly the re-normalized static version

$$H(\pi|R_{01}^k)/k \to \min; \qquad \pi \in P(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1, \tag{$\text{S}^k$}$$

satisfies $\Gamma\text{-}\lim_{k\to\infty} (\text{S}^k) = (\text{MK})$. Recall that this implies that under some compactness requirements, the values converge: $\lim_{k\to\infty} \inf (\text{S}_{\text{dyn}}^k) = \inf (\text{MK}_{\text{dyn}})$ and

any limit point of the sequence of minimizers $(\widehat{P}^k)_{k\geq 1}$ of $(S_{dyn}^k)$ solves $(MK_{dyn})$. A similar statement holds with the static problems.

In particular, the time-marginal flow

$$\mu_t^k := \widehat{P}_t^k, \quad t \in [0,1],$$

of the solution to $(S_{dyn}^k)$ converges as $k$ tends to infinity to the displacement interpolation $[\mu_0, \mu_1]$ when $(MK)$ admits a unique solution, for instance when both $\mu_0$ and $\mu_1$ are absolutely continuous. Denoting and calling the *entropic interpolation* $[\mu_0, \mu_1]^k := (\mu_t^k)_{t \in [0,1]}$ of order $k$, we have

$$\lim_{k \to \infty} [\mu_0, \mu_1]^k = [\mu_0, \mu_1], \tag{12}$$

with respect to the topology of uniform convergence on $C([0,1], P_2(\mathcal{X}))$ where $P_2(\mathcal{X})$ is equipped with $W_2$.

### 1.3. The Schrödinger problem is a regular approximation of the Monge-Kantorovich problem.
Now, we explain informally why in some sense, $(S_{dyn}^k)$ is a regularization of its limiting Monge-Kantorovich problem $(MK_{dyn})$. Unlike $(MK_{dyn})$, for each $k \geq 1$, $(S_{dyn}^k)$ admits a *unique* solution $\widehat{P}^k \in P(\Omega)$. It can be proved that $\widehat{P}^k$ is a Markov diffusion whose semigroup generator $(A_t^k)_{0 \leq t \leq 1}$ is of the following form

$$A_t^k = \nabla \psi_t^k \cdot \nabla + \Delta/(2k), \quad 0 \leq t < 1,$$

with $\psi^k$ the smooth function on $[0,1) \times \mathcal{X}$ which is the unique classical solution of the Hamilton-Jacobi-Bellman equation

$$\begin{cases} \partial_t \psi^k + \frac{1}{2}|\nabla \psi^k|^2 + \Delta \psi^k/(2k) = 0, & t \in [0,1) \\ \psi_1^k, & t = 1 \end{cases}$$

for some measurable function $\psi_1^k$ designed for recovering[2] $\mu_1^k = \mu_1$ as the final distribution of the weak solution to

$$\begin{cases} \int_{[0,1] \times \mathcal{X}} (\partial_t + A_t^k) u(t,x) \, \mu_t^k(dx) dt = 0, & \forall u \in \mathcal{C}_o^\infty((0,1) \times \mathcal{X}) \\ \mu_0^k = \mu_0, & t = 0. \end{cases}$$

which is the evolution equation of the *entropic interpolation* $[\mu_0, \mu_1]^k$ of order $k$.

Remark that the current equation (8)-(a): $\begin{cases} \partial_t \mu + \nabla \cdot (\mu \nabla \psi) = 0, & t \in (0,1] \\ \mu_0, & t = 0 \end{cases}$ signifies

$$\begin{cases} \int_{[0,1] \times \mathcal{X}} (\partial_t + A_t) u(t,x) \, \mu_t(dx) dt = 0, & \forall u \in \mathcal{C}_o^\infty((0,1) \times \mathcal{X}) \\ \mu_0, & t = 0 \end{cases}$$

with

$$A_t = \nabla \psi_t \cdot \nabla, \quad 0 \leq t < 1,$$

to be compared with the second order operator $A_t^k$ above. We see that, as a consequence of the smoothing and positivity-improving effects of the Laplace operator, the entropic interpolation of order $k$: $[\mu_0, \mu_1]^k$, is positive and regular on $(0,1) \times \mathcal{X}$. This is in contrast with the limiting displacement interpolation $[\mu_0, \mu_1]$.

---

[2]In fact, one can recover exactly $\mu_1$ if it has a regular density. Otherwise, one can build a sequence $\mu_1^k$ such that $\lim_{k \to \infty} \mu_1^k = \mu_1$.

*Extension of the framework.* We have chosen $R^k$ to be attached to the Brownian motion, but taking $R^k$ to be any Markov measure on a Polish state space $\mathcal{X}$ satisfying a large deviation principle with some rate function $C$ leads to limiting Monge-Kantorovich problems associated to alternate cost functions $C$ and $c$ which are still linked by the contraction formula (4). Such extensions based on continuous random paths are considered in [40]. Extensions where the reference measure $R$ is a random walk on a discrete graph are investigated in [33], see also Sections 4 and 5 below.

1.4. **New results.** Although this article is mainly a survey, we have obtained some new results. Theorem 2.5 recollects several sufficient conditions on the reference path measure $R$ and the prescribed marginal measures $\mu_0$ and $\mu_1$, for the unique solution $\widehat{P}$ of $(\mathrm{S_{dyn}})$ to admit the following *product*-shaped Radon-Nikodym derivative

$$\widehat{P} = f_0(X_0)g_1(X_1)\, R \in \mathrm{P}(\Omega), \tag{13}$$

where $f_0$ and $g_1$ are measurable positive functions on $\mathcal{X}$. The slight innovation is due to the possibility that $R$ might have an infinite mass, e.g. the reversible Brownian motion on $\mathbb{R}^n$. Theorem 2.6 is a significant improvement of Theorem 2.5 in the special important case where $R$ is assumed to be Markov. Under some additional requirement on $R$, it states that (13) holds where $f_0$ and $g_1$ may vanish on some sets.

Proposition 5 simply states that, if $R$ is Markov, then the solution $\widehat{P}$ of $(\mathrm{S_{dyn}})$ is also Markov. Although this is intuitively clear, the author couldn't find in the literature any proof of this result. Finally, the Benamou-Brenier type formulas that are stated at Propositions 6 and 7, are new results.

1.5. **Outline of the paper.** In Section 2, the dynamical and static Schrödinger problems are rigorously stated, their main properties of existence and uniqueness are discussed and the shape of their minimizers is described. This specific shape, given by (13), suggests to introduce at Section 3 the notion of $(f, g)$-transform of a Markov measure $R$ which is a time-symmetric version of Doob's $h$-transform. In particular, the classical analogue of Born's formula, which was Schrödinger's motivation in [65, 66], is derived at Theorem 3.3. Then we illustrate at Section 4 the general results of Sections 2 and 3. First, we revisit the case where $R$ is the reversible Brownian motion. Then, we consider a discrete setting where the reference measure is a reversible random walk on a graph. At Section 5, we see that slowing the reference Markov process down to a complete absence of motion, is the right asymptotic to consider for recovering optimal transport from minimal entropy. Technically, this is expressed in terms of $\Gamma$-convergence results in the spirit of (11). In Section 6, by means of basic large deviation results, we present the motivation for addressing the entropy minimization problem $(\mathrm{S_{dyn}})$. This leads us naturally to the *lazy gas experiment*, a starting point to the Lott-Sturm-Villani theory. Literature is discussed at Section 7.

2. **Schrödinger's problem.** We begin fixing some notation and describing the general framework. Then, Schrödinger's problem is stated and its main properties are discussed in a general setting.

2.1. **Path measures.** Depending on the context, we denote by the same letter the set $\Omega = C([0,1], \mathcal{X})$ of all continuous paths from the unit time interval $[0,1]$ to the topological state space $\mathcal{X}$, or $\Omega = D([0,1], \mathcal{X})$ the set of all càdlàg (right-continuous and left-limited) paths. We furnish $\mathcal{X}$ with its Borel $\sigma$-field and $\Omega$ with the canonical $\sigma$-field $\sigma(X_t; 0 \le t \le 1)$ which is generated by the time projections

$$X_t(\omega) := \omega_t \in \mathcal{X}, \qquad \omega = (\omega_s)_{0 \le s \le 1} \in \Omega, \ t \in [0,1].$$

The mapping $X = (X_t)_{0 \le t \le 1} : \Omega \to \Omega$ which is the identity on $\Omega$, is usually called the canonical process. We call a *path measure*, any positive measure $Q \in \mathrm{M}_+(\Omega)$ on $\Omega$. Its time-marginals are the push-forward measures

$$Q_t := (X_t)_{\#} Q \in \mathrm{M}_+(\mathcal{X}), \quad t \in [0,1].$$

This means that for any Borel subset $A \subset \mathcal{X}$, $Q_t(A) = Q(X_t \in A)$. If $Q$ describes the behaviour of the random path $(X_t)_{0 \le t \le 1}$ of some particle, then $Q_t$ describes the behaviour of the random position $X_t$ of the particle at time $t$. Remark that the flow $(Q_t)_{0 \le t \le 1} \in \mathrm{M}_+(\mathcal{X})^{[0,1]}$ contains less information than the path measure $Q \in \mathrm{M}_+(\Omega)$. In particular, $(Q_t)_{0 \le t \le 1}$ doesn't tell us anything about the correlations between two positions at different times $s$ and $t$ which are encoded in $Q_{st} := (X_s, X_t)_{\#} Q \in \mathrm{M}_+(\mathcal{X}^2)$. We shall be primarily concerned with the endpoint marginal measure

$$Q_{01} := (X_0, X_1)_{\#} Q \in \mathrm{M}_+(\mathcal{X}^2),$$

meaning that for any Borel subsets $B \subset \mathcal{X}^2$, $Q_{01}(B) = Q((X_0, X_1) \in B)$. We also denote

$$Q^{xy} = Q(\cdot \mid X_0 = x, X_1 = y) \in \mathrm{P}(\Omega),$$

the bridge of $Q$ between $x$ and $y$. For each $Q \in \mathrm{M}_+(\Omega)$, the disintegration formula (71) with $\phi = (X_0, X_1)$ writes as follows:

$$Q(\cdot) = \int_{\mathcal{X}^2} Q^{xy}(\cdot) \, Q_{01}(dxdy) \in \mathrm{M}_+(\Omega).$$

We assume that the topological state space $\mathcal{X}$ is a Polish (separable and complete metric) space and equip $\Omega = D([0,1], \mathcal{X})$ with the corresponding Skorokhod topology. It is well-known [4] that this topology turns $\Omega$ into a Polish space and that the corresponding Borel $\sigma$-field is precisely the canonical one: $\sigma(X_t; 0 \le t \le 1)$. Moreover, in restriction to $C([0,1], \mathcal{X})$, the Skorokod topology is the topology of uniform convergence which also turns $C([0,1], \mathcal{X})$ into a Polish space. We still have the coincidence of the Borel $\sigma$-field and the canonical one.

The path space $\Omega$ is furnished with this topology.

2.2. **Why unbounded path measures.** One may wonder why a random behaviour should be described by an unbounded measure rather than a probability measure. We have in mind as a particular but important application, the *reversible Brownian motion* on $\mathcal{X} = \mathbb{R}^n$. It is the Brownian motion whose forward dynamics is driven by the heat semigroup as usual, but its random initial position $X_0$ is uniformly distributed on $\mathbb{R}^n$. Denoting $R \in \mathrm{M}_+(\Omega)$ the corresponding path measure on $\Omega = C([0,1], \mathbb{R}^n)$, $R_0(dx) = dx$ is the Lebesgue measure[3] on $\mathbb{R}^n$

---

[3] Although this paper is not concerned with the interpretation of such a description, one should note that a "frequencist" interpretation fails unless one introduces an infinite system of independent particles initially distributed according to a Poisson point process with a uniform spatial frequency. An alternate information viewpoint is also relevant: the Lebesgue measure (or any of its positive multiples) is the less informative a priori measure for modelling our complete lack of knowledge

and $R(\cdot) = \int_{\mathcal{X}} \mathcal{W}_x(\cdot) \, dx$ where $\mathcal{W}_x$ is the Wiener probability measure with initial marginal $\delta_x$. Clearly, $R$ has the same infinite mass as $R_0$.

Similarly, the simple random walk on a countably infinite graph $\mathcal{X}$ admits an unbounded reversing measure $R_0$ so that the corresponding reversible simple random walk is described by an unbounded measure $R \in \mathrm{M}_+(\Omega)$ with $\Omega = D([0,1], \mathcal{X})$.

Considering such reversible path measures $R \in \mathrm{M}_+(\Omega)$ as reference measures usually simplifies computations.

2.3. **Relative entropy.** Let $r$ be some $\sigma$-finite positive measure on some space $Y$. The relative entropy of the probability measure $p$ with respect to $r$ is loosely defined by

$$H(p|r) := \int_Y \log(dp/dr) \, dp \in (-\infty, \infty], \qquad p \in \mathrm{P}(Y) \tag{14}$$

if $p \ll r$ and $H(p|r) = \infty$ otherwise. The rigorous definition of the relative entropy and its basic properties are recalled at the appendix section A.

2.4. **Statement of Schrödinger's problem.** The main data is a given reference path measure $R \in \mathrm{M}_+(\Omega)$. In this section any (non-zero) $\sigma$-finite path measure in $\mathrm{M}_+(\Omega)$ can serve as a reference measure.

We first state a dynamical version ($\mathrm{S}_{\mathrm{dyn}}$) of Schrödinger's problem which is associated to $R$. Then, we define Schrödinger's problem ($\mathrm{S}$) as a static projection of ($\mathrm{S}_{\mathrm{dyn}}$) and the connections between the solutions of ($\mathrm{S}$) and ($\mathrm{S}_{\mathrm{dyn}}$) are described at Proposition 1.

**Definition 2.1** (Dynamical Schrödinger problem). The dynamical Schrödinger problem associated with the *reference path measure* $R \in \mathrm{M}_+(\Omega)$ is the following entropy minimization problem

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

where $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$ are prescribed initial and final marginals.

Considering the projection $R_{01} = (X_0, X_1)_{\#} R \in \mathrm{M}_+(\mathcal{X}^2)$ of $R$ on the product space $\mathcal{X}^2$ as a reference measure, leads us to Schrödinger's (static) problem.

**Definition 2.2** (Schrödinger's problem). The (static) Schrödinger problem associated with the *reference measure* $R_{01} \in \mathrm{M}_+(\mathcal{X}^2)$ is the following entropy minimization problem

$$H(\pi|R_{01}) \to \min; \qquad \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \ \pi_1 = \mu_1 \tag{$\mathrm{S}$}$$

where $\pi_0 := \pi(\cdot \times \mathcal{X})$ and $\pi_1 := \pi(\mathcal{X} \times \cdot) \in \mathrm{P}(\mathcal{X})$ denote respectively the first and second marginals of $\pi \in \mathrm{P}(\mathcal{X}^2)$ and $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$ are prescribed marginals.

These optimization problems are highly connected. This is the content of next proposition.

**Proposition 1** (Föllmer, [24]). *The Schrödinger problems* ($\mathrm{S}_{\mathrm{dyn}}$) *and* ($\mathrm{S}$) *admit respectively at most one solution* $\widehat{P} \in \mathrm{P}(\Omega)$ *and* $\widehat{\pi} \in \mathrm{P}(\mathcal{X}^2)$.

*If* ($\mathrm{S}_{\mathrm{dyn}}$) *admits the solution* $\widehat{P}$, *then* $\widehat{\pi} = \widehat{P}_{01}$ *is the solution of* ($\mathrm{S}$).

---

about the initial position. Indeed, it is invariant under isometries and translations, and the entropic problems to be considered below are insensitive to homotheties (up to an additive constant).

*Conversely, if $\widehat{\pi}$ solves* (S), *then* (S$_{\text{dyn}}$) *admits the solution*

$$\widehat{P}(\cdot) = \int_{\mathcal{X}^2} R^{xy}(\cdot)\,\widehat{\pi}(dxdy) \in \mathrm{P}(\Omega) \tag{15}$$

*which means that*

$$\widehat{P}_{01} = \widehat{\pi} \in \mathrm{P}(\mathcal{X}^2)$$

*and that $\widehat{P}$ shares its bridges with $R$:*

$$\widehat{P}^{xy} = R^{xy}, \qquad \forall (x,y)\ \widehat{\pi}\text{-}a.e.$$

*Proof.* Being *strictly* convex problems, (S$_{\text{dyn}}$) and (S) admit respectively at most one solution.

Let us particularize the consequences of the additive property formula (72) to $r = R$, $p = P$ and $\phi = (X_0, X_1)$. We have for all $P \in \mathrm{P}(\Omega)$,

$$H(P|R) = H(P_{01}|R_{01}) + \int_{\mathcal{X}^2} H(P^{xy}|R^{xy})\,P_{01}(dxdy)$$

which implies that $H(P_{01}|R_{01}) \leq H(P|R)$ with equality (when $H(P|R) < \infty$) if and only if $P^{xy} = R^{xy}$ for $P_{01}$-almost every $(x,y) \in \mathcal{X}^2$, see (73) and (74). Note that this additive property formula is available since both $\mathcal{X}^2$ and $\Omega$ are Polish spaces. Therefore $\widehat{P}$ is the (unique) solution of (S$_{\text{dyn}}$) if and only if it disintegrates as (15). □

2.5. **Existence results.** We present below at Proposition 2 a simple criterion for (S) and (S$_{\text{dyn}}$) to have a solution. We first need a preliminary result.

**Lemma 2.3.** *We have:*    $\inf(\mathrm{S}_{\text{dyn}}) = \inf(\mathrm{S}) \in (-\infty, \infty]$.
*Let $B : \mathcal{X} \to [0, \infty)$ be a measurable function such that*

$$\int_{\mathcal{X}^2} e^{-B(x)-B(y)} R_{01}(dxdy) < \infty \tag{16}$$

*and take $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$ such that*

$$\int_{\mathcal{X}} B\,d\mu_0,\ \int_{\mathcal{X}} B\,d\mu_1 < \infty. \tag{17}$$

*The static and dynamical Schrödinger problems* (S) *and* (S$_{\text{dyn}}$) *admit a (unique) solution if and only if* $\inf(\mathrm{S}_{\text{dyn}}) = \inf(\mathrm{S}) < \infty$ *or equivalently if and only if the prescribed marginals $\mu_0$ and $\mu_1$ are such that*

*there exists some $\pi^o \in \mathrm{P}(\mathcal{X}^2)$ such that $\pi_0^o = \mu_0, \pi_1^o = \mu_1$ and $H(\pi^o|R_{01}) < \infty$.*
$$\tag{18}$$

*Proof.* The first identity comes from the proof of Proposition 1.

Since $\mathcal{X}$ is Polish, the probability measures $\mu_0$ and $\mu_1$ are tight measures on $\mathcal{X}$ and it follows with the Prokhorov criterion on $\mathcal{X}^2$ that the closed constraint set $\Pi(\mu_0, \mu_1) := \left\{ \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1 \right\}$ is uniformly tight and therefore compact in $\mathrm{P}(\mathcal{X}^2)$.

