FROM THE SCHRÖDINGER PROBLEM TO THE MONGE-KANTOROVICH PROBLEM

CHRISTIAN LÉONARD

Abstract. The aim of this article is to show that the Monge-Kantorovich problem is the limit of a sequence of entropy minimization problems when a fluctuation parameter tends down to zero. We prove the convergence of the entropic values to the optimal transport cost as the fluctuations decrease to zero, and we also show that the limit points of the entropic minimizers are optimal transport plans. We investigate the dynamic versions of these problems by considering random paths and describe the connections between the dynamic and static problems. The proofs are essentially based on convex and functional analysis. We also need specific properties of Γ-convergence which we didn't find in the literature. Hence we prove these Γ-convergence results which are interesting in their own right.

Contents

1. Introduction

The aim of this article is to describe a link between the Monge-Kantorovich optimal transport problem and a sequence of entropy minimization problems. We show that the Monge-Kantorovich problem is the limit of this sequence when a fluctuation parameter tends down to zero. More precisely, we prove that the entropic values tend to the optimal cost as the fluctuations decrease to zero, and also that the limit points of the entropic minimizers are optimal transport plans. We also investigate the dynamic versions of these problems by considering random paths.

Our main results are stated at Section 3, they are Theorems 3.3, 3.6 and 3.7.

Although the assumptions of these results are in terms of large deviation principle, it is not necessary to be acquainted to this theory or even to probability theory to read this article. It is written for analysts and we tried as much as possible to formulate the

Date: November 2010.

2000 *Mathematics Subject Classification.* Primary: 46N10, 49J45. Secondary: 60F10.

Key words and phrases. Optimal transport, Monge-Kantorovich problem, relative entropy, large deviations, Gamma-convergence.

probabilistic notions in terms of analysis and measure theory. A short reminder of the basic definitions and results of large deviation theory is given at the Appendix.

In its Kantorovich form, the optimal transport problem dates back to the 40's, see [Kan42, Kan48]. It appears that its entropic approximation has its roots in an even older problem which was addressed by Schrödinger in the early 30's in connection with the newly born wave mechanics, see [Sch32].

The Monge-Kantorovich optimal transport problem is about finding the best way of transporting some given mass distribution onto another one. We describe these mass distributions by means of two probability measures on a state space $\mathcal{X}$: the initial one is called $\mu_0 \in \mathrm{P}(\mathcal{X})$ and the final one $\mu_1 \in \mathrm{P}(\mathcal{X})$ where $\mathrm{P}(\mathcal{X})$ is the set of all probability measures on $\mathcal{X}$. The rules of the game are (i): it costs $c(x,y) \in [0,\infty]$ to transport a unit mass from x to y and (ii): it is possible to transport infinitesimal portions of mass from the x-configuration μ_0 to the y-configuration μ_1. The resulting minimization problem is the celebrated Monge-Kantorovich problem

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

where $\mathrm{P}(\mathcal{X}^2)$ is the set of all probability measures on $\mathcal{X}^2$ and $\pi_0, \pi_1 \in \mathrm{P}(\mathcal{X})$ are respectively the first and second marginal measures of the joint probability measure $\pi \in \mathrm{P}(\mathcal{X}^2)$. Optimal transport is an active field of research. For a remarkable overview of this exciting topic, see Villani's textbook [Vil09] and the references therein.

Now, let us have a look at Schrödinger's problem. Suppose that you observe a very large number of non-interacting indistinguishable particles which are approximately distributed around a probability measure $\mu_0 \in \mathrm{P}(\mathcal{X})$ on the state space $\mathcal{X}$. We view μ_0 as the initial configuration of the whole particle system. Suppose that you know that the dynamics of each individual particle is driven by a stochastic process whose law is $R^k \in \mathrm{P}(\Omega)$: i.e. a probability measure on the space

$$\Omega = \mathcal{X}^{[0,1]}$$

of all paths[1] from the time interval $[0,1]$ to the state space $\mathcal{X}$. The parameter k describes the fluctuation level $1/k$. As k tends to infinity, R^k tends to some deterministic dynamics: R^∞ describes a deterministic flow. As a typical example, one can think of R^k as the law of a Brownian motion with diffusion coefficient $1/k$. Knowing this dynamics, the law of large numbers tells you that you should expect to see the configuration of the large particle system at the final time $t = 1$ not very far from some expected configuration, with a very high probability. Now, suppose that you observe the system in a configuration close to some $\mu_1 \in \mathrm{P}(\mathcal{X})$ which is far from the expected one. Schrödinger's question is : *"Conditionally on this very rare event, what is the most likely path of the whole system between the times $t = 0$ and $t = 1$?"* As will be seen at Section 2, the answer to this question is related to the entropy minimization problem

$$\frac{1}{k} H(P|R^k) \to \min; \quad P \in \mathrm{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1 \tag{2}$$

where $H(P|R^k)$ is the relative entropy of P with respect to the reference stochastic process R^k and the renormalization factor $1/k$ is here to prevent the entropy from exploding as the fluctuations of R^k decrease. Recall that $H(P|R) := \int_\Omega \log\left(\frac{dP}{dR}\right) dP \in [0,\infty]$, for any $P, R \in \mathrm{P}(\Omega)$. Schrödinger's problem looks like the Monge-Kantorovich one not only because of μ_0 and μ_1, but also because of some cost of transportation. Indeed, if the

[1] During the rigorous treatment, we shall only consider subspaces Ω of $\mathcal{X}^{[0,1]}$, for instance the subspace of all continuous paths.

random dynamics creates a trend to move in some direction rather than in another one, it costs less to the particle system to end up at some configurations μ_1 than others. Even if no direction is favoured, we shall see that the structure of the random fluctuations which is described by the sequence $(R^k)_{k\geq 1}$ encodes some zero-fluctuation cost function c on $\mathcal{X}^2$.

Remark that, although $1/k$ should be of the order of Planck's constant $\hbar$ to build a Euclidean analogue of the quantum dynamics, in [Sch32] Schrödinger isn't concerned with the semiclassical limit $k \to \infty$. Let us also mention that Schrödinger's paper is the starting point of the Euclidean quantum mechanics which was developed by Zambrini [CZ08].

An informal presentation of the convergence result. Our assumption is that $(R^k)_{k\geq 1}$ satisfies a large deviation principle in the path space Ω. This roughly means that

$$R^k(A) \underset{k\to\infty}{\asymp} \exp\left(-k \inf_{\omega\in A} C(\omega)\right), \tag{3}$$

for some rate function $C : \Omega \to [0,\infty]$ and a large class of measurable subsets $A \in \Omega$. For a rigorous definition of a large deviation principle and basic results about large deviation theory, see the Appendix. Very informally, the most likely paths ω correspond to high values $R^k(d\omega)$ and therefore to low values of $C(\omega)$. Under endpoint constraints, it shouldn't be surprising to meet the following family of geodesic problems

$$C(\omega) \to \min; \quad \omega \in \Omega : \omega_0 = x, \omega_1 = y$$

where x, y describe $\mathcal{X}$ and ω_0 and ω_1 are the initial and final positions of the path ω. We see that the large deviation behavior of the sequence $(R^k)_{k\geq 1}$ brings us a family of geodesic paths. It will be shown that the limit (in some sense to be made precise) of the problems (2) is the Monge-Kantorovich problem with the "geodesic" cost function

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

For instance, if R^k is the law of a Brownian motion on $\mathcal{X} = \mathbb{R}^d$ with diffusion coefficient $1/k$, the rate function C is given by Schilder's theorem, a standard large deviation result which tells us that C is the classical kinetic action functional which is given for any path ω by

$$C(\omega) = \frac{1}{2}\int_{[0,1]} |\dot\omega_t|^2\, dt \in [0,\infty]$$

if $\omega = (\omega_t)_{0\leq t\leq 1}$ is absolutely continuous ($\dot\omega$ is its time derivative), and $+\infty$ otherwise. The corresponding static cost is the standard quadratic cost

$$c(x,y) = \frac{1}{2}|y-x|^2, \quad x, y \in \mathbb{R}^d.$$

As a consequence of our general results, we obtain that if the quadratic cost transport problem admits a unique solution, for instance when $\int_{\mathcal{X}} |x|^2\mu_0(dx), \int_{\mathcal{X}} |y|^2\mu_1(dy) < \infty$ and μ_0 is absolutely continuous, then the sequence $(\widehat{P}^k)_{k\geq 1}$ built with the diffusion processes which are the unique solutions to (2) as k varies, converges to the deterministic process

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

where for each $x, y \in \mathcal{X}$, σ^{xy} is the constant velocity geodesic between x and y, $\delta_{\sigma^{xy}}$ is the Dirac measure at σ^{xy} and $\widehat{\pi} \in \mathrm{P}(\mathcal{X}^2)$ is the unique solution to the quadratic cost Monge-Kantorovich transport problem (1). The marginal flow of $\widehat{P}$ is defined to be $(\widehat{P}_t)_{0\leq t\leq 1}$ where for each $0 \leq t \leq 1$, $\widehat{P}_t = \int_{\mathcal{X}^2} \delta_{\sigma^{xy}_t}(\cdot)\, \widehat{\pi}(dxdy) \in \mathrm{P}(\mathcal{X})$ is the law of the random position at time t when the law of the whole random path is $\widehat{P} \in \mathrm{P}(\Omega)$. This flow is precisely the *displacement interpolation* between μ_0 and μ_1 with respect to the quadratic cost transport problem, see [Vil09, Chapter 7] for this notion.

Presentation of the results. The quadratic cost is an important instance of transport cost, but our results are valid for any cost functions c and C satisfying (3) and (4), plus some coercivity properties. For each $k \geq 1$, denote $\rho^k \in \mathrm{P}(\mathcal{X}^2)$ the law of the couple of initial and final positions of the random path driven by $R^k \in \mathrm{P}(\Omega)$. Then,

$$\frac{1}{k} H(\pi|\rho^k) \to \min; \quad \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1 \tag{5}$$

is the static "projection" of (2).

In the sequel, any limit of sequences of probability measures is understood with respect to the usual narrow topology. Theorem 3.7 states that, as k tends to infinity, there exists a sequence $(\mu_1^k)_{k\geq 1}$ in $\mathrm{P}(\mathcal{X})$ such that $\lim_{k\to\infty} \mu_1^k = \mu_1$ and the modified minimization problem

$$\frac{1}{k} H(\pi|\rho^k) \to \min; \quad \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1^k \tag{6}$$

verifies the following two assertions:

- The minimal value of (6) tends to the minimal value of (1), where c is given by (4), which is precisely the optimal transport cost $T_c(\mu_0, \mu_1)$;
- If $T_c(\mu_0, \mu_1)$ is finite, for all large enough k, (6) admits a unique minimizer $\widehat{\pi}^k$, the sequence $(\widehat{\pi}^k)_{k\geq 1}$ admits limit points in $\mathrm{P}(\mathcal{X}^2)$ and any such limit point is a solution to the Monge-Kantorovich problem (1), i.e. an optimal transport plan.

It is not necessary that c is derived from a dynamical cost C via (4). A similar result holds in this more general setting, this is the content of Theorem 3.3. The dynamical analogue of this convergence result is stated at Theorem 3.6 and the connection between the dynamic and static minimizers is described at Theorem 3.7.

Examples of random dynamics $(R^k)_{k\geq 1}$ are introduced. They are mainly based on random walks so that one can compute the corresponding cost functions C and c. In particular, we propose dynamics which generate the standard costs $c_p(x, y) := |y - x|^p$, $x, y \in \mathbb{R}^d$ for any $p > 0$, see Examples 4.6 for such dynamics based on the Brownian motion.

We also prove technical results about Γ-convergence which we didn't find in the literature. They are efficient tools for the proofs of the above mentioned convergence results. A typical result about the Γ-convergence of a sequence of convex functions $(f_k)_{k\geq 1}$ is: If the sequence of the convex conjugates $(f_k^*)_{k\geq 1}$ converges pointwise, then $(f_k)_{k\geq 1}$ Γ-converges. Known results of this type are usually stated in separable reflexive Banach spaces, which is a natural setting when working with PDEs. But here, we need to work with the narrow topology on the set of probability measures. Theorem 6.2 is such a result in this weak topology setting.

Finally, we also proved Theorem 7.1 which tells us that if one adds a continuous constraint to an equi-coercive sequence of Γ-converging minimization problems, then the minimal values and the minimizers of the new problems still enjoy nice convergence properties.

Literature. The connection between large deviation and optimal transport has already been done by Mikami [Mik04] in the context of the quadratic transport. Although no relative entropy appears in [Mik04] where an optimal control approach is performed, our results might be seen as extensions of Mikami's ones. In the same spirit, still using optimal control, Mikami and Thieullen [MT06, MT08] obtained Kantorovich type duality results.

Recently, Adams, Dirr, Peletier and Zimmer [ADPZ] have shown that the small time large deviation behavior of a large particle system is equivalent up to the second order to a single step of the Jordan-Kinderleher-Otto gradient flow algorithm. This is reminiscent of the Schrödinger problem, but the connection is not completely understood by now.

The connection between the Monge-Kantorovich and the Schrödinger problems is also exploited implicitly in some works where (1) 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 $\rho^k = Z_k^{-1} e^{-kc}\rho$ 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 (5). See for instance the papers by Rüschendorf and Thomsen [RT93, RT98] and the references therein. Also interesting is the recent paper by Galichon and Salanie [GS] with an applied point of view.

Proposition 3.4 below is an important technical step on the way to our main results. A variant of this proposition, under more restrictive assumptions than ours, was proved by Dawson and Gärtner [DG94, Thm 2.9] in a context which is different from optimal transport and with no motivation in this direction. Indeed, [DG94] is aimed at studying the large deviations of a large number of diffusion processes subject to a hierarchy of mean-field interactions, by means of random variables which live in $\mathrm{P}(\mathrm{P}(\Omega))$: the set of probability measures on the set of probability measures on the path space Ω. The proofs of Proposition 3.4 in the present paper and in [DG94] differ significantly. Dawson-Gärtner's proof is essentially probabilistic while the author's one is analytic. The strategy of the proofs are also separate: Dawson-Gärtner's proof is based on rather precise probability estimates which partly rely on the specific structure of the problem, while the present one takes place in the other side of convex duality, using the Laplace-Varadhan principle and Γ-convergence. Because of these significantly different proofs and of the weakening of the hypotheses in the present paper, we provide a complete analytic proof of Proposition 3.4 at Section 5.

Organization of the paper. Section 2 is devoted to the presentations of the Monge-Kantorovich and Schrödinger problems. We also show informally that they are tightly connected. Our main results are stated at Section 3. We also give here a simple illustration of these abstract results by means of Schrödinger's original example based on the Brownian motion. Since our primary object is the sequence of random processes $(R^k)_{k\geq 1}$, it is necessary to connect it with the cost functions C and c. This is the purpose of Section 4 where these costs functions are derived for a large family of random dynamics. The proofs of our main results are done at Section 5. They are partly based on two Γ-convergence results which are stated and proved at Sections 6 and 7. Finally, we recall some basic notions about large deviation theory at the Appendix.

Notation. Let us introduce our main notations.

Measures. For any topological space X, we denote $\mathrm{P}(X)$ the set of all Borel probability measures on X and we endow it with the usual *narrow topology* $\sigma(\mathrm{P}(X), \mathrm{C}_b(X))$ weakened by the space $\mathrm{C}_b(X)$ of all continuous bounded functions on X. We also furnish $\mathrm{P}(X)$ with the corresponding Borel σ-field.
The push-forward of the measure m by the measurable application f is denoted by $f_\# m$ and defined by $f_\# m(A) = m(f^{-1}(A))$ for any measurable set A.
The Dirac measure at a is denoted by δ_a.

Measures on a path space. We take a polish space $\mathcal{X}$ which is furnished with its Borel σ-field. The relevant space of paths from the time interval $[0,1]$ to the state space $\mathcal{X}$ is either the space $\Omega = C([0,1],\mathcal{X})$ of all continuous paths, or the space $\Omega = D([0,1],\mathcal{X})$ of paths which are left continuous and right limited (càdlàg[2]) paths. We denote $X = (X_t)_{t\in[0,1]}$ the canonical process which is defined for all $t \in [0,1]$ by

$$X_t(\omega) := \omega_t, \quad \omega = (\omega_t)_{t\in[0,1]} \in \Omega.$$

For each $t \in [0,1]$, X_t is the position at time t which is seen as an application on Ω. Of course, X is the identity on Ω. The set Ω is endowed with the σ-field $\sigma(X_t, t \in [0,1])$ which is generated by the canonical process. It is known that it matches with the Borel σ-field of Ω when Ω is furnished with the Skorokhod topology[3] which turns Ω into a polish space, see [Bil68]. We denote $\mathrm{P}(\mathcal{X})$, $\mathrm{P}(\mathcal{X}^2)$ and $\mathrm{P}(\Omega)$ the set of all probability measures on $\mathcal{X}$, $\mathcal{X}^2 = \mathcal{X}\times\mathcal{X}$ and Ω respectively. For any $P \in \mathrm{P}(\Omega)$, i.e. P is the law of a random path, we denote

$$P_t := (X_t)_\# P \in \mathrm{P}(\mathcal{X}), \quad t \in [0,1].$$

In particular, P_0 and P_1 are the laws of the initial and final random positions under P. Also useful is the joint law of the initial and final positions

$$P_{01} := (X_0, X_1)_\# P \in \mathrm{P}(\mathcal{X}^2).$$

Of course, P_{01} carries more information than the couple (P_0, P_1) because of the correlation structure. A similar remark holds for $P \in \mathrm{P}(\Omega)$ and $(P_t; t \in [0,1]) \in \mathrm{P}(\mathcal{X})^{[0,1]}$. We denote the disintegration of P with respect to (X_0, X_1) : $P(d\omega) = \int_{\mathcal{X}^2} P^{xy}(d\omega)\, P_{01}(dxdy)$ where

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

is the conditional law of X knowing that $X_0 = x$ and $X_1 = y$ under P. Its is usually called the bridge of P between x and y.
When working with the product space $\mathcal{X}^2$, one sees the first and second factors $\mathcal{X}$ as the sets of initial and final states respectively. Therefore, the canonical projections are denoted $X_0(x,y) := x$ and $X_1(x,y) := y$, $(x,y) \in \mathcal{X}^2$. We denote the marginals of the probability measure $\pi \in \mathrm{P}(\mathcal{X}^2)$ by $\pi_0 := (X_0)_\#\pi \in \mathrm{P}(\mathcal{X})$ and $\pi_1 := (X_1)_\#\pi \in \mathrm{P}(\mathcal{X})$.

