# Optimal transport: introduction, applications and derivation
# Lecture 4: The Schrödinger problem

## Contents

The goal of this lecture is to introduce the Schrödinger problem of statistical mechanics, and to show its links with the optimal transport problem. Just like for the optimal transport problem, this problem has three version: a dynamical one, that is, in term of processes, a Benamou-Brenier like, that is, in terms of curves in the space of probabilities, and a static one, in terms of transport plans. We will start with the dynamical one, and show how it translates into its other versions. Unfortunately, due to the lack of time, we will only be sketchy concerning some results of probability theory. We will give as much as possible references to books or articles that provide the necessary information to prove everything rigorously.

On the other hand, we will prove rigorously the convergence towards optimal transport at the Benamou-Brenier level.

If time allows, we will also show the interest of working with the Schrödinger problem instead of the optimal transport problem: the Sinkhorn algorithm provides an efficient algorithm for computing solutions.

# 1 Introduction

The Schrödinger problem falls into the class of large deviation problems for a large population of random particles. A priori, we expect the law of large number to govern the evolution of such a system. But under the very rare event when it is not the case, what can we say?

## 1.1 Large deviations of the empirical process and the relative entropy functional

Let $\mathcal{X}$ be a separable and complete metric space, and $R \in \mathcal{P}(\mathcal{X})$ be a Borel probability measure on $\mathcal{X}$. We are interested in the sequence $(\mu_n)_{n\in\mathbb{N}^*}$ of random probability measures on $\mathcal{X}$ built as follows. Let $X_1, X_2, \dots$ be a sequence of i.i.d. random variables of law $R$, defined on a probability space $(\Omega, \mathbb{P})$. For $n \in \mathbb{N}^*$, we call

$$\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i} \in \mathcal{P}(\mathcal{X}).$$

The law of large numbers gives a convergence result of $(\mu_n)$ converges towards $R$.

**Proposition 1.** *The family $(\mu_n)$ converges almost surely towards $R$ in the narrow topology.*

*Some elements of the proof.* For the first convergence, we want to show that almost surely, for all $\varphi \in C_b(\mathcal{X})$,

$$\lim_{n\to+\infty} \int \varphi \,\mathrm{d}\mu_n = \int \varphi \,\mathrm{d}R.$$

But for all $n \in \mathbb{N}$ and $\varphi \in C_b(\mathcal{X})$,

$$\int \varphi \,\mathrm{d}\mu_n = \frac{1}{n}\Big(\varphi(X_1) + \cdots + \varphi(X_n)\Big).$$

But $\varphi(X_1), \varphi(X_2), \dots$ is a sequence of i.i.d. bounded (therefore integrable) random variables, of mean value $\mathbb{E}[\varphi(X_1)] = \int \varphi \,\mathrm{d}R$. So by the strong law of large numbers, its empirical average

$$\frac{1}{n}\Big(\varphi(X_1) + \cdots + \varphi(X_n)\Big)$$

converges towards its mean value $\int \varphi \,\mathrm{d}R$. Hence, the result consist in swapping "for all $\varphi$" and "almost surely". This is a density argument that I do not want to write. □

In this context, the large deviation questions are interested in answering the following type of questions:

- If $\mathcal{E}$ is a family of test functions and for all $\varphi \in \mathcal{E}$, a number $c_\varphi$ is given, what is the probability to observe for a given large $n$ and for all $\varphi \in \mathcal{E}$

$$\int \varphi \,\mathrm{d}\mu_n \approx c_\varphi.$$

- If we consider $P \in \mathcal{P}(\mathcal{X})$ different from $R$, what is the probability to have $\mu_n \approx P$.

The answer involves the *relative entropy functional* defined as follows.

**Definition 2.** The relative entropy of $P \in \mathcal{P}(\mathcal{X})$ w.r.t. $R \in \mathcal{P}(\mathcal{X})$ is given by

$$H(P|R) := \begin{cases} \displaystyle\int \frac{\mathrm{d}P}{\mathrm{d}R} \log \frac{\mathrm{d}P}{\mathrm{d}R}\, \mathrm{d}R = \int \log \frac{\mathrm{d}P}{\mathrm{d}R}\, \mathrm{d}P & \text{if } P \ll R, \\ +\infty & \text{otherwise.} \end{cases}$$

*Remark* 3. In this context when $R$ is a probability measure, the problems that we had in the previous lecture to define the entropy does not exist: it is always well defined with values in $[0, +\infty]$. Indeed, the negative part of $\frac{\mathrm{d}P}{\mathrm{d}R} \log \frac{\mathrm{d}P}{\mathrm{d}R}$ is below bounded, and therefore $R$ is integrable. In addition, calling $h : x \in \mathbb{R}_+ \mapsto x \log x + 1 - x$, $h$ is nonnegative and we have

$$H(P|R) = \int h\left(\frac{\mathrm{d}P}{\mathrm{d}R}\right) \mathrm{d}R \geq 0.$$

The relative entropy w.r.t. $R$ has good properties for minimization problems.

**Proposition 4.** *Given $R \in \mathcal{P}(\mathcal{X})$, $H(\cdot|R)$ is strictly convex and has compact sublevels in $\mathcal{P}(\mathcal{X})$ endowed with the topology of narrow convergence.*

*Proof.* The strict convexity is a direct consequence of the strict convexity of $s \mapsto s \log s$ and of the linearity of the Radon-Nikodym derivative.

For the lower semi-continuity, remark the important formula:

$$H(P|R) = \sup_{\varphi \in L_b(\mathcal{X})} \int \varphi \, \mathrm{d}P - \int \exp(\varphi - 1)\, \mathrm{d}R = \sup_{\varphi \in C_b(\mathcal{X})} \int \varphi \, \mathrm{d}P - \int \exp(\varphi - 1)\, \mathrm{d}R,$$

where $L_b(\mathcal{X})$ stands for the set of measurable and bounded measures on $\mathcal{X}$. To see that, remark that for the first equality, the "$\geq$" part comes from the convex inequality $uz \leq z \log z + \exp(u-1)$ holding for all $u \in \mathbb{R}$ and $z \in \mathbb{R}_+$. The "$\geq$" part of the second inequality is obvious. And the fact that the r.h.s. is larger than the l.h.s. follows from approximating $1 + \log \frac{\mathrm{d}P}{\mathrm{d}R}$ by continuous and bounded functions. The lower semicontinuity follows.