Taking (16) into account, (67) and (68) give us $H(\pi|R_{01}) = H(\pi|R_{01}^B) - \int_{\mathcal{X}^2} W\,d\pi - z_B$, $\pi \in \mathrm{P}(\mathcal{X}^2)$ with $W(x,y) = B(x) + B(y)$, $x,y \in \mathcal{X}$, $z_B := \int_{\mathcal{X}^2} e^{-B(x)-B(y)} R_{01}(dxdy) < \infty$ and $R_{01}^B := z_B^{-1} e^{-B\oplus B} R_{01} \in \mathrm{P}(\mathcal{X}^2)$. In restriction to $\Pi(\mu_0, \mu_1)$, we obtain

$$H(\pi|R_{01}) = H(\pi|R_{01}^B) - \int_{\mathcal{X}} B\,d\mu_0 - \int_{\mathcal{X}} B\,d\mu_1 - z_B, \quad \pi \in \Pi(\mu_0, \mu_1). \tag{19}$$

Together with (17), this implies that $H(\cdot|R_{01})$ is lower bounded and lower semi-continuous on the compact set $\Pi(\mu_0, \mu_1)$. Hence, (S) admits a solution if and only inf (S) $< \infty$. We already remarked that ($S_{dyn}$) has a solution if and only (S) has a solution that is: inf (S) $< \infty$, or equivalently if and only (18) is satisfied. $\qquad\square$

**Proposition 2.** *Suppose that $R_0 = R_1 = m \in M_+(\mathcal{X})$ (this is satisfied in particular when $R$ is reversible with $m$ as its reversing measure).*

*(a) For ($S_{dyn}$) and (S) to have a solution, it is necessary that $H(\mu_0|m), H(\mu_1|m) < \infty$.*

*(b) Let us give a set of sufficient conditions for an existence result. Suppose that there exist some nonnegative measurable functions $A$ and $B$ on $\mathcal{X}$ such that*
   *(i) $R_{01}(dxdy) \geq e^{-A(x)-A(y)} m(dx)m(dy)$;*
   *(ii) $\int_{\mathcal{X}^2} e^{-B(x)-B(y)} R_{01}(dxdy) < \infty$*
   *Let $\mu_0$ and $\mu_1$ satisfy*
   *(iii) $\int_{\mathcal{X}}(A+B) \, d\mu_0, \int_{\mathcal{X}}(A+B) \, d\mu_1 < \infty$*
   *(iv) $H(\mu_0|m), H(\mu_1|m) < \infty$*
   *Then, ($S_{dyn}$) and (S) admit a unique solution.*

*(c) If we have (iv) and*
   *(v) $\int_{\mathcal{X}} e^{\alpha(A+B)} \, dm < \infty$ for some $\alpha > 0$,*
   *then (iii) is satisfied.*

Remark that for (v) to be satisfied, it is necessary that $m$ is a bounded measure.

*Proof.* Statement (a) follows directly from Lemma 2.3 and $H(\mu_0|m), H(\mu_1|m) \leq H(\pi^o|R_{01}) < \infty$, see (73).

Let us look at statement (b). Testing (18) with $\pi^o = \mu_0 \otimes \mu_1$, one easily observes that when (i) and (iii) are satisfied, it suffices that $H(\mu_0|m), H(\mu_1|m) < \infty$ for (18) to hold true, and consequently by Lemma 2.3, for ($S_{dyn}$) and (S) to have a (unique) solution.

Let us look at statement (c). With the variational representation formula (69), one sees that (iv) and (v) imply (iii). $\qquad\square$

2.6. **The dual problem.** Take a measurable function $B : \mathcal{X} \to [1, \infty)$ that satisfies (16). Define $C_B(\mathcal{X})$ to be the space of all continuous functions $u : \mathcal{X} \to \mathbb{R}$ such that $\sup |u|/B < \infty$ and $P_B(\mathcal{X}) := \left\{ \mu \in P(\mathcal{X}); \int_{\mathcal{X}} B \, d\mu < \infty \right\}$. Based on the variational representation of the relative entropy (70) and on the observation that for each $\pi \in P(\mathcal{X}^2)$, $(\pi_0, \pi_1) = (\mu_0, \mu_1)$ if and only if $\int_{\mathcal{X}^2}[\varphi(x) + \psi(y)] \pi(dxdy) = \int_{\mathcal{X}} \varphi \, d\mu_0 + \int_{\mathcal{X}} \psi \, d\mu_1$, for all $\varphi, \psi \in C_b(\mathcal{X})$ (the space of all bounded continuous numerical functions on $\mathcal{X}$), it can be proved that a dual problem to the Schrödinger problem (S) is

$$\int_{\mathcal{X}} \varphi \, d\mu_0 + \int_{\mathcal{X}} \psi \, d\mu_1 - \log \int_{\mathcal{X}^2} e^{\varphi \oplus \psi} \, dR_{01} \to \max; \qquad \varphi, \psi \in C_B(\mathcal{X}) \qquad \text{(D)}$$

where $\varphi \oplus \psi : (x, y) \in \mathcal{X}^2 \mapsto \varphi(x) + \psi(y) \in \mathbb{R}$ and it is assumed that the prescribed marginals satisfy $\mu_0, \mu_1 \in P_B(\mathcal{X})$.

In particular, the dual equality inf (S) $=$ sup (D) $\in (-\infty, \infty]$ is satisfied. This is proved, for instance, in [37] when the reference measure is a probability measure. In the general case, take (19) into account to get back to a reference measure with a finite mass. Of course, there is no reason for the dual attainment to hold in general in a space of regular functions such as $C_B(\mathcal{X})^2$. Suppose however that $\mu_0$ and $\mu_1$

are such that (D) is attained at $(\hat{\varphi}, \hat{\psi})$. Then, the dual equality: $\int_{\mathcal{X}^2} \hat{\varphi} \oplus \hat{\psi} \, d\hat{\pi} - \log \int_{\mathcal{X}^2} e^{\hat{\varphi} \oplus \hat{\psi}} \, dR_{01} = H(\hat{\pi}|R_{01})$ and the case of equality in (70) lead us to

$$d\hat{\pi}/dR_{01} = \exp(\hat{\varphi} \oplus \hat{\psi}), \qquad (20)$$

at least when $d\hat{\pi}/dR_{01} > 0$. The shape of the minimizer $\hat{\pi}$ of (S) will be discussed further in next subsection.

Similarly, a dual problem to the dynamical version $(S_{dyn})$ of (S) is

$$\int_{\mathcal{X}} \varphi \, d\mu_0 + \int_{\mathcal{X}} \psi \, d\mu_1 - \log \int_{\Omega} e^{\varphi(X_0)+\psi(X_1)} \, dR \to \max; \qquad \varphi, \psi \in C_B(\mathcal{X}) \quad (D_{dyn})$$

We observe that $(D)=(D_{dyn})$.

### 2.7. Some properties of the minimizer $\hat{\pi}$ of (S).

We give some detail about the structure of the unique minimizer $\hat{\pi}$ of (S) which is assumed to exist; for instance under the hypotheses of Proposition 2.

It is proved in [38, Thm. 5.1 & (5.9)][4] that there exist two functions $\varphi, \psi : \mathcal{X} \to \mathbb{R}$ such that

(i)  $d\hat{\pi}/dR_{01} = \exp(\varphi \oplus \psi)$, $\hat{\pi}$-a.e.
(ii) $\varphi \oplus \psi : \mathcal{X}^2 \to \mathbb{R}$ is $R_{01}$-measurable.

It is tempting to write that $\hat{\pi}$ has the following shape

$$\hat{\pi}(dxdy) = f(x)g(y) \, R_{01}(dxdy)$$

where $f = e^{\varphi}$ and $g = e^{\psi}$ are such that the marginal constraints

$$\begin{cases} f(x)E_R[g(X_1) \mid X_0 = x] & = & d\mu_0/dR_0(x), & R_0\text{-a.e.} \\ g(y)E_R[f(X_0) \mid X_1 = y] & = & d\mu_1/dR_1(y), & R_1\text{-a.e.} \end{cases} \qquad (21)$$

are satisfied. This was already suggested by (20). But this is not allowed in the general case. Indeed, two obstacles have to be avoided. Some comments are necessary.

*Obstacle (i).* Firstly, the identity (i) is only valid $\hat{\pi}$-almost everywhere and it might happen that it doesn't hold true $R_{01}$-almost everywhere. Otherwise stated, there exists some measurable subset $S \subset \mathcal{X}^2$ such that:

(i)'  $\hat{\pi} = \mathbf{1}_S \exp(\varphi \oplus \psi) \, R_{01}$

and it is not true in general that the set $S \subset \mathcal{X}^2$ is the product $S = S_0 \times S_1$ of two measurable subsets of $\mathcal{X}$, see [25, §2] or [38, §5].

*Obstacle (ii).* Secondly, statement (ii) does not imply that $\varphi$ and $\psi$ are respectively $R_0$ and $R_1$-measurable on $\mathcal{X}$. Only the tensor sum $\varphi \oplus \psi$ is $R_{01}$-measurable on the product space $\mathcal{X}^2$. Hence, one is not allowed to consider the conditional expectations in (21).

_____

[4] The assumptions of [38] require that $R_{01}$ is a probability measure. In the general case where $R_{01}$ is unbounded, use (19) to go back to the unit mass setting.

*Avoiding obstacle (i).* To avoid the obstacle (i), it is enough to slightly modify the prescribed marginals $\mu_0$ and $\mu_1$ as follows. It is shown in [38] that (i) is satisfied $R_{01}$-a.e. (rather than $\widehat{\pi}$-a.e.) if $(\mu_0, \mu_1)$ is in the intrinsic core: $\mathrm{icor}\,\mathcal{C}$, of the set of all admissible constraints

$$\mathcal{C} := \{(\mu_0, \mu_1) \in \mathrm{P}(\mathcal{X})^2; \inf \mathrm{(S)}_{(\mu_0, \mu_1)} < \infty\}.$$

Recall that for any convex set $C$, its intrinsic core is defined as $\mathrm{icor}\,C := \{y \in C; \exists x, z \in C : y \in ]x, z[\}$ where $]x, z[:= \{(1-t)x + tz; 0 < t < 1\} \subset C$. It is also shown in [38] that $\mathcal{C} = \{\Lambda^* < \infty\}$ where $\Lambda^*$ is the convex conjugate of the extended real valued function $\Lambda$ which is defined for any measurable functions $\varphi, \psi$ by $\Lambda(\varphi, \psi) = \log \int_{\mathcal{X}^2} e^{\varphi \oplus \psi} \, dR_{01} \in (-\infty, \infty]$. Therefore $\mathcal{C}$ is a convex subset of $\mathrm{P}(\mathcal{X})^2$. In particular, considering

$$\left\{ \begin{array}{rcl} \mu_0^\epsilon & := & (1-\epsilon)\mu_0 + \epsilon R_0^w \\ \mu_1^\epsilon & := & (1-\epsilon)\mu_1 + \epsilon R_1^w \end{array} \right. \tag{22}$$

with $R_0^w, R_1^w \in \mathrm{P}(\mathcal{X})$ the marginals of $R_{01}^w = z_w^{-1} e^{-w} R_{01} \in \mathrm{P}(\mathcal{X}^2)$ where the function $w$ is chosen[5] such that $\int_{\mathcal{X}^2} w e^{-w} \, dR_{01} < \infty$ for $H(R_{01}^w | R_{01}) < \infty$ to be satisfied with $\int_{\mathcal{X}^2} w \, dR_{01}^w < \infty$, see (68), we observe that for any admissible $(\mu_0, \mu_1) \in \mathcal{C}$, for any arbitrarily small $\epsilon > 0$, $(\mu_0^\epsilon, \mu_1^\epsilon) \in \mathrm{icor}\,\mathcal{C}$. Therefore, $(\mu_0^\epsilon, \mu_1^\epsilon)$ is arbitrarily close to $(\mu_0, \mu_1)$ in total variation norm and the corresponding solution $\widehat{\pi}^\epsilon$ of $\mathrm{(S)}_{(\mu_0^\epsilon, \mu_1^\epsilon)}$ satisfies

$$\widehat{\pi}^\epsilon = \exp(\varphi^\epsilon \oplus \psi^\epsilon) \, R_{01}$$

for some functions $\varphi^\epsilon$ and $\psi^\epsilon$ such that $\varphi^\epsilon \oplus \psi^\epsilon$ is jointly $R_{01}$-measurable.

**Proposition 3.** *We say that the constraint $(\mu_0, \mu_1)$ is* internal *if it is in the intrinsic core of the set of all admissible constraints: $(\mu_0, \mu_1) \in \mathrm{icor}\,\mathcal{C}$. In this case, we have*

$$\widehat{\pi} = \exp(\varphi \oplus \psi) \, R_{01}$$

*for some jointly $R_{01}$-measurable function $\varphi \oplus \psi$ on $\mathcal{X}^2$.*

*Overcoming obstacle (ii).* To overcome the measurability obstacle (ii), it is necessary to impose some restriction on the reference measure $R_{01}$. It is proved in [5, Proposition 6.1] that when the function $\varphi \oplus \psi$ is measurable with respect to some product measure $\alpha \otimes \beta$ on the product space $\mathcal{X}^2$, the functions $\varphi$ and $\psi$ are respectively $\alpha$-measurable and $\beta$-measurable. Therefore, if it is assumed that

$$R_0 \otimes R_1 \ll R_{01},$$

any $R_{01}$-measurable function is $R_0 \otimes R_1$-measurable. As $\varphi \oplus \psi$ is $R_{01}$-measurable, $\varphi$ and $\psi$ are respectively $R_0$ and $R_1$-measurable.

Choosing another way, it is also possible to obtain the desired measurability property when assuming

$$R_{01} \ll R_0 \otimes R_1.$$

To see this, let $\widehat{\pi}$, $\varphi$ and $\psi$ satisfy (i). It is proved in [11] that $\varphi \oplus \psi$ stands in the $L^1(\widehat{\pi})$-closure of $L^1(\mu_0) \oplus L^1(\mu_1)$, denoted by $\Lambda_{\widehat{\pi}}$. In addition, with [63, Proposition 2], we know that when $\pi \in \mathrm{P}(\mathcal{X}^2)$ with first and second marginals $\pi_0 = \mu_0$ and $\pi_1 = \mu_1$, is such that $\pi \ll \pi_0 \otimes \pi_1 = \mu_0 \otimes \mu_1$, then for any $\theta \in \Lambda_\pi$, there exist two *measurable* functions $\theta_0$ and $\theta_1$ such that $\theta = \theta_0 \oplus \theta_1$, $\pi$-a.s. Therefore, if

$$\widehat{\pi} \ll \mu_0 \otimes \mu_1,$$

---

[5] In the important special case where $R$ is a probability measure, just take $w = 0$ in order that $R_{01}^w = R_{01}$.

then $\varphi$ and $\psi$ are respectively $\mu_0$ and $\mu_1$-measurable functions.

But this implicit criterion is not easy to check. The following result is more practical.

**Proposition 4.** [63, Thm. 3] *If $R_{01} \ll R_0 \otimes R_1$, then $\varphi$ and $\psi$ in (i): $d\widehat{\pi}/dR_{01} = \exp(\varphi \oplus \psi)$, $\widehat{\pi}$-a.e., can be chosen respectively as $\mu_0$ and $\mu_1$-measurable functions.*

*Proof.* By assumption $\widehat{\pi} \ll R_{01} \ll R_0 \otimes R_1$. But $\widehat{\pi} \ll R_0 \otimes R_1$ implies that $\widehat{\pi} \ll \widehat{\pi}_0 \otimes \widehat{\pi}_1 = \mu_0 \otimes \mu_1$ and we conclude as above.                    $\square$

Extending the functions $\varphi$ and $\psi$ to their $R_0$ and $R_1$-measurable versions: $\mathbf{1}_{\{d\mu_0/dR_0>0\}}\varphi$ and $\mathbf{1}_{\{d\mu_1/dR_1>0\}}\psi$, we see that $\varphi$ and $\psi$ can be taken respectively $R_0$ and $R_1$-measurable.

*Summing up.* Putting Propositions 2, 3 and the above considerations together with (21), we obtain the following

**Theorem 2.4.** *Suppose that $R$ satisfies*

(i) $R_0 = R_1 = m \in M_+(\mathcal{X})$;

(ii) $R_{01}(dxdy) \geq e^{-A(x)-A(y)}\, m(dx)m(dy)$ *for some nonnegative measurable function $A$ on $\mathcal{X}$;*

(iii) $\int_{\mathcal{X}^2} e^{-B(x)-B(y)} R_{01}(dxdy) < \infty$ *for some nonnegative measurable function $B$ on $\mathcal{X}$;*

(iv) $m^{\otimes 2} \ll R_{01}$ *or* $R_{01} \ll m^{\otimes 2}$.

*Suppose also that the constraint $(\mu_0, \mu_1)$ satisfies*

(v) $H(\mu_0|m), H(\mu_1|m) < \infty$;

(vi) $\int_{\mathcal{X}}(A+B)\, d\mu_0, \int_{\mathcal{X}}(A+B)\, d\mu_1 < \infty$ *where $A$ and $B$ appear at (ii) and (iii) above;*

(vii) $(\mu_0, \mu_1)$ *is internal, see the statement of Proposition 3.*
*This is the case for instance when $m$ is a probability measure and $\mu_0, \mu_1 \geq \epsilon m$, for some $\epsilon > 0$.*

*Then, (S) admits a unique solution $\widehat{\pi}$ and*

$$\widehat{\pi}(dxdy) = f_0(x)g_1(y)\, R_{01}(dxdy) \tag{23}$$

*where the positive functions $f_0$ and $g_1$ are $m$-measurable and solve :*

$$\begin{cases} f_0(x)E_R[g_1(X_1) \mid X_0 = x] &= d\mu_0/dm(x), \quad m\text{-}a.e. \\ g_1(y)E_R[f_0(X_0) \mid X_1 = y] &= d\mu_1/dm(y), \quad m\text{-}a.e. \end{cases} \tag{24}$$

*which is called the* Schrödinger system[6].

- It is not necessary for $E_R[g_1(X_1) \mid X_0]$ and $E_R[f_0(X_0) \mid X_1]$ to be well-defined that $f_0(X_0)$ and $g_1(X_1)$ are $R$-integrable, since $f_0$ and $g_1$ are positive measurable functions. Only a notion of integration of *nonnegative* functions is required, see [35].
- The assumption (vii) is here to make sure that $d\widehat{\pi}/dR_{01} > 0$. If it is not satisfied, $d\widehat{\pi}/dR_{01}$ may not have a product form and its structure may be quite complex. The complete description of $d\widehat{\pi}/dR_{01}$ in this case is given in [38].

---

[6]The article [63] refers to (24) as the Schrödinger equation, but this is misleading. After Fortet and Beurling [26, 3], we prefer calling (24) the Schrödinger system.