Functions. Recall that a function $f : X \to (-\infty, \infty]$ is said to be lower semicontinuous on the topological space X if all its sublevel sets $\{f \le a\}$, $a \in \mathbb{R}$ are closed. It is said to be *coercive* if X is assumed to be Hausdorff and its sublevel sets are compact.
Let X and Y be two topological vector spaces equipped with a duality bracket $\langle x, y\rangle \in \mathbb{R}$, that is a bilinear form on $X \times Y$. The convex conjugate f^* of $f : X \to (-\infty, \infty]$ with respect to this duality bracket is defined by

$$f^*(y) := \sup_{x\in X}\{\langle x, y\rangle - f(x)\} \in [-\infty, \infty], \quad y \in Y.$$

It is a convex $\sigma(Y,X)$-lower semicontinuous function.
The relative entropy of the probability P with respect to the probability R is

$$H(P|R) := \begin{cases} \int \log\left(\frac{dP}{dR}\right) dP \in [0,\infty] & \text{if } P \ll R \\ \infty & \text{otherwise.} \end{cases}$$

[2]This is the french acronym for *continu à droite et limité à gauche*.

[3]In the special case where the paths are continuous: $\Omega = C([0,1],\mathcal{X})$, this topology reduces to the topology of uniform convergence.

2. Monge-Kantorovich and Schrödinger problems

In this section we present the Monge-Kantorovich optimal transport problem and the Schrödinger entropy minimization problem. Then, we show informally with the aid of Schrödinger's original example that they are connected to each other, by letting some fluctuation coefficient tend to zero.
Warning. This informal section contains probabilistic material. The probability-allergic reader can skip it without harm. Nevertheless, it also contains some very clever ideas of Schrödinger which acted as a guide for the author.

The Monge-Kantorovich optimal transport problem. Let $c : \mathcal{X}^2 \to [0, \infty]$ be a lower semicontinuous function on $\mathcal{X}^2$ with possibly infinite values. For any $x, y \in \mathcal{X}$, $c(x, y)$ is interpreted as the cost for transporting a unit mass from location x to location y. Let $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$ be two prescribed probability measures on $\mathcal{X}$. An admissible transport plan from μ_0 to μ_1 is any probability measure $\pi \in \mathrm{P}(\mathcal{X}^2)$ which has its first and second marginals equal to $\pi_0 = \mu_0$ and $\pi_1 = \mu_1$, respectively. For such a π,

$$\int_{\mathcal{X}^2} c\, d\pi \in [0, \infty]$$

is interpreted as the total cost for transporting μ_0 to μ_1 when choosing the plan π. The Monge-Kantorovich optimal transport problem is the corresponding minimization problem, i.e.

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

A minimizer $\widehat{\pi} \in \mathrm{P}(\mathcal{X}^2)$ is called an optimal plan and the minimal value $\inf(\mathrm{MK}) \in [0, \infty]$ is the optimal transport cost. Remark that (MK) is a convex minimization problem. But, as it is *not* a *strictly* convex problem, it might admit several solutions.

The Schrödinger entropy minimization problem. Take a reference process R on Ω (the unusual letter R is chosen as a reminder of ***r**eference*). By this, it is meant a positive σ-finite measure on Ω which is not necessarily bounded. Consider n *independent* random dynamic particles $(Y^i; 1 \le i \le n)$ where each random realization of Y^i lives in Ω. More specifically, $(Y^i; 1 \le i \le n)$ is a collection of independent random paths where for each i, the law of Y^i is

$$\mathrm{Law}(Y^i \mid Y^i_0) = R(\cdot \mid X_0 = Y^i_0) \in \mathrm{P}(\Omega) \tag{7}$$

and $(Y^i_0; 1 \le i \le n)$ should be interpreted as the random initial positions.

Example 2.1 (Schrödinger's heat bath). As a typical example, one can take R to be the law of the Brownian motion (Wiener process) on $\mathcal{X} = \mathbb{R}^d$ with diffusion coefficient σ^2 and the Lebesgue measure as its initial distribution. The random motions are described by

$$Y^i_t = Y^i_0 + \sigma B^i_t, \quad 1 \le i \le n, t \in [0, 1]$$

where $Y^1_0, \dots, Y^n_0$ are random independent initial positions, $B^1, \dots, B^n$ are independent Brownian motions with initial position $B^i_0 = 0 \in \mathbb{R}^d$, $i = 1, \dots, n$, and $\sigma > 0$ is the square root of the temperature σ^2.

Schrödinger's original problem [Sch32] which is based on this specific example can be stated as follows. Suppose that at time $t = 0$ you observe a large particle system approximately in the configuration $\mu_0 \in \mathrm{P}(\mathbb{R}^d)$. The law of large numbers tells you that with a very high probability you observe the system at time $t = 1$ in a configuration very near the convolution $\mu_0 * \mathcal{N}(0, \sigma^2 \mathrm{I})$ of μ_0 and the normal density with mean 0 and covariance matrix $\sigma^2 \mathrm{Id}$ in $\mathbb{R}^d$. But, since the number n of particle is *finite*, it is still possible (with a tiny probability of order e^{-an} with $a > 0$) to observe the system at time $t = 1$ in a configuration which is significantly far form the expected profile of distribution $\mu_0 * \mathcal{N}(0, \sigma^2 \mathrm{I})$. Now, Schrödinger's question is[4]: *Suppose that you observe the system at time* $t = 1$ *in a configuration which is approximately* $\mu_1 \in \mathrm{P}(\mathbb{R}^d)$ *and that* μ_1 *is significantly different from* $\mu_0 * \mathcal{N}(0, \sigma^2 \mathrm{I})$, *what is the most likely path of the whole system from* μ_0 *to* μ_1 *during the time interval* $[0, 1]$? In [Sch32], Schrödinger gave the complete answer to this question with a proof based on Stirling's formula. Although proved informally, there is nothing significant to be added today to his answer.

The modern way of addressing this problem is in terms of large deviations, see [DZ98] for an excellent overview of the large deviation theory (a short reminder about large deviation theory is also given at the Appendix). This has been done by Föllmer in his Saint-Flour lecture notes [Föl88]. Denoting δ_a the unit mass Dirac measure at a, the whole system is described by its empirical measure

$$L^n := \frac{1}{n} \sum_{i=1}^n \delta_{Y^i} \in \mathrm{P}(\Omega).$$

It is a $\mathrm{P}(\Omega)$-valued random variable[5] which contains all the information about the dynamic system up to any permutation of the labels of the particles. Nothing is lost when the particles are indistinguishable. It also contains more information than the random path

$$t \in [0, 1] \mapsto L^n_t := \frac{1}{n} \sum_{i=1}^n \delta_{Y^i_t} \in \mathrm{P}(\mathcal{X})$$

which describes the evolution of the configurations. The observed initial and final configurations are the empirical measures

$$L^n_0 = \frac{1}{n} \sum_{i=1}^n \delta_{Y^i_0}, \ L^n_1 = \frac{1}{n} \sum_{i=1}^n \delta_{Y^i_1} \in \mathrm{P}(\mathcal{X}).$$

Now, we give an informal presentation of the answer to Schrödinger's question. For a rigorous treatment, one can have a look at [Föl88]. Take $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$ and C^δ_0, C^δ_1 two δ-neighborhoods in $\mathrm{P}(\mathcal{X})$ (with respect to a distance on $\mathrm{P}(\mathcal{X})$ compatible with the narrow topology $\sigma(\mathrm{P}(\mathcal{X}), \mathrm{C}_b(\mathcal{X}))$) of μ_0 and μ_1 respectively. One can recast Schrödinger's problem as follows: Find the measurable sets $A \subset \mathrm{P}(\Omega)$ such that the conditional probability

$$\mathbb{P}(L^n \in A \mid L^n_0 \in C^\delta_0, L^n_1 \in C^\delta_1) := \frac{\mathbb{P}(L^n \in A, L^n_0 \in C^\delta_0, L^n_1 \in C^\delta_1)}{\mathbb{P}(L^n_0 \in C^\delta_0, L^n_1 \in C^\delta_1)}$$

is maximal when n is large and δ is small. We introduced the δ-blowups C^δ_0 and C^δ_1 to prevent from dividing by zero. A slight variant of Sanov's theorem[6], a standard large deviation result whose exact statement is in terms of large deviation principle, see Theorem A.1 at the Appendix, states that

$$\mathbb{P}(L^n \in A) \underset{n\to\infty}{\asymp} \exp\left(-n \inf\left\{H(P|R); P \in A\right\}\right), \quad A \subset \mathrm{P}(\Omega)$$

[4]Schrödinger's french words are: "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 trouvé 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 ?"

[5]Strictly speaking L^n is a measurable function with its values in $\mathrm{P}(\Omega)$ and the statement "$L^n \in \mathrm{P}(\Omega)$" is an abuse of notation. Nevertheless it is a useful shorthand which will be used below without warning.

[6]If R is a *probability* measure, then this is really Sanov's theorem and $H(P|R) \in [0, \infty]$.

for any measurable subset A of $\mathrm{P}(\Omega)$ (by the way, one must define a σ-field on $\mathrm{P}(\Omega)$) where

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

is the relative entropy of P with respect to R. Under integrability conditions on μ_0 and μ_1 which insure that there exists some $P\in\mathrm{P}(\Omega)$ such that $P_0=\mu_0$, $P_1=\mu_1$ and $H(P|R)<\infty$, one deduces that

$$\mathbb{P}(L^n\in A \mid L^n_0\in C^\delta_0, L^n_1\in C^\delta_1) \underset{n\to\infty}{\asymp} \frac{\exp\left(-n\inf\left\{H(P|R); P: P\in A, P_0\in C^\delta_0, P_1\in C^\delta_1\right\}\right)}{\exp\left(-n\inf\left\{H(P|R); P: P_0\in C^\delta_0, P_1\in C^\delta_1\right\}\right)}.$$

It follows that we have the conditional law of large numbers

$$\lim_{n\to\infty}\mathbb{P}(L^n\in A \mid L^n_0\in C^\delta_0, L^n_1\in C^\delta_1) = \begin{cases} 1 & \text{if } A\ni \widehat{P}^\delta \\ 0 & \text{otherwise}\end{cases}$$

where $\widehat{P}^\delta$ is the unique solution of the entropy minimization problem

$$H(P|R)\to \min; \quad P\in\mathrm{P}(\Omega): P_0\in C^\delta_0, P_1\in C^\delta_1.$$

The uniqueness comes directly from the *strict* convexity of the relative entropy $H(\cdot|R)$. We finally see that, under some conditions on the limits $C^\delta_0 \underset{\delta\downarrow 0}{\to} \{\mu_0\}$ and $C^\delta_1 \underset{\delta\downarrow 0}{\to} \{\mu_1\}$, the solution to the Schrödinger problem is

$$\lim_{\delta\downarrow 0}\lim_{n\to\infty}\mathbb{P}(L^n\in A \mid L^n_0\in C^\delta_0, L^n_1\in C^\delta_1) = \begin{cases} 1 & \text{if } A\ni \widehat{P} \\ 0 & \text{otherwise}\end{cases}$$

where $\widehat{P}$ is the unique solution of the *Schrödinger entropy minimization problem*:

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

If one prefers relative entropies with respect to probability measures, $\widehat{P}$ is also the unique solution to

$$H(P|R^{\mu_0})\to\min; \quad P\in\mathrm{P}(\Omega): P_0=\mu_0, P_1=\mu_1 \tag{8}$$

where the σ-finite measure R has been replaced by the probability measure

$$R^{\mu_0}(d\omega) := \int_{\mathcal{X}} R(d\omega \mid X_0=x)\,\mu_0(dx)\in\mathrm{P}(\Omega) \tag{9}$$

which is the law of the process with initial distribution $\mu_0\in\mathrm{P}(\mathcal{X})$ and the same dynamics as R. This last formulation is analytically suitable, but it introduces an artificial time asymmetry. Nevertheless, we keep it because it will be useful. Remark that

$$H(P|R^{\mu_0})\in[0,\infty]$$

is nonnegative and the minimization problems (S) and (8) share the same minimizer under the constraint $P_0=\mu_0$ since $H(P|R^{\mu_0}) = H(P|R) - \int_{\mathcal{X}} \frac{d\mu_0}{dR_0}\, dP_0 = H(P|R) - \int_{\mathcal{X}} \frac{d\mu_0}{dR_0}\, d\mu_0$.

The minimization problem (S) looks like the Monge-Kantorovich problem (MK), but we can do better in this direction, relying on the tensorization property of the relative entropy. Namely, for any measurable function $\Phi:\Omega\to\mathcal{Z}$ where $\mathcal{Z}$ is any polish space with its Borel σ-field, we have

$$H(P|R) = H(\phi_\# P|\phi_\# R) + \int_{\mathcal{Z}} H\Big(P(\cdot\mid\phi=z)\Big|R(\cdot\mid\phi=z)\Big)\,\phi_\# P(dz). \tag{10}$$

With $\phi = (X_0, X_1)$, this gives us

$$H(P|R) = H(P_{01}|R_{01}) + \int_{\mathcal{X}^2} H\Big(P^{xy}\Big|R^{xy}\Big)\, P_{01}(dxdy). \tag{11}$$

Now, decomposing the marginal constraint $P_0 = \mu_0$, $P_1 = \mu_1$ into $P_{01} = \pi \in \mathrm{P}(\mathcal{X}^2)$ and $(X_0)_\# \pi = \mu_0$, $(X_1)_\# \pi = \mu_1$ we obtain

$$\begin{aligned} & \inf\{H(P|R); P \in \mathrm{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\} \\ = \ & \inf\{\inf[H(P|R); P \in \mathrm{P}(\Omega) : P_{01} = \pi]; \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1\}. \end{aligned}$$

With (11), we see that the inner term is

$$\begin{aligned} & \inf[H(P|R); P \in \mathrm{P}(\Omega) : P_{01} = \pi] \\ = \ & H(\pi|R_{01}) + \inf\left[\int_{\mathcal{X}^2} H\Big(P^{xy}\Big|R^{xy}\Big)\, \pi(dxdy); P \in \mathrm{P}(\Omega) : P_{01} = \pi\right] \\ = \ & H(\pi|R_{01}) \end{aligned}$$

where the inf is uniquely attained when $P^{xy} = R^{xy}$, for π-almost every (x, y), since in this case $0 = H\Big(P^{xy}\Big|R^{xy}\Big)$ which is the minimal value of the relative entropy. This also shows that for each $\pi \in \mathrm{P}(\mathcal{X}^2)$,

$$\inf[H(P|R); P \in \mathrm{P}(\Omega) : P_{01} = \pi] = H(R^\pi|R) = H(\pi|R_{01}) \tag{12}$$

where

$$R^\pi(\cdot) := \int_{\mathcal{X}^2} R^{xy}(\cdot)\, \pi(dxdy) \tag{13}$$

is the mixture of the bridges R^{xy} with π as a mixing measure. Hence, the solution of (S) is

$$\widehat{P} = R^{\widehat{\pi}}$$

where $\widehat{\pi} \in \mathrm{P}(\mathcal{X}^2)$ is the unique solution of

$$H(\pi|R_{01}) \to \min; \quad \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1. \tag{14}$$

Connecting (S) **and** (MK)**.** The problem (14) is similar to (MK), but it remains something to do in order to connect R_{01} with some cost function c. We are going to do it in the special case which is described at Example 2.1, by letting the temperature

$$1/k := \sigma^2$$

tend to zero where $k \geq 1$ describes the positive integers. The general situation will be investigated later at Section 3.

Let us make the k-dependence explicit in our notation. We denote R^k the law of the process Y^k which is defined by

$$Y^k_t = Y_0 + \sqrt{1/k} B_t, \quad 0 \leq t \leq 1 \tag{15}$$

with Y_0 having the Lebesgue measure as its distribution. In particular, the joint law of the initial and final positions under this reference process at positive temperature $1/k$ is

$$R^k_{01}(dxdy) = dx\, (2\pi/k)^{-d/2} \exp\left(-k\frac{|y-x|^2}{2}\right)\, dy.$$

Rewriting (14) with the k-dependence made explicit, we get

$$\frac{1}{k} H(\pi|R^k_{01}) \to \min; \quad \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1. \tag{$\widetilde{\mathcal{S}}^k_{01}$}$$

We "tilde" the name of this problem, because there will be another "untilded" problem later:

$$\frac{1}{k}H(\pi|R_{01}^{\mu_0,k}) \to \min; \quad \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1^k, \tag{$\mathcal{S}_{01}^k$}$$

with better convergence properties, where the constraint μ_1 is replaced by the "moving" constraint μ_1^k which is indexed by k and satisfies $\lim_{k\to\infty} \mu_1^k = \mu_1$ and R_{01}^k is replaced by R_{01}^{k,μ_0}, see (9).