To show that the sublevels of the entropy are precompact for the topology of narrow convergence, we use the Prokhorov theorem. Given $M > 0$, let us show that $\{P \in \mathcal{P}(\mathcal{X}) \text{ such that } H(P|R) \leq M\}$ is tight. Let us take $\varepsilon > 0$ and $P$ with $H(P|R) \leq M$. Considering $\varphi := 1 + \lambda 1_A$ in the previous formula, for some measurable $A \subset \mathcal{X}$ and $\lambda \in \mathbb{R}$,

$$1 + \lambda P(A) \leq H(P|R) + \exp(\lambda) R(A) + R({}^c A) \leq M + \exp(\lambda) R(A) + 1.$$

Therefore, symplifying the 1, dividing by $\lambda$ and choosing $\lambda := -\log R(A)$, we find

$$P(A) \leq \frac{1+M}{-\log R(A)}.$$

Now, consider $A = {}^c K$ where $K \subset \mathcal{X}$ is a compact set that is so big that the r.h.s. of the previous inequality (which is independent of $P$) is below $\varepsilon$ we find that $P(A) \leq \varepsilon$, and hence the result. □

To answer the question that we stated, the main theorem is the following. Its proof can be found in [2].

**Theorem 5.** *The family of laws of $(\mu_n)$ satisfies a large deviation principle for the topology of narrow convergence, of good rate function $H(\cdot|R)$.*

*In other words, for all closed set $F \in \mathcal{P}(\mathcal{X})$,*

$$\limsup_{n \to +\infty} \frac{1}{n} \log(\mathbb{P}(\mu_n \in F)) \leq - \inf_{P \in F} H(P|R),$$

*and for all open set $U \in \mathcal{P}(\mathcal{X})$,*

$$\liminf_{n\to+\infty} \frac{1}{n}\log(\mathbb{P}(\mu_n \in U)) \geq -\inf_{P\in U} H(P|R).$$

*Remark* 6. This theorem has the following consequence: if the set $A \subset \mathcal{X}$ is such that $\inf_{\overline{A}} H(\cdot|R) = \inf_{\mathring{A}} H(\cdot|R)$ (which is generally true for a great variety of sets), then

$$\mathbb{P}(\mu_n \in A) = \exp\left(-n\inf_{P\in A} H(P|R) + \underset{n\to+\infty}{o}(n)\right).$$

The probability to observe $\mu_n$ in $A$ is exponentially low, with a rate given by the smallest value of $H(P|R)$ for $P \in A$. Assume that the minimizer $P^*$ is unique. In this case, conditionally to the event $P \in A$, $\mu_n$ is close to $P^*$ with very high probability. Therefore, conditionally to the event $\mu_n \in A$, we conclude the stronger statement that $\mu_n \approx P^*$. This is the main idea of the Schrödinger problem. Computing the minimizers of the entropy on some sets gives the most probable configuration conditionally to a very rare event.

## 1.2 The Schrödinger problem and its interpretation

Now, $\mathcal{X} = C([0,1];\mathbb{T}^d)$, and $R$ is the law of the Brownian motion on $\mathbb{T}^d$, starting from the Lebesgue measure, of diffusivity $\nu$ (this is the reversible Brownian motion on the torus). It is well known that for this choice, for all time $t \in [0,1]$, $X_{t\#}R = \mathrm{d}x$.

The Schrödinger problem states as follows.

**Definition 7.** Let $\rho_0, \rho_1 \in \mathcal{P}(\mathbb{T}^d)$. The Schrödinger problem consists in computing the value

$$\inf\left\{H(P|R)\,\middle|\, P \in \mathcal{P}(C([0,1];\mathbb{T}^d)) \text{ with } X_{0\#}P = \rho_0 \text{ and } X_{1\#}P = \rho_1\right\}.$$

Due to the previous section, this problem asks the following question: imagine that a large population of Brownian particles are originally uniformly distributed on $\mathbb{T}^d$. At a first later time that we call $t = 0$, we expect them to be still uniformly distributed. But instead, we observe a large deviation from this expected result, and realize that they are distributed according to $\rho_0$. At time $t = 1$, we expect the density of particles to be $\rho_0 * \tau_\nu$, $(\tau_s)$ being the heat kernel. But instead, we observe $\rho_1$. By the Sanov theorem, we conclude that under this observation, the most probable is that this population of particles is described by $P$, the solution of the Schrödinger problem.

*Remark* 8. The relative entropy is strictly convex w.r.t. $P$ and the constraints are affine. Therefore, uniqueness is obvious. Concerning existence, as the entropy is lower semicontinuous and the constraints are closed, it is equivalent to the existence of a competitor with finite entropy.

## 1.3 Additivity of the entropy, static formulation and formal link with optimal transport

Our next task is to find a necessary and sufficient conditions on $\rho_0$ and $\rho_1$ for the previous problem to admit a (unique) solution/ This will be done in the next section, relying on an important property of the relative entropy that we state now: it behaves well under disintegration. The proposition is the following.

**Proposition 9.** *Let $\mathcal{X}, \mathcal{Y}$ be two complete and separable metric spaces, $P, R$ two Borel probability measures on $\mathcal{X}$ and $T : \mathcal{X} \to \mathcal{Y}$ be a measurable map.*

*Disintegrating $P, R$ by $T$, we know that there are two families of probability measures on $\mathcal{X}$ that we call $(P^y)_{y\in\mathcal{Y}}$ and $(R^y)_{y\in\mathcal{Y}}$ that are respectively well defined for $T_\#P$ and $T_\#R$ almost all $y \in \mathcal{Y}$, that are concentrated on the set $\{x\,|\,T(x) = y\}$ on the $y$s where they are well defined, and such that for all measurable function $\varphi$ nonnegative or bounded*

$$\int \varphi\,\mathrm{d}P = \int_{\mathcal{Y}}\left(\int \varphi\,\mathrm{d}P^y\right)\mathrm{d}T_\#P(y) \quad \textit{and} \quad \int \varphi\,\mathrm{d}R = \int_{\mathcal{Y}}\left(\int \varphi\,\mathrm{d}R^y\right)\mathrm{d}T_\#R(y).$$