- In view of (22), for the assumption (vii) to hold, it is enough that

$$
\left\{
\begin{array}{ccc}
\mu_0 & \geq & \epsilon R_0^w \\
\mu_1 & \geq & \epsilon R_1^w
\end{array}
\right. ,
$$

for some $\epsilon > 0$, and we can choose $w = 0$ when $m$ is a probability measure.

### 2.8. The solution $\widehat{P}$ of ($S_{\mathrm{dyn}}$).

We deduce from this theorem the characterization of $\widehat{P}$.

**Theorem 2.5.** *Suppose that the hypotheses of Theorem 2.4 are satisfied. Then, ($S_{\mathrm{dyn}}$) admits the unique solution*

$$
\widehat{P} = f_0(X_0)g_1(X_1)\, R \in \mathrm{P}(\Omega) \tag{25}
$$

*where $f_0$ and $g_1$ are the measurable positive functions which appear at (23) and solve (24).*

*Proof.* The existence of the solution $\widehat{P}$ and its representation by the Radon-Nikodym formula (25) are direct consequences of Proposition 1, Theorem 2.4, (15) and (23).
□

### 2.9. The special case where $R$ is Markov.

We are going to assume that the reference path measure $R$ is Markov. Under this restriction, we obtain at Theorem 2.6 below a more efficient version of Theorem 2.4. Let us recall the time-symmetric definition of the Markov property.

*Markov property.* One says that $R \in \mathrm{M}_+(\Omega)$ is *Markov* if its time marginal $R_0$ (or any other time marginal $R_{t_o}$ with $0 \leq t_o \leq 1$) is $\sigma$-finite[7] and for each $0 \leq t \leq 1$,

$$
R(X_{[0,t]} \in \cdot, X_{[t,1]} \in \cdot\cdot \mid X_t) = R(X_{[0,t]} \in \cdot \mid X_t)R(X_{[t,1]} \in \cdot\cdot \mid X_t)
$$

signifying that under $R$, for any $t$, conditionally on the present state $X_t$ at time $t$, past and future are independent. This is equivalent to the usual forward time-oriented Markov property

$$
R(X_{[t,1]} \in \cdot | X_{[0,t]}) = R(X_{[t,1]} \in \cdot | X_t), \quad \forall t \in [0,1].
$$

**Proposition 5.** *Suppose that the reference measure $R \in \mathrm{M}_+(\Omega)$ is Markov. If it exists, the solution $\widehat{P}$ of ($S_{\mathrm{dyn}}$) is also Markov.*

*Proof.* We need some notation. For each $0 \leq t \leq 1$, we set $\Omega_{[0,t]} := \left\{\omega_{|[0,t]}; \omega \in \Omega\right\}$ and $\Omega_{[t,1]} := \left\{\omega_{|[t,1]}; \omega \in \Omega\right\}$ the set of all paths on $[0,t]$ and $[t,1]$ respectively. For any $Q \in \mathrm{M}_+(\Omega)$, $Q^{t,z} := Q(\cdot|X_t = z) \in \mathrm{P}(\Omega)$, $Q_{[0,t]}^{t,z} := Q(X_{[0,t]} \in \cdot | X_t = z) \in \mathrm{P}(\Omega_{[0,t]})$ and $Q_{[t,1]}^{t,z} := Q(X_{[t,1]} \in \cdot | X_t = z) \in \mathrm{P}(\Omega_{[t,1]})$.

**Claim 1.** *We fix $0 \leq t \leq 1$. Among all the $P \in \mathrm{P}(\Omega)$ such that $P_t = \mu$, $P_{[0,t]}^{tz} = Q_<^{tz}$ and $P_{[t,1]}^{tz} = Q_>^{tz}$, $z \in \mathcal{X}$, where $\mu \in \mathrm{P}(\mathcal{X})$, $Q_<^{tz} \in \mathrm{P}(\Omega_{[0,t]} \cap \{X_t = z\})$ and $Q_>^{tz} \in \mathrm{P}(\Omega_{[t,1]} \cap \{X_t = z\})$ are prescribed, the relative entropy $H(\cdot|R)$ attains its unique minimum at $P^*(\cdot) = \int_{\mathcal{X}} Q_<^{tz} \otimes Q_>^{tz}(\cdot)\, \mu(dz)$. In particular, $P_{[t,1]}^*(\cdot|X_{[0,t]}) = P_{[t,1]}^*(\cdot|X_t)$.*

---

[7]This assumption is necessary for defining conditional versions of $R$, such as $R(\cdot \mid X_t)$ or $R(\cdot \mid X_{[0,t]})$, see [35].

Accept this claim for a while and suppose, ad absurdum, that $\widehat{P}$ is not Markov. Then, there exists some $0 \le t \le 1$ such that $\widehat{P}(\cdot|X_t) \ne \widehat{P}_{[0,t]}(\cdot|X_t) \otimes \widehat{P}_{[t,1]}(\cdot|X_t)$. Choosing $\mu = \widehat{P}_t$, $Q_<^{tz} = \widehat{P}_{[0,t]}^{tz}$ and $Q_>^{tz} = \widehat{P}_{[t,1]}^{tz}$ in the above claim, we see that the time marginals are unchanged: $P_s^* = \widehat{P}_s$, for all $s \in [0,1]$ and that $H(P^*|R) < H(\widehat{P}|R) : \widehat{P}$ is not the solution to $(\mathrm{S_{dyn}})$, a contradiction.

It remains to prove the claim. With (72), we see that

$$H(P|R) = H(\mu|R_t) + \int_{\mathcal{X}} H(P^{tz}|R^{tz}) \, \mu(dz)$$

and

$$H(P^{tz}|R^{tz}) = H(Q_<^{tz}|R_{[0,t]}^{tz}) + \int_{\Omega_{[0,t]}} H(P_{[t,1]}^{tz}(\cdot|X_{[0,t]})|R_{[t,1]}^{tz}(\cdot|X_{[0,t]})) \, dQ_<^{tz}.$$

Noting that $R_{[t,1]}^{tz}(\cdot|X_{[0,t]}) = R_{[t,1]}^{tz}$ and $Q_>^{tz} = \int_{\Omega_{[0,t]}} P_{[t,1]}^{tz}(\cdot|X_{[0,t]}) \, dQ_<^{tz}$, we obtain with Jensen's inequality that

$$
\begin{aligned}
H\big(Q_>^{tz}\big|R_{[t,1]}^{tz}\big) &= H\Big( \int_{\Omega_{[0,t]}} P_{[t,1]}^{tz}(\cdot|X_{[0,t]}) \, dQ_<^{tz} \Big| R_{[t,1]}^{tz}(\cdot|X_{[0,t]}) \Big) \\
&\le \int_{\Omega_{[0,t]}} H(P_{[t,1]}^{tz}(\cdot|X_{[0,t]})|R_{[t,1]}^{tz}(\cdot|X_{[0,t]})) \, dQ_<^{tz}
\end{aligned}
$$

with equality if and only if $P_{[t,1]}^{tz}(\cdot|X_{[0,t]}) = Q_>^{tz}$, $\quad Q_<^{tz}$-a.e. Since this holds for $\mu$-almost every $z$, this amounts to say that $P(\cdot) = \int_{\mathcal{X}} Q_<^{tz} \otimes Q_>^{tz}(\cdot) \, P_t(dz)$ and it also means that $P_{[t,1]}^{tz}(\cdot|X_{[0,t]}) = P_{[t,1]}^{tz}$, $P_{[0,t]}^{tz}$-a.e. which is the desired forward Markov property at time $t$. This completes the proofs of the claim and the proposition. $\qquad\square$

*Reversibility.* A path measure $R \in \mathrm{M}_+(\Omega)$ is said to be *reversible* with $m \in \mathrm{M}_+(\mathcal{X})$ as its reversing measure (*m-reversible* for short), if $R_0 = m$ and for any $0 \le u \le v \le 1$, $R$ is invariant with respect to the time reversal mapping $\mathrm{rev}^{uv}$ defined by: $\mathrm{rev}_t^{uv} := X_{(u+v-t)^+}, u \le t \le v$, meaning that $(\mathrm{rev}^{uv})_\# R = (X_{[u,v]})_\# R$.

Clearly, this implies that $R_u = R_v$ for any $u, v$. In other words, $R$ is *m-stationary* i.e. $R_t = m$, for all $0 \le t \le 1$.

This notion is invoked at statement (c) of the following result.

**Theorem 2.6** (The Markov case).

*Suppose that the reference measure $R \in \mathrm{M}_+(\Omega)$ satisfies*

(i) *$R$ is Markov;*

(ii) *there exist some $0 < t_o < 1$ and some measurable $\mathcal{X}_o \subset \mathcal{X}$ such that $R_{t_o}(\mathcal{X}_o) > 0$ and*

$$R_{01} \ll R\big((X_0, X_1) \in \cdot | X_{t_o} = z\big), \quad \forall z \in \mathcal{X}_o.$$

(iii) *$R_0 = R_1 = m \in \mathrm{M}_+(\mathcal{X})$;*

(iv) *$R_{01}(dxdy) \ge e^{-A(x)-A(y)} \, m(dx)m(dy)$ for some nonnegative measurable function $A$ on $\mathcal{X}$;*

(v) *$\int_{\mathcal{X}^2} e^{-B(x)-B(y)} R_{01}(dxdy) < \infty$ for some nonnegative measurable function $B$ on $\mathcal{X}$;*

*Suppose also that the constraint $(\mu_0, \mu_1)$ satisfies*

(vi) *$H(\mu_0|m), H(\mu_1|m) < \infty$;*

(vii) *$\int_{\mathcal{X}} (A + B) \, d\mu_0, \int_{\mathcal{X}} (A + B) \, d\mu_1 < \infty$ where $A$ and $B$ appear at (iv) and (v) above.*

(a) *Then, the unique solution $\widehat{P}$ of* ($S_{\text{dyn}}$) *is also Markov and*

$$\widehat{P} = f_0(X_0)g_1(X_1)\, R \in \mathrm{P}(\Omega) \tag{26}$$

   *where $f_0$ and $g_1$ are the $m$-measurable* nonnegative *functions which appear at* (23) *and solve the Schrödinger system* (24).

(b) *Conversely, let $\widehat{P}$ be defined by* (26) *with $f_0$ and $g_1$ two $m$-measurable nonnegative functions solving the Schrödinger system* (24). *Then, $\widehat{P}$ is Markov and it is the unique solution of* ($S_{\text{dyn}}$).

(c) *For the properties (i), (ii) and (iii) to hold, it is enough that $R$ is a $m$-reversible Markov measure which admits a regenerative set in the following sense: There exists a measurable subset $\mathcal{X}_o \subset \mathcal{X}$ with $m(\mathcal{X}_o) > 0$ such that for each $0 < h < 1$ and all measurable subset $A \subset \mathcal{X}$ with $m(A) > 0$, we have: $R(X_h \in A | X_0 = x) > 0$, for all $x \in \mathcal{X}_o$.*

*Proof.* • Proof of (a). By Proposition 2, the properties (iii)–(vii) ensure the existence of the unique solution $\widehat{P}$ of ($S_{\text{dyn}}$).

With Proposition 1, we have $\widehat{P}^{xy} = R^{xy}$, for all $(x, y)$, $\widehat{\pi}$-a.e. This means that

$$\frac{d\widehat{P}}{dR} = \frac{d\widehat{\pi}}{dR_{01}}(X_0, X_1).$$

On the other hand, we have just seen at Proposition 5 that $\widehat{P}$ is Markov. But, it is proved in [42] that under the assumptions (i) and (ii), if $P \in \mathrm{P}(\Omega)$ is a Markov measure such that $dP/dR = h(X_0, X_1)$ for some measurable function $h$, then there exist two measurable nonnegative functions $f$ and $g$ such that $P = f(X_0)g(X_1)\, R$. This proves statement (a).

• Proof of (b). The fact that $\widehat{P}$ is the solution of ($S_{\text{dyn}}$) is proved in [11] by a geometric approach or in [38] by a functional analytic approach. We easily see that $\widehat{P}$ inherits the Markov property of $R$ using the time-symmetric definition of the Markov property together with the product shape of (26).

• Proof of (c). Statement (c) is an easy exercise.                                    □

Remark that, unlike Theorem 2.4, Theorem 2.6 does not require that the constraint $(\mu_0, \mu_1)$ is internal. Also remark that the functions $f_0$ and $g_1$ are nonnegative (in contrast with Theorem 2.4 where they are positive) and that it may happen that $R_0(f_0 = 0)$ or $R_1(g_1 = 0)$ is positive.

Theorem 2.6 extends a similar result by Föllmer and Gantert in [25], where it is required for the product shape formula (26) to hold, that $R \ll \widehat{P}$ and also that $\mathcal{X}_o$ has full measure.

**3. $(f, g)$-transform of a Markov measure.** Motivated by Theorems 2.5 and 2.6, we introduce the transform $f_0(X_0)g_1(X_1)\, R$ of a Markov measure $R$ and call it an $(f, g)$-transform. It was already noticed by Föllmer [24, 25] and Nagasawa [56], that it is a time symmetric version of Doob's usual $h$-transform [19, 20].

**3.1. $(f, g)$-transform of a reversible Markov measure.** Let us first state an assumption for simplicity which will hold for the remainder of the paper.

**Assumption 1.** *The reference path measure $R \in \mathrm{M}_+(\Omega)$ is Markov and $m$-reversible with $m \in \mathrm{M}_+(\mathcal{X})$.*

The Markov property of the reference measure will turn out to be crucial for the description of the dynamics of the solution $\widehat{P}$ of ($S_{dyn}$). Indeed, we have already seen at Proposition 5 that $\widehat{P}$ inherits the Markov property from $R$. It follows that its dynamics is characterized by its stochastic derivatives. On the other hand, reversibility is only assumed for simplicity.

**Definition 3.1** (($f, g$)-transform). Let $f_0$, $g_1 : \mathcal{X} \to [0, \infty)$ be two nonnegative measurable functions such that $E_R(f_0(X_0)g_1(X_1)) = 1$. The path measure

$$P := f_0(X_0)g_1(X_1)\, R \in \mathrm{P}(\Omega) \tag{27}$$

is called an ($f, g$)-transform of $R$.

This definition is motivated by Theorems 2.5 and 2.6 which assert that the solution $\widehat{P}$ of ($S_{dyn}$) is an ($f, g$)-transform of $R$. Note that under Theorem 2.5's assumptions, $f_0$ and $g_1$ are positive, while they are allowed to vanish under Theorem 2.6's assumptions, as in Definition 3.1.

Let us introduce for each $t \in [0, 1]$, the functions $f_t, g_t : \mathcal{X} \to [0, \infty)$ defined by

$$\begin{cases} f_t(z) & := & E_R(f_0(X_0) \mid X_t = z) \\ g_t(z) & := & E_R(g_1(X_1) \mid X_t = z) \end{cases}, \quad \text{for } P_t\text{-a.e. } z \in \mathcal{X}. \tag{28}$$

Remark that although we have $E_R(f_0(X_0)g_1(X_1)) < \infty$, this does not ensure that $f_0(X_0)$ and $g_1(X_1)$ are integrable. We have to use positive integration to give a meaning to the conditional expectations $E_R(f_0(X_0) \mid X_t), E_R(g_1(X_1) \mid X_t) \in [0, \infty]$, see [35].

Next result is a kind of converse of Theorems 2.5 and 2.6.

**Theorem 3.2.** *If the functions $f_0$ and $g_1$ entering the definition of the ($f, g$)-transform $P$ of $R$ given at (27) satisfy*

$$\begin{cases} \int_{\mathcal{X}} g_0 f_0 \log f_0 \, dm < \infty \\ \int_{\mathcal{X}} f_1 g_1 \log g_1 \, dm < \infty \end{cases}$$

*(as a convention $0 \log 0 = 0$), then $P_{01}$ and $P$ are the unique solutions of* (S) *and* ($S_{dyn}$) *respectively, where the prescribed constraints $\mu_0$, $\mu_1 \in \mathrm{P}(\mathcal{X})$ are chosen to satisfy* (24)*, i.e. using notation* (28)

$$\begin{cases} \mu_0 = f_0 g_0 \, m, \\ \mu_1 = f_1 g_1 \, m. \end{cases} \tag{29}$$

*Proof.* The pair of finite integrals is equivalent to $H(P|R) < \infty$. Hence, the statement about (S) is a direct consequence of [38, Thm. 5.1]. Its corollary about ($S_{dyn}$) follows as for Theorem 2.5.    □

Note that there may exist solutions of ($S_{dyn}$) which are not ($f, g$)-transforms of $R$. This happens when the support of the solution is not a rectangle (i.e. the product of Borel subsets), see [25, §2] or [38, §5].

Next result extends the product formulas (29) to all $t \in [0, 1]$.

**Theorem 3.3** (Euclidean analogue of Born's formula). *The path measure $P = f_0(X_0)g_1(X_1)\, R$ is Markov and for each $0 \le t \le 1$, its time marginal $P_t \in \mathrm{P}(\mathcal{X})$ is given by*

$$P_t = f_t g_t \, m. \tag{30}$$

**Remark 1.** It follows with (30) that for all $t \in [0, 1]$, $0 < E_R(f_0(X_0) \mid X_t)$, $E_R(g_1(X_1) \mid X_t) < \infty$, $P$-a.e., but not $R$-a.e. in general.

*Proof of Theorem 3.3.* The Markov property of $P$ is a direct consequence of Theorem 3.2 and Proposition 5.

We propose an alternate simple proof. To prove that $P$ is Markov, we show that for each $0 \leq t \leq 1$ and any bounded measurable functions $a \in \sigma(X_{[0,t]})$, $b \in \sigma(X_{[t,1]})$, we have

$$E_P(ab \mid X_t) = E_P(a \mid X_t)E_P(b \mid X_t).$$

Indeed, we have

$$
E_P(ab \mid X_t) \stackrel{\text{(i)}}{=} \frac{E_R(f_0(X_0)abg_1(X_1) \mid X_t)}{E_R(f_0(X_0)g_1(X_1) \mid X_t)}
$$

$$
\stackrel{\text{(ii)}}{=} \frac{E_R(f_0(X_0)a \mid X_t)E_R(bg_1(X_1) \mid X_t)}{E_R(f_0(X_0) \mid X_t)E_R(g_1(X_1) \mid X_t)}
$$

$$
\stackrel{\text{(iii)}}{=} E_P(a \mid X_t)E_P(b \mid X_t), \quad P\text{-a.e.,}
$$

which is the desired result. Equality (i) is a general result about conditioning; note that we do not divide by zero $P$-a.s. Equality (ii) uses crucially the assumed Markov property of $R$ and one obtains (iii) by considering separately the cases when $b \equiv 1$ and $a \equiv 1$ in the just obtained identity to see that $E_P(a \mid X_t) = \frac{E_R(f_0(X_0)a|X_t)}{E_R(f_0(X_0)|X_t)}$ and $E_P(b \mid X_t) = \frac{E_R(bg_1(X_1)|X_t)}{E_R(g_1(X_1)|X_t)}$.