We have also introduced the renormalization $\frac{1}{k}H(\cdot|R_{01}^k)$. To see that $\frac{1}{k}$ is the right multiplying factor, suppose that π is such that $H(\pi|R_{01}^k) < \infty$. This implies that $\pi \ll R_{01}^k \ll \lambda$ where $\lambda(dxdy) = dxdy$ stands for the Lebesgue measure on $\mathbb{R}^d \times \mathbb{R}^d$. We see that

$$H(\pi|R_{01}^k) = H(\pi|\lambda) + \frac{d}{2}\log(2\pi/k) + k\int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{|y-x|^2}{2}\,\pi(dxdy).$$

If we assume that π satisfies $\int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{|y-x|^2}{2}\,\pi(dxdy) < \infty$, we obtain that the Boltzmann entropy $H(\pi|\lambda)$ is finite and

$$\lim_{k\to\infty} \frac{1}{k}H(\pi|R_{01}^k) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{|y-x|^2}{2}\,\pi(dxdy)$$

which is the cost for transporting π_0 to π_1 with respect to the quadratic cost

$$c(x,y) = |y-x|^2/2, \quad x,y \in \mathbb{R}^d.$$

This indicates that ($\mathcal{S}_{01}^k$) *might converge to* (MK) *in some sense, as k tends to infinity.* Indeed, this will be made precise and proved in the subsequent pages.

The renormalized problem (S) with the dependence on k made explicit is

$$\frac{1}{k}H(P|R^k) \to \min; \quad P \in \mathrm{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1. \tag{$\widetilde{\mathcal{S}}^k$}$$

Because of (13), its solution is $R^{k,\tilde{\pi}^k}$ where $\tilde{\pi}^k$ is the solution to ($\widetilde{\mathcal{S}}_{01}^k$). Remark that for two distinct $k, k' > 0$, the supports of R^k and $R^{k'}$ are disjoint subsets of Ω. Therefore, at the level of the process laws $P \in \mathrm{P}(\Omega)$, we see that for all $P \in \mathrm{P}(\Omega)$, $H(P|R^k) = \infty$ for every $k \geq 1$ except possibly one. It appears that the pointwise limit of $\frac{1}{k}H(\cdot|R^k)$ as k tends to infinity is irrelevant. We shall see that the good notion of convergence is that of Γ-convergence. Also, we shall need the following "untilded" variant of ($\widetilde{\mathcal{S}}^k$):

$$\frac{1}{k}H(P|R^{k,\mu_0}) \to \min; \quad P \in \mathrm{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1^k, \tag{$\mathcal{S}^k$}$$

where $\lim_{k\to\infty} \mu_1^k = \mu_1$.

3. Statement of the main results

The statements of our results is in terms of Γ-convergence and large deviation principle. We start introducing their definitions.

Γ-convergence. We refer to the monograph by Dal Maso [Mas93] for a clear exposition of the subject. Recall that if it exists, the Γ-limit of the sequence $(f_k)_{k\geq 1}$ of $(-\infty,\infty]$-valued functions on a topological space X is given for all x in X by

$$\Gamma\text{-}\lim_{k\to\infty} f_k(x) = \sup_{V\in\mathcal{N}(x)} \lim_{k\to\infty} \inf_{y\in V} f_k(y)$$

where $\mathcal{N}(x)$ is the set of all neighborhoods of x. In a metric space X, this is equivalent to:

(i) For any sequence $(x_k)_{k\geq 1}$ such that $\lim_{k\to\infty} x_k = x$,

$$\liminf_{k\to\infty} f_k(x_k) \geq f(x)$$

(ii) and there exits a sequence $(\tilde{x}_k)_{k\geq 1}$ such that $\lim_{k\to\infty} \tilde{x}_k = x$ and

$$\liminf_{k\to\infty} f_k(\tilde{x}_k) \leq f(x).$$

Item (i) is called the *lower bound* and the sequence $(\tilde{x}_k)_{k\geq 1}$ in item (ii) is the *recovery sequence.*

Large deviation principle. We refer to the monograph by Dembo and Zeitouni [DZ98] for a clear exposition of the subject. Let X be a polish space furnished with its Borel σ-field. One says that the sequence $(\gamma_n)_{n\geq 1}$ of probability measures on X satisfies the large deviation principle (LDP for short) with scale n and rate function I, if for each Borel measurable subset A of X we have

$$-\inf_{x\in \operatorname{int} A} I(x) \overset{\text{(i)}}{\leq} \liminf_{n\to\infty} \frac{1}{n}\log \gamma_n(A) \leq \limsup_{n\to\infty} \frac{1}{n}\log \gamma_n(A) \overset{\text{(ii)}}{\leq} -\inf_{x\in \operatorname{cl} A} I(x) \qquad (16)$$

where $\operatorname{int} A$ and $\operatorname{cl} A$ are respectively the topological interior and closure of A in X and the rate function $I : X \to [0,\infty]$ is lower semicontinuous. The inequalities (i) and (ii) are called respectively the *LD lower bound* and *LD upper bound*, where LD is an abbreviation for large deviation. The LDP is the exact statement of what was meant in previous section when writing

$$\gamma_n(A) \underset{n\to\infty}{\asymp} \exp\left(-n \inf_{x\in A} I(x)\right)$$

for "all" $A \subset X$.

The main results. For any topological space X, we denote $\mathrm{P}(X)$ the set of all Borel probability measures on X and we endow it with the usual weak topology $\sigma(\mathrm{P}(X), \mathrm{C}_b(X))$ weakened by the space $\mathrm{C}_b(X)$ of all continuous bounded functions on X. We also furnish $\mathrm{P}(X)$ with the corresponding Borel σ-field.
One says that a function $f : X \to (-\infty, \infty]$ is *coercive* if for any real $a \geq \inf f$, the sublevel set $\{f \leq a\}$ is compact. This implies that f is lower semicontinuous if X is Hausdorff, which will be the case of all the topological spaces in the sequel.
The convex analysis indicator of a set $A \subset X$ is defined by

$$\iota_{\{x\in A\}} = \iota_A(x) = \begin{cases} 0 & \text{if } x \in A \\ \infty & \text{otherwise} \end{cases}, \qquad x \in X.$$

We keep the notation of Section 2. In particular, $\mathcal{X}$ is a polish space (metric complete and separable) with its Borel σ-field and

$$\Omega = D([0,1], \mathcal{X})$$

is the set of all càdlàg $\mathcal{X}$-valued paths endowed with the Skorokhod metric, which turns it into a polish space.

Static version. For each integer $k \geq 1$, we take a measurable kernel

$$(\rho^{k,x} \in \mathrm{P}(\mathcal{X}); x \in \mathcal{X})$$

of probability measures on $\mathcal{X}$. We also take $\mu_0 \in \mathrm{P}(\mathcal{X})$, denote

$$\rho^{k,\mu_0}(dxdy) := \mu_0(dx)\rho^{k,x}(dy) \in \mathrm{P}(\mathcal{X}^2)$$

and define the functions

$$\mathcal{C}_{01}^{k,\mu_0}(\pi) := \frac{1}{k}H(\pi|\rho^{k,\mu_0}) + \iota_{\{\pi_0=\mu_0\}},\ k \geq 1; \quad \mathcal{C}_{01}^{\mu_0}(\pi) := \int_{\mathcal{X}^2} c\, d\pi + \iota_{\{\pi_0=\mu_0\}}, \quad \pi \in \mathrm{P}(\mathcal{X}^2).$$

Proposition 3.1. *We assume that for each $x \in \mathcal{X}$, the sequence $((X_1)_\# \rho^{k,x})_{k\geq 1}$ satisfies the LDP in $\mathcal{X}$ with scale k and the* coercive *rate function*

$$c(x,\cdot) : \mathcal{X} \to [0,\infty]$$

where $c : \mathcal{X}^2 \to [0,\infty]$ is a lower semicontinuous function.
Then, for any $\mu_0 \in \mathrm{P}(\mathcal{X})$ we have: $\Gamma\text{-}\lim_{k\to\infty} \mathcal{C}_{01}^{k,\mu_0} = \mathcal{C}_{01}^{\mu_0}$ in $\mathrm{P}(\mathcal{X}^2)$.

Let us define the functions

$$\begin{aligned} T_{01}^k(\mu_0,\nu) &:= \inf\left\{\frac{1}{k}H(\pi|\rho^{k,\mu_0}); \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \nu\right\} \\ &= \inf\{\mathcal{C}_{01}^k(P); \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_1 = \nu\}, \qquad \nu \in \mathrm{P}(\mathcal{X}) \end{aligned}$$

and

$$\begin{aligned} T_{01}(\mu_0,\nu) &:= \inf\left\{\int_{\mathcal{X}^2} c\, d\pi; \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \nu\right\} \\ &= \inf\{\mathcal{C}_{01}^{\mu_0}(\pi); \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_1 = \nu\}, \qquad \nu \in \mathrm{P}(\mathcal{X}). \end{aligned}$$

The subsequent result follows easily from Proposition 3.1.

Corollary 3.2. *Under the assumptions of Proposition 3.1, for any $\mu_0 \in \mathrm{P}(\mathcal{X})$ we have*

$$\Gamma\text{-}\lim_{k\to\infty} T_{01}^k(\mu_0,\cdot) = T_{01}(\mu_0,\cdot)$$

on $\mathrm{P}(\mathcal{X})$. In particular, for any $\mu_1 \in \mathrm{P}(\mathcal{X})$, there exists a sequence $(\mu_1^k)_{k\geq 1}$ such that $\lim_{k\to\infty} \mu_1^k = \mu_1$ in $\mathrm{P}(\mathcal{X})$ and $\lim_{k\to\infty} T_{01}^k(\mu_0,\mu_1^k) = T_{01}(\mu_0,\mu_1) \in [0,\infty]$.

Now, let us consider a sequence of minimization problems which is a generalization of $(\mathcal{S}_{01}^k)_{k\geq 1}$ at Section 2. It is given for each $k \geq 1$, by

$$\frac{1}{k}H(\pi|\rho^{k,\mu_0}) \to \min; \quad \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1^k \tag{S_{01}^k}$$

where $(\mu_1^k)_{k\geq 1}$ is a sequence in $\mathrm{P}(\mathcal{X})$ as in Corollary 3.2. The Monge-Kantorovich problem associated with $(\mathrm{S}_{01}^k)_{k\geq 1}$ is

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

The main result of the paper is the following theorem.

Theorem 3.3. *Under the assumptions of Proposition 3.1, for any $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$ we have* $\lim_{k\to\infty} \inf(\mathrm{S}_{01}^k) = \inf(\mathrm{MK}) \in [0,\infty]$.
Suppose that in addition $\inf(\mathrm{MK}) < \infty$, *then for each large enough k, (S_{01}^k) admits a* unique *solution* $\widehat{\pi}^k \in \mathrm{P}(\mathcal{X}^2)$.
Moreover, any limit point of the sequence $(\widehat{\pi}^k)_{k\geq 1}$ in $\mathrm{P}(\mathcal{X}^2)$ is a solution to (MK). *In particular, if* (MK) *admits a unique solution $\widehat{\pi} \in \mathrm{P}(\mathcal{X}^2)$, then $\lim_{k\to\infty} \widehat{\pi}^k = \widehat{\pi}$ in $\mathrm{P}(\mathcal{X}^2)$.*

Remark that $\lim_{k\to\infty} \inf(\mathrm{S}^k_{01}) = \inf(\mathrm{MK})$ is a restatement of $\lim_{k\to\infty} T^k_{01}(\mu_0,\mu_1^k) = T_{01}(\mu_0,\mu_1)$ in Corollary 3.2.

Proposition 3.1 and Theorem 3.3 admit a dynamic version.

Dynamical version. For each integer $k\geq 1$, we take a measurable kernel

$$(R^{k,x}\in \mathrm{P}(\Omega^x); x\in\mathcal{X})$$

of probability measures on Ω, with

$$\Omega^x := \{X_0 = x\}.$$

We have in mind the situation where $R^k \in \mathrm{P}(\Omega)$ is the law of stochastic process and $R^{k,x} = R^k(\cdot \mid X_0 = x)$ is its conditional law knowing that $X_0 = x$, see (7). For any $\mu_0 \in \mathrm{P}(\mathcal{X})$, denote

$$R^{k,\mu_0}(\cdot) := \int_{\mathcal{X}} R^{k,x}(\cdot)\,\mu_0(dx) \in \mathrm{P}(\Omega), \quad R^{k,\mu_0}_{01}(\cdot) := \int_{\mathcal{X}} R^{k,x}_{01}(\cdot)\,\mu_0(dx)\in \mathrm{P}(\mathcal{X}^2).$$

We see that R^{k,μ_0} is the law of a stochastic process with initial law μ_0 and its dynamics determined by $(R^{k,x}; x\in\mathcal{X})$ where x must be interpreted as an initial position, while $R^{k,\mu_0}_{01} = (X_0,X_1)_\# R^{k,\mu_0}$ is the joint law of the initial and final positions under R^{k,μ_0}. Let us define the functions

$$\mathcal{C}^{k,\mu_0}(P) := \frac{1}{k} H(P|R^{k,\mu_0}) + \iota_{\{P_0=\mu_0\}},\ k\geq 1, \quad \mathcal{C}^{\mu_0}(P) := \int_\Omega C\,dP + \iota_{\{P_0=\mu_0\}}, \quad P\in \mathrm{P}(\Omega)$$

where $C:\Omega\to[0,\infty]$ is a lower semicontinuous function.

Proposition 3.4. *We assume that for each $x\in\mathcal{X}$, the sequence $(R^{k,x})_{k\geq 1}$ satisfies the LDP in Ω with scale k and the* coercive *rate function*

$$C^x = C + \iota_{\{X_0=x\}} : \Omega\to[0,\infty]$$

where $C:\Omega\to[0,\infty]$ is a lower semicontinuous function.
Then, for any $\mu_0\in\mathrm{P}(\mathcal{X})$ we have: $\Gamma\text{-}\lim_{k\to\infty}\mathcal{C}^{k,\mu_0} = \mathcal{C}^{\mu_0}$ in $\mathrm{P}(\Omega)$.

Let us define the functions

$$\begin{aligned} T^k(\mu_0,\nu) &:= \inf\left\{\frac{1}{k}H(P|R^{k,\mu_0}); P\in\mathrm{P}(\Omega): P_0=\mu_0, P_1=\nu\right\} \\ &= \inf\{\mathcal{C}^{k,\mu_0}(P); P\in\mathrm{P}(\Omega): P_1=\nu\}, \quad \nu\in\mathrm{P}(\mathcal{X}) \end{aligned}$$

and

$$\begin{aligned} T(\mu_0,\nu) &:= \inf\left\{\int_\Omega C\,dP; P\in\mathrm{P}(\Omega): P_0=\mu_0, P_1=\nu\right\} \\ &= \inf\{\mathcal{C}^{\mu_0}(P); P\in\mathrm{P}(\Omega): P_1=\nu\}, \quad \nu\in\mathrm{P}(\mathcal{X}). \end{aligned}$$

Corollary 3.5. *Under the assumptions of Proposition 3.4, we have*

$$\Gamma\text{-}\lim_{k\to\infty} T^k(\mu_0,\cdot) = T(\mu_0,\cdot)$$

on $\mathrm{P}(\mathcal{X})$. In particular, for any $\mu_1\in\mathrm{P}(\mathcal{X})$, there exists a sequence $(\mu_1^k)_{k\geq 1}$ such that

$$\lim_{k\to\infty}\mu_1^k = \mu_1$$

in $\mathrm{P}(\mathcal{X})$ and $\lim_{k\to\infty} T^k(\mu_0,\mu_1^k) = T(\mu_0,\mu_1)\in[0,\infty]$.

Now, let us consider a sequence of minimization problems which is a generalization of $(\mathcal{S}^k)_{k\geq 1}$ at Section 2. It is given for each $k \geq 1$, by

$$\frac{1}{k}H(P|R^{k,\mu_0}) \to \min; \quad P \in \mathrm{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1^k \tag{S^k}$$

where $(\mu_1^k)_{k\geq 1}$ is a sequence in $\mathrm{P}(\mathcal{X})$ as in Corollary 3.5. The dynamic Monge-Kantorovich problem associated with $(\mathrm{S}^k)_{k\geq 1}$ is

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

Indeed, next result states that $(\mathrm{S}^k)_{k\geq 1}$ tends to $(\mathrm{MK_{dyn}})$ in the sense that not only the values of $(\mathrm{S}^k)_{k\geq 1}$ tend to $\inf(\mathrm{MK_{dyn}})$, but also the minimizers of $(\mathrm{S}^k)_{k\geq 1}$ tend to some minimizers of the limiting problem $(\mathrm{MK_{dyn}})$.

Theorem 3.6. *Under the assumptions of Proposition 3.4, for any $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$ we have* $\lim_{k\to\infty} \inf(\mathrm{S}^k) = \inf(\mathrm{MK_{dyn}}) \in [0, \infty]$.
Suppose that in addition $\inf(\mathrm{MK_{dyn}}) < \infty$, *then for each large enough k,* (S^k) *admits a* unique *solution* $\widehat{P}^k \in \mathrm{P}(\Omega)$.
Moreover, any limit point of the sequence $(\widehat{P}^k)_{k\geq 1}$ *in* $\mathrm{P}(\Omega)$ *is a solution to* $(\mathrm{MK_{dyn}})$. *In particular, if* $(\mathrm{MK_{dyn}})$ *admits a unique solution* $\widehat{P} \in \mathrm{P}(\Omega)$, *then* $\lim_{k\to\infty} \widehat{P}^k = \widehat{P}$ *in* $\mathrm{P}(\Omega)$.