*Then, we have*

$$H(P|R) = H(T_\# P|T_\# R) + \int H(P^y|R^y)\,\mathrm{d}T_\# P(y). \tag{1}$$

*In particular,*

$$H(T_\# P|T_\# R) \leq H(P|R).$$

*Pushing forward measures reduces the value of the entropy.*

*Proof.* First, let us compute $\frac{\mathrm{d}T_\# P}{\mathrm{d}T_\# R}$ in terms of $F := \frac{\mathrm{d}P}{\mathrm{d}R}$. Let us consider $\psi$ a measurable and nonnegative function on $\mathcal{Y}$. We have

$$\int \psi\,\mathrm{d}T_\# P = \int \psi \circ T\,\mathrm{d}P = \int \psi \circ T \times F\,\mathrm{d}R = \int \left( \int \psi \circ T \times F\,\mathrm{d}R^y \right) \mathrm{d}T_\# R(y).$$

But for $T_\# R$-almost all $y$, for $R^y$-almost all $x$, $T(x) = y$. Therefore,

$$\int \psi\,\mathrm{d}T_\# P = \int \psi(y) \left( \int F\,\mathrm{d}R^y \right) \mathrm{d}T_\# R(y).$$

We conclude that $P \ll R$ and for $R$-almost all $y$,

$$\frac{\mathrm{d}T_\# P}{\mathrm{d}T_\# R}(y) = \int F\,\mathrm{d}R^y =: G(y).$$

Then, let us compute $\frac{\mathrm{d}P^y}{\mathrm{d}R^y}(x)$, for $T_\# P$-almost all $y$ and $R^y$ almost all $x$. Let $\varphi$ be a nonnegative measurable function on $\mathcal{X}$. We have

$$\int \varphi\,\mathrm{d}P = \int \varphi \times F\,\mathrm{d}R = \int \left( \int \varphi \times F\,\mathrm{d}R^y \right) \mathrm{d}T_\# R(y).$$

Now, observe that for $T_\# R$-almost all $y$, whenever $G(y) = 0$, then we also have $\int \varphi \times F\,\mathrm{d}R^y = 0$. So using as a convention $\frac{0}{0} = 0$, we find

$$\int \varphi\,\mathrm{d}P = \int \left( \int \varphi \times \frac{F}{G(y)}\,\mathrm{d}R^y \right) G(y)\,\mathrm{d}T_\# R(y) = \int \left( \int \varphi \times \frac{F}{G(y)}\,\mathrm{d}R^y \right) \mathrm{d}T_\# P(y).$$

This is enough to conclude that for $T_\# P$-almost all $y$, $P^y \ll R^y$, and for $R^y$-almost all $x$,

$$\frac{\mathrm{d}P^y}{\mathrm{d}R^y}(x) = \frac{F(x)}{G(y)}.$$

We end up with the following formula: for $T_\# P$-almost all $y$, $R^y$-almost all $x$,

$$\frac{\mathrm{d}P}{\mathrm{d}R}(x) = \frac{\mathrm{d}T_\# P}{\mathrm{d}T_\# R}(y) \times \frac{\mathrm{d}P^y}{\mathrm{d}R^y}(x).$$

In particular, for $P$-almost all $x$,

$$\frac{\mathrm{d}P}{\mathrm{d}R}(x) = \frac{\mathrm{d}T_\# P}{\mathrm{d}T_\# R}(T(x)) \times \frac{\mathrm{d}P^{T(x)}}{\mathrm{d}R^{T(x)}}(x).$$

Taking the log of this formula and integrating it against $P$, we find

$$\begin{aligned} H(P|R) &= \int \left( \log \frac{\mathrm{d}T_\# P}{\mathrm{d}T_\# R}(T(x)) + \log \frac{\mathrm{d}P^{T(x)}}{\mathrm{d}R^{T(x)}}(x) \right) \mathrm{d}P(x) \\ &= H(T_\# P|T_\# R) + \int H(P^y|R^y)\,\mathrm{d}T_\# P(y). \end{aligned}$$

The last point of the proposition is obvious. $\square$

Back to the context where the Schrödinger problem is set, the following proposition gives a problem equivalent to the Schrödinger problem, but much simpler to study. In the following, we call $R_{01} := (X_0, X_1)_\# R$. We have

$$\mathrm{d}R_{01}(x,y) = \tau_\nu(y-x)\,\mathrm{d}x\,\mathrm{d}y. \tag{2}$$

Finally, for all $x, y \in \mathbb{T}^d$, we call $R^{x,y} := R(\cdot|X_0 = x, X_1 = y)$ the Brownian bridge between $x$ and $y$. A priori, it is only defined for almost every $x, y$, but classical results of stochastic processes let us define it for all $x, y$.

**Proposition 10.** *For all $P \in \mathcal{P}(C([0,1];\mathbb{T}^d))$, we have*

$$H(P|R) \geq H\Big((X_0,X_1)_\# P \Big| R_{01}\Big).$$

*Conversely, for all $\gamma \in \mathcal{P}(\mathbb{T}^d \times \mathbb{T}^d)$, define $P^\gamma := \int R^{x,y}\,\mathrm{d}\gamma(x,y)$. We have*

$$H(\gamma|R_{01}) = H(P|R).$$

*Consequently, for $\rho_0, \rho_1 \in \mathcal{P}(\mathbb{T}^d)$,*

$$\inf\Big\{H(P|R)\,\Big|\,P \in \mathcal{P}(C([0,1];\mathbb{T}^d)) \text{ with } X_{0\#}P = \rho_0 \text{ and } X_{1\#}P = \rho_1\Big\} = \inf\Big\{H(\gamma|R_{01})\,\Big|\,\gamma \in \Pi(\rho_0,\rho_1)\Big\},$$

*and the existence of a competitor or equivalently of a minimizer in both problems is equivalent.*

*Proof.* The first point is a direct consequence the fact that the entropy is nonincreasing under the push-forward operation.

The second point follows from the observation that $(X_0, X_1)_\# P^\gamma = \gamma$, and for $\gamma$-almost all $x, y$, $(P^\gamma)^{x,y} = R^{x,y}$. Therefore, identity (1) lets us conclude.