Finally, to prove (30), remark that

$$
\frac{dP_t}{dm}(X_t) = \frac{dP_t}{dR_t}(X_t) = E_R\left(\frac{dP}{dR} \mid X_t\right) = E_R(f_0(X_0)g_1(X_1) \mid X_t)
$$

$$
\stackrel{\checkmark}{=} E_R(f_0(X_0) \mid X_t)E_R(g_1(X_1) \mid X_t) =: f_t(X_t)g_t(X_t)
$$

where we used the Markov property of $R$ at the marked equality. $\qquad\square$

### 3.2. Forward and backward generators.
Let $Q \in \mathrm{M}_+(\Omega)$ be a Markov measure. Its *forward stochastic derivative* $\partial + \overrightarrow{L}^Q$ is defined by

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

for any measurable function $u : [0,1] \times \mathcal{X} \to \mathbb{R}$ in the set dom $\overrightarrow{L}^Q$ for which this limit exists $Q_t$-a.e. for all $0 \leq t < 1$. In fact this definition is only approximate, we give it here as a support for understanding the relations between the forward and backward generators. For a precise statement see [36, §2]. Since the time reversed $Q^*$ of $Q$ is still Markov, $Q$ admits a *backward stochastic derivative* $-\partial + \overleftarrow{L}^Q$ which is defined by

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

for any measurable function $u : [0,1] \times \mathcal{X} \to \mathbb{R}$ in the set dom $\overleftarrow{L}^Q$ for which this limit exists $Q_t$-a.e. for all $0 < t \leq 1$. Remark that $\overleftarrow{L}_t^Q = \overrightarrow{L}_{1-t}^{Q^*}$, $0 \leq t \leq 1$.

It is proved in [36, §2] that these stochastic derivatives are extensions of the extended forward and backward generators of $Q$ in the sense of semimartingales. In particular, they offer us a natural way for computing generators. Later on, we shall call $\overrightarrow{L}^Q$ and $\overleftarrow{L}^Q$ generators, rather than stochastic derivatives.

For simplicity, we denote $\overrightarrow{L}^R = \overleftarrow{L}^R = L$ without the superscript $R$ and without the time arrows, since $R$ is assumed to be reversible. We also write $\overrightarrow{A} = \overrightarrow{L}^P$ and $\overleftarrow{A} = \overleftarrow{L}^P$ the generators of the $(f_0, g_1)$-transform $P$ defined by (27).

Stochastic derivatives have been introduced by E. Nelson in [58] while studying the dynamical properties of the Brownian motion. The above definition (more precisely the one of [36]), which is an extension of Nelson's one, is necessary for technical reasons.

3.3. **The dynamics of the $(f, g)$-transform.** To give the expressions of the generators $\overrightarrow{A}$ and $\overleftarrow{A}$, we need to introduce the carré du champ of $R$. It is defined for any functions $u, v$ on $\mathcal{X}$ such that $u, v$ and $uv$ are in dom $L$, by

$$\Gamma(u, v) := L(uv) - uLv - vLu.$$

In general, the forward and backward generators $(\partial_t + \overrightarrow{A}_t)_{0 \leq t \leq 1}$ and $(-\partial_t + \overleftarrow{A}_t)_{0 \leq t \leq 1}$ of $P$ depend explicitly on $t$. The following informal statement is known for long in specific situations. In the important examples which are discussed below at Section 4, these claims are easy consequences of Itô's formula (for instance see [62, Ch. 8, § 3] in the continuous diffusion case).

**Informal statement 1.** *Under some hypotheses on $R$, the forward and backward generators of $P$ are given for any function $u : [0, 1] \times \mathcal{X} \to \mathbb{R}$ belonging to some class $\mathcal{U}_R$ of regular functions, by*

$$\begin{cases} \overrightarrow{A}_t u(x) &= Lu(x) + \dfrac{\Gamma(g_t, u)(x)}{g_t(x)}, \quad (t, x) \in [0, 1) \times \mathcal{X} \\ \overleftarrow{A}_t u(x) &= Lu(x) + \dfrac{\Gamma(f_t, u)(x)}{f_t(x)}, \quad (t, x) \in (0, 1] \times \mathcal{X} \end{cases} \tag{31}$$

*where $f_t, g_t$ are defined at (28).*

*Because of (30), for any $t$ no division by zero occurs $P_t$-a.e.*

Rigorous statement and proof are given in [36] for instance.

*Idea of proof of Statement 1.* To obtain the forward generator of $P$, we are going to compute the stochastic derivative $\overrightarrow{A}_t u(x) := \lim_{h \downarrow 0} h^{-1} E_P[u(X_{t+h}) - u(X_t) \mid X_t = x]$. Let us denote for simplicity $F_0 = f_0(X_0)$, $G_t = g_t(X_t)$ and $U_t = u(X_t)$. We have

$$\begin{aligned} E_P[U_{t+h} - U_t \mid X_t = x] &= \frac{E_R[F_0(U_{t+h} - U_t)G_1 \mid X_t = x]}{E_R[F_0 G_1 \mid X_t = x]} \\ &= \frac{E_R[F_0 \mid X_t = x]E_R[(U_{t+h} - U_t)G_1 \mid X_t = x]}{E_R[F_0 \mid X_t = x]E_R[G_1 \mid X_t = x]} \\ &= \frac{E_R[(U_{t+h} - U_t)G_1 \mid X_t = x]}{g_t(x)}, \end{aligned}$$

where the Markov property of $R$ is used at second identity. But,

$$\begin{aligned} & E_R[(U_{t+h} - U_t)G_1 \mid X_t = x] \\ =\ & E_R[(U_{t+h} - U_t)G_{t+h} \mid X_t = x] \\ =\ & g_t(x)E_R[(U_{t+h} - U_t) \mid X_t = x] + E_R[(U_{t+h} - U_t)(G_{t+h} - G_t) \mid X_t = x], \end{aligned}$$

where the first equality is a martingale identity. We conclude by means of the definition of $L$: $\lim_{h \downarrow 0} h^{-1} E_R[(U_{t+h} - U_t) \mid X_t = x] =: Lu(x)$, and with the following identity $\lim_{h \downarrow 0} h^{-1} E_R[(U_{t+h} - U_t)(G_{t+h} - G_t) \mid X_t = x] = \Gamma(u, g_t)(x)$. □

One sees that it is necessary that the functions $f_t$ and $g_t$ are regular enough to be in the domains of the carré du champ operators. For instance, choosing $f_0, g_1 \in \mathrm{dom}\, L$ insures that $f \in \mathrm{dom}(-\partial + L)$ and $g \in \mathrm{dom}(\partial + L)$ and also that $f$ and $g$ are *classical solutions* of the following parabolic PDEs

$$\begin{cases} (-\partial_t + L)f(t,x) = 0, & 0 < t \leq 1, \\ f_0, & t = 0, \end{cases} \qquad \begin{cases} (\partial_t + L)g(t,x) = 0, & 0 \leq t < 1, \\ g_1, & t = 1. \end{cases}$$
(32)

Even better, since $R$ is assumed to be $m$-reversible, its Markov generator $L$ is self-adjoint on $L^2(m)$ and for any $f_0, g_1$ in $L^2(m)$ we have $f \in \mathrm{dom}(-\partial + L)$ and $g \in \mathrm{dom}(\partial + L)$.

It is worthwhile describing the dynamics (31) in terms of

$$\begin{cases} \varphi := \log f, \\ \psi := \log g. \end{cases}$$
(33)

Remark that because of (30), for any $0 \leq t \leq 1$, $\varphi_t$ and $\psi_t$ are well defined $P_t$-a.e. In analogy with the Kantorovich potentials which appear in the optimal transport theory, we call $\varphi$ and $\psi$ the *Schrödinger potentials*. They are solutions of the "second order"[8] Hamilton-Jacobi-Bellman equations

$$\begin{cases} (-\partial_t + B)\varphi(t,x) = 0, & 0 < t \leq 1, & P_t\text{-a.e.} \\ \varphi_0 = \log f_0, & t = 0, \end{cases}$$
(34)

and

$$\begin{cases} (\partial_t + B)\psi(t,x) = 0, & 0 \leq t < 1, & P_t\text{-a.e.} \\ \psi_1 = \log g_1, & t = 1, \end{cases}$$
(35)

where the non-linear operator $B$ is defined by

$$Bu := e^{-u} L e^u$$

for any function $u$ such that $e^u \in \mathrm{dom}\, L$.

4. **Standard examples.** We present two well-known reference processes: the reversible Brownian motion and a reversible random walk on a graph. We also apply the above general results to these important examples.

4.1. **Reversible Brownian motion.** The reversible Brownian motion $R$ on $\mathcal{X} = \mathbb{R}^n$ is specified by

$$\begin{cases} L = \Delta/2, \\ R_0(dx) = m(dx) = dx \end{cases}$$
(36)

where the Markov generator $L = \Delta/2$ is defined on $\mathcal{C}^2(\mathbb{R}^n)$. It is easily checked that $R$ is $m$-reversible.

Let $P = f_0(X_0)g_1(X_1)\, R \in \mathrm{P}(\Omega)$ be any $(f,g)$-transform of $R$. By the regularity improving property of the heat kernel, $(f_0, g_1)$ is such that $f \in \mathrm{dom}(-\partial_t + \Delta/2)$ for $t \in (0,1]$ and $g \in \mathrm{dom}(\partial_t + \Delta/2)$ for $t \in [0,1)$. We have $Bu = \Delta u/2 + |\nabla u|^2/2$, $\Gamma(u,v) = \nabla u \cdot \nabla v$ for any $u, v \in \mathcal{C}^2(\mathcal{X})$. The expressions (31)

$$\begin{cases} \overrightarrow{A}_t & = & \Delta/2 + \nabla\psi_t \cdot \nabla \\ \overleftarrow{A}_t & = & \Delta/2 + \nabla\varphi_t \cdot \nabla \end{cases}$$
(37)

---

[8]When $R$ is a random walk on a discrete space for instance, then the term "second order" is only justified in analogy with the continuous case, see (48).

of the forward and backward generators tell us that the density $\mu_t(x) := dP_t/dx$ solves the following parabolic PDEs

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

where $\psi$ solves (35):

$$\begin{cases} (\partial_t + \Delta/2)\psi_t(x) + |\nabla\psi_t(x)|^2/2 = 0, & (t,x) \in [0,1) \times \mathcal{X} \\ \psi_1 = \log g_1, & t = 1, \end{cases} \tag{38}$$

and in the reversed sense of time

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

where $\varphi$ solves (34):

$$\begin{cases} (-\partial_t + \Delta/2)\varphi_t(x) + |\nabla\varphi_t(x)|^2/2 = 0, & (t,x) \in (0,1] \times \mathcal{X} \\ \varphi_0 = \log f_0, & t = 0. \end{cases}$$

It is important to note the smoothing effect of the semigroup of $R$ which allows us to define the *classical gradients* $\nabla\psi_t$ and $\nabla\varphi_t$ for all $t$ in $[0,1)$ and $(0,1]$ respectively. They are the forward and backward drift vector fields of the canonical process under $P$. Also recall that, as a direct consequence of (30), no logarithm of zero is taken, $P_t$-almost surely, i.e. almost everywhere, and we have the time-reversal formula

$$\nabla\psi_t + \nabla\varphi_t = \nabla \log \mu_t, \quad 0 < t < 1.$$

*Back to Schrödinger problem.* Under the assumption of Theorem 3.2, $P = f_0(X_0)g_1(X_1)R$ solves ($S_{dyn}$) with the prescribed marginals given at (29). It is a $(f,g)$-transform of $R$, and we have just seen that there exist $\varphi$ and $\psi$ such that its forward and backward generators are given by (37).

*Minimal action.* We derive the analogue of Benamou-Brenier formula (7). Consider the problem of minimizing the average kinetic action

$$\int_{[0,1] \times \mathbb{R}^n} \frac{|v_t(x)|^2}{2} \, \nu_t(dx)dt \to \min \tag{39}$$

among all $(\nu, v)$ where $\nu = (\nu_t)_{0 \le t \le 1}$ is a measurable path in $P(\mathbb{R}^n)$, $(v_t(x))_{(t,x) \in [0,1] \times \mathbb{R}^n}$ is a measurable $\mathbb{R}^n$-valued vector field and the following constraints are satisfied:

$$\begin{cases} (\partial_t - \Delta/2)\nu + \nabla \cdot (\nu v) = 0, & \text{on } (0,1) \times \mathbb{R}^n \\ \nu_0 = \mu_0, \ \nu_1 = \mu_1 \end{cases} \tag{40}$$

where this evolution equation is meant in the following weak sense: for any function $u \in \mathcal{C}_o^{1,2}((0,1) \times \mathbb{R}^n)$, we have $\int_{(0,1) \times \mathbb{R}^n}(\partial_t + \Delta/2 + v \cdot \nabla)u(t,x) \, \nu_t(dx)dt = 0$.

**Proposition 6.** *Let $\mu_0, \mu_1 \in P_2(\mathbb{R}^n)$ be such that $H(\mu_0|\text{Leb}), H(\mu_1|\text{Leb}) < \infty$. Then, ($S_{dyn}$) has a unique solution $\widehat{P} \in P(\Omega)$ and the unique solution to the minimal action problem (39) is $((\mu_t)_{t \in [0,1]}, \nabla\psi)$ where*

$$\mu_t = \widehat{P}_t, \quad t \in [0,1],$$

*and*

$$\psi_t(x) = \log E_R[g_1(X_1) \mid X_t = x], \quad (t,x) \in [0,1) \times \mathbb{R}^n$$

*with $g_1$ a solution of* (24). *Moreover, $\psi$ is the unique classical solution of the Hamilton-Jacobi-Bellman equation* (38) *and*

$$\inf\left\{H(\pi|R_{01}); \pi \in \mathrm{P}(\mathcal{X}^2): \pi_0 = \mu_0, \pi_1 = \mu_1\right\} - H(\mu_0|m)$$

$$= \inf\left\{\int_{\mathcal{X}} H(\pi^x \mid R_1^x)\,\mu_0(dx); (\pi^x)_{x\in\mathcal{X}}: \int_{\mathcal{X}} \pi^x(\cdot)\,\mu_0(dx) = \mu_1\right\}$$

$$= \int_{[0,1]\times\mathbb{R}^n} \frac{|\nabla\psi_t(x)|^2}{2}\,\mu_t(dx)dt \tag{41}$$

$$= \inf\left\{\int_{[0,1]\times\mathbb{R}^n} \frac{|v_t(x)|^2}{2}\,\nu_t(dx)dt; (\nu,v):(\nu,v)\ \text{satisfies}\ (40)\right\}.$$

*where $R_1^x := R(X_1 \in \cdot|X_0 = x) \in \mathrm{P}(\mathcal{X})$ and $(\pi^x)_{x\in\mathcal{X}} \in \mathrm{P}(\mathcal{X})^{\mathcal{X}}$ is a measurable Markov kernel.*

*Proof.* The integrability assumptions on $\mu_0$ and $\mu_1$ ensure that the hypotheses of Proposition 2 are satisfied. Therefore, ($\mathrm{S_{dyn}}$) admits a unique solution.

As a consequence of Girsanov's theory, for any $Q \in \mathrm{P}(\Omega)$ such that $H(Q|R) < \infty$, there exists some predictable $\mathbb{R}^n$-valued drift field $\beta$ such that

(i) $E_Q \int_{[0,1]} |\beta_t|^2\,dt < \infty$,
(ii) $Q$ solves the martingale problem associated with the forward generator$(\partial_t + \Delta/2 + \beta_t \cdot \nabla)_{0\le t\le 1}$
(iii) and $H(Q|R)$ is given by

$$H(Q|R) = H(Q_0|m) + E_Q \int_{[0,1]} \frac{|\beta_t|^2}{2}\,dt. \tag{42}$$

For a proof with an analytic flavour, see for instance [41].

It is proved in [6] (see also [46] for a related result) that when $(\nu,v)$ satisfies $(\partial_t - \Delta/2)\nu + \nabla\cdot(\nu v) = 0$ on $(0,1)\times\mathbb{R}^n$, and $\int_{[0,1]\times\mathbb{R}^n} |v_t(x)|^2/2\,\nu_t(dx)dt < \infty$, then there exists a path measure $Q \in \mathrm{P}(\Omega)$ such that

$$Q_t = \nu_t, \quad \forall 0 \le t \le 1, \tag{43}$$

$Q$ solves the martingale problem associated with the forward stochastic derivative $(\partial_t + \Delta/2 + v_t \cdot \nabla)_{0\le t\le 1}$ and, by (42)

$$H(Q|R) = H(\mu_0|m) + \int_{[0,1]\times\mathbb{R}^n} \frac{|v_t(x)|^2}{2}\,\nu_t(dx)dt < \infty \tag{44}$$

with $Q_0 = \mu_0$ and $Q_1 = \mu_1$. Therefore, to any $(\nu,v)$ which satisfies $\int_{[0,1]\times\mathbb{R}^n} |v_t(x)|^2/2\,\nu_t(dx)dt < \infty$ and (40), one can associate some $Q \in \mathrm{P}(\Omega)$ which verifies (43) and (44). Since

$$\inf(\mathrm{S}) = \inf(\mathrm{S_{dyn}}) = H(\widehat{P}|R) = H(\mu_0|m) + \int_{[0,1]\times\mathcal{X}} \frac{|\nabla\psi_t(x)|^2}{2}\,\mu_t(dx)dt$$

with $\mu_t := \widehat{P}_t$ and $\psi$ given at (37), and since $\widehat{P}$ is the unique solution to ($\mathrm{S_{dyn}}$), we have proved that identity (41) holds true, $(\mu, \nabla\psi)$ solves (39) and $\mu$ is unique in the sense that if $(\nu,v)$ and $(\nu',v')$ are solutions, then $\nu = \nu' = \mu$. It remains to check that $\nabla\psi$ is also unique. The following property of $\mu$ holds with $v = \nabla\psi$:

$$\int_{(0,1)\times\mathbb{R}^n} (\partial_t + \Delta/2 + v \cdot \nabla)u\,d\mu_t dt = 0, \quad \forall u \in \mathcal{C}_o^\infty((0,1)\times\mathbb{R}^n).$$

Suppose that it is also verified with $v'$. Subtracting, we obtain $\int_{(0,1)\times\mathbb{R}^n}(v'-\nabla\psi)\cdot\nabla u\,d\mu_t dt = 0$, $\forall u \in \mathcal{C}_o^\infty((0,1)\times\mathbb{R}^n)$, meaning that $v' - \nabla\psi$ is orthogonal in $L^2_{\mathbb{R}^n}((0,1)\times\mathbb{R}^n, d\mu_t dt)$ to $\nabla\mathcal{C}_o^\infty((0,1)\times\mathbb{R}^n)$. It follows that the squared norm $\|v'\|^2 := \int_{[0,1]\times\mathbb{R}^n}|v'_t(x)|^2\,\mu_t(dx)dt$ in $L^2_{\mathbb{R}^n}((0,1)\times\mathbb{R}^n, d\mu_t dt)$ is minimal at the orthogonal projection of $\nabla\psi$ on the closure of $\nabla\mathcal{C}_o^\infty((0,1)\times\mathbb{R}^n)$. Of course, this projection is $\nabla\psi$ itself. $\qquad\square$

Clearly, there is also a backward version of Proposition 6.