From the dynamic to a static version. Once we have the dynamic results, the static ones can be derived by means of the continuous mapping $P \in \mathrm{P}(\Omega) \mapsto (X_0, X_1)_\# P = P_{01} \in \mathrm{P}(\mathcal{X}^2)$. The LD tool which is behind this transfer is the contraction principle which is recalled at Theorem A.2 below. The connection between the dynamic cost C and the static cost c is

$$c(x, y) := \inf\{C(\omega); \omega \in \Omega : \omega_0 = x, \omega_1 = y\} \in [0, \infty], \quad x, y \in \mathcal{X}^2. \tag{17}$$

This identity is connected to the *geodesic problem*:

$$C(\omega) \to \min; \quad \omega \in \Omega : \omega_0 = x, \omega_1 = y. \tag{G^{xy}}$$

Since C^x is coercive for all $x \in \mathcal{X}$, there exists at least one solution to this problem, called a *geodesic path*, provided that its value $c(x, y)$ is finite.

The above static results hold true for any $[0, \infty]$-valued function c satisfying the assumptions of Proposition 3.1 even if it is not derived from a dynamic rate function C via the identity (17). Note also that the coerciveness of C^x for all $x \in \mathcal{X}$, implies that $y \in \mathcal{X} \mapsto c(x, y)$ is coercive (the sublevel sets of $c(x, \cdot)$ are continuous projections of the sublevel sets of C^x which are assumed to be compact). Nevertheless, it is not clear at first sight that c is jointly (on $\mathcal{X}^2$) measurable. Next result tells us that it is jointly lower semicontinuous.
The coerciveness of C^x also guarantees that the set of all geodesic paths from x to y :

$$\Gamma^{xy} := \{\omega \in \Omega; \omega_0 = x, \omega_1 = y, C(\omega) = c(x, y)\}$$

is a compact subset of Ω which is nonempty as soon as $c(x, y) < \infty$. In particular, it is a Borel measurable subset.

Theorem 3.7. *Suppose that the assumptions of Proposition 3.4 are satisfied.*

(1) Then, not only the dynamic results Corollary 3.5 and Theorem 3.6 are satisfied with the cost function C, but also the static results Proposition 3.1, Corollary 3.2 and Theorem 3.3 hold with the cost function c which is derived from C by means of

(17). *It is also true that c is lower semicontinuous and* $\inf(\mathrm{MK}_{\mathrm{dyn}}) = \inf(\mathrm{MK}) \in [0,\infty]$.

Suppose in addition that $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$ satisfy $\inf(\mathrm{MK}) := T_{01}(\mu_0,\mu_1) < \infty$, *so that both* (MK) *and* ($\mathrm{MK}_{\mathrm{dyn}}$) *admit a solution.*

(2) *Then, for all large enough $k \ge 1$,* (S^k_{01}) *and* (S^k) *admit respectively a unique solution $\widehat{\pi}^k \in \mathrm{P}(\mathcal{X}^2)$ and $\widehat{P}^k \in \mathrm{P}(\Omega)$. Furthermore,*

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

which means that $\widehat{P}^k$ is the $\widehat{\pi}^k$-mixture of the bridges $R^{k,xy}$ of R^k.

(3) *The sets of solutions to* (MK) *and* ($\mathrm{MK}_{\mathrm{dyn}}$) *are nonempty convex compact subsets of* $\mathrm{P}(\mathcal{X}^2)$ *and* $\mathrm{P}(\Omega)$ *respectively.*

A probability $\widehat{P} \in \mathrm{P}(\Omega)$ is a solution to ($\mathrm{MK}_{\mathrm{dyn}}$) *if and only if $\widehat{P}_{01}$ is a solution to* (MK) *and*

$$\widehat{P}^{xy}(\Gamma^{xy}) = 1, \quad \forall (x,y) \in \mathcal{X},\ \widehat{P}_{01}\text{-a.e.} \tag{18}$$

In particular, if (MK) *admits a* unique *solution $\widehat{\pi} \in \mathrm{P}(\mathcal{X}^2)$ and for $\widehat{\pi}$-almost every $(x,y) \in \mathcal{X}^2$, the geodesic problem* (G^{xy}) *admits a* unique *solution $\gamma^{xy} \in \Omega$. Then,* ($\mathrm{MK}_{\mathrm{dyn}}$) *admits the unique solution*

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

which is the $\widehat{\pi}$-mixture of the Dirac measures at the geodesics γ^{xy} and

$$\lim_{k\to\infty} \widehat{P}^k = \widehat{P}$$

in $\mathrm{P}(\Omega)$.

Remarks 3.8.

(1) Formula (18) simply means that $\widehat{P}$ only charges geodesic paths. But we didn't write $\widehat{P}(\Gamma) = 1$ since it is not clear that the set $\Gamma := \bigcup_{x,y\in\mathcal{X}} \Gamma^{xy}$ of all geodesic paths is measurable.

(2) In case of uniqueness as in the last statement of this theorem, the marginal flow of $\widehat{P}$ is

$$\mu_t := \widehat{P}_t = \int_{\mathcal{X}^2} \delta_{\gamma^{xy}_t}\,\widehat{\pi}(dxdy) \in \mathrm{P}(\mathcal{X}), \quad t \in [0,1].$$

It is the *displacement interpolation* between μ_0 and μ_1.

As a consequence of the abstract disintegration result of the probability measures on a polish space, the kernel $(x,y) \mapsto \delta_{\gamma^{xy}}$ is measurable. This also means that $(x,y) \mapsto \gamma^{xy}$ is measurable.

(3) If no uniqueness requirement is verified, then

$$\mu_t = (X_t)_\#\widehat{P} \in \mathrm{P}(\mathcal{X}), \quad t \in [0,1]$$

is also a good candidate for being called a displacement interpolation between μ_0 and μ_1.

(4) The problem of knowing if $(\widehat{\pi}^k)_{k\ge1}$ converges even if (MK) admits several solutions is left open in this article. It might be possible that this holds true and that the entropy minimization approximation selects a "viscosity solution" of (MK).

Back to Schrödinger's heat bath. We illustrate these general results by means of Example 2.1. A well-known LD result is about the large deviations of the $\mathbb{R}^d$-valued process which we have already met at (15) and is defined by

$$Y_t^{k,x} = x + \sqrt{1/k}B_t, \quad 0 \le t \le 1, \tag{19}$$

where the initial condition $Y_0^{k,x} = x$ is deterministic, $B = (B_t)_{0\le t\le 1}$ is the Wiener process on $\mathbb{R}^d$ and we decided to take $\sigma^2 = 1/k$. with $k \ge 1$ an integer.

Theorem 3.9 (Schilder's theorem)**.** *The sequence of random processes $(Y^{k,x})_{k\ge 1}$ satisfies the LDP in $\Omega = C([0,1],\mathbb{R}^d)$ equipped with the topology of uniform convergence with scale k and rate function*

$$C^x(\omega) = \int_{[0,1]} \frac{|\dot\omega_t|^2}{2}\,dt \in [0,\infty], \quad \omega \in \Omega$$

if $\omega_0 = x$ and ω is an absolutely continuous path (its derivative is denoted by $\dot\omega$) and $C^x(\omega) = \infty$, otherwise.

For a proof, see [DZ98, Thm 5.2.3].

With our notation, this corresponds to

$$C(\omega) = \begin{cases} \int_{[0,1]} \frac{|\dot\omega_t|^2}{2}\,dt \in [0,\infty] & \text{if } \omega \in \Omega_{\mathrm{ac}} \\ \infty & \text{otherwise} \end{cases}, \quad \omega \in \Omega$$

where Ω_{ac} is the space of all absolutely continuous paths $\omega : [0,1] \to \mathbb{R}^d$. By Jensen's inequality, (17) leads us to

$$c(x,y) = |y-x|^2/2, \quad x,y \in \mathbb{R}^d$$

which is the well-known quadratic transport cost. Let $R^{k,x} \in \mathrm{P}(\Omega)$ denote the law of $Y^{k,x}$. Then $R^{k,\mu_0}(\cdot) = \int_{\mathbb{R}^d} R^{k,x}(\cdot)\,\mu_0(dx) \in \mathrm{P}(\Omega)$ is the law of

$$Y_t^k = Y_0 + \sqrt{1/k}B_t, \quad 0 \le t \le 1,$$

with initial law: $\mathrm{Law}(Y_0) = \mu_0 \in \mathrm{P}(\mathbb{R}^d)$. Also denote $\rho^{k,\mu_0} = (X_0,X_1)_\# R^{k,\mu_0} \in \mathrm{P}(\mathbb{R}^d \times \mathbb{R}^d)$, i.e.

$$\rho^{k,\mu_0}(dxdy) = \mu_0(dx)(2\pi/k)^{-d/2}\exp\left(-k\frac{|y-x|^2}{2}\right)\,dy.$$

The above results tell us that if there exists some $\pi^* \in \mathrm{P}(\mathbb{R}^d \times \mathbb{R}^d)$ such that $\pi_0^* = \mu_0$, $\pi_1^* = \mu_1$ and $\int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^2\,\pi^*(dxdy) < \infty$, then $T_{01}(\mu_0,\mu_1) < \infty$, there exists a sequence $(\mu_1^k)_{k\ge 1}$ such that $\lim_{k\to\infty}\mu_1^k = \mu_1$ in $\mathrm{P}(\mathbb{R}^d)$ and for any large enough $k \ge 1$, the entropy minimization problem

$$\frac{1}{k}H(\pi|\rho^{k,\mu_0}) \to \min; \quad \pi \in \mathrm{P}(\mathbb{R}^d \times \mathbb{R}^d), \pi_0 = \mu_0, \pi_1 = \mu_1^k$$

admits a unique solution $\widehat\pi^k \in \mathrm{P}(\mathbb{R}^d \times \mathbb{R}^d)$, the sequence $(\widehat\pi^k)_{k\ge 1}$ admits at least a limit point in $\mathrm{P}(\mathbb{R}^d \times \mathbb{R}^d)$ and any such limit point is a solution of the Monge-Kantorovich quadratic transport problem

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{|y-x|^2}{2}\,\pi(dxdy) \to \min; \quad \pi \in \mathrm{P}(\mathbb{R}^d \times \mathbb{R}^d), \pi_0 = \mu_0, \pi_1 = \mu_1.$$

In addition, we have $\lim_{k\to\infty}\frac{1}{k}H(\widehat\pi^k|\rho^{k,\mu_0}) = T_{01}(\mu_0,\mu_1)$.
Moreover, for all large enough $k \ge 1$, the corresponding dynamic problem

$$\frac{1}{k}H(P|R^{k,\mu_0}) \to \min; \quad P \in \mathrm{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1^k$$

has a unique solution $\widehat{P}^k \in \mathrm{P}(\Omega)$ which is given by

$$\widehat{P}^k = R^{k,\widehat{\pi}^k} = \int_{\mathbb{R}^d\times\mathbb{R}^d} R^{k,xy}(\cdot)\,\widehat{\pi}^k(dxdy) \in \mathrm{P}(\Omega),$$

the sequence $(\widehat{P}^k)_{k\geq 1}$ admits at least a limit point in $\mathrm{P}(\Omega)$ and any such limit point is a solution of the dynamic Monge-Kantorovich quadratic transport problem

$$\int_{\Omega_{\mathrm{ac}}} \left[\int_{[0,1]} \frac{|\dot\omega_t|^2}{2}\,dt\right] P(d\omega) \to \min; \quad P \in \mathrm{P}(\Omega_{\mathrm{ac}}), P_0 = \mu_0, P_1 = \mu_1.$$

In the case where μ_0 or μ_1 is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, it is well known [Bre91, McC95] that (MK) admits a unique solution $\widehat{\pi}$. By Theorem 3.7, we obtain $\lim_{k\to\infty} \widehat{P}^k = \widehat{P}$ in $\mathrm{P}(\Omega)$ where

$$\widehat{P}(\cdot) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \delta_{[t\mapsto(1-t)x+ty]}(\cdot)\,\widehat{\pi}(dxdy) \in \mathrm{P}(\Omega)$$

and the corresponding displacement interpolation is the marginal flow of $\widehat{P}$ which is

$$\mu_t(\cdot) = \widehat{P}_t(\cdot) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \delta_{(1-t)x+ty}(\cdot)\,\widehat{\pi}(dxdy) \in \mathrm{P}(\mathbb{R}^d), \quad t \in [0,1].$$

Moreover, the marginal flow of $\widehat{P}^k$ is

$$\mu^k_t(\cdot) = \widehat{P}^k_t(\cdot) = \int_{\mathbb{R}^d\times\mathbb{R}^d} R^{k,xy}_t(\cdot)\,\widehat{\pi}^k(dxdy) \in \mathrm{P}(\mathbb{R}^d), \quad t \in [0,1]$$

and for each $t \in [0,1]$, $\lim_{k\to\infty} \mu^k_t = \mu_t$ in $\mathrm{P}(\mathbb{R}^d)$.

Mikami's paper [Mik04] is in the context of Schrödinger's heat bath based on the Wiener process as above. Although no relative entropy nor Γ-convergence enter the statements of [Mik04]'s results, some of the previous results about Schrödinger's heat bath are close to the main results of [Mik04] which are proved by means of cyclical monotonicity with a stochastic optimal control point of view. Theorems 3.3, 3.6, 3.7 apply to a large class of optimal transport costs, see Section 4. They shed a new light on [Mik04]'s results and extend them in the sense that the reference process R is not restricted to be the Wiener process and the LD principle which is satisfied by $(R^k)_{k\geq 1}$ is not restricted to the setting of Schilder's theorem.

4. From stochastic processes to transport cost functions

We have just seen that Schilder's theorem leads to the quadratic cost function. The aim of this section is to present a series of examples of LD sequences $(R^k)_{k\geq 1}$ in $\mathrm{P}(\Omega)$ which give rise to cost functions c on $\mathcal{X}^2$.

Simple random walks on $\mathbb{R}^d$. Instead of (19), let us consider

$$Y^{k,x}_t = x + W^k_t, \quad 0 \leq t \leq 1, \tag{20}$$

where for each $k \geq 1$, W^k is a random walk. The law of $Y^{k,x}$ is our $R^{k,x} \in \mathrm{P}(\Omega)$.

To build these random walks, one needs a sequence of independent copies $(Z_m)_{m\geq 1}$ of a random variable Z in $\mathbb{R}^d$. For each integer $k \geq 1$, W^k is the rescaled random walk defined for all $0 \leq t \leq 1$, by

$$W^k_t = \frac{1}{k}\sum_{j=1}^{\lfloor kt\rfloor} Z_j \tag{21}$$

where $\lfloor kt \rfloor$ is the integer part of kt. This sequence satisfies a LDP which is given by Mogulskii's theorem. As a pretext to set some notations, we recall its statement. The logarithm of the Laplace transform of the law $m_Z \in \mathrm{P}(\mathbb{R}^d)$ of Z is $\log \int_{\mathbb{R}^d} e^{\zeta \cdot z}\, m_Z(dz)$. Its convex conjugate is

$$c_Z(v) := \sup_{\zeta \in \mathbb{R}^d} \left\{ \zeta \cdot v - \log \int_{\mathbb{R}^d} e^{\zeta \cdot z}\, m_Z(dz) \right\}, \quad v \in \mathbb{R}^d. \tag{22}$$

One can prove, see [DZ98], that c_Z is a convex $[0,\infty]$-valued function which attains its minimum value 0 at $v = \mathbb{E}Z = \int_{\mathbb{R}^d} z\, m_Z(dz)$. Moreover, the closure of its effective domain $\mathrm{cl}\{c_Z < \infty\}$ is the closed convex hull of the topological support $\mathrm{supp}\, m_Z$ of the probability measure m_Z. Under the assumption (23) below, it is also *strictly* convex.
For each initial value $x \in \mathbb{R}^d$, we define the action functional

$$C_Z^x(\omega) := \begin{cases} \int_{[0,1]} c_Z(\dot\omega_t)\, dt & \text{if } \omega \in \Omega_{\mathrm{ac}} \text{ and } \omega_0 = x \\ +\infty & \text{otherwise} \end{cases}, \quad \omega \in \Omega.$$

Theorem 4.1 (Mogulskii's theorem)**.** *Under the assumption*

$$\int_{\mathbb{R}^d} e^{\zeta \cdot z}\, m_Z(dz) < +\infty, \quad \forall \zeta \in \mathbb{R}^d, \tag{23}$$

for each $x \in \mathbb{R}^d$ the sequence $(R^{k,x})_{k\ge 1}$ of the laws of $(Y^{k,x})_{k\ge1}$ specified by (20) *satisfies the LDP in $\Omega = D([0,1],\mathbb{R}^d)$, equipped with its natural σ-field and the topology of uniform convergence, with scale k and the coercive rate function C_Z^x.*

For a proof see [DZ98, Thm 5.1.2]. This result corresponds to our general setting with

$$C(\omega) = C_Z(\omega) := \begin{cases} \int_{[0,1]} c_Z(\dot\omega_t)\, dt & \text{if } \omega \in \Omega_{\mathrm{ac}} \\ +\infty & \text{otherwise} \end{cases}, \quad \omega \in \Omega. \tag{24}$$

Since c_Z is a strictly convex function, the geodesic problem (G^{xy}) admits as unique solution the *constant velocity geodesic*

$$\sigma^{xy} : t \in [0,1] \mapsto (1-t)x + ty \in \mathbb{R}^d. \tag{25}$$

Now, let us only consider the final position

$$Y_1^{k,x} = x + \frac{1}{k} \sum_{j=1}^k Z_j.$$

Denote $\rho^{k,x} = (X_1)_\# R^{k,x} \in \mathrm{P}(\mathcal{X})$ the law of $Y_1^{k,x}$. By the contraction principle, see Theorem A.2 at the Appendix, one deduces immediately from Mogulskii's theorem the simplest result of LD theory which is the Cramér theorem.