The last point follows directly. $\square$

*Remark* 11. Hence, we have reformulated the Schrödinger problem as

$$\inf\Big\{H(\gamma|R_{01})\,\Big|\,\gamma \in \Pi(\rho_0,\rho_1)\Big\}.$$

But using (2), we find that for all $\gamma \in \mathcal{P}(\mathbb{T}^d \times \mathbb{T}^d)$,

$$\begin{aligned}
H(\gamma|R_{01}) &= \int \log \frac{\mathrm{d}\gamma}{\tau_\nu(x,y)\,\mathrm{d}x\,\mathrm{d}y}\,\mathrm{d}\gamma(x,y) = \int \log \frac{\mathrm{d}\gamma}{\mathrm{d}x\,\mathrm{d}y}\,\mathrm{d}\gamma(x,y) - \int \log \tau_\nu(y-x)\,\mathrm{d}\gamma(x,y) \\
&= H(\gamma|\,\mathrm{d}x\,\mathrm{d}y) - \int \log \tau_\nu(y-x)\,\mathrm{d}\gamma(x,y).
\end{aligned}$$

Multiplying this identity by $\nu$ and using that for almost all $x, y \in \mathbb{T}^d$

$$-\nu \log \tau_\nu(y-x) = \frac{1}{2} d(x,y)^2 + \frac{d}{2}\nu \log(2\pi\nu) + \underset{\nu\to 0}{o}(\nu),$$

we find that

$$\nu H(\gamma|R_{01}) = \frac{1}{2}\int d(x,y)^2\,\mathrm{d}\gamma(x,y) + \nu H(\gamma|\,\mathrm{d}x\,\mathrm{d}y) + \frac{d}{2}\nu\log(2\pi\nu) + \underset{\nu\to 0}{o}(\nu).$$

Therefore, the Schrödinger problem consists in minimizing the optimal transport cost while penalizing the entropy of the transport plans. Formally, as $\nu \to 0$, we expect to recover the optimal transport problem. This is done rigorously in $\mathbb{R}^d$ in [1].

## 1.4 Existence of solutions

With Proposition 10 at hand, we can give a necessary and sufficient condition for existence of solutions in the Schrödinger problem.

**Theorem 12.** *Let $\rho_0, \rho_1 \in \mathcal{P}(\mathbb{T}^d)$. The Schrödinger problem between $\rho_0$ and $\rho_1$ admits a solution if and only if $H(\rho_0|\,\mathrm{d}x)$ and $H(\rho_1|\,\mathrm{d}x)$ are both finite.*

*Proof.* By Proposition 10, we can work with the static problem instead of the dynamic one. Assume that the Schrödinger problem admits a solution $\gamma$. Projecting $H(\gamma|R_{01})$ on the first coordinate, we find $H(\rho_0|\,\mathrm{d}x) < +\infty$. In the same way, $H(\rho_1|\,\mathrm{d}x) < +\infty$.

Now, let us assume that $H(\rho_0|\,\mathrm{d}x) < +\infty$ and $H(\rho_1|\,\mathrm{d}x) < +\infty$. Let us show that $\rho_0 \otimes \rho_1$ is an admissible competitor for the static Schrödinger problem. Of course, $\rho_0 \otimes \rho_1 \in \Pi(\rho_0, \rho_1)$. In addition, identifying $\rho_0$ and $\rho_1$ with their densities w.r.t. the Lebesgue measure, we find

$$\begin{aligned} H(\rho_0 \otimes \rho_1 | R_{01}) &= \int \log \frac{\rho_0(x)\rho_1(y)}{\tau_\nu(y-x)} \rho_0(x)\rho_1(y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int \rho_0 \log \rho_0 \,\mathrm{d}x + \int \rho_1 \log \rho_1 \,\mathrm{d}y - \int \log \tau_\nu(y-x)\rho_0(x)\rho_1(y)\,\mathrm{d}x\,\mathrm{d}y \\ &\leq H(\rho_0|\,\mathrm{d}x) + H(\rho_1|\,\mathrm{d}x) + \max_{z\in\mathbb{T}^d} \Big( -\log \tau_\nu(z) \Big) < +\infty. \end{aligned}$$

The result follows. □

## 1.5 The Benamou-Brenier formulation

Just like the optimal transport problem, the Schrödinger problem admits a formulation of Benamou-Brenier type. Deriving it involves several results of probability theory, and more particularly of the theory of stochastic processes. We will state them without a proof.

**The Girsanov theorem.** The first important thing is the Girsanov theorem. It states the following: a law $P$ that has finite entropy w.r.t. the law of the Brownian motion is the law of a Brownian motion to which a drift is added, and this drift is $L^2(P \otimes \mathrm{d}t)$. The theorem, proved for instance in [4], states as follows.

**Theorem 13** (Girsanov under finite entropy)**.** *Let $R$ be the law of the reversible Brownian motion on the torus, of diffusivity $\nu$. For all $P \in \mathcal{P}(C([0,1];\mathbb{T}^d))$, $H(P|R) < +\infty$ if and only if there exists a unique progressive vector valued process $(\overrightarrow{b}_t)_{t\in[0,1]} \in L^2(P\otimes\mathrm{d}t)$ such that under $P$, the canonical process solves the SDE*

$$\mathrm{d}X_t = \overrightarrow{b}_t\,\mathrm{d}t + \mathrm{d}B_t, \tag{3}$$

*where $B$ is a Brownian motion of diffusivity $\nu$ under $P$.*

*In this case, the following identity holds*

$$H(P|R) = H(X_{0\#}P|R) + \frac{1}{2\nu} E_P\left[\int_0^1 |\overrightarrow{b}_t|^2\,\mathrm{d}t\right]. \tag{4}$$

*Remark* 14. Knowing that $(\overrightarrow{b}_t) \in L^2(P\otimes\mathrm{d}t)$, the SDE (3) can be reformulated as follows.

- Under $P$, the process

  $$t \mapsto X_t - \int_0^t \overrightarrow{b}_s\,\mathrm{d}s$$

  is a Brownian motion of diffusivity $\nu$.

- For all $\varphi \in C^2([0,1]\times\mathbb{T}^d)$, under $P$, the following process, which is clearly bounded in $L^2(P)$,

$$t \mapsto \varphi(t,X_t) - \int_0^t \left\{ \partial_t\varphi(s,X_s) + \nabla\varphi(s,X_s)\cdot \overrightarrow{b}_s + \frac{\nu}{2}\Delta\varphi(s,X_s)\right\} \mathrm{d}s \tag{5}$$

  is a martingale.

As shown in the next proposition, there is a PDE binding $\rho$ and $\overrightarrow{b}$.