4.2. **Reversible random walk on a graph.** Let $R$ be a random walk on a countable connected graph $(\mathcal{X},\sim)$ where $x \sim y$ means that $x$ and $y$ are next neighbours. Its generator is given for any finitely supported function $u$ by

$$Lu(x) = \sum_{y:y\sim x}[u(y)-u(x)]\,J_x(y), \quad x \in \mathcal{X} \qquad (45)$$

where $J_x(y) > 0$ for all $x \sim y$ is interpreted as the average frequency of jumps from $x$ to $y$. Its dual formulation is the current equation

$$\partial_t\rho_t(x) = \sum_{y:y\sim x}[\rho_t(y)J_y(x)-\rho_t(x)J_x(y)], \qquad x\in\mathcal{X}, 0<t<1$$

where $\rho_t(x) := R(X_t = x)$. Clearly, the path measure $R$ admits the stationary measure $m \in \mathrm{M}_+(\mathcal{X})$ if and only if $\sum_{y:y\sim x}[m_y J_y(x) - m_x J_x(y)] = 0$, for all $x \in \mathcal{X}$. It admits $m$ as a reversing measure if this global equilibrium property is reinforced into the following detailed one:

$$m_x J_x(y) = m_y J_y(x), \quad \forall x,y \in \mathcal{X}, x \sim y.$$

The special case where the jump measure

$$J_x := \sum_{y:y\sim x} J_x(y)\,\delta_y \in \mathrm{M}_+(\mathcal{X})$$

is equal to $J_x^o = \frac{1}{n_x}\sum_{y:y\sim x}\delta_y$ where

$$n_x := \#\{y : y \sim x\} < \infty$$

is the number of neighbours of $x$ which is assumed to be finite for all $x$, corresponds to the *simple random walk* on the graph $(\mathcal{X},\sim)$. It is easily checked that its reversing measure is $m^o = \sum_{x\in\mathcal{X}} n_x\delta_x \in \mathrm{M}_+(\mathcal{X})$ which is unbounded when $\mathcal{X}$ is infinite.

For simplicity we assume in the general case that

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

where $J_x(\mathcal{X}) := \sum_{y:y\sim x} J_x(y)$ is the global average frequency of jumps at $x$. This ensures that for any initial marginal $R_0 \in \mathrm{M}_+(\mathcal{X})$, there exists a Markov measure $R \in \mathrm{M}_+(\Omega)$ with generator $L$. Moreover, any bounded function $u : \mathcal{X} \to \mathbb{R}$ is in the domain of $L$.

Viewing $L = (L_{xy})_{x,y\in\mathcal{X}}$ as a matrix with $L_{xy} = \begin{cases} J_x(y), & \text{if } x \sim y; \\ -J_x(\mathcal{X}), & \text{if } x = y; \\ 0, & \text{otherwise} \end{cases}$ and

the functions as column vectors indexed by $\mathcal{X}$, we observe that the solutions of the heat equations (32) are

$$f_t = e^{tL}f_0 \qquad \text{and} \qquad g_t = e^{(1-t)L}g_1, \quad 0 \le t \le 1. \qquad (46)$$

It follows that for any couple $(f_0, g_1)$ of nonnegative functions and any $0 \leq t \leq 1$, $f_t \in \mathrm{dom}(-\partial_t + L)$ and $g_t \in \mathrm{dom}(\partial_t + L)$.

We have $Bu(x) = Lu(x) + \sum_{y:y \sim x} \theta(u(y) - u(x)) J_x(y)$ with

$$\theta(a) := e^a - a - 1, \quad a \in \mathbb{R}, \tag{47}$$

$\Gamma(u, v)(x) = \sum_{y:y \sim x} [u(y) - u(x)][v(y) - v(x)] J_x(y)$, for any bounded functions $u, v$ and any $x \in \mathcal{X}$. Let $P = f_0(X_0) g_1(X_1) R$ be any $(f, g)$-transform of the random walk $R$. Applying (31), the forward and backward generators of $P$ turn out to be the jump generators associated respectively with the jump measures

$$\begin{cases} \overrightarrow{J}_x &= \displaystyle\sum_{y:y \sim x} \frac{g_t(y)}{g_t(x)} J_x(y)\, \delta_y &= \displaystyle\sum_{y:y \sim x} \exp\big(\psi_t(y) - \psi_t(x)\big) J_x(y)\, \delta_y \\[2mm] \overleftarrow{J}_x &= \displaystyle\sum_{y:y \sim x} \frac{f_t(y)}{f_t(x)} J_x(y)\, \delta_y &= \displaystyle\sum_{y:y \sim x} \exp\big(\varphi_t(y) - \varphi_t(x)\big) J_x(y)\, \delta_y \end{cases}$$

Again, no division by zero occurs $P_t$-a.e. for every $0 \leq t \leq 1$, i.e. everywhere for each $0 < t < 1$. The functions $f$ and $g$ satisfy (46) and the Schrödinger potentials $\psi$ and $\varphi$ satisfy (35):

$$\begin{cases} (\partial_t + L)\psi_t(x) + \sum_{y:y \sim x} \theta\big(\psi_t(y) - \psi_t(x)\big) J_x(y) = 0, & (t, x) \in [0, 1) \times \mathcal{X} \\ \psi_1 = \log g_1, & t = 1, \end{cases} \tag{48}$$

and (34):

$$\begin{cases} (-\partial_t + L)\varphi_t(x) + \sum_{y:y \sim x} \theta\big(\varphi_t(y) - \varphi_t(x)\big) J_x(y) = 0, & (t, x) \in (0, 1] \times \mathcal{X} \\ \varphi_0 = \log f_0, & t = 0. \end{cases}$$

*Minimal action.* Let $\theta^*$ be the convex conjugate of $\theta$ defined at (47), i.e.

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

Consider the problem of minimizing

$$\int_{[0,1] \times \mathcal{X}^2} \theta^*\big(j(t, x; y) - 1\big)\, \nu_t(dx) J_x(dy) dt \to \min \tag{49}$$

among all $(\nu, j)$ where $\nu = (\nu_t)_{0 \leq t \leq 1}$ is a measurable path in $\mathrm{P}(\mathcal{X})$, $j : [0, 1] \times \mathcal{X}^2 \to [0, \infty)$ is a measurable nonnegative function and the following constraints are satisfied:

$$\begin{cases} \partial_t \nu_t(x) = \sum_{y:y \sim x} \big\{ \nu_t(y) j(t, y; x) J_y(x) - \nu_t(x) j(t, x; y) J_x(y) \big\} = 0, & \text{on } (0, 1) \times \mathcal{X} \\ \nu_0 = \mu_0, \nu_1 = \mu_1 \end{cases}$$

$$\tag{50}$$

where we write $\nu_t = \sum_{x \in \mathcal{X}} \nu_t(x)\, \delta_x$.

**Proposition 7.** *Let $\mu_0, \mu_1 \in \mathrm{P}(\mathbb{R}^n)$ be such that $\inf (\mathrm{S}) < \infty$, for instance when the assumptions of Proposition 2 are satisfied.*

*The unique solution to the minimal action problem (49) is $(\mu, j^\psi)$ where $\mu = (\mu_t)_{t \in [0,1]}$ with*

$$\mu_t = \widehat{P}_t, \quad t \in [0, 1],$$

*and $\widehat{P} \in \mathrm{P}(\Omega)$ is the unique solution of $(\mathrm{S_{dyn}})$, and*

$$j^\psi(t, x; y) = \exp(\psi_t(y) - \psi_t(x))$$

*with*

$$\psi_t(x) = \log E_R[g_1(X_1) \mid X_t = x], \quad (t, x) \in [0, 1) \times \mathbb{R}^n$$

*with $g_1$ a solution of (24). Moreover, $\psi$ is the unique classical solution of the Hamilton-Jacobi-Bellman equation (48) and*

$$
\begin{aligned}
&\inf \left\{ H(\pi | R_{01}); \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1 \right\} - H(\mu_0 | m) \\
=\ & \inf \left\{ \int_{\mathcal{X}} H(\pi^x \mid R_1^x) \, \mu_0(dx); (\pi^x)_{x \in \mathcal{X}} : \int_{\mathcal{X}} \pi^x(\cdot) \, \mu_0(dx) = \mu_1 \right\} \\
=\ & \int_{[0,1] \times \mathcal{X}^2} \theta^* \big( j^\psi(t, x; y) - 1 \big) \, \mu_t(dx) J_x(dy) dt \\
=\ & \inf \left\{ \int_{[0,1] \times \mathcal{X}^2} \theta^* \big( j(t, x; y) - 1 \big) \, \nu_t(dx) J_x(dy) dt; (\nu, j) : (\nu, j) \ \textit{satisfies} \ (50) \right\}
\end{aligned}
\tag{51}
$$

*where $R_1^x := R(X_1 \in \cdot | X_0 = x) \in \mathrm{P}(\mathcal{X})$ and $(\pi^x)_{x \in \mathcal{X}} \in \mathrm{P}(\mathcal{X})^{\mathcal{X}}$ is any measurable Markov kernel.*

Remark that $\theta^* \big( j^\psi(t, x; y) - 1 \big) = \theta^* \big( \frac{g_t(y) - g_t(x)}{g_t(x)} \big)$ where $\frac{g_t(y) - g_t(x)}{g_t(x)}$ looks like a discrete logarithmic derivative which is analogous to $\nabla \psi_t(x) = \nabla \log g_t(x)$.

*Proof.* The proof follows the same line as Proposition 6's one. $\qquad \square$

5. **Slowing down.** In this section, we describe an efficient way to recover optimal transport as a limit of Schrödinger problems. The main idea consists in slowing the reference process down to a no-motion process. In the following lines, we present some heuristics and refer the reader to [40] for a rigorous treatment in the case where $\mathcal{X}$ is a real vector space and [33] in the alternate case where $\mathcal{X}$ is a discrete graph. The specific case of the reversible Brownian motion has been investigated by T. Mikami in [49] with a stochastic control approach which differs from what is presented below.

Let $R$ be Markov with generator $L$. The slowed down process is represented by the sequence $(R^k)_{k \geq 1}$ in $\mathrm{M}_+(\Omega)$ of Markov measures associated with the generators

$$L^k := L/k, \qquad k \geq 1.$$

Remark that slowing the process down doesn't modify its reversible measure $m$; one converges more slowly towards the same equilibrium. Suppose also that the sequence $(R^k)_{k \geq 1}$ in $\mathrm{M}_+(\Omega)$ obeys the large deviation principle in $\Omega$ with speed $\alpha_k$ and rate function $C$, meaning approximately that for a "large class" of measurable subsets $A$ of $\Omega$, we have

$$R^k(A) \underset{k \to \infty}{\asymp} \exp \left[ -\alpha_k \inf_{\omega \in A} C(\omega) \right]. \tag{52}$$

For instance, in the case (36) when $R$ is the reversible Brownian motion on $\mathbb{R}^n$, Schilder's theorem tells us that $C$ is the kinetic action (3) and $\alpha_k = k$. In the case (45) when $R$ is a reversible random walk, it is proved in [33] that $\alpha_k = \log k$ and the rate function is

$$C(\omega) := \sum_{0 \leq t \leq 1} \mathbf{1}_{\{\omega_{t^-} \neq \omega_t\}}, \qquad \omega \in \Omega, \tag{53}$$

where in this situation $\Omega$ is the space of all right-continuous paths with finitely many jumps. Remark that $C$ is simply the total number of jumps of the path.

At a heuristic level, the Γ-convergence of

$$H(P|R^k)/\alpha_k \to \min; \qquad P \in \mathrm{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1 \qquad (\mathrm{S}^k_{\mathrm{dyn}})$$

as $k$ tends to infinity to

$$\int_\Omega C \, dP \to \min; \qquad P \in \mathrm{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1 \qquad (\mathrm{MK}_{\mathrm{dyn}})$$

is best seen with the dual problems. Without getting into detail, to show this Γ-convergence, it is enough to prove that the objective functions of the dual problems converge pointwise, see [40] for detail. Let us check this pointwise convergence. Recall that the dual problem of (S) is

$$\Delta_R(\varphi_0, \psi_1) \to \max; \qquad \varphi_0, \psi_1 \in C_b(\mathcal{X}) \qquad (\mathrm{D})$$

with $\Delta_R(\varphi_0, \psi_1) := \int_\mathcal{X} \varphi_0 \, d\mu_0 + \int_\mathcal{X} \psi_1 \, d\mu_1 - \log \int_\Omega e^{\varphi_0(X_0) + \psi_1(X_1)} \, dR$. Consequently, for each $k \geq 1$, the dual problem of ($\mathrm{S}^k_{\mathrm{dyn}}$) consists in maximizing $\alpha_k^{-1} \Delta_{R^k}(\varphi_0, \psi_1)$ or equivalently: $\alpha_k^{-1} \Delta_{R^k}(\alpha_k \varphi_0, \alpha_k \psi_1)$. This leads us to

$$\int_\mathcal{X} \varphi_0 \, d\mu_0 + \int_\mathcal{X} \psi_1 \, d\mu_1 - \alpha_k^{-1} \log \int_\Omega e^{\alpha_k[\varphi_0(X_0) + \psi_1(X_1)]} \, dR^k \to \max; \quad \varphi_0, \psi_1 \in C_b(\mathcal{X})$$
$$(\mathrm{D}^k)$$

The pointwise limit, as $k$ tends to infinity, of $\alpha_k^{-1} \Delta_{R^k}(\alpha_k \varphi_0, \alpha_k \psi_1)$ is a direct consequence of the large deviation principle (52) and the Laplace-Varadhan integral lemma [16, Thm. 4.3.1], which provide us with

$$\lim_{k \to \infty} \alpha_k^{-1} \log \int_\Omega e^{\alpha_k[\varphi_0(X_0) + \psi_1(X_1)]} \, dR^k = \sup_\Omega \{\varphi_0(X_0) + \psi_1(X_1) - C\}.$$

Here, we took advantage of $\varphi_0(X_0) + \psi_1(X_1) \in C_b(\Omega)$ to apply the Laplace-Varadhan lemma. We see that the pointwise limit of the objective function of ($\mathrm{D}^k$) is

$$\lim_{k \to \infty} \alpha_k^{-1} \Delta_{R^k}(\alpha_k \varphi_0, \alpha_k \psi_1) = \int_\mathcal{X} \varphi_0 \, d\mu_0 + \int_\mathcal{X} \psi_1 \, d\mu_1 - \sup_\Omega \{\varphi_0(X_0) + \psi_1(X_1) - C\}$$

and the limit dual problem is (this is informal)

$$\int_\mathcal{X} \varphi_0 \, d\mu_0 + \int_\mathcal{X} \psi_1 \, d\mu_1 - \sup_\Omega \{\varphi_0(X_0) + \psi_1(X_1) - C\} \to \max; \qquad \varphi_0, \psi_1 \in C_b(\mathcal{X}).$$

But this problem is equivalent to

$$\int_\mathcal{X} \varphi_0 \, d\mu_0 + \int_\mathcal{X} \psi_1 \, d\mu_1 \to \max; \qquad \varphi_0, \psi_1 \in C_b(\mathcal{X}) : \varphi_0 \oplus \psi_1 \leq c \qquad (\mathrm{D}^\infty)$$

where

$$c(x, y) := \inf \{C(\omega); \omega \in \Omega : \omega_0 = x, \omega_1 = y\}, \qquad (54)$$

recall (4). To see this, first remark that for any $\lambda \in \mathbb{R}$, transforming $\varphi_0 \oplus \psi_1$ into $\varphi_0 \oplus \psi_1 + \lambda$ doesn't modify the value of $\int_\mathcal{X} \varphi_0 \, d\mu_0 + \int_\mathcal{X} \psi_1 \, d\mu_1 - \sup_\Omega \{\varphi_0(X_0) + \psi_1(X_1) - C\}$. Hence, when $\sup_\Omega \{\varphi_0(X_0) + \psi_1(X_1) - C\} < \infty$, we can normalize $\varphi_0 \oplus \psi_1$ so that $\sup_\Omega \{\varphi_0(X_0) + \psi_1(X_1) - C\} = 0$ and we obtain the equivalent problem

$$\int_\mathcal{X} \varphi_0 \, d\mu_0 + \int_\mathcal{X} \psi_1 \, d\mu_1 \to \max; \qquad \varphi_0, \psi_1 \in C_b(\mathcal{X}) : \varphi_0(X_0) + \psi_1(X_1) - C \leq 0.$$

But, $\varphi_0(X_0) + \psi_1(X_1) - C \leq 0$ on $\Omega$ if and only if $\varphi_0 \oplus \psi_1 \leq c$ on $\mathcal{X}^2$.

We have informally shown that $\lim_{k \to \infty} (\mathrm{D}^k) = (\mathrm{D}^\infty)$ (in some insufficiently specified sense) which is the usual Kantorovich problem, dual to (MK). Consequently,

we must have $\lim_{k \to \infty} (\mathrm{S}^k_{\mathrm{dyn}}) = (\mathrm{MK})$ in some sense. Similarly, the static analogue of $(\mathrm{S}^k_{\mathrm{dyn}})$ which is

$$H(\pi | R^k_{01})/\alpha_k \to \min; \qquad \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1 \qquad (\mathrm{S}^k)$$

converges to

$$\int_{\mathcal{X}^2} c \, d\pi \to \min; \qquad \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1 \qquad (\mathrm{MK})$$

with $c$ as above.

It happens that this convergence is in terms of $\Gamma$-convergence.

**Informal statement 2** (See [40, 33]). *Suppose that the slowed down Markov measure $R^k \in \mathrm{M}_+(\Omega)$ associated with the generator $L^k := L/k$ satisfies the large deviation principle (52) with speed $\alpha_k$ and rate function $C$ in $\Omega$. Then,*

$$\Gamma\text{-}\lim_{k \to \infty} (\mathrm{S}^k_{\mathrm{dyn}}) = (\mathrm{MK}_{\mathrm{dyn}}) \qquad and \qquad \Gamma\text{-}\lim_{k \to \infty} (\mathrm{S}^k) = (\mathrm{MK}).$$

*In particular:*

1. *In the reversible Brownian motion case (36) in $\mathcal{X} = \mathbb{R}^n$, we have:*
   $\alpha_k = k$, $C(\omega) = \int_{[0,1]} |\dot{\omega}_t|^2/2 \, dt$, $c(x,y) = |y - x|^2/2$ *and we recover (11).*
2. *In the case of a random walk on a graph (45), we have:*
   $\alpha_k = \log k$, $C(\omega) = \sum_{0 \le t \le 1} \mathbf{1}_{\{\omega_t \ne \omega_{t^-}\}}$ *and $c = d_\sim$ : the graph distance on $(\mathcal{X}, \sim)$.*

The diffusion case is treated in detail in [40]. In the specific Brownian case (1), the Schrödinger problem converges to the quadratic Monge-Kantorovich problem and, as already remarked at (10), the bridges converge as follows:

$$\lim_{k \to \infty} R^{k,xy} = \delta_{\gamma^{xy}} \in \mathrm{P}(\Omega)$$

where $\gamma^{xy}_t = (1 - t)x + ty$, $0 \le t \le 1$ is the constant speed geodesic path between $x$ and $y$. In case $(\mathrm{MK}_{\mathrm{dyn}})$ has a unique solution $\widehat{P}$, we also have $\lim_{k \to \infty} \widehat{P}^k = \widehat{P} \in \mathrm{P}(\Omega)$.