Corollary 4.2 (A complicated version of Cramér's theorem)**.** *Under the assumption* (23)*, for each $x \in \mathbb{R}^d$ the sequence $(\rho^{k,x})_{k\ge1}$ of the laws of $(Y_1^{k,x})_{k\ge1}$ satisfies the LDP in $\mathbb{R}^d$ with scale k and the coercive rate function*

$$y \in \mathcal{X} \mapsto c_Z(y-x) \in [0,\infty] \quad y \in \mathcal{X}$$

where c_Z is given at (22)*.*
Furthermore, $c_Z(v) = \inf\{C_Z(\omega); \omega \in \Omega : \omega_0 = x, \omega_1 = x+v\}$ for all $x, v \in \mathbb{R}^d$.

Last identity is a simple consequence of Jensen's inequality which also lead us to (25) a few lines earlier. Cramér's theorem corresponds to the case when $x = 0$ and only the deviations of $Y_1^{k,0} = \frac{1}{k}\sum_{j=1}^k Z_j$ in $\mathbb{R}^d$ are considered.

Theorem 4.3 (Cramér's theorem). *Under the assumption* (23), *the sequence* $(\frac{1}{k}\sum_{j=1}^k Z_j)_{k\geq 1}$ *satisfies the LDP in* $\mathbb{R}^d$ *with scale* k *and the coercive rate function* c_Z *given at* (22).

For a proof, see [DZ98, Thm 2.2.30].

We have just described a general procedure which converts the law $m_Z \in \mathrm{P}(\mathbb{R}^d)$ into the cost functions C_Z and c_Z. Here are some examples with explicit computations.

Examples 4.4. We recall some well-known examples of Cramér transform c_Z.

(1) To obtain the quadratic cost function $c_Z(v) = |v|^2/2$, choose Z as a standard normal random vector in $\mathbb{R}^d$: $m_Z(dz) = (2\pi)^{-d/2}\exp(-|z|^2/2)\,dz$.
(2) Taking Z such that $\mathbb{P}(Z=+1) = \mathbb{P}(Z=-1) = 1/2$, i.e. $m_Z = (\delta_{-1}+\delta_{+1})/2$ leads to
$$c_Z(v) = \begin{cases} [(1+v)\log(1+v) + (1-v)\log(1-v)]/2, & \text{if } -1<v<+1 \\ \log 2, & \text{if } v\in\{-1,+1\} \\ +\infty, & \text{if } v\not\in[-1,+1]. \end{cases}$$
(3) If Z has an exponential law with expectation 1, i.e. $m_Z(dz) = \mathbf{1}_{\{z\geq 0\}}e^{-z}\,dz$, then $c_Z(v) = v-1-\log v$ if $v>0$ and $c_Z(v)=+\infty$ if $v\leq 0$.
(4) If Z has a Poisson law with expectation 1, i.e. $m_Z(dz) = e^{-1}\sum_{n\geq 0}\frac{1}{n!}\delta_n(dz)$, then $c_Z(v) = v\log v - v + 1$ if $v>0$, $c_Z(0)=1$ and $c_Z(v)=+\infty$ if $v<0$.

We have $c_Z(0)=0$ if and only if $\mathbb{E}Z := \int_{\mathbb{R}^d} z\,m_Z(dz) = 0$. More generally, $c_Z(v)\in[0,+\infty]$ and $c_Z(v)=0$ if and only if $v=\mathbb{E}Z$. We also have
$$c_{aZ+b}(u) = c_Z\big(a^{-1}(v-b)\big)$$
for all invertible linear operator $a:\mathbb{R}^d\to\mathbb{R}^d$ and all $b\in\mathbb{R}^d$.

If $\mathbb{E}Z=0$, c_Z is quadratic at the origin since $c_Z(v) = v\cdot\Gamma_Z^{-1}v/2 + o(|v|^2)$ where Γ_Z is the covariance matrix of Z. This rules out the usual costs $c(v)=|v|^p$ with $p\neq 2$.

Nevertheless, taking Z a real valued variable with density $C\exp(-|z|^p/p)$ with $p\geq 1$ leads to $c_Z(v) = |v|^p/p(1+o_{|v|\to\infty}(1))$. The case $p=1$ follows from Example 4.4-(3) above. To see that the result still holds with $p>1$, compute by means of the Laplace method the principal part as ζ tends to infinity of $\int_0^\infty e^{-z^p/p}e^{\zeta z}\,dz = \sqrt{2\pi(q-1)}\zeta^{1-q/2}e^{\zeta^q/q}(1+o_{\zeta\to+\infty}(1))$ where $1/p+1/q=1$.
Of course, we deduce a related d-dimensional result considering Z with the density $C\exp(-|z|_p^p/p)$ where $|z|_p^p = \sum_{i\leq d}|z_i|^p$. This gives $c_Z(v) = |v|_p^p/p(1+o_{|v|\to\infty}(1))$.

Remark 4.5. Let R^k be defined by (20) and (21) where Z is only allowed to take isolated values as Examples 4.4-(2) and (4). Suppose that μ_0 has a discrete support, then R_1^{k,μ_0} has also a discrete support. It follows that any $P\in\mathrm{P}(\Omega)$ which is absolutely continuous with respect to R_1^{k,μ_0} is such that P_1 has a discrete support. Now, if you choose a diffuse measure for μ_1, there is no solution to the non-modified minimization problem (2). We see that it is necessary to introduce a sequence $(\mu_1^k)_{k\geq 1}$ of discrete measures such that $\lim_{k\to\infty}\mu_1^k = \mu_1$ for the sequences of entropy minimization modified problems $(\mathrm{S}_{01}^k)_{k\geq 1}$ and $(\mathrm{S}^k)_{k\geq 1}$ to admit solutions.

Nonlinear transformations. By means of the contraction principle (Theorem A.2), we can twist the cost functions which have been obtained earlier. We only present some examples to illustrate this technique.

The static case. Here, we only consider the LD of the final position Y_1^k. We have just remarked that the cost functions c_Z as above are necessarily quadratic at the origin. This drawback will be partly overcome by means of continuous transformations.

We are going to look at an example

$$Y_1^{k,x} = x + V^k$$

where $(V^k)_{k\geq 1}$ satisfies a LDP which is not given by Cramér's theorem. Let $(Z_j)_{j\geq 1}$ be as above and let α be any continuous mapping on $\mathbb{R}^d$. Consider

$$V^k = \alpha\left(\frac{1}{k}\sum_{1\leq j\leq k} Z_j\right).$$

We obtain $c(v) = \inf\{c_Z(u); u \in \mathbb{R}^d, \alpha(u) = v\}, v \in \mathbb{R}^d$ as a consequence of the contraction principle. In particular if α is a continuous injective mapping, then

$$c = c_Z \circ \alpha^{-1}. \tag{26}$$

For instance, if Z is a standard normal vector as in Example 4.4-(1), we know that the empirical mean of independent copies of $Z : \frac{1}{k}\sum_{1\leq j\leq k} Z_j$, is a centered normal vector with variance Id/k. Taking $\alpha = \alpha_p$ which is given for each $p > 0$ and $v \in \mathbb{R}^d$ by $\alpha_p(v) = 2^{-1/p}|v|^{2/p-1}v$, leads us to

$$V^k \overset{\text{Law}}{=} (2k)^{-1/p}|Z|^{2/p-1}Z, \tag{27}$$

the equality in law $\overset{\text{Law}}{=}$ simply means that both sides of the equality share the same distribution. The mapping α_p has been chosen to obtain with (26):

$$c(v) := c_p(v) = |v|^p, \quad v \in \mathbb{R}^d.$$

Note that V^k has the same law as $k^{-1/p}Z_p$ where the density of the law of Z_p is $\kappa|z|^{p/2-1}e^{-|z|^p}$ for some normalizing constant κ.

The dynamic case. We now look at an example where

$$Y_t^{k,x} = x + V_t^k, \quad 0 \leq t \leq 1 \tag{28}$$

where $(V^k)_{k\geq 1}$ satisfies a LDP in Ω which is not given by Mogulskii's theorem.

We present examples of dynamics V^k based on the standard Brownian motion $B = (B_t)_{0\leq t\leq 1}$ in $\mathbb{R}^d$. In these examples, one can restrict the path space to be the space $\Omega = C([0,1],\mathbb{R}^d)$ equipped with the uniform topology. The item (1) is already known to us, we recall it for the comfort of the reader.

Examples 4.6.

(1) An important example is given by

$$V_t^k = k^{-1/2}B_t, \quad 0 \leq t \leq 1.$$

Schilder's theorem states that $(V^k)_{k\geq 1}$ satisfies the LDP in Ω with the coercive rate function

$$C^0(\omega) = \begin{cases} \int_0^1 |\dot\omega_t|^2/2\,dt & \text{if } \omega \in \Omega_{\mathrm{ac}}, \omega_0 = 0 \\ +\infty & \text{otherwise.} \end{cases}$$

As in Example 4.4-(1), it corresponds to the quadratic cost function $|v|^2/2$, but with a different dynamics.

(2) More generally, with $p > 0$, we have just seen that

$$V_t^k = (2k)^{-1/p}|B_t|^{2/p-1}B_t, \quad 0 \leq t \leq 1$$

corresponds to the power cost function $c_p(v)=|v|^p$, $v\in\mathbb{R}^d$, since $V_1^k \overset{\text{Law}}{=} V^k$ as in (27). The associated dynamic cost is given for all $\omega\in\Omega$ by

$$C^0(\omega)=\begin{cases} p^2/4\int_{[0,1]}|\omega_t|^{p-2}|\dot\omega_t|^2\,dt & \text{if } \omega\in\Omega_{\mathrm{ac}},\omega_0=0\\ +\infty & \text{otherwise.}\end{cases}$$

(3) Similarly, with $p>0$, the dynamics

$$V_t^k=(2k)^{-1/p}|B_t/t|^{2/p-1}B_t,\quad 0<t\le 1$$

also corresponds to the power cost function $c_p(v)=|v|^p$, $v\in\mathbb{R}^d$, since $V_1^k \overset{\text{Law}}{=} V^k$ as in (27). But, this time the associated dynamic cost is given for all $\omega\in\Omega$ by

$$C^0(\omega)=\begin{cases} \frac14\int_{(0,1]}\mathbf{1}_{\{\omega_t\neq 0\}}|\omega_t/t|^p\Big|(2-p)\omega_t/|\omega_t|+pt\dot\omega_t/|\omega_t|\Big|^2\,dt & \text{if } \omega\in\Omega_{\mathrm{ac}},\omega_0=0\\ +\infty & \text{otherwise.}\end{cases}$$

Recall that a geodesic path from x to y is some $\omega\in\Omega_{\mathrm{ac}}$ which solves the minimization problem (G^{xy}). It is well known that the geodesic paths for Item (1) are the constant velocity paths σ^{xy}, see (25). The geodesic paths for Item (2) are still straight lines but with a time dependent velocity (except for $p=2$). On the other hand, the geodesic paths for Item (3) are the constant velocity paths.

Modified random walks on $\mathbb{R}^d$. Simple random walks correspond to (28) with $V^k=W^k$ given by (21). We introduce a generalization which is defined by (28) with

$$V_t^k=\alpha_t(W_t^k),\quad 0\le t\le 1$$

where $\alpha:(t,v)\in[0,1]\times\mathbb{R}^d\mapsto\alpha_t(v)\in\mathbb{R}^d$ is a continuous application such that $\alpha_0(0)=0$ (remark that $W_0^k=0$ almost surely) and α_t is injective for all $0<t\le 1$.
For all $x\in\mathbb{R}^d$ and all $k\ge 1$, the random path $Y^{k,x}=x+V^k$ satisfies

$$Y^{k,x}=\Phi(W^{k,x})$$

where $W^{k,x}=x+W^k$ and $\Phi:\Omega\to\Omega$ is the bicontinuous injective mapping given for all $\omega\in\Omega$ by $\Phi(\omega)=(\Phi_t(\omega))_{0\le t\le 1}$ where

$$\Phi_t(\omega)=\omega_0+\alpha_t(\omega_t-\omega_0),\quad 0\le t\le 1.$$

As for (26), the LD rate function of $(Y^{k,x})_{k\ge 1}$ is $C^x=C+\iota_{\{X_0=x\}}$ where

$$C=C_Z\circ\Phi^{-1}$$

and C_Z is given at (24). It is easy to see that for all $\phi\in\Omega$, $\Phi^{-1}(\phi)=(\Phi_t^{-1}(\phi))_{0\le t\le 1}$ where for all $0<t\le 1$, $\Phi_t^{-1}(\phi)=\phi_0+\beta_t(\phi_t-\phi_0)$ with $\beta_t:=\alpha_t^{-1}$. Assuming that β is differentiable on $(0,1]\times\mathbb{R}^d$, we obtain

$$C(\omega)=\begin{cases}\int_{[0,1]}c_Z\big(\partial_t\beta_t(\omega_t-\omega_0)+\nabla\beta_t(\omega_t-\omega_0)\cdot\dot\omega_t\big)\,dt & \text{if } \omega\in\Omega_{\mathrm{ac}}\\ +\infty & \text{otherwise}\end{cases},\quad \omega\in\Omega.$$

For each $x,y\mathbb{R}^d$, (G^{xy}) admits a unique solution γ^{xy} which is given by the equation $\Phi^{-1}(\gamma^{xy})=\sigma^{x,x+\beta_1(y-x)}$ where σ^{xy} is the constant velocity geodesic, see (25). That is

$$\gamma_t^{xy}=x+\alpha_t(t\beta_1(y-x)),\quad 0\le t\le 1.$$

The corresponding static cost function c which is specified by (17), i.e.

$$c(x,y)=C(\gamma^{xy}),\quad x,y\in\mathbb{R}^d.$$

In the case when α doesn't depend on t, we see that for all $x,y\in\mathbb{R}^d$,

$$c(x,y)=C(\gamma^{xy})=C_Z(\sigma^{x,x+\beta(y-x)})=c_Z(\alpha^{-1}(y-x)),$$

which is (26), but the velocity of the geodesic path

$$\dot{\gamma}_t^{xy} = \nabla\alpha\big(t\alpha^{-1}(y-x)\big)\cdot\alpha^{-1}(y-x)$$

is not constant in general.

5. Proofs of the results of Section 3

The main technical result is Proposition 3.4.

It will be used at several places that $X_0, X_1 : \Omega \to \mathcal{X}$ are continuous. This is clear when $\Omega = C([0,1],\mathcal{X})$ since it is furnished with the topology of uniform convergence. In the general case where $\Omega = D([0,1],\mathcal{X})$ is furnished with the Skorokhod topology, it is known that X_t is not continuous in general. But, it remains true that X_0 and X_1 are continuous, due to the specific form of the metric at the endpoints.

Proof of Proposition 3.4. The space $\mathrm{C}_b(\Omega)$ is furnished with the supremum norm $\|f\| = \sup_\Omega |f|$, $f\in \mathrm{C}_b(\Omega)$ and $\mathrm{C}_b(\Omega)'$ is its topological dual space. Let $\mathrm{M}_b(\Omega)$, resp. $\mathrm{M}_b^+(\Omega)$ denote the spaces of all bounded, resp. bounded positive, Borel measures on Ω. Of course, $\mathrm{M}_b(\Omega)\subset \mathrm{C}_b(\Omega)'$ with the identification $\langle f,Q\rangle_{\mathrm{C}_b(\Omega),\mathrm{C}_b(\Omega)'} = \int_\Omega f\,dQ$ for any $Q\in \mathrm{M}_b(\Omega)$. We write $\langle f,Q\rangle := \langle f,Q\rangle_{\mathrm{C}_b(\Omega),\mathrm{C}_b(\Omega)'}$ for simplicity.
Dropping the superscript k for a moment, we have $(R^x\in \mathrm{P}(\Omega); x\in\mathcal{X})$ a measurable kernel and $R^{\mu_0} := \int_\mathcal{X} R^x(\cdot)\,\mu_0(dx)$ where $\mu_0\in\mathrm{P}(\mathcal{X})$ is the initial law.

Lemma 5.1. *For all $Q\in \mathrm{C}_b(\Omega)'$,*

$$H(Q|R^{\mu_0}) + \iota_{\{Q\in\mathrm{P}(\Omega):Q_0=\mu_0\}} = \sup_{f\in\mathrm{C}_b(\Omega)}\left\{\langle f,Q\rangle - \int_\mathcal{X}\log\langle e^f,R^x\rangle\,\mu_0(dx)\right\}.$$

This identity should be compared with the well-known variational representation of the relative entropy

$$H(Q|R) + \iota_{\mathrm{P}(\Omega)}(Q) = \sup_{f\in\mathrm{C}_b(\Omega)}\left\{\langle f,Q\rangle - \log\langle e^f,R\rangle\right\},\quad Q\in\mathrm{M}_b(\Omega) \tag{29}$$

which holds for any reference *probability* measure $R\in\mathrm{P}(\Omega)$ on any polish space Ω.