**Proposition 15.** *In the context of the Girsanov theorem, calling for all $t \in [0,1]$ $\rho(t) := X_{t\#}P$ and for $\mathrm{d}t$-almost all $t$ and $\rho(t)$-almost all $x \in \mathbb{T}^d$ $\overrightarrow{v}(t,x) := E_P[\overrightarrow{b}_t | X_t = x]$, we have distributionally*

$$\partial_t\rho + \mathrm{div}(\rho\overrightarrow{v}) = \frac{\nu}{2}\Delta\rho. \tag{6}$$

*Proof.* Considering $\varphi \in C^2_c(]0,1[\times\mathbb{T}^d)$ and taking the expectation of the martingale in (5) at time $t=1$, we find

$$E_P\left[\int_0^1 \left\{ \partial_t\varphi(s,X_s) + \nabla\varphi(s,X_s)\cdot \overrightarrow{b}_s + \frac{\nu}{2}\Delta\varphi(s,X_s)\right\} \mathrm{d}s\right] = 0.$$

But on the other hand

$$\begin{aligned} &E_P\left[\int_0^1 \left\{ \partial_t\varphi(s,X_s) + \nabla\varphi(s,X_s)\cdot v(s,X_s) + \frac{\nu}{2}\Delta\varphi(s,X_s)\right\} \mathrm{d}s\right] \\ &\qquad = \int_0^1\!\!\int \left\{ \partial_t\varphi(s,x) + \nabla\varphi(s,x)\cdot v(s,x) + \frac{\nu}{2}\Delta\varphi(s,x)\right\} \mathrm{d}\rho(s,x)\,\mathrm{d}s, \end{aligned}$$

so the result follows. □

**Time reversal.** Let $P$ be a law such that $H(P|R) < +\infty$, and

$$T : \omega \in C([0,1];\mathbb{T}^d) \mapsto \Big(t \mapsto \omega(1-t)\Big) \in C([0,1];\mathbb{T}^d).$$

By Proposition 9, $H(T_\# P + T_\# R) \leq H(P|R) < +\infty$ (actually, as $T\circ T = \mathrm{Id}$, applying once again this inequality, we see that it is actually an equality). On the other hand, by a famous property of the Brownian motion, $T_\# R = R$. Therefore , $T_\# P$ has finite entropy w.r.t. $R$, and the Girsanov theorem applies. We can apply the same reasoning as before but with $T_\# P$ instead of $P$. This provides.

**Proposition 16.** *Let $P$ be such that $H(P|R) < +\infty$, $(\overleftarrow{b}_t)$ be the drift associated with the law $T_\# P$ by the Girsanov theorem and $\rho(t) := X_{t\#}P$. Then $(\overleftarrow{b}_t \circ T) \in L^2(P\otimes \mathrm{d}t)$, and*

$$H(P|R) = H(\rho(1)|R) + \frac{1}{2\nu}E_P\left[\int_0^1 |\overleftarrow{b}_t|^2\,\mathrm{d}t\right]. \tag{7}$$

*Moreover, calling for $\mathrm{d}t$-almost all $t$ and $\rho(t)$-almost all $x \in \mathbb{T}^d$, $\overleftarrow{v}(t,x) := -E_P[\overleftarrow{b}_{1-t}\circ T | X_t = x]$, we have*

$$\partial_t\rho + \mathrm{div}(\rho\overleftarrow{v}) = -\frac{\nu}{2}\Delta\rho. \tag{8}$$

*Proof.* If $\varphi \in C^2_c(]0,1[\times\mathbb{T}^d)$, calling $\psi(t,x) := \varphi(1-t,x)$, we have

$$E_{T_\# P}\left[\int_0^1 \left\{ \partial_t\psi(s,X_s) + \nabla\psi(s,X_s)\cdot \overleftarrow{b}_s + \frac{\nu}{2}\Delta\psi(s,X_s)\right\} \mathrm{d}s\right] = 0.$$

But we have

$$\begin{aligned}
E_{T_\# P}&\left[\int_0^1 \left\{\partial_t \psi(s,X_s) + \nabla\psi(s,X_s)\cdot \overleftarrow{b}_s + \frac{\nu}{2}\Delta\psi(s,X_s)\right\}\mathrm{d}s\right] \\
&= E_P\left[\int_0^1 \left\{\partial_t \psi(s,X_{1-s}) + \nabla\psi(s,X_{1-s})\cdot \overleftarrow{b}_s\circ T + \frac{\nu}{2}\Delta\psi(s,X_{1-s})\right\}\mathrm{d}s\right] \\
&= E_P\left[\int_0^1 \left\{-\partial_t \varphi(s,X_s) + \nabla\varphi(s,X_s)\cdot \overleftarrow{b}_{1-s}\circ T + \frac{\nu}{2}\Delta\varphi(s,X_s)\right\}\mathrm{d}s\right],
\end{aligned}$$

and the result follows. $\square$

The next remark gathers the main ideas leading to the Benamou-Brenier formulation of the Schrödinger problem.

*Remark* 17. Notice that taking the half sum of (6) and (8), and calling

$$c := \frac{\overrightarrow{v} + \overleftarrow{v}}{2} \tag{9}$$

$\rho$ solves the continuity equation with velocity $c$. We call this $c$ the *current velocity* associated with $P$.

Also, taking the difference between (6) and (8) and calling

$$w := \frac{\overrightarrow{v} - \overleftarrow{v}}{2}, \tag{10}$$

we find

$$\operatorname{div}\Big(\rho w\Big) = \frac{\nu}{2}\operatorname{div}\Big(\nabla\rho\Big), \tag{11}$$

We call this $w$ the *osmotic velocity* associated with $P$.

By the Jensen's inequality, $\overrightarrow{v}$, $\overleftarrow{v}$, $c$ and $w$ all belong to $L^2(\mathrm{d}t\otimes\rho(t))$.

Actually, we know more than (11) concerning the osmotic velocity. For a proof, see [3].

**Theorem 18.** *Let $P$ be such that $H(P|R) < +\infty$. With the same notations as above, the spatial distributional gradient of $\rho$ satisfies*

$$\frac{\nu}{2}\nabla\rho = \rho w$$

*in $\mathcal{D}'(]0,1[\times\mathbb{T}^d)$.*

*Remark* 19. Whenever a measure $\rho\in\mathcal{P}(\mathbb{T}^d)$ is such that $\nabla\rho$ is a Radon measure with $\nabla\rho \ll \rho$ (in the following, we simply write $\nabla\rho\ll\rho$), we call $\nabla\log\rho$ its Radon-Nikodym derivative. Therefore, in the context of the theorem above, we have

$$w = \frac{\nu}{2}\nabla\log\rho.$$

**Benamou-Brenier formulation.** We have now all the ingredients to state and prove the Benamou-Brenier formula for the Schrödinger problem.