Since the rigorous version of the Informal Statement 2 is simpler to state in the second case (2) of a random walk on a graph, we refer the reader to [40] for detail about (1) and we restrict our attention to (2). In this random walk case, the rigorous version of the Informal Statement 2 is stated below at Theorem 5.1. Some preparation is needed. In particular, let us recall basic facts about $\Gamma$-convergence.

$\Gamma$-*convergence.* Recall that $\Gamma\text{-}\lim_{k \to \infty} f^k = f$ on the metric space $Y$ if and only if for any $y \in Y$,

(a) $\liminf_{k \to \infty} f^k(y_k) \ge f(y)$ for any convergent sequence $y_k \to y$,
(b) $\lim_{k \to \infty} f^k(y^o_k) = f(y)$ for some sequence $y^o_k \to y$.

A function $f$ is said to be coercive if for any $a \ge \inf f$, $\{f \le a\}$ is a compact set.

The sequence $(f^k)_{k \ge 1}$ is said to be equi-coercive if for any real $a$, there exists some compact set $K_a$ such that $\cup_k \{f^k \le a\} \subset K_a$.

If in addition to $\Gamma\text{-}\lim_{k \to \infty} f^k = f$, the sequence $(f^k)_{k \ge 1}$ is equi-coercive, then:

- $\lim_{k \to \infty} \inf f^k = \inf f$,
- if $\inf f < \infty$, any limit point $y^*$ of a sequence $(y^*_k)_{k \ge 1}$ of approximate minimizers i.e.: $f^k(y^*_k) \le \inf f^k + \epsilon_k$ with $\epsilon_k \ge 0$ and $\lim_{k \to \infty} \epsilon_k = 0$, minimizes $f$ i.e.: $f(y^*) = \inf f$.

For more detail about $\Gamma$-convergence, see [13] for instance.

The convex indicator $\iota_A$ of any subset $A$, is defined to be equal to $0$ on $A$ and to $\infty$ outside $A$. We denote for each $k \geq 2$, (we drop $k = 1$ not to divide by $\log(1)$ below),

$$I^k(P) := H(P|R^k)/\log k + \iota_{\{P:P_0=\mu_0,P_1=\mu_1\}}, \qquad P \in \mathrm{P}(\Omega),$$

so that $(\mathrm{S}^k_{\mathrm{dyn}})$ is simply: $(I^k \to \min)$. We also define

$$I(P) = E_P C + \iota_{\{P:P_0=\mu_0,P_1=\mu_1\}}, \qquad P \in \mathrm{P}(\Omega)$$

with $C$ given at (53). The dynamical Monge-Kantorovich problem $(\mathrm{MK}_{\mathrm{dyn}})$ rewrites as $(I \to \min)$.

Similarly, we denote for each $k \geq 2$,

$$J^k(\pi) := H(\pi|R^k_{01})/\log k + \iota_{\{\pi:\pi_0=\mu_0,\pi_1=\mu_1\}}, \qquad \pi \in \mathrm{P}(\mathcal{X}^2),$$

so that $(\mathrm{S}^k)$ is simply: $(J^k \to \min)$. We also define

$$J(\pi) = \int_{\mathcal{X}^2} c \, d\pi + \iota_{\{\pi:\pi_0=\mu_0,\pi_1=\mu_1\}}, \qquad \pi \in \mathrm{P}(\mathcal{X}^2)$$

with $c$ given at (54). The Monge-Kantorovich problem $(\mathrm{MK})$ rewrites as $(J \to \min)$.

The spaces $\mathrm{P}(\Omega)$ and $\mathrm{P}(\mathcal{X}^2)$ are equipped with the topologies of narrow convergence: weakened by the spaces of all numerical continuous and bounded functions. The $\Gamma$-convergences on $\mathrm{P}(\Omega)$ and $\mathrm{P}(\mathcal{X}^2)$ are related to these topologies.

**Theorem 5.1** ([33]). *Assume that the random walk $R$ and the prescribed marginal measures $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$ satisfy the hypotheses of Theorem 2.6.*

*For each $k \geq 2$, let $\widehat{P}^k \in \mathrm{P}(\Omega)$ and $\widehat{\pi}^k \in \mathrm{P}(\mathcal{X}^2)$ be the respective solutions of $(\mathrm{S}^k_{\mathrm{dyn}})$ and $(\mathrm{S}^k)$.*

1. *The sequence $(I^k)_{k \geq 2}$ is equi-coercive and $\Gamma\text{-}\lim_{k \to \infty} I^k = I$ in $\mathrm{P}(\Omega)$.*

   *In particular, $\lim_{k \to \infty} \inf(\mathrm{S}^k_{\mathrm{dyn}}) = \inf(\mathrm{MK}_{\mathrm{dyn}})$ and any limit point of $(\widehat{P}^k)_{k \geq 2}$ is a solution of $(\mathrm{MK}_{\mathrm{dyn}})$.*

   *A more careful study allows to show that there is a unique limit point $\widehat{P}$, so that $\lim_{k \to \infty} \widehat{P}^k = \widehat{P} \in \mathrm{P}(\Omega)$, and that $\widehat{P}$ is the only solution of the auxiliary entropy minimization $(\widetilde{\mathrm{S}}_{\mathrm{dyn}})$ which is stated below.*

2. *The sequence $(J^k)_{k \geq 2}$ is equi-coercive and $\Gamma\text{-}\lim_{k \to \infty} J^k = J$ in $\mathrm{P}(\mathcal{X}^2)$.*

   *In particular, $\lim_{k \to \infty} \inf(\mathrm{S}^k) = \inf(\mathrm{MK})$ and any limit point of $(\widehat{\pi}^k)_{k \geq 2}$ is a solution of $(\mathrm{MK})$ which is the Monge-Kantorovich problem associated with the metric cost $c = d_\sim$ : the usual graph distance .*

   *Furthermore, this sequence admits the unique limit point $\widehat{P}_{01}$, so that $\lim_{k \to \infty} \widehat{\pi}^k = \widehat{P}_{01} \in \mathrm{P}(\mathcal{X}^2)$.*

It is also proved in [33] that for any distinct $x, y \in \mathcal{X}$,

$$\lim_{k \to \infty} R^{k,xy} = \widetilde{R}^{xy} \in \mathrm{P}(\Omega)$$

where $\widetilde{R}^{xy}$ is the $(x, y)$-bridge of

$$\widetilde{R} := \mathbf{1}_G \exp\left(\int_{[0,1]} J_{X_t}(\mathcal{X}) \, dt\right) R \in \mathrm{M}_+(\Omega)$$

with $G$ the set of all geodesic paths on $(\mathcal{X}, d_\sim)$. Remark that the set $G^{xy}$ of all geodesic paths between any two distinct states $x$ and $y$ is infinite since the instants

of jump are not specified: only the ordered enumeration of the visited states is relevant. Let us denote $\mathcal{M}(\mu_0, \mu_1) \subset P(\mathcal{X}^2)$ the set of all solutions of the Monge-Kantorovich problem (MK) with $c = d_\sim$, and introduce the subsequent auxiliary entropic minimization problem

$$H(P|\widetilde{R}) \to \min; \qquad P_{01} \in \mathcal{M}(\mu_0, \mu_1) \qquad\qquad (\widetilde{S}_{\mathrm{dyn}})$$

The set of all solutions of ($\mathrm{MK}_{\mathrm{dyn}}$) consists of all $P \in P(\Omega)$ concentrated on $G$, i.e. $P(G) = 1$, and such that the endpoint marginal $P_{01} \in P(\mathcal{X}^2)$ solves (MK). Although ($\mathrm{MK}_{\mathrm{dyn}}$) has always infinitely many solutions (for any distinct $x, y$, $G^{xy}$ is infinite), the sequence of Schrödinger problems ($\mathrm{S}^k_{\mathrm{dyn}}$) selects a unique limit point:

$$\lim_{k \to \infty} \widehat{P}^k = \widehat{P} \in P(\Omega)$$

where $\widehat{P}$ is the unique solution of ($\widetilde{S}_{\mathrm{dyn}}$). We obtain the corresponding results about the static problems ($\mathrm{S}^k$) and (MK) by considering the push-forward mapping $P \in P(\Omega) \to P_{01} \in P(\mathcal{X}^2)$.

6. **The statistical physics motivation of Schrödinger's problem.** We consider a large number $n$ of independent (non-interacting) moving random particles in the state space $\mathcal{X}$. They are described by the independent stochastic processes $Y^1, \ldots, Y^n$ taking their random values in $\Omega$ with the laws

$$\mathrm{Law}(Y^i) = R(\cdot \mid X_0 = y_0^i) \in P(\Omega), \qquad 1 \le i \le n \qquad\qquad (55)$$

where $R \in \mathrm{M}_+(\Omega)$ is a path measure and $y_0^i \in \mathcal{X}$ is the deterministic initial position of the $i$-th particle. It is also assumed that the particles are indistinguishable. Therefore, one doesn't loose information considering the empirical probability measure

$$L^n := \frac{1}{n} \sum_{1 \le i \le n} \delta_{Y^i}$$

which is a random element of $P(\Omega)$. At each time $t$, the empirical measure of the particle system is the following random element of $P(\mathcal{X})$,

$$L^n_t := \frac{1}{n} \sum_{1 \le i \le n} \delta_{Y^i_t}, \quad 0 \le t \le 1.$$

Suppose that the initial positions are close to a profile $\mu_0 \in P(\mathcal{X})$, i.e.

$$L^n_0 = \frac{1}{n} \sum_{1 \le i \le n} \delta_{y_0^i} \underset{n \to \infty}{\to} \mu_0 \in P(\mathcal{X})$$

with respect to the narrow topology $\sigma(P(\mathcal{X}), C_b(\mathcal{X}))$. The law of large numbers tells us that, as $n$ tends to infinity, $L^n$ converges in law to the deterministic limit ${}^{\mu_0}R := \int_{\mathcal{X}} R(\cdot \mid X_0 = x)\, \mu_0(dx)$ in $P(\Omega)$ and in particular that at time $t = 1$,

$$L^n_1(dy) \underset{n \to \infty}{\to} ({}^{\mu_0}R)_1(dy) := \int_{\mathcal{X}} R(X_1 \in dy \mid X_0 = x)\, \mu_0(dx).$$

Schrödinger addressed the following problem. Suppose that at the final time $t = 1$, you observe the system in a profile $L^n_1$ far away from the expected profile $({}^{\mu_0}R)_1$ : for all large enough $n$, $L^n_1$ is in a very small neighbourhood of some $\mu_1 \in P(\mathcal{X})$ which doesn't contain $({}^{\mu_0}R)_1$. This may happen since $n$ is finite, but this is a very rare event, i.e. with an exponentially small probability, see (59) below. *Nevertheless,*

*conditionally on this rare event, what is the most likely dynamical behaviour of the whole random system described by $L^n$?*

Before stating this rigorously at Problem 6 below, take a distance d on $P(\mathcal{X})$ compatible with the narrow topology and denote $B(\mu, \epsilon) = \{\nu \in P(\mathcal{X}); d(\nu, \mu) < \epsilon\}$ the open ball centred at $\mu$ with radius $\epsilon > 0$.

**Problem** (Schrödinger's question [66]). *Let $\mu_0, \mu_1 \in P(\mathcal{X})$ be given. What is the limit*

$$\lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \text{Prob}(L^n \in \cdot | L_1^n \in B(\mu_1, \epsilon)) \tag{56}$$

*in $P(P(\Omega))$?*

6.1. **Solving the problem without getting into detail.** Schrödinger's approximate proof contains the main ideas. It is based on a statistical physics approach. The main tool for obtaining the limiting behavior as $n$ tends to infinity of the combinatoric terms is Stirling's formula. As pointed out by Föllmer in [24], its modern counterpart, which is available in a much more general setting, is Sanov's theorem.

**Informal statement 3** (Informal statement of Sanov's theorem). *Let $Y^1, \ldots, Y^n$, ... be a sequence of independent identically distributed $\Omega$-valued random variables with common law $R \in P(\Omega)$[9]. Define $L^n := \frac{1}{n} \sum_{i=1}^n \delta_{Y^i}$ its empirical measure. Then, for a "large class" of measurable subsets $A$ of $P(\Omega)$, we have*

$$\text{Prob}(L^n \in A) \underset{n \to \infty}{\asymp} \exp\left[-n \inf_{P \in A} H(P|R)\right]. \tag{57}$$

The rigorous statement of this result is in terms of a large deviation principle. It is valid for a general class of spaces $\Omega$, not necessarily a path space. For a comprehensive introduction to the theory of large deviations including Sanov's theorem, a good textbook is [16]. One says that $H(\cdot|R)$ is the rate function of the large deviations of $\{L^n\}_{n \geq 1}$ as $n$ tends to infinity.

*Idea of proof (a hint to agree with this statement).* We consider informally the situation where $\Omega$ is replaced by a three-point set. Take $\Omega = \{a, b, c\}$, $R = \alpha \delta_a + \beta \delta_b + \gamma \delta_c$ and $P = p \delta_a + q \delta_b + r \delta_c$ with $\alpha, \beta, \gamma, p, q, r > 0$ and $\alpha + \beta + \gamma = p + q + r = 1$. Then,

$\quad \text{Prob}(L^n \approx P)$

$= \text{Prob}(L^n(a) \approx p, L^n(b) \approx q, L^n(c) \approx r)$

$\approx \dfrac{n!}{(np)!(nq)!(nr)!} \alpha^{np} \beta^{nq} \gamma^{nr}$

$\approx \exp[n \log n - np \log(np) - nq \log(nq) - nr \log(nr) + np \log \alpha + nq \log \beta + nr \log \gamma]$

$= \exp[-n(p \log(p/\alpha) + q \log(q/\beta) + r \log(r/\gamma))]$

$= \exp[-nH(P|R)]$

where we used Stirling's formula: $k! \approx \exp[k \log k - k]$ as $k$ tends to infinity.    □

This hint is very much in the spirit of Schrödinger's derivation in [66].

Since $P \mapsto H(P|R) \in [0, \infty]$ is strictly convex and $H(P|R) = 0$ if and only if $P = R$, see [66], one observes that if $R \in A$, (57) leads to the law of large numbers: $\lim_{n \to \infty} L^n = R$, with an exponential rate of convergence.

We need a slight modification of Sanov's theorem.

---

[9]We take a probability measure, rather than $R \in M_+(\Omega)$, for the simplicity of exposition.

**Informal statement 4** (Informal statement of Sanov's modified theorem). *Let $Y^1, \dots, Y^n, \dots$ be the sequence of independent $\Omega$-valued random variables specified by* (55). *Define $L^n := \frac{1}{n} \sum_{i=1}^n \delta_{Y^i}$ its empirical measure. Then, for a "large class" of measurable subsets $A$ of $\mathrm{P}(\Omega)$, we have*

$$\mathrm{Prob}(L^n \in A) \underset{n\to\infty}{\asymp} \exp\left[-n \inf_{P \in A: P_0 = \mu_0} \left\{ H(P|R) - H(\mu_0|R_0) \right\} \right]. \qquad (58)$$

This statement is proved by D. Dawson and J. Gärtner in [15, Thm. 3.5] where the Schrödinger problem, when $R$ is a diffusion process, is re-discovered[10] and investigated by means of large deviations of large non-interacting particle systems. More precisely, [15, Thm. 3.5] only states a variational formula in the spirit of (70) for the rate function. An alternate proof is given in [6, Thm. 2.1] which states that the rate function is $H(P|\mu_0 R) + \iota_{\{P:P_0=\mu_0\}}$, where $\mu_0 R := \frac{d\mu_0}{dR_0}(X_0)R(\cdot|X_0)$. But it is easily seen with the additive property (72) of the relative entropy that for any $P$ such that $P_0 = \mu_0$, we have $H(P|\mu_0 R) = \int_{\mathcal{X}} H\big(P(\cdot|X_0 = x)|R(\cdot|X_0 = x)\big)\,\mu_0(dx)$ and $H(P|R) = H(\mu_0|R_0) + \int_{\mathcal{X}} H\big(P(\cdot|X_0 = x)|R(\cdot|X_0 = x)\big)\,\mu_0(dx)$, which implies that $H(P|\mu_0 R) = H(P|R) - H(\mu_0|R_0)$, the desired result.

The conditional probability in (56) has the form $\mathrm{Prob}(L^n \in A | L^n \in C_\epsilon)$ where

$$C_\epsilon = \{P \in \mathrm{P}(\Omega); P_0 = \mu_0, P_1 \in B(\mu_1, \epsilon)\}$$

and $\epsilon > 0$ is introduced to guarantee that $\mathrm{Prob}(L^n \in C_\epsilon)$ doesn't vanish. With (58) one sees that for each $\epsilon > 0$ and "all" $A \in \mathrm{P}(\Omega)$,

$$\mathrm{Prob}(L^n \in A | L^n \in C_\epsilon) \underset{n\to\infty}{\asymp} \exp\left[-n \left\{ \inf_{P \in A \cap C_\epsilon} H(P|R) - \inf_{P \in C_\epsilon} H(P|R) \right\} \right]$$

Some analytical work (formally, think of $A$ as an arbitrarily small neighbourhood of a generic $P \in \mathrm{P}(\Omega)$) allows us to show that this implies that $\lim_{n\to\infty} \mathrm{Prob}(L^n \in \cdot | L^n \in C_\epsilon) = \delta_{\widehat{P}_\epsilon}$ where $\widehat{P}_\epsilon$ is the solution of the convex minimization problem

$$H(P|R) \to \min; \qquad P \in C_\epsilon.$$

Note that this problem admits a unique solution since $H(\cdot|R)$ is a *strictly* convex function on the convex set $C_\epsilon$. Existence is obtained as usual showing that $H(\cdot|R)$ has compact sublevel sets. Finally, as $\epsilon$ decreases to zero, $C_\epsilon$ decreases to $C = \{P \in \mathrm{P}(\Omega); P_0 = \mu_0, P_1 = \mu_1\}$ and the objective functions of the minimization problems on $C_\epsilon$ : $H(P|R) + \iota_{C_\epsilon}(P)$ where $\iota_{C_\epsilon}(P) = \begin{cases} 0 & \text{if } P \in C_\epsilon \\ \infty & \text{otherwise} \end{cases}$ increase towards $H(P|R) + \iota_C(P)$. Together with some compactness, this monotonicity allows to prove easily that $\lim_{\epsilon\to 0} \widehat{P}_\epsilon = \widehat{P}$ where $\widehat{P}$ is the unique solution to the limiting minimization problem:

$$H(P|R) \to \min; \qquad P \in C.$$

Therefore, we have informally obtained the answer to Schrödinger's question.