Proof. Denote

$$\Theta(f) = \int_\mathcal{X}\log\langle e^f,R^x\rangle\,\mu_0(dx)\in(-\infty,\infty],\quad f\in\mathrm{C}_b(\Omega)$$

Its convex conjugates with respect to the duality $\langle \mathrm{C}_b(\Omega),\mathrm{C}_b(\Omega)'\rangle$ is given for all $Q\in \mathrm{C}_b(\Omega)'$ by $\Theta^*(Q) := \sup_{f\in\mathrm{C}_b(\Omega)}\{\langle f,Q\rangle - \Theta(f)\}$. It will be proved at Lemma 5.2 that any $Q\in\mathrm{C}_b(\Omega)'$ such that $\Theta^*(Q)<\infty$ is in $\mathrm{M}_b^+(\Omega)$. Let us admit this for a while, and take $Q\in\mathrm{M}_b^+(\Omega)$ such that $\Theta^*(Q)<\infty$. Taking $f=\phi(X_0)$ with $\phi\in\mathrm{C}_b(\mathcal{X})$, we see that $\sup_{\phi\in\mathrm{C}_b(\mathcal{X})}\int_\mathcal{X}\phi\,d(Q_0-\mu_0)\le\Theta^*(Q)$. Hence, $\Theta^*(Q)<\infty$ implies that $Q_0=\mu_0$. This shows us that if $\Theta^*(Q)<\infty$, then Q is a probability measure with $Q_0=\mu_0$.
It remains to prove that for such a $Q\in\mathrm{P}(\Omega)$, we have $\Theta^*(Q)=H(Q|R^{\mu_0})$. Since Ω is a polish space, any $Q\in\mathrm{P}(\Omega)$ such that $Q_0=\mu_0$ disintegrates as

$$Q(\cdot) = \int_\mathcal{X} Q^x(\cdot)\,\mu_0(dx)$$

where $(Q^x; x\in\mathcal{X})$ is a measurable kernel of probability measures. We see that

$$\Theta^*(Q) = \sup_{f\in\mathrm{C}_b(\Omega)}\int_\mathcal{X}[\langle f,Q^x\rangle - \log\langle e^f,R^x\rangle]\,\mu_0(dx).$$

We obtain

$$\begin{aligned}\Theta^*(Q) &\leq \int_{\mathcal{X}} \sup_{f\in \mathrm{C}_b(\Omega)} [\langle f, Q^x\rangle - \log\langle e^f, R^x\rangle]\, \mu_0(dx)\\ &\overset{\checkmark}{=} \int_{\mathcal{X}} H(Q^x|R^x)\,\mu_0(dx)\\ &= H(Q|R^{\mu_0})\end{aligned}$$

where (29) is used at the marked equality and last equality follows from the tensorization property (10). Note that $x\mapsto H(Q^x|R^x)$ is measurable. Indeed, $(Q,R)\mapsto H(Q|R)$ is lower semicontinuous being the supremum of continuous functions, see (29). Hence, it is Borel measurable. On the other hand, $x\mapsto R^x$ and $x\mapsto Q^x$ are also measurable, being the disintegration kernels of Borel measures on a polish space.

Let us prove the converse inequality. By Jensen's inequality: $\int_{\mathcal{X}} \log\langle e^f, R^x\rangle\,\mu_0(dx) \leq \log \int_{\mathcal{X}} \langle e^f, R^x\rangle\,\mu_0(dx) = \log\langle e^f, R^{\mu_0}\rangle$, so that

$$\Theta^*(Q) \geq \sup_{f\in \mathrm{C}_b(\Omega)} \left\{ \int_\Omega f\,dQ - \log \int_\Omega e^f\,dR^{\mu_0}\right\} = H(Q|R^{\mu_0})$$

where the equality is (29) again. This completes the proof of the lemma. □

During the proof of Lemma 5.1, we used a result which is stated at Lemma 5.2. Denote also

$$\Lambda(f) := \int_{\mathcal{X}} \sup_{\Omega}\{f - C^x\}\,\mu_0(dx) = \int_{\mathcal{X}} \sup_{\Omega^x}\{f - C\}\,\mu_0(dx), \quad f\in \mathrm{C}_b(\Omega)$$

where $\Omega^x := \{X_0 = x\} \subset \Omega$. It will appear later that the function Λ is the convex conjugate of the Γ-limit $\mathcal{C}$. Its convex conjugate with respect to the duality $\langle \mathrm{C}_b(\Omega), \mathrm{C}_b(\Omega)'\rangle$ is given for all $Q\in \mathrm{C}_b(\Omega)'$ by $\Lambda^*(Q) := \sup_{f\in\mathrm{C}_b(\Omega)}\{\langle f, Q\rangle - \Lambda(f)\}$.

Lemma 5.2.

(1) $\{\Theta^* < \infty\} \subset \mathrm{M}_b^+(\Omega)$;
(2) $\{\Lambda^* < \infty\} \subset \mathrm{M}_b^+(\Omega)$.

Proof. For a positive element $Q\in \mathrm{C}_b(\Omega)'$ to be in $\mathrm{M}_b(\Omega)$, it necessary and sufficient that it is σ-additive. That is, for all *decreasing* sequence $(f_n)_{n\geq 1}$ in $\mathrm{C}_b(\Omega)$ such that $\lim_{n\to\infty} f_n = 0$ *pointwise*, we have $\lim_{n\to\infty}\langle f_n, Q\rangle = 0$.

• Proof of (1). Let us prove that $\{\Theta^* < \infty\} \subset \mathrm{M}_b^+(\Omega)$.
Let us show that $Q\geq 0$ if $\Theta^*(Q) < \infty$. Let $f\in \mathrm{C}_b(\Omega)$ be such that $f\geq 0$. As $\Theta(af)\leq 0$ for all $a\leq 0$,

$$\begin{aligned}\Theta^*(Q) &\geq \sup_{a\leq 0}\{a\langle f, Q\rangle - \Theta(af)\}\\ &\geq \sup_{a\leq 0}\{a\langle f, Q\rangle\}\\ &= \begin{cases} 0, & \text{if } \langle f, Q\rangle \geq 0\\ +\infty, & \text{otherwise.}\end{cases}\end{aligned}$$

Therefore, if $\Theta^*(Q) < \infty$, $\langle f, Q\rangle \geq 0$ for all $f\geq 0$, which is the desired result.

Let us take a decreasing sequence $(f_n)_{n\geq 1}$ in $\mathrm{C}_b(\Omega)$ which converges pointwise to zero. By the dominated convergence theorem, we have

$$\lim_{n\to\infty} \Theta(af_n) = 0, \quad \forall a\geq 0.$$

It follows that for all $Q \in \mathrm{C}_b(\Omega)'$,

$$\begin{aligned}\Theta^*(Q) &\geq \sup_{a\geq 0} \limsup_{n\to\infty} \{a\langle f_n, Q\rangle - \Theta(af_n)\} \\ &\geq \sup_{a\geq 0} \left(\limsup_{n\to\infty} a\langle f_n, Q\rangle - \lim_{n\to\infty} \Theta(af_n) \right) \\ &= \sup_{a\geq 0} a \limsup_{n\to\infty} \langle f_n, Q\rangle \\ &= \begin{cases} 0 & \text{if } \limsup_{n\to\infty}\langle f_n, Q\rangle \leq 0 \\ +\infty & \text{otherwise.} \end{cases}\end{aligned}$$

Therefore, if $\Theta^*(Q) < \infty$, we have $\limsup_{n\to\infty}\langle f_n, Q\rangle \leq 0$. Since we have just seen that $Q \geq 0$, we have the desired result.

• Proof of (2). Let us prove that $\{\Lambda^* < \infty\} \subset \mathrm{M}_b^+(\Omega)$.
Let us show that $Q \geq 0$ if $\Lambda^*(Q) < \infty$. Let $f \in \mathrm{C}_b(\Omega)$ be such that $f \geq 0$. As $\inf C = 0$, $\Lambda(af) \leq 0$ for all $a \leq 0$, and we conclude as at item (1).

Let us take a decreasing sequence $(f_n)_{n\geq 1}$ in $\mathrm{C}_b(\Omega)$ which converges pointwise to zero. By Lemma 5.3 below, for all $x \in \mathcal{X}$, $\big(\sup_\Omega\{f_n - C^x\}\big)_{n\geq 1}$ is a decreasing sequence and $\lim_{n\to\infty} \sup_\Omega\{f_n - C^x\} = 0$. As $|\sup_\Omega\{f_n - C^x\}| \leq \sup_\Omega |f_1| < \infty$ for all n and x, we can apply the dominated convergence theorem to obtain that $\lim_{n\to\infty} \Lambda(af_n) = 0$, for all $a \geq 0$ and we conclude as at item (1).

Finally, one must be careful with the measurability of $x \in \mathcal{X} \mapsto u_n(x) := \inf_\Omega\{C^x - f_n\} = -\sup_\Omega\{f_n - C^x\} \in \mathbb{R}$. Since Ω and $\mathcal{X}$ are assumed to be polish, we can apply a general result by Beiglböck and Schachermayer [BS09, Lemmas 3.7, 3.8] which tells us that for each $n \geq 1$ and each Borel probability measure μ on $\mathcal{X}$, there exists a Borel measurable function $\tilde{u}_n$ on $\mathcal{X}$ such that $\tilde{u}_n \leq u_n$ and $\tilde{u}_n(x) = u_n(x)$ for μ-a.e. $x \in \mathcal{X}$. □

During the proof of the previous lemma we have invoked the following result.

Lemma 5.3. *Let J be a coercive $[0,\infty]$-valued function on Ω and $(f_n)_{n\geq 1}$ a decreasing sequence of continuous bounded functions on Ω which converges pointwise to some bounded upper semicontinuous function f. Then, $(\sup_\Omega\{f_n - J\})_{n\geq 1}$ is a decreasing sequence and*

$$\lim_{n\to\infty} \sup_\Omega \{f_n - J\} = \sup_\Omega \{f - J\}.$$

Proof. Changing sign and denoting $g_n = J - f_n$, $g = J - f$, we want to prove that $\lim_{n\to\infty} \inf_\Omega g_n = \inf_\Omega g$.

We see that $(g_n)_{n\geq 1}$ is an increasing sequence of lower semicontinuous functions. It follows by the Proposition 5.4 of [Mas93] that it is a Γ-convergent sequence and

$$\Gamma\text{-}\lim_{n\to\infty} g_n = \lim_{n\to\infty} g_n = g. \tag{30}$$

Let us admit for a while that there exists some compact set K which satisfies

$$\inf_\Omega g_n = \inf_K g_n \tag{31}$$

for all n. This and the convergence (30) allow to apply Theorem 7.4 of [Mas93] to obtain $\lim_{n\to\infty} \inf_\Omega g_n = \inf_\Omega \Gamma\text{-}\lim_{n\to\infty} g_n = \inf_\Omega g$ which is the desired result.

It remains to check that (31) is true. Let $\omega_* \in \Omega$ be such that $J(\omega_*) < \infty$ (if $J \equiv +\infty$, there is nothing to prove). Then, $\inf_\Omega g_n \leq g_n(\omega_*) = J(\omega_*) - f_n(\omega_*) \leq J(\omega_*) - f(\omega_*) \leq J(\omega_*) - \inf_\Omega f$. On the other hand, for all n, $f_n \leq f_1 \leq A := \sup f_1$. Let $B := A + 1 +$

$J(\omega_*) - \inf_\Omega f$. For all ω such that $J(\omega) > B$, we have $g_n(\omega) > B - \sup_\Omega f_n \geq B - A \geq J(\omega_*) - \inf_\Omega f + 1$. We have just seen that for all n,

$$\inf_\Omega g_n \leq J(\omega_*) - \inf_\Omega f \qquad \text{and} \qquad \inf_{\omega; J(\omega)>B} g_n(\omega) \geq J(\omega_*) - \inf_\Omega f + 1.$$

This proves (31) with the compact level set $K = \{J \leq B\}$ and completes the proof of the lemma. □

Recall that for all $Q \in \mathrm{M}_b(\Omega)$, $\mathcal{C}^{k,\mu_0}(Q) = \frac{1}{k} H(Q|R^{k,\mu_0}) + \iota_{\{Q_0=\mu_0\}}$. With Lemma 5.1, we see that

$$\mathcal{C}^{k,\mu_0}(Q) = \Lambda_k^*(Q), \quad Q \in \mathrm{C}_b(\Omega)' \tag{32}$$

where Λ_k^* is the convex conjugate of

$$\Lambda_k(f) = \int_{\mathcal{X}} \frac{1}{k} \log\langle e^{kf}, R^{k,x}\rangle\, \mu_0(dx), \quad f \in \mathrm{C}_b(\Omega)$$

with respect to the duality $\langle \mathrm{C}_b(\Omega), \mathrm{C}_b(\Omega)'\rangle$. The keystone of the proof of Proposition 3.4 is the following consequence of the Laplace-Varadhan principle.

Lemma 5.4. *Under the assumptions of Proposition 3.4, for all $f \in \mathrm{C}_b(\Omega)$, we have*

(1) $\lim_{k\to\infty} \Lambda_k(f) = \Lambda(f)$;

(2) $\sup_{k\geq 1} |\Lambda_k(f)| \leq \|f\|, \quad |\Lambda(f)| \leq \|f\| := \sup_\Omega |f|$.

The functions Λ_k and Λ are convex.

Proof. Our assumptions allow us to apply the Laplace-Varadhan principle, see Theorem A.3. It tells us that for each $x \in \mathcal{X}$,

$$\lim_{k\to\infty} \frac{1}{k} \log\langle e^{kf}, R^{k,x}\rangle = \sup_\Omega \{f - C^x\}.$$

On the other hand, it is clear that for each $k \geq 1$, $|\frac{1}{k}\log\langle e^{kf}, R^{k,x}\rangle| \leq \|f\|$. Passing to the limit, we also get $|\sup_\Omega\{f - C^x\}| \leq \|f\|$. Now by the Lebesgue dominated convergence theorem, we obtain the statements (1) and (2).
Note that $x \mapsto \sup_\Omega\{f - C^x\}$ is measurable as a pointwise limit of measurable functions. It is standard to prove with Hölder's inequality that $f \mapsto \frac{1}{k}\log\langle e^{kf}, R^{k,x}\rangle$ is convex. It follows that Λ_k and Λ are also convex. □

We are in position to apply Corollary 6.4. Let us equip $\mathrm{C}_b(\Omega)'$ with the $*$-weak topology $\sigma(\mathrm{C}_b(\Omega)', \mathrm{C}_b(\Omega))$. By Corollary 6.4, we have

$$\Gamma\text{-}\lim_{k\to\infty} \Lambda_k^* = \Lambda^* \tag{33}$$

where

$$\Lambda^*(Q) = \sup_{f\in \mathrm{C}_b(\Omega)} \left\{ \langle f, Q\rangle_{\mathrm{C}_b(\Omega),\mathrm{C}_b(\Omega)'} - \int_{\mathcal{X}} \sup_\Omega\{f - C^x\}\, \mu_0(dx) \right\}, \quad Q \in \mathrm{C}_b(\Omega)'.$$

This limit still holds in $\mathrm{M}_b(\Omega) \subset \mathrm{C}_b(\Omega)'$, by Lemma 5.2.

Because of (32), (33) and Lemma 5.2, to complete the proof of Proposition 3.4, it remains to prove the subsequent lemma.

Lemma 5.5. *Let C be a lower semicontinuous $[0,\infty]$-valued function on the polish space Ω. Denote $C^x = C + \iota_{\{\theta=x\}}$ for each $x \in \mathcal{X}$, where $\theta : \Omega \to \mathcal{X}$ is a continuous application with its values in polish space $\mathcal{X}$. Take $\mu \in \mathrm{P}(\mathcal{X})$ and suppose that*

$$\inf_\Omega C^x = 0$$

for μ-almost every $x \in \mathcal{X}$. Then, we have

$$\sup_{f \in \mathrm{C}_b(\Omega)} \left\{ \langle f, Q \rangle - \int_{\mathcal{X}} \sup_{\Omega} \{f - C^x\} \, \mu(dx) \right\} = \int_{\Omega} C \, dQ + \iota_{\{Q \in \mathrm{P}(\Omega): \theta_\# Q = \mu\}}, \quad Q \in \mathrm{M}_b(\Omega). \tag{34}$$

Note that since $C \geq 0$ and C is measurable, the integral $\int_\Omega C \, dP$ makes sense in $[0, \infty]$ for any $P \in \mathrm{P}(\Omega)$.
As the function C of Proposition 3.4 is such that C^x is a LD rate function for all $x \in \mathcal{X}$, it satisfies the assumption $\inf_\Omega C^x = 0$ for μ-almost every $x \in \mathcal{X}$.

Proof. Let us first check that if $Q \in \mathrm{M}_b(\Omega)$ satisfies $\Lambda^*(Q) < \infty$, then $Q \in \mathrm{P}(\Omega)$ and $\theta_\# Q = \mu \in \mathrm{P}(\mathcal{X})$. We already know by Lemma 5.2 that $Q \in \mathrm{M}_b^+(\Omega)$. Choosing $f = \phi \circ \theta$ with $\phi \in \mathrm{C}_b(\mathcal{X})$, since $\inf_\Omega C^x = 0$, we see that $\sup_\Omega \{\phi \circ \theta - C^x\} = \phi(x)$. Hence, $\sup_{\phi \in \mathrm{C}_b(\mathcal{X})} \int_{\mathcal{X}} \phi \, d(\theta_\# Q - \mu) \leq \Lambda^*(Q) < \infty$ which implies that $\theta_\# Q = \mu$. This proves the desired result.