**Theorem 20.** *Let $\rho_0,\rho_1\in\mathcal{P}(\mathbb{T}^d)$. We have*

$$\begin{aligned}
\inf&\left\{\nu H(P|R)\,\middle|\, X_{0\#}P = \rho_0 \text{ and } X_{1\#}P=\rho_1\right\} \\
&= \nu\frac{H(\rho_0|\,\mathrm{d}x) + H(\rho_1|\,\mathrm{d}x)}{2} + \inf\left\{\frac12\int_0^1\!\!\int\left\{|c|^2 + \left|\frac{\nu}{2}\nabla\log\rho\right|^2\right\}\mathrm{d}\rho(t)\,\mathrm{d}t\,\middle|\, \begin{array}{c}(\rho,c) \text{ sol of (CE)},\\ \rho(0)=\rho_0 \text{ and } \rho(1)=\rho_1\end{array}\right\},
\end{aligned}$$

*where the integral in the r.h.s. is set to $+\infty$ unless $\nabla\rho(t)\ll\rho(t)$ for almost every $t$.*

*Proof.* The "$\geq$" part is a consequence of what we already did. Take $\rho(t)$, $\overrightarrow{v}$, $\overleftarrow{v}$, $c$ and $w$ as before. Taking the half sum of (4) and (7), we find that

$$\begin{aligned}
\nu H(P|R) &= \nu \frac{H(\rho_0|\,\mathrm{d}x) + H(\rho_1|\,\mathrm{d}x)}{2} + \frac{1}{2} E_P\left[\int_0^1 \frac{|\overrightarrow{b}_t|^2 + |\overleftarrow{b}_{1-t}|^2}{2}\,\mathrm{d}t\right]\\
&\geq \nu \frac{H(\rho_0|\,\mathrm{d}x) + H(\rho_1|\,\mathrm{d}x)}{2} + \frac{1}{2} E_P\left[\int_0^1 \frac{|\overrightarrow{v}(t,X_t)|^2 + |\overleftarrow{v}(t,X_t)|^2}{2}\,\mathrm{d}t\right]\\
&= \nu \frac{H(\rho_0|\,\mathrm{d}x) + H(\rho_1|\,\mathrm{d}x)}{2} + \frac{1}{2} E_P\left[\int_0^1 \left\{|c(t,X_t)|^2 + |w(t,X_t)|^2\right\}\mathrm{d}t\right]\\
&= \nu \frac{H(\rho_0|\,\mathrm{d}x) + H(\rho_1|\,\mathrm{d}x)}{2} + \frac{1}{2} \int_0^1\!\!\int \left\{|c|^2 + |w|^2\right\}\mathrm{d}\rho(t)\,\mathrm{d}t.
\end{aligned}$$

As $(\rho, c)$ solves the continuity equation and $w = \frac{\nu}{2}\nabla \log \rho$, our claims follows.

The proof of the "$\leq$" part is very similar to the proof of existence of a generalized flow associated with a solution of the continuity equation. First, if $H(\rho_0|\,\mathrm{d}x) = +\infty$ or $H(\rho_1|\,\mathrm{d}x) = +\infty$, then by Proposition 9, both sides of the inequality is infinite. Otherwise, let $(\rho, c)$ be a competitor for the r.h.s.. We use the standard regularization for a given $\varepsilon > 0$

$$\forall t \in [0,1], \quad \rho^\varepsilon(t) := \rho(t) * \tau_\varepsilon, \quad c^\varepsilon(t) := \frac{(\rho(t)c(t)) * \tau_\varepsilon}{\rho^\varepsilon(t)}.$$

Notice that with this definition, we also have

$$\nabla \log \rho^\varepsilon(t) = \frac{(\rho(t)\nabla \log \rho(t)) * \tau_\varepsilon}{\rho^\varepsilon(t)},$$

so that not only

$$\int_0^1\!\!\int |c^\varepsilon|^2\,\mathrm{d}\rho^\varepsilon(t)\,\mathrm{d}t \leq \int_0^1\!\!\int |c|^2\,\mathrm{d}\rho(t)\,\mathrm{d}t,$$

but also

$$\int_0^1\!\!\int |\nabla \log \rho^\varepsilon|^2\,\mathrm{d}\rho^\varepsilon(t)\,\mathrm{d}t \leq \int_0^1\!\!\int |\nabla \log \rho|^2\,\mathrm{d}\rho(t)\,\mathrm{d}t.$$

Now, call $\overrightarrow{v}^\varepsilon := c^\varepsilon + \frac{\nu}{2}\nabla \log \rho^\varepsilon$, and take $P^\varepsilon$ the unique law solution of the SDE

$$\begin{cases} \mathrm{d}X_t = \overrightarrow{v}^\varepsilon(t, X_t)\,\mathrm{d}t + \mathrm{d}B_t,\\ \qquad X_0 \sim \rho^\varepsilon(0), \end{cases}$$

$(B_t)$ being a Brownian motion of diffusivity $\nu$. The law $P^\varepsilon$ has forward drift $\overrightarrow{v}^\varepsilon$ which is clearly $L^2(P^\varepsilon \otimes \mathrm{d}t)$, so $H(P^\varepsilon|R) < +\infty$. Therefore, its osmotic velocity is $\frac{\nu}{2}\nabla \log \rho^\varepsilon$, and hence its current velocity is $c^\varepsilon$ and its backward drift is $\overleftarrow{v}^\varepsilon = \overrightarrow{v}^\varepsilon - \nu \nabla \log \rho^\varepsilon$. A computation similar to the one of the beginning of the proof gives

$$\begin{aligned}
\nu H(P^\varepsilon|R) &= \nu \frac{H(\rho_0^\varepsilon|\,\mathrm{d}x) + H(\rho_1^\varepsilon|\,\mathrm{d}x)}{2} + \frac{1}{2}\int_0^1\!\!\int \left\{|c^\varepsilon|^2 + |w^\varepsilon|^2\right\}\mathrm{d}\rho^\varepsilon(t)\,\mathrm{d}t\\
&\leq \nu \frac{H(\rho_0|\,\mathrm{d}x) + H(\rho_1|\,\mathrm{d}x)}{2} + \frac{1}{2}\int_0^1\!\!\int \left\{|c|^2 + |w|^2\right\}\mathrm{d}\rho(t)\,\mathrm{d}t,
\end{aligned}$$

by argument that we have already seen in the previous lectures. Therefore, as the relative entropy has compact sublevels for the topology of narrow convergence, the family $(P^\varepsilon)$ admits limit points, and for any limit point $P$,

$$\nu H(P|R) \leq \liminf_{\varepsilon \to 0} \nu H(P^\varepsilon|R) \leq \nu \frac{H(\rho_0|\,\mathrm{d}x) + H(\rho_1|\,\mathrm{d}x)}{2} + \frac{1}{2}\int_0^1\!\!\int \left\{|c|^2 + |w|^2\right\}\mathrm{d}\rho(t)\,\mathrm{d}t.$$