**Informal statement 5** (The answer to Schrödinger's question). *The limit (56) is*

$$\lim_{\epsilon\downarrow 0} \lim_{n\to\infty} \mathrm{Prob}(L^n \in \cdot | L_1^n \in B(\mu_1, \epsilon)) = \delta_{\widehat{P}} \in \mathrm{P}(\mathrm{P}(\Omega))$$

*where $\widehat{P}$ is the unique solution to the entropy minimization problem*

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

---

[10]No reference to the original papers by Schrödinger is given in [15].

Loosely speaking, this means that conditionally on $L_0^n \approx \mu_0$ and $L_1^n \approx \mu_1$, the whole system $L^n$ tends in law as $n$ tends to infinity towards $\widehat{P}$. In fact, the rigorous proof of this theorem [39, Thm. 7.3] uses large deviation principles and shows that this convergence is exponentially fast. Therefore, Borel-Cantelli lemma allows us to state an almost sure version of this conditional law of large numbers.

The same line of reasoning leads to the following evaluation of the probability that the system evolves spontaneously from the prepared initial profile $\mu_0$ to the unexpected profile final profile $\mu_1$ :

$$\lim_{\epsilon\downarrow 0}\lim_{n\to\infty}\frac{1}{n}\log\mathrm{Prob}(L_1^n \in B(\mu_1,\epsilon)|L_0^n \in B(\mu_0,\epsilon)) = -\big\{\inf(\mathrm{S_{dyn}}) - H(\mu_0|m)\big\}. \tag{59}$$

These considerations show that solving Schrödinger's problem amounts to solve the convex minimization problem $(\mathrm{S_{dyn}})$ which, in statistical physics, enters the class of Boltzmann-Gibbs conditioning principles.

For a variation on this theme, with killed particles, see [14].

6.2. **The lazy gas experiment.** In his textbook [72, pp. 445-446], C. Villani writes in a section entitled *"A fluid mechanics feeling for Ricci curvature - The lazy gas experiment"*, the following sentences.

*Take a perfect gas in which particles do not interact, and ask him to move from a certain prescribed density field at time $t = 0$, to another prescribed density field at time $t = 1$. Since the gas is lazy, he will find a way to do so that needs a minimal amount of work (least action principle). Measure the entropy[11] of the gas at each time, and check that it always lies* above *the line joining the final and initial entropies. If such is the case, then we know that we live in a nonnegatively curved space.*

This is clearly Schrödinger's thought experiment. As [72] is only concerned with optimal transport, this lazy gas experiment must be understood at the level of the displacement interpolations. It refers to the important discovery by R. McCann [44, 45] that entropy along displacement interpolations enjoys convexity properties. Further developments by Cordero Erausquin, McCann, Schmuckenschläger, Otto, Villani, Sturm and von Renesse [45, 61, 8, 69] have put in light the tight connection between these convexity properties and Ricci curvature lower bounds. Indeed, Otto's heuristic calculus (see [30, 60] and [72, Ch. 15]) allows us to guess that, along the displacement interpolations $[\mu_0, \mu_1]$ with respect to quadratic optimal transport on a Riemannian manifold $\mathcal{X}$, the second derivative of the entropy as a function of time: $t \in [0, 1] \mapsto h_{[\mu_0,\mu_1]}(t) := H(\mu_t|\mathrm{vol})$ satisfies

$$h''_{[\mu_0,\mu_1]}(t) = \langle\Gamma_2(\psi_t), \mu_t\rangle, \quad 0 \le t \le 1, \tag{60}$$

where $\psi$ solves (8) and $\Gamma_2(\psi) := L\Gamma(\psi) - 2\Gamma(\psi, L\psi)$ is the iterated carré du champ, with $L = \Delta/2$. Bochner's formula, relates $\Gamma_2$ and the Ricci curvature:

$$\Gamma_2(\psi) = \|\nabla\psi\|_{\mathrm{HS}}^2 + \mathrm{Ric}(\nabla\psi).$$

This is the starting point of the Lott-Sturm-Villani theory [67, 68, 43].

Schrödinger problem suggests a slight (more realistic :-) variant of this thought experiment where displacement interpolations are replaced with entropic interpolations, see (12). This is really a lazy gas experiment, while in some sense, the above mentioned lazy gas experiment in [72] is a *very* lazy gas experiment. Indeed, in the

---

[11]Here, the entropy is standard Boltzmann's one: $p \mapsto -H(p|\mathrm{vol})$, which is a concave function.

displacement interpolation setting, not only the particles need to find a cooperative lazy behaviour (the transport mapping $x \mapsto y$) but also each individual particle must find an economic way to travel (the minimizing geodesic path) as a result of its intrinsic laziness: it is very slow and at the limit $k \to \infty$, it doesn't want to move at all, recall Statement 2.

It is interesting to know that, without slowing down, along the entropic interpolation $[\mu_0, \mu_1]$,

$$h''_{[\mu_0,\mu_1]}(t) = \frac{1}{2}\langle \Gamma_2(\varphi_t) + \Gamma_2(\psi_t), \mu_t \rangle, \quad 0 \le t \le 1, \tag{61}$$

where the functions $\varphi$ and $\psi$ are given at (33): $\varphi_t(z) = \log E_R(f_0(X_0) \mid X_t = z)$, $\psi_t(z) = \log E_R(g_1(X_1) \mid X_t = z)$ with $f_0, g_1$ such that $\mu_t = (f_0(X_0)g_1(X_1)\,R)_t$. This is proved by the author in [34]. Remark that (61) is an extension of (60) where $\psi = -\varphi$. Unlike this asymptotic case ($k = \infty$) where the convexity estimate (60) is obtained informally by means of Otto's heuristic calculus, the entropic interpolation formula (61) is rigorous since $h_{[\mu_0,\mu_1]}$ is genuinely second differentiable on $(0,1)$ and the stochastic calculus result (31) is rigorous.

In the general setting of a reference reversible measure $R$, and in particular with a reversible random walk on a graph, it is shown in [34] that stochastic calculus for the $(f,g)$-transforms as developed at Statement 1 leads us to the following rigorous formula

$$h''_{[\mu_0,\mu_1]}(t) = \langle \Theta_2(\varphi_t) + \Theta_2(\psi_t), \mu_t \rangle, \quad 0 < t < 1, \tag{62}$$

where

$$\Theta_2 \psi := L\Theta\psi + e^{-\psi}\Gamma\left(e^{\psi}, \Theta\psi\right) + e^{-\psi}\Gamma(e^{\psi}, \psi)B\psi - e^{-\psi}\Gamma(e^{\psi}B\psi, \psi) \tag{63}$$

with

$$\Theta\psi := e^{-\psi}\Gamma(e^{\psi}, \psi) - B\psi + L\psi.$$

In the special case where $L$ is a diffusion operator, then $\Theta = \Gamma/2$ and $\Theta_2 = \Gamma_2/2$. One may expect that formulae (62) and (63) could lead to some results about the curvature of graphs, in the same spirit as (61), which is related to curvature via Bochner's formula, carries information about the curvature of the underlying Riemannian manifold.

7. **A short history of Schrödinger's problem and related literature.** Schrödinger's problem was first addressed by E. Schrödinger in a German written article [65] which was published in 1931 and entitled *"Über die Umkehrung der Naturgesetze"* [12], then in a French written article [66] which was published in 1932 and entitled *"Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique"*[13]. The entropy minimization problem appears at the last section VII of the 1932 article which can be read independently of the preceding sections and is entitled: *"Une analogie entre la mécanique ondulatoire et quelques problèmes de probabilités en physique classique"*[14].

Let us quote Schrödinger's introduction to this section.

*Il s'agit d'un problème classique : problème de probabilités dans la théorie du mouvement brownien. Mais en fin de compte, il ressortira une analogie avec la mécanique ondulatoire, qui fut si frappante pour moi lorsque je l'eus trouvée, qu'il m'est difficile de la croire purement accidentelle.*

---

[12]On the reversibility of the laws of nature.

[13]On the relativistic theory of the electron and the interpretation of quantum mechanics.

[14]An analogy between wave mechanics and some probabilistic problems in classical physics.

*À titre d'introduction, je voudrais citer une remarque que j'ai trouvée dans les "Glifford lectures" de A.S. Eddington (Cambridge, 1928, p. 216 et sqq). Eddington, en parlant de l'interprétation de la mécanique ondulatoire, fait dans une note en bas de page la remarque suivante :* [15]

*"The whole interpretation is very obscure, but it seems to depend on wether you are considering the probability* after you know what has happened *or the probability for the purposes of prediction. The $\psi\bar{\psi}$ is obtained by introducing two symmetrical systems of $\psi$ waves travelling in opposite directions in time; one of these must presumably correspond to probable inference from what is known (or is stated) to have been the condition at a later time."*

In 1931, wave mechanics is newly born and many physicists are puzzled by its possible interpretations. Based on Eddington's remark, one may wonder at first sight if in the quantum world knowledge from the far future is available. Of course, this is not so, but why? In his 1931-32 papers, Schrödinger solves this paradox by providing an amazingly close analogue of the quantum wave function propagation in the classical world, by means of the entropy minimization problem ($\mathrm{S_{dyn}}$). In particular, formula (30) in Theorem 3.3: $P_t(dx) = f_t(x)g_t(x)\,m(dx)$, must be interpreted as the classical analogue of Born's formula: $P_t(dx) = \psi_t(x)\bar{\psi}_t(x)\,dx$. Let us quote [66] again (this quotation also appears in [24]) to emphasize that, although derived in a heuristic manner in [65, 66], the system (24) and Born's formula (30) are motivated by the following question of large deviations in the framework of the lazy gas experiment: *Imaginez que vous observez un système de particules en diffusion, qui soient en équilibre thermodynamique. Admettons qu'à un instant donné $t_0$ vous les ayez trouvées en répartition à peu près uniforme et qu'à $t_1 > t_0$ vous ayez trouvé un écart spontané et* considérable *par rapport à cette uniformité. On vous demande de quelle manière cet écart s'est produit. Quelle en est la manière la plus probable ?* [16]

As a concluding comment in his 1932 article, Schrödinger writes : *La fonction [d'onde] complexe $\psi$ correspond à* deux *fonctions réelles, de sorte qu'il suffit de définir les conditions aux limites en se donnant la valeur de $\psi$ à* un seul *instant déterminé ; c'est la façon de voir généralement admise en mécanique quantique. Est-elle la seule admissible ? Dans notre problème, cela reviendrait à regarder comme données les valeurs de $f$ et $g$*[17] *à un instant déterminé (au lieu des valeurs de leur produit à deux instants différents), chose inadmissible et absolument dénuée de sens.*

*Doit-on interpréter la remarque d'Eddington, citée plus haut, comme signalant la nécessité de modifier cette manière de voir en mécanique ondulatoire et prendre comme conditions aux limites les valeurs d'une seule probabilité réelle à deux instants différents ?* [18]

---

[15]This is a classical problem: a probability problem in the theory of Brownian motion. But eventually an analogy with the wave mechanics will appear. This analogy stroke me so hard once I discovered it, that it is difficult for me to believe that it is purely accidental.

As an introduction, let me quote a remark that I found in the "Glifford lectures" of A. S. Eddington (Cambridge, 1928, p. 216 et sqq). Discussing the interpretation of wave mechanics, Eddington writes in a footnote the following remark: "The whole interpretation is very obscure, ..."

[16]Imagine that you observe a system of diffusing particles which is in thermal equilibrium. Suppose that at a given time $t_0$ you see that their repartition is almost uniform and that at $t_1 > t_0$ you find a spontaneous and *significant* deviation from this uniformity. You are asked to explain how this deviation occurred. What is its most likely behaviour?

[17]With the notation of the present article.

[18] The complex [wave] function $\psi$ corresponds to *two* real functions. Therefore, it is enough to define the limit conditions by prescribing the value of $\psi$ at a *unique* given time. This is the

This has been performed in 1942 by R. Feynman in his PhD thesis [22], without knowing Schrödinger's contribution. Feynman's thesis is entitled: *The principle of least action in quantum mechanics*. Based on a seminal article by Dirac [17], entitled *The Lagrangian in quantum mechanics* (also reproduced in [22]), and in contrast with the regular Hamiltonian approach, Feynman's thesis proposes a Lagrangian approach to quantum mechanics which will be further developed in several directions, see [23].

*Föllmer's contribution.* Although Schrödinger obtains the classical Born formula (30), he does not write explicitly the problems ($S_{dyn}$) and (S). Their explicit formulation is due to H. Föllmer in his Saint-Flour lecture notes [24, pp. 154-167]. Proposition 1 which is based on the additive property of the relative entropy (72), also appears in [24].

7.1. **Early mathematical developments.** Although this part of Schrödinger's work has been forgotten for some decades, it had influenced leading mathematicians soon after its publishing.

*Reciprocal processes, 1932.* Very soon after Schrödinger's 1931 article, S. Bernstein published in 1932 an article [2] about the general problem of deriving limit theorems for sequences of dependent random variables. Among other notions, he explored the Markov property and, motivated by [65], proposed a type of time-correlation which is less restrictive than the Markov property and is still symmetric with respect to time reversal[19]. He suggested that such stochastic processes could be called reciprocal process. While a Markov measure $Q \in M_+(\Omega)$ satisfies for any $0 \le s \le t \le 1$,

$$Q(X_{[0,s]} \in \cdot, X_{[t,1]} \in \cdot \cdot \mid X_{[s,t]}) = Q(X_{[0,s]} \in \cdot \mid X_s)Q(X_{[t,1]} \in \cdot \cdot \mid X_t),$$

a path measure $Q \in M_+(\Omega)$ is reciprocal if for any $0 \le s \le t \le 1$,

$$Q(X_{[s,t]} \in \cdot \mid X_{[0,s]}, X_{[t,1]}) = Q(X_{[s,t]} \in \cdot \mid X_s, X_t).$$

Any Markov measure is reciprocal, but the converse is false.

The theory of reciprocal processes has been forgotten for a while after Bernstein's article and was eventually developed by B. Jamison in 1974, [28, 29]. A significant contribution of Jamison to the theory of Schrödinger problem was that its solution $\widehat{P}$ is not only reciprocal, but also Markov and it is indeed an $h$-transform of the Markov reference process. This is performed without any entropy, but solely by means of reciprocal transitions. Föllmer recovered these results in [24] using the entropy minimization problem ($S_{dyn}$). For more information about the relations between reciprocal and Markov measures, see [42].

*Time-reversal, 1936.* In the very first lines of his celebrated paper [32] about Markov processes and time-reversal, A. Kolmogorov quotes Schrödinger's 1931 paper as a motivation. This has been surprisingly forgotten afterwards.

---

regular practice in quantum mechanics. Is it the only admissible one? In our problem, this would correspond to considering that the values of $f$ and $g$ [with the notation of the present article] are prescribed at a given time (instead of the values of their product at two distinct times). This is inadmissible and meaningless.

Should one interpret the previously quoted remark of Eddington, as a hint for the necessity of modifying our usual way of looking at quantum mechanics by defining the limit conditions in terms of the values of a single real probability at two distinct times?

[19] It is not clear that Bernstein was aware of the time-symmetry of the Markov property. This symmetry has clearly been identified twenty years later by J.L. Doob in his textbook [18].

*Schrödinger system, 1940.* Schrödinger had left open the problem of finding criteria for the system (24) to have a solution $(f_0, g_1)$. In 1940, R. Fortet [26] proposed a partial solution and in 1960, A. Beurling [3] gave a solution close to the statement of Theorem 2.4. Beurling's proof also relies upon an entropy argument. Beurling's result was improved by Jamison in [29] who obtained the complete solution of Schrödinger's system.

7.2. **Stochastic deformations of mechanics.** The aim of Euclidean quantum mechanics (EQM), which is mainly developed by J.-C. Zambrini since 1986 [73, 10, 9, 7], is to transfer by analogy, known results from quantum mechanics to the theory of stochastic processes and the other way round[20]. The starting paper [73] of this program relies on Schrödinger's discovery and adapts Jamison's results for an appropriate class of reciprocal processes (unlike Zambrini, Jamison doesn't use the time-reversed filtration in his construction of reciprocal processes). Then the EQM program was extended to the derivation of rigorous results about various kinds of stochastic processes which are suggested by the textbook *Quantum mechanics and path integrals* by Feynman and Hibbs [23]. This textbook presents, indeed, a time-symmetric (Lagrangian) approach to quantum mechanics which extends Feynman's early works and in particular his PhD thesis [22].

Feynman's approach is an enlightening, efficient and intuitive guideline for physicists, but unfortunately it is impossible to put it on a rigorous mathematical ground: it is proved that Feynman's integral is an oddly defined object. However, replacing Feynman's integration by stochastic calculus suggests interesting results about diffusion processes. The first of these results was the celebrated Feynman-Kac's formula [31]. EQM viewpoint, however, is that there is much more in Feynman's method than this time-asymmetric measure theoretic perturbative formula. EQM uses Kac's strategy in a systematic manner and its basic program is to obtain rigorous stochastic analogues of several intuitive statements from [23]; intuitive, but highly efficient since they are corroborated by experiments. In EQM, the natural stochastic processes to work with are reciprocal processes. However, in several important situations, it appears that the critical (solving some variational problem) reciprocal processes are Markov. In this case, it is sufficient to work with $(f, g)$-transforms of Markov reference processes (see [29, 24] for an $h$-transform representation) and their extensions: $f_0(X_0) \exp\left(\int_0^1 U(X_t)\, dt\right) g_1(X_1)\, R$ with the additional Feynman-Kac integral term $(x, y) \mapsto \exp\left(\int_0^1 U(X_t)\, dt\right) R^{xy}$ which is the classical analogue of Feynman's propagator. These extensions of $h$-transforms are also used by M. Nagasawa in [56, 57] who also explores connections between stochastic processes and quantum physics which are highly inspired by the Schrödinger problem.

It is also possible to stochastically deform all the mathematical tools of classical mechanics to derive new results about diffusion processes. For instance, M. Thieullen designed a second order calculus for reciprocal processes in [70] and without referring to $(S_{dyn})$ or entropy in general, M. Thieullen and J.-C. Zambrini have obtained a stochastic deformation of Noether theorem [71].