It remains to prove the equality for a fixed $P \in \mathrm{P}(\Omega)$ which satisfies $\theta_\# P = \mu$. Because Ω and $\mathcal{X}$ are polish spaces, we know that P disintegrates as follows: $P(\cdot) = \int_{\mathcal{X}} P^x(\cdot) \, \mu(dx)$, with $x \in \mathcal{X} \mapsto P^x(\cdot) := P(\cdot \mid \theta = x) \in \mathrm{P}(\Omega)$ Borel measurable. For any $f \in \mathrm{C}_b(\Omega)$,

$$\begin{aligned}
\langle f, P \rangle - \int_{\mathcal{X}} \sup_{\Omega} \{f - C^x\} \, \mu(dx) &= \int_{\mathcal{X}} [\langle f, P^x \rangle - \sup_{\Omega} \{f - C^x\}] \, \mu(dx) \\
&= \int_{\mathcal{X}} [\langle C^x, P^x \rangle + \langle f - C^x - \sup_{\Omega} \{f - C^x\}, P^x \rangle] \, \mu(dx) \\
&\leq \int_{\mathcal{X}} \langle C^x, P^x \rangle \, \mu(dx) \\
&= \int_{\Omega} C \, dP.
\end{aligned}$$

Optimizing, we obtain

$$\sup_{f \in \mathrm{C}_b(\Omega)} \left\{ \langle f, P \rangle - \int_{\mathcal{X}} \sup_{\Omega} \{f - C^x\} \, \mu(dx) \right\} \leq \int_{\Omega} C \, dP.$$

If C is in $\mathrm{C}_b(\Omega)$, the case of equality is obtained with $f = C$, P-a.e. and in this situation we see that the identity (34) is valid. This will be invoked very soon.

In the general case, C is only assumed to be lower semicontinuous. By means of the Moreau-Yosida approximation procedure which is implementable since Ω is a metric space, one can build an increasing sequence $(C_n)_{n \geq 1}$ of functions in $\mathrm{C}_b(\Omega)$ which converges pointwise to C. Therefore,

$$\begin{aligned}
&\sup_{f\in \mathrm{C}_b(\Omega)}\left\{\langle f,P\rangle - \int_{\mathcal{X}} \sup_{\Omega}\{f-C^x\}\,\mu(dx)\right\}\\
\le\ & \int_\Omega C\,dP\\
\overset{\text{(i)}}{=}\ & \sup_{n\ge 1}\int_\Omega C_n\,dP\\
\overset{\text{(ii)}}{=}\ & \sup_{n\ge 1}\sup_{f\in \mathrm{C}_b(\Omega)}\left\{\langle f,P\rangle - \int_{\mathcal{X}} \sup_{\Omega}\{f-C_n^x\}\,\mu(dx)\right\}\\
=\ & \sup_{f\in \mathrm{C}_b(\Omega)}\left\{\langle f,P\rangle + \sup_{n\ge 1}\int_{\mathcal{X}} \inf_{\Omega}\{C_n^x-f\}\,\mu(dx)\right\}\\
\overset{\text{(iii)}}{\le}\ & \sup_{f\in \mathrm{C}_b(\Omega)}\left\{\langle f,P\rangle + \int_{\mathcal{X}} \inf_{\Omega}\{C^x-f\}\,\mu(dx)\right\}\\
=\ & \sup_{f\in \mathrm{C}_b(\Omega)}\left\{\langle f,P\rangle - \int_{\mathcal{X}} \sup_{\Omega}\{f-C^x\}\,\mu(dx)\right\},
\end{aligned}$$

which proves the desired identity (34).
Equality (i) follows from the monotone convergence theorem. Since C_n stands in $\mathrm{C}_b(\Omega)$, equality (ii) is valid (this has been proved a few lines earlier) and the inequality (iii) is a direct consequence of $C_n \le C$ for all $n\ge 1$. Note that $x\in\mathcal{X}\mapsto \inf_\Omega\{C_n^x-f\}\in\mathbb{R}$ is upper semicontinuous and it is a fortiori Borel measurable. □

Proofs of the remaining results. The keystone of the proofs of the remaining results is Proposition 3.4.

Proposition 3.1. Proposition 3.1 is a particular case of Proposition 3.4. Indeed, choosing $\Omega=\mathcal{X}^2$ which can be interpreted as the space of all $\mathcal{X}$-valued paths on the two-point time interval $\{0,1\}$, and taking $C(\omega)=c(\omega_0,\omega_1)$ where c is assumed to be lower semicontinuous, with $\omega=(x,y)$ we see that $C^x(x',y)=c(x,y)+\iota_{\{x'=x\}}$ for all $x,x',y\in\mathcal{X}$. The assumption that $c(x,\cdot)$ is coercive on $\mathcal{X}$ is equivalent to the coerciveness of C^x on $\mathcal{X}^2$.

Corollary 3.5 and Theorem 3.6. With Proposition 3.4 in hand, Corollary 3.5 and Theorem 3.6 are immediate consequences of Theorem 7.1 and of the equi-coerciveness with respect to the $*$-weak topology $\sigma(\mathrm{P}(\Omega),\mathrm{C}_b(\Omega))$ of $\{\mathcal{C},\mathcal{C}^k;k\ge 1\}$. This equi-coerciveness follows from Corollary 6.4 and Lemma 5.4. The uniqueness of the solution to (S^k) follows from the strict convexity of the relative entropy.

Corollary 3.2 and Theorem 3.3. Similarly, once we have Proposition 3.1 in hand, Corollary 3.2 and Theorem 3.3 are immediate consequences of Theorem 7.1 and of the equi-coerciveness with respect to the $*$-weak topology $\sigma(\mathrm{P}(\mathcal{X}),\mathrm{C}_b(\mathcal{X}))$ of $\{\mathcal{C}_{01},\mathcal{C}_{01}^k;k\ge 1\}$. This equi-coerciveness follows from the fact that the set of all probability measures $\pi\in\mathrm{P}(\mathcal{X}^2)$ such that $\pi_0=\mu_0$ and $\pi_1\in\{\mu_1,\mu_1^k;k\ge 1\}$ is relatively compact since $\lim_{k\to\infty}\mu_1^k=\mu_1$; a consequence of Prokhorov's theorem in a polish space.
Again, the uniqueness of the solution to (S_{01}^k) follows from the strict convexity of the relative entropy.
Note that, when C and c are linked by (17), one can also derive the equi-coerciveness of $\{\mathcal{C}_{01},\mathcal{C}_{01}^k;k\ge 1\}$ from the equi-coerciveness of $\{\mathcal{C},\mathcal{C}^k;k\ge 1\}$, as in the proof of Theorem 7.1.

Theorem 3.7. The proof of Theorem 3.7 relies upon the subsequent lemma.

Lemma 5.6. *Under the assumptions of Proposition 3.4, the function* c *defined by* (17) *is lower semicontinuous and*

$$\inf\left\{\int_\Omega C\,dP; P\in \mathrm{P}(\Omega), P_{01}=\pi\right\} = \int_{\mathcal{X}} c\,d\pi \in [0,\infty],$$

for all $\pi\in \mathrm{P}(\mathcal{X}^2)$.

Proof. Let us define the function

$$\Psi(\pi) := \inf\left\{\int_\Omega C\,dP; P\in \mathrm{P}(\Omega): P_{01}=\pi\right\},\quad \pi\in \mathrm{P}(\mathcal{X}^2).$$

As C is assumed to be lower semicontinuous on Ω, Ψ satisfies the Kantorovich type dual equality:

$$\Psi(\pi) = \sup_{f\in\mathcal{F}}\int_{\mathcal{X}^2} f\,d\pi,\quad \pi\in \mathrm{P}(\mathcal{X}^2) \tag{35}$$

where $\mathcal{F} := \{f\in \mathrm{C}_b(\mathcal{X}^2); f(X_0,X_1)\le C\}$. For a proof of (35), one can rewrite mutatis mutandis the proof of the Kantorovich dual equality. See for instance [Léo, Thm 3.2] and note that this result takes into account cost functions which may take infinite values as in the present case.
This shows that Ψ is a lower semicontinuous function on $\mathrm{P}(\mathcal{X}^2)$, being the supremum of continuous functions. Define the function

$$\psi(x,y) := \Psi(\delta_{(x,y)}),\quad x,y\in\mathcal{X}.$$

We deduce immediately from the lower semicontinuity of Ψ that ψ is lower semicontinuous on $\mathcal{X}^2$. Hence it is Borel measurable. Since it is $[0,\infty]$-valued, the integral $\int_{\mathcal{X}}\psi\,d\pi$ is meaningful for all $\pi\in \mathrm{P}(\mathcal{X}^2)$. We are going to prove that

$$\Psi(\pi) = \int_{\mathcal{X}^2}\psi\,d\pi,\quad \pi\in \mathrm{P}(\mathcal{X}^2). \tag{36}$$

For any $\pi\in \mathrm{P}(\mathcal{X}^2)$, we obtain

$$\begin{aligned}
\Psi(\pi) &= \inf\left\{\int_{\mathcal{X}^2}\left(\int_\Omega C\,dP^{xy}\right)\pi(dxdy); P\in \mathrm{P}(\Omega)\right\}\\
&\ge \int_{\mathcal{X}^2}\inf\left\{\int_\Omega C\,dP; P\in \mathrm{P}(\Omega): P_{01}=\delta_{(x,y)}\right\}\pi(dxdy)\\
&= \int_{\mathcal{X}^2}\psi\,d\pi.
\end{aligned}$$

Let us show the converse inequality. With (35), we see that for each $f\in\mathcal{F}$ and all $(x,y)\in\mathcal{X}^2$, $\psi(x,y)=\Psi(\delta_{(x,y)})\ge\int_{\mathcal{X}^2} f\,d\delta_{(x,y)} = f(x,y)$. That is $f\le\psi$, for all $f\in\mathcal{F}$. Therefore, $\Psi(\pi)=\sup_{f\in\mathcal{F}}\int_{\mathcal{X}^2} f\,d\pi\le\int_{\mathcal{X}^2}\psi\,d\pi$, completing the proof of (36).

It remains to establish that $\psi=c$. With (35), we get $\psi=\sup\mathcal{F}$. But it is clear that $f\in\mathcal{F}$ if and only if for all $x,y\in\mathcal{X}$, $f(x,y)\le\inf\{C(\omega);\omega\in\Omega:\omega_0=x,\omega_1=y\}:=c(x,y)$. Hence, ψ is the upper envelope of the set of all functions $f\in \mathrm{C}_b(\mathcal{X}^2)$ such that $f\le c$. In other words ψ is the lower semicontinuous envelope $\mathrm{ls}\,c$ of c. Finally, for all $x,y\in\mathcal{X}$, $\mathrm{ls}\,c(x,y)=\psi(x,y)=\inf\left\{\int_\Omega C\,dP^{xy}; P\in \mathrm{P}(\Omega)\right\}\ge c(x,y)\ge \mathrm{ls}\,c(x,y)$. This implies the desired result: $\psi=\mathrm{ls}\,c=c$. ☐

With this result at hand, let us prove Theorem 3.7. It is assumed that for any $x \in \mathcal{X}$, $(R^{k,x})_{k\geq 1}$ satisfies the LDP with scale k and rate function C^x. We have $\rho^{k,x} = (X_1)_\# R^{k,x}$. Taking the continuous image $X_1 : \Omega \to \mathcal{X}$, by means of the contraction principle, see Theorem A.2 at the Appendix, we obtain that for any $x \in \mathcal{X}$, $(\rho^{k,x})_{k\geq 1}$ satisfies the LDP with scale k and rate function

$$y \in \mathcal{X} \mapsto \inf\{C^x(\omega); \omega \in \Omega : \omega_1 = y\} = c(x, y) \in [0, \infty].$$

• Proof of (1). The first assertion of Theorem 3.7 follows from the lower semicontinuity of c which was obtained at Lemma 5.6. Indeed, this shows that the assumptions of Proposition 3.1 are fulfilled. The identity $\inf(\mathrm{MK_{dyn}}) = \inf(\mathrm{MK})$ is a direct consequence of Lemma 5.6.

• Proof of (2).The second assertion follows from $\inf(\mathrm{MK_{dyn}}) = \inf(\mathrm{MK})$, the convergence of the minimal values which was obtained at item (1) together with the strict convexity (for the uniqueness) and the coerciveness (for the existence) of the relative entropy. The relation between $\widehat{P}^k$ and $\widehat{\pi}^k$ is (13).

• Proof of (3). Let us first show that $P \mapsto \langle C, P\rangle + \iota_{\{P_0=\mu_0\}}$ is coercive on $\mathrm{P}(\Omega)$. By (34) and the proof of Corollary 6.4, we see that its sublevel sets are relatively compact. Since C is lower semicontinuous, it is also lower semicontinuous. Therefore, it is coercive and so is $P \mapsto \langle C, P\rangle + \iota_{\{P_0=\mu_0, P_1=\mu_1\}}$. In particular, if $\inf(\mathrm{MK_{dyn}}) < \infty$, the set of minimizers of $(\mathrm{MK_{dyn}})$ is a nonempty convex compact subset of $\mathrm{P}(\Omega)$.
Let $\widehat{P}$ be such a minimizer. It disintegrates as $\widehat{P}(\cdot) = \int_{\mathcal{X}^2} \widehat{P}^{xy}(\cdot)\, \widehat{P}_{01}(dxdy)$ and with Lemma 5.6, we see that $\widehat{P}_{01} := \widehat{\pi}$ is a solution to (MK). Moreover, $\int_{\mathcal{X}^2} c\, d\widehat{\pi} = \psi(\widehat{\pi}) = \int_\Omega C\, d\widehat{P} = \int_{\mathcal{X}^2} \left(\int_\Omega C\, d\widehat{P}^{xy}\right) \widehat{\pi}(dxdy)$ and $\int_\Omega C\, d\widehat{P}^{xy} \geq c(x, y)$ for $\widehat{\pi}$-a.e. (x, y). Hence, $\int_\Omega C\, d\widehat{P}^{xy} = c(x, y)$ for $\widehat{\pi}$-a.e. (x, y). This means that for $\widehat{\pi}$-a.e. (x, y), $\widehat{P}^{xy}(\Gamma^{xy}) = 1$ where $\Gamma^{xy} := \{\omega \in \Omega; \omega_0 = x, \omega_1 = y, C(\omega) = c(x, y)\}$ is the set of all geodesic paths from x to y. Remark that Γ^{xy} is a compact subset of Ω which is nonempty as soon as $c(x, y) < \infty$. In particular, it is a Borel measurable subset. Following the cases of equality, it is clear that if, conversely $P \in \mathrm{P}(\Omega)$ satisfies $P^{xy}(\Gamma^{xy}) = 1$ for P_{01}-a.e. (x, y), then P minimizes $Q \mapsto \int_\Omega C\, dQ$ subject to $Q_{01} = P_{01}$. This completes the proof of the theorem.

6. Γ-convergence of convex functions on a weakly compact space

A typical result about the Γ-convergence of a sequence of convex functions $(f_k)_{k\geq 1}$ is: If the sequence of the convex conjugates $(f_k^*)_{k\geq 1}$ converges in some sense, then $(f_k)_{k\geq 1}$ Γ-converges. Known results of this type are usually stated in separable reflexive Banach spaces. For instance Corollary 3.13 of H. Attouch's monograph [Att84] is

Theorem 6.1. *Let X be a separable reflexive Banach space and $(f_k)_{k\geq 1}$ a sequence of closed convex functions from X into $(-\infty, +\infty]$ satisfying the equi-coerciveness assumption: $f_k(x) \geq \alpha(\|x\|)$ for all $x \in X$ and $k \geq 1$ with $\lim_{r\to+\infty} \alpha(r)/r = +\infty$. Then, the following statements are equivalent*

(1) $f = \mathrm{seq}X_w\text{-}\Gamma\text{-}\lim_{k\to\infty} f_k$
(2) $f^* = X_s^*\text{-}\Gamma\text{-}\lim_{n\to\infty} f_k^*$
(3) $\forall y \in X^*,\ f^*(y) = \lim_{k\to\infty} f_k^*(y)$

where X^* is the dual space of X, $\mathrm{seq}X_w$ *refers to the weak sequential convergence in X and X_s^* to the strong convergence in X^*.*

Going beyond the reflexivity assumption is not so easy, as can be seen in Beer's monograph [Bee93].

In some applications in probability, the reflexive Banach space setting is not as natural as it is for the usual applications of variational convergence to PDEs. For instance when dealing with random measures on $\mathcal{X}$, the narrow topology $\sigma(\mathrm{P}(\mathcal{X}), C_b(\mathcal{X}))$ doesn't fit the above framework since $C_b(\mathcal{X})$ endowed with the uniform topology may not be separable (unless $\mathcal{X}$ is compact) and is not reflexive.

The next result is an analogue of Theorem 6.1 which agrees with applications for random probability measures. Since we didn't find it in the literature, we give its detailed proof.

Let X and Y be two vector spaces in separating duality. The space X is furnished with the weak topology $\sigma(X, Y)$.

We denote ι_C the indicator function of the subset C of X which is defined by $\iota_C(x) = 0$ if x belongs to C and $\iota_C(x) = +\infty$ otherwise. Its convex conjugate is the support function of $C : \iota_C^*(y) = \sup_{x \in C} \langle x, y \rangle$, $y \in Y$.

Theorem 6.2. *Let $(g_k)_{k \geq 1}$ be a sequence of functions on Y such that*

(a) *for all k, g_k is a real-valued convex function on Y,*
(b) *$(g_k)_{k \geq 1}$ converges pointwise to $g := \lim_{k \to \infty} g_k$,*
(c) *g is real-valued and*
(d) *in restriction to any finite dimensional vector subspace Z of Y, $(g_k)_{k \geq 1}$ Γ-converges to g, i.e. $\Gamma\text{-}\lim_{k \to \infty}(g_k + \iota_Z) = g + \iota_Z$, where ι_Z is the indicator function of Z.*

Denote the convex conjugates on X : $f_k = g_k^*$ and $f = g^*$.