As taking the marginals is continuous for the topology of narrow convergence, $P$ is a competitor of the Schrödinger problem, and therefore the result follows. □

*Remark* 21. • Tracking the inequality in the first part of the proof, we see that for the optimizers, $P$-almost everywhere, for almost every $t$,

$$\overrightarrow{b}_t = \overrightarrow{v}(t, X_t) \quad \text{and} \quad \overleftarrow{b}_t = \overleftarrow{v}(t, X_t).$$

Otherwise stated, the drift only depends of the current position. It means that the law $P$ is Markov.

- Therefore, given $\rho_0, \rho_1 \in \mathcal{P}(\mathbb{T}^d)$, the Benamou-Brenier formulation of the Schrödinger problem is

$$\inf \left\{ \frac{1}{2}\int_0^1\!\!\int \left\{ |c|^2 + \left|\frac{\nu}{2}\nabla \log \rho\right|^2 \right\} \mathrm{d}\rho(t)\,\mathrm{d}t \;\middle|\; \begin{array}{c} (\rho, c) \text{ sol of (CE)}, \\ \rho(0) = \rho_0 \text{ and } \rho(1) = \rho_1 \end{array} \right\}.$$

By the proof of Theorem 20, whenever $\rho_0$ and $\rho_1$ have finite entropy w.r.t. Lebesgue, this minimization admits a solution (the density/current velocity of the solution of the dynamical Schrödinger problem), and this solution is unique (otherwise, they would give rise to two different solutions in the dynamical Schrödinger problem).

# 2 Γ-convergence towards optimal transport

In this paragraph, the goal is to show the following theorem.

**Theorem 22.** *Let $\rho_0, \rho_1 \in \mathcal{P}(\mathbb{T}^d)$ have finite entropy w.r.t. the Lebesgue measure. We have*

$$\begin{aligned}\lim_{\nu \to 0} \quad \inf \left\{ \frac{1}{2}\int_0^1\!\!\int \left\{ |c|^2 + \left|\frac{\nu}{2}\nabla \log \rho\right|^2 \right\} \mathrm{d}\rho(t)\,\mathrm{d}t \;\middle|\; \begin{array}{c} (\rho, c) \textit{ sol of (CE)}, \\ \rho(0) = \rho_0 \textit{ and } \rho(1) = \rho_1 \end{array} \right\} \\ = \inf \left\{ \frac{1}{2}\int_0^1\!\!\int |c|^2 \,\mathrm{d}\rho(t)\,\mathrm{d}t \;\middle|\; \begin{array}{c} (\rho, c) \textit{ sol of (CE)}, \\ \rho(0) = \rho_0 \textit{ and } \rho(1) = \rho_1 \end{array} \right\}.\end{aligned}$$

*Moreover, if $(\rho^\nu, c^\nu)_{\nu > 0}$ is the unique solution of the l.h.s., then the family $(\rho^\nu, \rho^\nu c^\nu)$ admit limit points in $\mathcal{P}([0,1] \times \mathbb{T}^d) \times \mathcal{M}([0,1] \times \mathbb{T}^d; \mathbb{R}^d)$, if $(\rho, m)$ is such a limit point, $m \ll \rho$, and calling $c$ its Radon-Nikodym derivative, $(\rho, c)$ is a solution of the r.h.s..*

In this sense, we can say that the Schrödinger problem converges towards the optimal transport problem, at least in Benamou-Brenier formulation. As often in this kind of result, the only hard thing to do is to build a *recovery sequence*, that is, given $(\rho, c)$ a competitor of the r.h.s., to build a sequence $(\rho^\nu, c^\nu)$ for which the l.h.s. functional evaluated at $(\rho^\nu, c^\nu)$ converges towards the r.h.s. functional evaluated at $(\rho, c)$. This is linked to the theory of Γ-convergence that we do not have the time to formalize, but which is the good notion of convergence for minimization problems.

*Proof of Theorem 22.* Given $\rho \in C([0,1]; \mathcal{P}(\mathbb{T}^d))$ and $c$ a measurable vector field, we write

$$\mathcal{A}(\rho, c) := \frac{1}{2}\int_0^1\!\!\int |c|^2 \,\mathrm{d}\rho(t)\,\mathrm{d}t \qquad \text{and} \qquad \mathcal{F}(\rho, c) := \frac{1}{2}\int_0^1\!\!\int \left|\frac{1}{2}\nabla \log \rho\right|^2 \mathrm{d}\rho(t)\,\mathrm{d}t,$$

with value $+\infty$ if $\nabla\rho$ is not absolutely continuous w.r.t. $\rho$. Also, we call $\mathcal{CE}(\rho_0, \rho_1)$ the set of pairs $(\rho, c)$ solution of the continuity equation between $\rho_0$ and $\rho_1$. We want to show

$$\inf_{\mathcal{CE}(\rho_0,\rho_1)} \mathcal{A} = \lim_{\nu \to 0} \inf_{\mathcal{CE}(\rho_0,\rho_1)} \mathcal{A} + \nu^2 \mathcal{F}.$$

Of course, as $\mathcal{F} \geq 0$, it suffices to prove that

$$\limsup_{\nu \to 0} \inf_{\mathcal{CE}(\rho_0,\rho_1)} \mathcal{A} + \nu^2 \mathcal{F} \leq \inf_{\mathcal{CE}(\rho_0,\rho_1)} \mathcal{A}.$$

To do so, we consider $(\rho, c) \in \mathcal{CE}(\rho_0, \rho_1)$ such that $\mathcal{A}(\rho, c) < +\infty$.