*An interesting problem.* This suggests that it would be also interesting to derive a type of Noether theorem for the Monge-Kantorovich dynamical problem. Let us

---

[20]Unlike Nelson's stochastic mechanics [59] or Nagasawa's interpretation of quantum mechanics [57], EQM is not aimed at giving a stochastic interpretation of quantum mechanics. Such a project still remains an open problem eighty years after the advent of this theory.

give some hint of what is meant. In the Euclidean case, the displacement interpolation $[\mu_0, \mu_1]$ is a solution of ($\text{MK}_{\text{dyn}}$) with $C = C_{\text{kin}} := \int_{[0,1]} |\dot{X}_t|^2/2 \, dt$, the kinetic action. It has a constant speed; this means that twice the average kinetic energy $\int_\Omega |\dot{X}_t|^2 \, d\widehat{P} = \int_\mathcal{X} |\nabla \psi_t(x)|^2 \, \mu_t(dx)$, with the notation of the Benamou-Brenier formula (7), doesn't depend on time $t$. What happens when considering, instead of $C_{\text{kin}}$, the action functional $C = \int_{[0,1]} (|\dot{X}_t|^2/2 + U(X_t)) \, dt$ which should be connected with some Newton equation? What are the quantities that are conserved along the minimizer $[\mu_0, \mu_1]$, in terms of the symmetries of the potential $U$?

7.3. **Stochastic optimal control.** Optimal transport can be deformed into a stochastic optimal control problem. This is mainly the contribution of T. Mikami, see [52] for an overview of this approach and some of its main developments. With (42), one obtains that the Brownian Schrödinger problem ($\text{S}_{\text{dyn}}$), i.e. taking $R$ to be the reversible Brownian motion, is also expressed as follows:

$$E_{P^u} \int_0^1 L(u_t) \, dt \to \min; \quad u \in \mathcal{A} : P_0^u = \mu_0, P_1^u = \mu_1 \qquad (64)$$

where $\mathcal{A}$ is the set (of admissible controls) which consists of all the $\mathbb{R}^n$-valued progressively measurable processes $u$ and $P^u$ (if it exists) is the law of the semi-martingale

$$X_t^u = X_0^u + \int_0^t u_s \, ds + W_t, \quad 0 \le t \le 1$$

where $W$ is a standard Brownian motion starting from 0 and

$$L(u) = |u|^2/2, \quad u \in \mathbb{R}^n.$$

This is a stochastic version of the quadratic Monge-Kantorovich problem ($\text{MK}_{\text{dyn}}$):

$$E_{P^u} \int_0^1 L(u_t) \, dt \to \min; \quad u \in \mathcal{A}_{\text{MK}} : P_0^u = \mu_0, P_1^u = \mu_1$$

which is obtained by replacing $\mathcal{A}$ with $\mathcal{A}_{\text{MK}}$, the set of all controls $u \in L^1_{\mathbb{R}^n}([0,1])$ and taking $P^u$ to be the law of

$$X_t^u = X_0^u + \int_0^t u_s \, ds, \quad 0 \le t \le 1,$$

a process with a random initial position and a deterministic evolution.

This theory extends naturally to the case where $L$ is strictly convex, regular and coercive enough: $\lim_{|u| \to \infty} L(u)/|u|^p = \infty$, for some $p > 1$. But results close to optimal transport are obtained with $L$ admitting a quadratic growth, i.e. in harmony with the Brownian motion $W$.

When $L$ is quadratic, if the Brownian motion $W$ is replaced with $\sqrt{\epsilon}W$ and $\epsilon$ tends to zero, then Mikami shows in [49] that ($\text{S}_{\text{dyn}}$) tends to ($\text{MK}_{\text{dyn}}$). Unfortunately, this type of convergence remains unclear unless $L$ is quadratic, i.e. unless the stochastic optimal control problem corresponds to ($\text{S}_{\text{dyn}}$).

T. Mikami and M. Thieullen have proved a Kantorovich-type dual equality in [54] for (64) and recovered related optimal transport results in [55]. T. Mikami has intensively studied the connections between stochastic control and optimal transport. In particular, soon after the discovery by Jordan, Kinderlehrer and Otto [30] of the relation between gradient flows, Wasserstein distance and dissipative evolution equations, he proposed in [47] a stochastic approach to the JKO approximation scheme. In addition to the already cited articles by Mikami, several other

works by the same author are related to a probabilistic approach to optimal transport: [48, 50, 51, 53]. Let us also quote the early contributions of Mikami [46] and P. Dai Pra [12] where the Schrödinger problem is translated in terms of stochastic control.

7.4. **Penalized Monge-Kantorovich problem.** The connection between the Monge-Kantorovich and the Schrödinger problems is also exploited implicitly in some works where (MK) is penalized by a relative entropy, leading to the minimization problem

$$\int_{\mathcal{X}^2} c \, d\pi + \frac{1}{k} H(\pi|\rho) \to \min; \quad \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1$$

where $\rho \in \mathrm{P}(\mathcal{X}^2)$ is a fixed reference probability measure on $\mathcal{X}^2$, for instance $\rho = \mu_0 \otimes \mu_1$. Putting $\rho^k(dxdy) = Z_k^{-1} e^{-kc(x,y)} \rho(dxdy)$ with $Z_k = \int_{\mathcal{X}^2} e^{-kc} \, d\rho < \infty$, up to the additive constant $\log(Z_k)/k$, this minimization problem rewrites as $(\mathrm{S}^k)$ with $\rho^k$ instead of $R_{01}^k$. See for instance the papers by Rüschendorf and Thomsen [63, 64] and the references therein. Also interesting are the papers by Dupuy, Galichon and Salanie [27, 21] with an applied point of view.

**Appendix** A. **Relative entropy with respect to an unbounded measure.** This appendix section is a short version of [35, § 2] which we refer to for more detail. Let $r$ be some $\sigma$-finite positive measure on some space $Y$. The relative entropy of the probability measure $p$ with respect to $r$ is loosely defined by

$$H(p|r) := \int_Y \log(dp/dr) \, dp \in (-\infty, \infty], \qquad p \in \mathrm{P}(Y) \tag{65}$$

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 (65) is meaningful. It follows from the strict convexity of $h$ that $H(\cdot|r)$ is also strictly convex. In addition, since $h(a) = \inf h = 0 \iff a = 1$, we also have for any $p \in \mathrm{P}(Y)$,

$$H(p|r) = \inf H(\cdot|r) = 0 \iff p = r. \tag{66}$$

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 $\sigma$-finite, there exist measurable functions $W : Y \to [1, \infty)$ such that

$$z_W := \int_Y e^{-W} \, dr < \infty. \tag{67}$$

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

$$H(p|r) := H(p|r_W) - \int_Y W \, dp - \log z_W \in (-\infty, \infty] \tag{68}$$

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 (67). It can be proved that

$$
H(p|r) \quad \overset{\text{(i)}}{=} \quad \sup\left\{ \int_Y u\,dp - \log\int_Y e^u\,dr; u : Y \to [-\infty, \infty), \int_Y e^u\,dr < \infty \right\} \tag{69}
$$

$$
\overset{\text{(ii)}}{=} \quad \sup\left\{ \int_Y u\,dp - \log\int_Y e^u\,dr; u \in C_W(Y) \right\}, \tag{70}
$$

where

(i) identity (i) is valid when $p$ is assumed to be absolutely continuous with respect to $r$;

(ii) identity (ii) is meaningful when $Y$ is a topological space equipped with its Borel $\sigma$-field since we have set $C_W(Y)$ to be the space of all continuous functions $u : Y \to \mathbb{R}$ such that $\sup |u|/W < \infty$, where $W$ is any nonnegative function satisfying (67). In this case, it follows that, being the supremum of affine continuous functions, $H(\cdot|r)$ is a convex lower semi-continuous function with respect to the weak topology $\sigma(\{p \in \mathrm{P}(Y); \int_Y W\,dp < \infty\}, C_W(Y))$.

Clearly, identity (i) entails that $H(p|r) = \infty$ whenever $p \in \mathrm{P}(Y)$ is such that $\int_Y W\,dp = \infty$

It follows from the strict convexity of $H(\cdot|r_W)$ and (68) that $H(\cdot|r)$ is also strictly convex.

Let $Y$ and $Z$ be two Polish spaces equipped with their Borel $\sigma$-fields. For any measurable function $\phi : Y \to Z$ and any measure $q \in \mathrm{M}_+(Y)$ we have the disintegration formula

$$
q(dy) = \int_Z q(dy|\phi = z)\,\phi_\# q(dz) \tag{71}
$$

where $z \in Z \mapsto q(\cdot|\phi = z) \in \mathrm{P}(Y)$ is measurable, and the following additive property

$$
H(p|r) = H(\phi_\# p|\phi_\# r) + \int_Z H\Big(p(\cdot \mid \phi = z)|r(\cdot \mid \phi = z)\Big)\,\phi_\# p(dz), \tag{72}
$$

is valid for any $p \in \mathrm{P}(Y)$ and any $\sigma$-finite $r \in \mathrm{M}_+(Y)$. In particular, as $r(\cdot \mid \phi = z)$ is a probability measure for each $z$, with (66) we see that

$$
H(\phi_\# p|\phi_\# r) \leq H(p|r), \quad \forall p \in \mathrm{P}(Y) \tag{73}
$$

with equality if and only if

$$
p(\cdot \mid \phi = z) = r(\cdot \mid \phi = z), \quad \forall z, \phi_\# p\text{-a.s.} \tag{74}
$$

## REFERENCES

[1] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., **84** (2000), 375–393.

[2] S. Bernstein, *Sur les liaisons entre les grandeurs aléatoires*, Verhand. Internat. Math. Kongr. Zürich, (1932), no. Band I.

[3] A. Beurling, *An automormhism of product measures*, Ann. Math., **72** (1960), 189–200.

[4] P. Billingsley, "Convergence of Probability Measures," John Wiley & Sons, Inc., New York-London-Sydney 1968 xii+253 pp.

[5] J. M. Borwein and A. S. Lewis, *Decomposition of multivariate functions*, Can. J. Math., **44** (1992), 463–482.

[6] P. Cattiaux and C. Léonard, *Large deviations and Nelson's processes*, Forum Math., **7** (1995), 95–115.

[7] K. L. Chung and J.-C. Zambrini, "Introduction to Random Time and Quantum Randomness," World Scientific, 2008.

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

[9] A. B. Cruzeiro, L. Wu and J.-C. Zambrini, *Bernstein processes associated with a Markov process*, Stochastic analysis and mathematical physics, ANESTOC'98. Proceedings of the Third International Workshop (Boston) (R. Rebolledo, ed.), Trends in Mathematics, Birkhäuser, 2000, pp. 41–71.

[10] A. B. Cruzeiro and J.-C. Zambrini, *Malliavin calculus and Euclidean quantum mechanics, I*, Functional calculus. J. Funct. Anal., **96** (1991), 62–95.

[11] I. Csiszár, *I-divergence geometry of probability distributions and minimization problems*, Annals of Probability, **3** (1975), 146–158.

[12] P. Dai Pra, *A stochastic control approach to reciprocal diffusion processes*, Appl. Math. Optim., **23** (1991), 313–329.

[13] G. Dal Maso, "An Introduction to $\gamma$-convergence," Progress in Nonlinear Differential Equations and Their Applications 8, Birkhäuser Boston, Inc., Boston, MA, 1993.

[14] D. Dawson, L. Gorostiza and A. Wakolbinger, *Schrödinger processes and large deviations*, J. Math. Phys., **31** (1990), 2385–2388.

[15] D. A. Dawson and J. Gärtner, *Large deviations from the McKean-Vlasov limit for weakly interacting diffusions*, Stochastics, **20** (1987), 247–308.

[16] A. Dembo and O. Zeitouni, "Large Deviations Techniques and Applications," Second edition, Applications of Mathematics 38, Springer Verlag, New York, 1998.

[17] P. A. M. Dirac, *The lagrangian in quantum mechanics*, Phys. Zeitsch. der Sowietunion, **3** (1933), 64–72.

[18] J. L. Doob, "Stochastic Processes," John Wiley & Sons, Inc., New York; Chapman & Hall, Limited, London, 1953. viii+654 pp.

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

[20] J. L. Doob, "Classical Potential Theory and Its Probabilistic Counterpart," 2nd ed., Classics in Mathematics, Springer, 2000, (reprint of the 1984 first edition).

[21] A. Dupuy and A. Galichon, *Personality traits and the marriage market*, Preprint. Available at SSRN: http://ssrn.com/abstract=2162812.

[22] R. Feynman, "Feynman's Thesis. A New Approach to Quantum Theory," World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005. xxii+121 pp.

[23] R. Feynman and A. Hibbs, "Quantum Mechanics and Path Integrals," McGraw-Hill, 1965.

[24] H. Föllmer, "Random Fields and Diffusion Processes, in école D'été De Probabilités De Saint-Flour xv-xvii-1985-87," Lecture Notes in Mathematics, vol. 1362, Springer, Berlin, 1988.

[25] H. Föllmer and N. Gantert, *Entropy minimization and Schrödinger processes in infinite dimensions*, Ann. Probab., **25** (1997), 901–926.

[26] R. Fortet, *Résolution d'un système d'équations de M. Schrödinger*, J. Math. Pures Appl. **9** (1940), 83.

[27] A. Galichon and B. Salanie, *Matching with trade-offs: Revealed preferences over competing characteristics*, Preprint. http://hal.archives-ouvertes.fr/hal-00473173/en/.

[28] B. Jamison, *Reciprocal processes*, Z. Wahrsch. verw. Geb., **30** (1974), 65–86.

[29] B. Jamison, *The Markov processes of Schrödinger*, Z. Wahrsch. verw. Geb., **32** (1975), 323–331.

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

[31] M. Kac, *On distributions of certain Wiener functionals*, Trans. Amer. Math. Soc., **65** (1949), 1–13.

[32] A. Kolmogorov, *Zur Theorie der Markoffschen Ketten*, Mathematische Annalen, **112** (1936).

[33] C. Léonard, *Lazy random walks and optimal transport on graphs*, Preprint.

[34] C. Léonard, *On the convexity of the entropy along entropic interpolations*, Preprint.

[35] C. Léonard, *Some properties of path measures*, Preprint.

[36] C. Léonard, *Stochastic derivatives and h-transforms of Markov processes*, Preprint.

[37] C. Léonard, *Minimization of energy functionals applied to some inverse problems*, J. Appl. Math. Optim., **44** (2001), 273–297.

[38] C. Léonard, *Minimizers of energy functionals*, Acta Math. Hungar., **93** (2001), 281–325.

[39] C. Léonard, *Entropic projections and dominating points*, ESAIM Probab. Stat., **14** (2010), 343–381.

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

[41] C. Léonard, *Girsanov theory under a finite entropy condition*, Séminaire de probabilités XLIV., Lecture Notes in Mathematics 2046. Springer, 2012, pp. 429–465.

[42] C. Léonard, S. Rœlly and J.-C. Zambrini, *On the time symmetry of some stochastic processes*, Preprint.

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

[44] R. McCann, "A Convexity Theory for Interacting Gases and Equilibrium Crystals," PhD thesis, Princeton Univ., 1994.

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

[46] T. Mikami, *Variational processes from the weak forward equation*, Comm. Math. Phys., **135** (1990), 19–40.

[47] T. Mikami, *Dynamical systems in the variational formulation of the Fokker-Planck equation by the Wasserstein metric*, Appl. Math. Optim., **42** (2000), 203–227.

[48] T. Mikami, *Optimal control for absolutely continuous stochastic processes and the mass transportation problem*, Electron. Comm. Probab., **7** (2002), 199–213.

[49] T. Mikami, *Monge's problem with a quadratic cost by the zero-noise limit of h-path processes*, Probab. Theory Relat. Fields, **129** (2004), 245–260.

[50] T. Mikami, *A simple proof of duality theorem for Monge-Kantorovich problem*, Kodai Math. J., **29** (2006), 1–4.

[51] T. Mikami, *Marginal problem for semimartingales via duality*, International Conference for the 25th Anniversary of Viscosity Solutions, Gakuto International Series. Mathematical Sciences and Applications, vol. 30, 2008, pp. 133–152.

[52] T. Mikami, *Optimal transportation problem as stochastic mechanics*, Selected papers on probability and statistics, 75–94, Amer. Math. Soc. Transl. Ser. 2, 227, Amer. Math. Soc., Providence, RI, 2009.

[53] T. Mikami, *A characterization of the Knothe-Rosenblatt processes by a convergence result*, SIAM J. Control and Optim., **50** (2012), 1903–1920.

[54] T. Mikami and M. Thieullen, *Duality theorem for the stochastic optimal control problem*, Stoch. Proc. Appl., **116** (2006), 1815–1835.

[55] T. Mikami and M. Thieullen, *Optimal transportation problem by stochastic optimal control*, SIAM J. Control Optim., **47** (2008), 1127–1139.

[56] M. Nagasawa, *Transformations of diffusion and Schrödinger processes*, Probab. Theory Related Fields, **82** (1989), 109–136.

[57] M. Nagasawa, "Stochastic Processes in Quantum Physics," Monographs in Mathematics, vol. 94, Birkhäuser Verlag, Basel, 2000.

[58] E. Nelson, "Dynamical Theories of Brownian Motion," Princeton University Press, Princeton, N.J. 1967 iii+142 pp.

[59] E. Nelson, "Quantum Fluctuations," Princeton Series in Physics, Princeton University Press, Princeton, NJ, 1985.

[60] F. Otto, *The geometry of dissipative evolution equations: The porous medium equation.*, Comm. Partial Differential Equations, **26** (2001), 101–174.

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

[62] D. Revuz and M. Yor, "Continuous Martingales and Brownian Motion," 3rd ed., Grundlehren der Mathematischen Wissenschaften, [Fundamental Principles of Mathematical Sciences], 293. Springer-Verlag, Berlin, 1999..

[63] L. Rüschendorf and W. Thomsen, *Note on the Schrödinger equation and I-projections*, Statist. Probab. Lett., **17** (1993), 369–375.

[64] L. Rüschendorf and W. Thomsen, *Closedness of sum spaces and the generalized Schrödinger problem*, Theory Probab. Appl., **42** (1998), 483–494.

[65] E. Schrödinger, *Über die umkehrung der naturgesetze*, Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math. **144** (1931), 144–153.

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

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

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

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

[70] M. Thieullen, *Second order stochastic differential equations and non Gaussian reciprocal diffusions*, Probab. Theory Related Fields, **97** (1993), 231–257.

[71] M. Thieullen and J.-C. Zambrini, *Symmetries in the stochastic calculus of variations*, Probab. Theory Related Fields, **107** (1997), 401–427.

[72] C. Villani, "Optimal Transport. Old and New," Grundlehren der mathematischen Wissenschaften, [Fundamental Principles of Mathematical Sciences], 338. Springer-Verlag, Berlin, 2009.

[73] J.-C. Zambrini, *Variational processes and stochastic versions of mechanics*, J. Math. Phys., **27** (1986), 2307–2330.

Received November 2012; revised March 2013.

*E-mail address*: christian.leonard@u-paris10.fr