For given $\nu, \delta > 0$ and $t \in [0,1]$, we define

$$\rho^{\nu,\delta}(t) = \rho(t) * \tau_{\delta + \nu t(1-t)}.$$

It solves the continuity equation with velocity field

$$c^{\nu,\delta} = c_1^{\nu,\delta} + c_2^{\nu,\delta}$$

where

$$c_1^{\nu,\delta}(t) = \frac{\big(\rho(t)c(t)\big) * \tau_{\delta+\nu t(1-t)}}{\rho^{\nu,\delta}(t)} \qquad \text{and} \qquad c_2^{\nu,\delta}(t) = -\nu\,(1-2t) \times \frac{1}{2}\nabla \log \rho^{\nu,\delta}(t).$$

A quick computation provides, by the same trick as in the previous lecture

$$\begin{aligned}
&\mathcal{A}(\rho^{\nu,\delta}, c^{\nu,\delta}) + \nu^2 \mathcal{F}(\rho^{\nu,\delta}, c^{\nu,\delta}) \\
&\quad = \frac{1}{2}\int_0^1\!\!\int |c_1^{\nu,\delta} + c_2^{\nu,\delta}|^2 \,\mathrm{d}\rho^{\nu,\delta}(t)\,\mathrm{d}t + \frac{\nu^2}{2}\int_0^1\!\!\int \left|\frac{1}{2}\nabla \log \rho^{\nu,\delta}\right|^2 \mathrm{d}\rho^{\nu,\delta}(t)\,\mathrm{d}t \\
&\quad = \frac{1}{2}\int_0^1\!\!\int |c_1^{\nu,\delta}|^2 \,\mathrm{d}\rho^{\nu,\delta}(t)\,\mathrm{d}t + \int_0^1\!\!\int c^{\nu,\delta}\cdot c_2^{\nu,\delta} \,\mathrm{d}\rho^{\nu,\delta}(t)\,\mathrm{d}t - \frac{1}{2}\int_0^1\!\!\int |c_2^{\nu,\delta}|^2 \,\mathrm{d}\rho^{\nu,\delta}(t)\,\mathrm{d}t \\
&\qquad\qquad + \frac{\nu^2}{2}\int_0^1\!\!\int \left|\frac{1}{2}\nabla \log \rho^{\nu,\delta}\right|^2 \mathrm{d}\rho^{\nu,\delta}(t)\,\mathrm{d}t \\
&\quad \le \mathcal{A}(\rho, c) - \frac{\nu}{2}\int_0^1 (1-2t)\,\frac{d}{dt} H(\rho^{\nu,\delta}(t)|\,\mathrm{d}x)\,\mathrm{d}t + \frac{\nu^2}{2}\int_0^1 \left(1-(1-2t)^2\right)\int \left|\frac{1}{2}\nabla\log\rho^{\nu,\delta}(t)\right|^2 \mathrm{d}\rho^{\nu,\delta}(t)\,\mathrm{d}t \\
&\quad = \mathcal{A}(\rho, c) + \nu\left(\frac{H(\rho^{\nu,\delta}(0)|\,\mathrm{d}x) + H(\rho^{\nu,\delta}(1)|\,\mathrm{d}x)}{2} - \int_0^1 H(\rho^{\nu,\delta}(t)|\,\mathrm{d}x)\,\mathrm{d}t\right) \\
&\qquad\qquad + \frac{\nu^2}{2}\int_0^1 4t(1-t)\int \left|\frac{1}{2}\nabla\log\rho^{\nu,\delta}(t)\right|^2 \mathrm{d}\rho^{\nu,\delta}(t)\,\mathrm{d}t \\
&\quad \le \mathcal{A}(\rho, c) + \nu\frac{H(\rho_0|\,\mathrm{d}x) + H(\rho_1|\,\mathrm{d}x)}{2} + \frac{\nu^2}{2}\int_0^1 4t(1-t)\int \left|\frac{1}{2}\nabla\log\rho^{\nu,\delta}(t)\right|^2 \mathrm{d}\rho^{\nu,\delta}(t)\,\mathrm{d}t.
\end{aligned}$$

Recalling the Li-Yau inequality, there is $K$ such that we have for all $\nu, \delta > 0$ and $t \in (0,1)$

$$\int \left|\frac{1}{2}\nabla \log \rho^{\nu,\delta}(t)\right|^2 \mathrm{d}\rho(t) \le \frac{K}{\delta + \nu t(1-t)} \le \frac{K}{\nu t(1-t)}.$$

Therefore, we end up with

$$\mathcal{A}(\rho^{\nu,\delta}, c^{\nu,\delta}) + \nu^2\mathcal{F}(\rho^{\nu,\delta}, c^{\nu,\delta}) \le \mathcal{A}(\rho, c) + \underset{\nu\to 0}{\mathcal{O}}(\nu).$$

By lower-semicontinuity arguments that we have already seen, we can let $\delta$ go to 0 and find for all $\nu > 0$

$$\mathcal{A}(\rho^{\nu}, c^{\nu}) + \nu^2\mathcal{F}(\rho^{\nu}, c^{\nu}) \le \mathcal{A}(\rho, c) + \underset{\nu\to 0}{\mathcal{O}}(\nu).$$

As $(\rho^\nu, c^\nu) \in \mathcal{CE}(\rho_0, \rho_1)$, the convergence of minimum values follows.

For the convergence of minimizers, observe that if $(\rho^\nu, c^\nu)$ is a minimizer for the Benamou-Brenier formulation of the Schrödinger problem, then $\mathcal{A}(\rho^\nu, c^\nu)$ is bounded uniformly in $\nu$ for $\nu$ sufficiently small (as its $\limsup$ is smaller than the optimal value in the optimal transport problem, as a consequence of the first part of the proof). Therefore, $(\rho^\nu, \rho^\nu c^\nu)$ has limit points in $\mathcal{P}([0,1]\times\mathbb{T}^d)\times\mathcal{M}([0,1]\times\mathbb{T}^d)$ and by lower semicontinuity of the optimal transport action, we get that any limit point is a minimizer in the optimal transport problem. $\square$

## References


[1] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM Journal on Mathematical Analysis*, 49(2):1385–1418, 2017.

[2] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38. Springer Science & Business Media, 2009.

[3] H. Föllmer. Time reversal on Wiener space. *Stochastic processes, mathematics and physics (Bielefeld, 1984). Lecture Notes in Mathematics*, 1158:119–129, 1986.

[4] C. Léonard. Girsanov theory under a finite entropy condition. In *Séminaire de Probabilités XLIV*, pages 429–465. Springer, 2012.