If in addition,

(e) *there exists a $\sigma(X, Y)$-compact set $K \subset X$ such that $\mathrm{dom}\, f_k \subset K$ for all $k \geq 1$ and $\mathrm{dom}\, f \subset K$*

then, $(f_k)_{k \geq 1}$ Γ-converges to f with respect to $\sigma(X, Y)$.

The proof of this theorem is postponed after the two preliminary Lemmas 6.5 and 6.6.

Remark 6.3. By ([Mas93], Proposition 5.12), under the assumption (a), assumption (d) is implied by:

(d') in restriction to any finite dimensional vector subspace Z of Y, $(g_k)_{k \geq 1}$ is equibounded, i.e. for all $y_o \in Z$, there exists $\delta > 0$ such that

$$\sup_{k \geq 1} \sup\{|g_k(y)|; y \in Z, |y - y_o| \leq \delta\} < \infty.$$

A useful consequence of Theorem 6.2 is

Corollary 6.4. *Let $(Y, \|\cdot\|)$ be a normed space and X its topological dual space. Let $(g_k)_{k \geq 1}$ be a sequence of functions on Y such that*

(a) *for all k, g_k is a real-valued convex function on Y,*
(b) *$(g_k)_{k \geq 1}$ converges pointwise to $g := \lim_{k \to \infty} g_k$ and*
(d") *there exists $c > 0$ such that $|g_k(y)| \leq c(1 + \|y\|)$ for all $y \in Y$ and $k \geq 1$.*

Then, $(f_k)_{k \geq 1}$ Γ-converges to f with respect to $\sigma(X, Y)$ where $f_k = g_k^*$ and $f = g^*$. Moreover, there exists a $\sigma(X, Y)$-compact set $K \subset X$ such that $\mathrm{dom}\, f_k \subset K$ for all $k \geq 1$ and $\mathrm{dom}\, f \subset K$.

Proof. Under (b), (d") implies (c). As (d") implies (d'), we have (d) by Remark 6.3. Finally, (d") implies (e) with $K = \{x \in X; \|x\|_* \leq c\}$ where $\|x\|_* = \sup_{y, \|y\| \leq 1} \langle x, y \rangle$ is the dual norm on X. Indeed, suppose that for all $y \in Y$, $g(y) \leq c + c\|y\|$ and take $x \in X$ such that $g^*(x) < +\infty$. As for all y, $\langle x, y \rangle \leq g(y) + g^*(x)$, we get $|\langle x, y \rangle| / \|y\| \leq$

$(g^*(x)+c)/\|y\|+c$. Letting $\|y\|$ tend to infinity gives $\|x\|_*\leq c$ which is the announced result.
The conclusion follows from Theorem 6.2. □

Lemma 6.5. *Let $f:X\to(-\infty,+\infty]$ be a lower semicontinuous convex function such that* $\operatorname{dom} f$ *is included in a compact set. Let V be a closed convex subset of X.*
Then, if V satisfies

$$V\cap\operatorname{dom} f\neq\emptyset \quad \textit{or} \quad V\cap\operatorname{cl}\operatorname{dom} f=\emptyset, \tag{37}$$

we have

$$\inf_{x\in V} f(x)=-\inf_{y\in Y}(f^*(y)+\iota_V^*(-y))\in(-\infty,\infty] \tag{38}$$

and if V doesn't satisfy (37), we have

$$\inf_{x\in W} f(x)=-\inf_{y\in Y}(f^*(y)+\iota_W^*(-y))=+\infty \tag{39}$$

for all closed convex set W such that $W\subset\operatorname{int} V$.

Proof. The proof is divided in two parts. We first consider the case where $V\cap\operatorname{dom} f\neq\emptyset$, then the case where $V\cap\operatorname{cl}\operatorname{dom} f=\emptyset$.

• *The case where $V\cap\operatorname{dom} f\neq\emptyset$.* As V is a nonempty closed convex set, its indicator function ι_V is a closed convex function so that its biconjugate satisfies $\iota_V^{**}=\iota_V$, i.e. $\iota_V(x)=\sup_{y\in Y}\{\langle x,y\rangle-\iota_V^*(y)\}$ for all $x\in X$. Consequently,

$$\inf_{x\in V} f(x)=\inf_{x\in X}\sup_{y\in Y}\{f(x)+\langle x,y\rangle-\iota_V^*(y)\}.$$

One wishes to invert $\inf_{x\in X}$ and $\sup_{y\in Y}$ by means of the following standard inf-sup theorem (see [Eke74] for instance). We have $\inf_{x\in X}\sup_{y\in Y}F(x,y)=\sup_{y\in Y}\inf_{x\in X}F(x,y)$ provided that $\inf_{x\in X}\sup_{y\in Y}F(x,y)\neq\pm\infty$ and

- $\operatorname{dom} F$ is a product of convex sets,
- $x\mapsto F(x,y)$ is convex and lower semicontinuous for all y,
- there exists y_o such that $x\mapsto F(x,y_o)$ is coercive and
- $y\mapsto F(x,y)$ is concave for all x.

Our assumptions on f allow us to apply this result with $F(x,y)=f(x)+\langle x,y\rangle-\iota_V^*(y)$. Note that

$$\inf_{x\in X} f(x)>-\infty \tag{40}$$

since f doesn't take the value $-\infty$ and is assumed to be lower semicontinuous on a compact set. Therefore, if $\inf_{x\in V} f(x)<+\infty$, we have

$$\inf_{x\in V} f(x)=\sup_{y\in Y}\inf_{x\in X}\{f(x)+\langle x,y\rangle-\iota_V^*(y)\}=-\inf_{y\in Y}\{f^*(y)+\iota_V^*(-y)\}.$$

• *The case where $V\cap\operatorname{cl}\operatorname{dom} f=\emptyset$.* As $\operatorname{cl}\operatorname{dom} f$ is assumed to be compact, by Hahn-Banach theorem $\operatorname{cl}\operatorname{dom} f$ and V are strictly separated: there exists $y_o\in Y$ such that $\iota_V^*(y_o)=\sup_{x\in V}\langle x,y_o\rangle<\inf_{\operatorname{cl}\operatorname{dom} f}\langle x,y_o\rangle\leq\inf_{x\in\operatorname{dom} f}\langle x,y_o\rangle$. Hence,

$$\inf_{x\in\operatorname{dom} f}\{\langle x,y_o\rangle-\iota_V^*(y_o)\}>0 \tag{41}$$

and

$$\begin{aligned} -\inf_{y\in Y}(f^*(y)+\iota_V^*(-y)) &= \sup_{y\in Y}\inf_{x\in X}\{f(x)+\langle x,y\rangle-\iota_V(y)\}\\ &= \sup_{y\in Y}\inf_{x\in\operatorname{dom} f}\{f(x)+\langle x,y\rangle-\iota_V(y)\}\\ &\geq \inf_{x\in X}f(x)+\sup_{a>0}\inf_{x\in\operatorname{dom} f}\{\langle x,ay_o\rangle-\iota_V^*(ay_o)\}\\ &= \inf_{x\in X}f(x)+\sup_{a>0}a\inf_{x\in\operatorname{dom} f}\{\langle x,y_o\rangle-\iota_V^*(y_o)\}\\ &= +\infty \end{aligned}$$

where the last equality follows from (40) and (41). This proves that (39) holds with $W = V$.

• Finally, if (37) isn't satisfied, taking W such that $W\subset \operatorname{int} V$ insures the strict separation of W and $\operatorname{cl}\operatorname{dom} f$ as above. □

Lemma 6.6. *Let the $\sigma(X,Y)$-closed convex neighbourhood V of the origin be defined by*

$$V=\{x\in X;\langle y_i,x\rangle\leq 1, 1\leq i\leq n\} \tag{42}$$

with $n\geq 1$ and $y_1,\dots,y_n\in Y$. Its support function ι_V^* is $[0,\infty]$-valued, coercive and its domain is the finite dimensional convex cone spanned by $\{y_1,\dots,y_n\}$. More precisely, its level sets are $\{\iota_V^*\leq b\}=b\operatorname{cv}\{y_1,\dots,y_n\}$ for each $b\geq 0$ where $\operatorname{cv}\{y_1,\dots,y_n\}$ is the convex hull of $\{y_1,\dots,y_n\}$.

Proof. The closed convex set V is the polar set of $N=\{y_1,\dots,y_n\}:V=N^\circ$. Let $x_1\in V$ and $x_o\in E:=\cap_{1\leq i\leq n}\ker y_i$. Then, $\langle y_i,x_1+x_o\rangle=\langle y_i,x_1\rangle\leq 1$. Hence, $x_1+x_o\in V$. Considering the factor space X/E, we now work within a finite dimensional vector space whose algebraic dual space is spanned by $\{y_1,\dots,y_n\}$.

We still denote by X and Y these finite dimensional spaces. We are allowed to apply the finite dimension results which are proved in the book [RW98] by Rockafellar and Wets. In particular, one knows that if C is a closed convex set in Y, then the gauge function $\gamma_C(y):=\inf\{\lambda\geq 0; y\in\lambda C\}, y\in Y$ is the support function of its polar set $C^\circ=\{x\in X;\langle x,y\rangle\leq 1,\forall y\in C\}$. This means that $\gamma_C=\iota_{C^\circ}^*$ (see [RW98], Example 11.19).

As $V=(N^{\circ\circ})^\circ$ and $N^{\circ\circ}$ is the closed convex hull of N, i.e. $N^{\circ\circ}=\operatorname{cv}(N)$: the convex hull of N, we get $V=\operatorname{cv}(N)^\circ$ and

$$\iota_V^*=\gamma_{\operatorname{cv}(N)}.$$

In particular, for all real b, $\iota_V^*(y)\leq b\Leftrightarrow\gamma_{\operatorname{cv}(N)}(y)\leq b\Leftrightarrow y\in b\operatorname{cv}(N)$. It follows that the effective domain of ι_V^* is the convex cone spanned by $y_1,\dots,y_n$ and ι_V^* is coercive. □

Proof of Theorem 6.2. Let $\mathcal{N}(x_o)$ denote the set of all the neighbourhoods of $x_o\in X$. We want to prove that $\Gamma\text{-}\lim_{k\to\infty}f_k(x_o):=\sup_{U\in\mathcal{N}(x_o)}\lim_{k\to\infty}\inf_{x\in U}f_k(x)=f(x_o)$. Since f is lower semicontinuous, we have $f(x_o)=\sup_{U\in\mathcal{N}(x_o)}\inf_{x\in U}f(x)$, so that it is enough to show that for all $U\in\mathcal{N}(x_o)$, there exists $V\in\mathcal{N}(x_o)$ such that $V\subset U$ and

$$\lim_{k\to\infty}\inf_{x\in V}f_k(x)=\inf_{x\in V}f(x). \tag{43}$$

The topology $\sigma(X,Y)$ is such that $\mathcal{N}(x_o)$ admits the sets

$$V=\{x\in X;|\langle y_i,x-x_o\rangle|\leq 1, i\leq n\}$$

as a base where $(y_1,\dots,y_n), n\geq 1$ describes the collection of all the finite families of vectors in Y. By Lemma 6.5, there exists such a $V\subset U$ which satisfies

$$\inf_{x\in V} f_k(x) = -\inf_{y\in Y} h_k(y) \text{ for all } k\geq 1 \text{ and } \inf_{x\in V} f(x) = -\inf_{y\in Y} h(y)$$

where we denote $h_k(y)=g_k(y)+\iota_V^*(-y)$ and $h(y)=g(y)+\iota_V^*(-y)$, $y\in Y$.

Let Z denote the vector space spanned by $(y_1,\dots,y_n)$ and h_k^Z, h^Z the restrictions to Z of h_k and h. For all $y\in Y$, we have

$$\iota_V^*(-y) = -\langle x_o, y\rangle + \iota_{V-x_o}^*(-y) \tag{44}$$

and by Lemma 6.6, the effective domain of ι_V^* is Z. Therefore, to prove (43) it remains to show that

$$\lim_{k\to\infty}\inf_{y\in Y} h_k^Z(y) = \inf_{y\in Y} h^Z(y). \tag{45}$$

By assumptions (b) and (d), (h_k^Z) Γ-converges and pointwise converges to h^Z. Note that this Γ-convergence is a consequence of the lower semicontinuity of the convex conjugate ι_V^* and Proposition 6.25 of [Mas93].

Because of assumptions (a) and (c), (h_k^Z) is also a sequence of finite convex functions which converges pointwise to the finite function h^Z. By ([Roc97], Theorem 10.8), (h_k^Z) converges to h^Z uniformly on any compact subset of Z and h^Z is convex.

We now consider three cases for x_o.

The case where $x_o\in\operatorname{dom} f$. We already know that (h_k^Z) Γ-converges to h^Z. To prove (45), it remains to check that the sequence (h_k^Z) is equicoercive (see [Mas93], ??).

For all $y\in Y$, $g(y)-\langle x_o,y\rangle\geq -f(x_o)$ and (44) imply $h^Z(y)\geq -f(x_o)+\iota_{V-x_o}^*(-y)$. Since, $-f(x_o)>-\infty$ and $\iota_{V-x_o}^*$ is coercive (Lemma 6.6), we obtain that h^Z is coercive. As (h_k^Z) converges to h^Z uniformly on any compact subset of Z, it follows that (h_k^Z) is equicoercive. This proves (45).

The case where $x_o\in\operatorname{cl}\operatorname{dom} f$. In this case, there exists $x_o'\in\operatorname{dom} f$ such that $V' = x_o'+(V-x_o)/2=\{x\in X; |\langle 2y_i, x-x_o'\rangle|\leq 1, i\leq k\}\in\mathcal{N}(x_o')$ satisfies $x_o\in V'\subset V\subset U$. One deduces from the previous case, that (45) holds true with V' instead of V.

The case where $x_o\notin\operatorname{cl}\operatorname{dom} f$. As (h_k^Z) Γ-converges to h^Z, by ([Bee93], Proposition 1.3.5) we have $\limsup_{n\to\infty}\inf_{y\in Y} h_k^Z(y)\leq\inf_{y\in Y} h^Z(y)$. As $x_o\notin\operatorname{cl}\operatorname{dom} f$, for any small enough $V\in\mathcal{N}(x_o)$, $\inf_{y\in Y} h^Z(y) = -\inf_{x\in V} f(x) = -\infty$. Therefore, $\lim_{k\to\infty}\inf_{y\in Y} h_k^Z(y) = \inf_{y\in Y} h(y) = -\infty$ which is (45).

This completes the proof of Theorem 6.2. □

7. Γ-convergence of minimization problems under constraints

As the subsequent theorem demonstrates, the notion of Γ-convergence is well-designed for minimization problems. Let $(f_k)_{k\geq 1}$ be a Γ-converging sequence of $(-\infty,\infty]$-valued functions on a metric space X. Let us denote its limit

$$\Gamma\text{-}\lim_{k\to\infty} f_k = f.$$

Let $\theta: X\to Y$ be a continuous function with values in another metric space Y. Assume that for each $k\geq 1$, f_k is coercive and also that the sequence $(f_k)_{k\geq 1}$ is *equi-coercive*, i.e. for all $a\geq 0$, $\bigcup_{k\geq 1}\{f_k\leq a\}$ is relatively compact in X.

Theorem 7.1. *Under the above assumptions, the sequence of functions* $(\psi_k)_{k\geq 1}$ *on* Y *which is defined by*

$$\psi_k(y) := \inf\{f_k(x); x\in X:\theta(x)=y\},\quad y\in Y, k\geq 1$$

Γ-*converges to*

$$\psi(y) := \inf\{f(x); x \in X : \theta(x) = y\}, \quad y \in Y.$$

In particular, for any $y^* \in Y$*, there exists a sequence* $(y_k^*)_{k\geq 1}$ *in* Y *such that* $\lim_{k\to\infty} y_k^* = y^*$ *and* $\lim_{k\to\infty} \inf\{f_k(x); x \in X : \theta(x) = y_k^*\} = \inf\{f(x); x \in X : \theta(x) = y^*\} \in (-\infty, \infty]$.

Moreover, if y^* *satisfies* $\inf\{f(x); x \in X : \theta(x) = y^*\} < \infty$*, then for each* $k \geq 1$*, the minimization problem*

$$f_k(x) \to \min; \quad x \in X : \theta(x) = y_k^*$$

admits at least a minimizer $\hat{x}_k \in X$*. Any sequence* $(\hat{x}_k)_{k\geq 1}$ *of such minimizers admits at least one limit point and any such limit point is a solution to the minimization problem*

$$f(x) \to \min; \quad x \in X : \theta(x) = y^*.$$

The proof of this result which is based on Lemmas 7.2 and 7.3 below, is postponed after the proofs of these lemmas.

Let Y be another metric space. We consider a Γ-convergent sequence $(g_k)_{k\geq 1}$ of $[0, \infty]$-valued functions on $X \times Y$ with

$$\Gamma\text{-}\lim_{k\to\infty} g_k = g.$$

Let us define for each $k \geq 1$ and $y \in Y$,

$$\psi_k(y) := \inf_{x\in X} g_k(x, y), \quad \psi(y) := \inf_{x\in X} g(x, y).$$

Assume that for each $k \geq 1$, g_k is coercive and also that the sequence $(g_k)_{k\geq 1}$ is equi-coercive on $X \times Y$.

Lemma 7.2. *Under the above assumptions on* $(g_k)_{k\geq 1}$*,* $\Gamma\text{-}\lim_{k\to\infty} \psi_k = \psi$ *in* Y*.*

Proof. Let us fix $y^* \in Y$ and prove that $\Gamma\text{-}\lim_{k\to\infty} \psi_k(y^*) = \psi(y^*)$. Since g_k is assumed to be coercive, for every $y \in Y$, there exists $\hat{x}_{k,y} \in X$ such that $\psi_k(y) = g_k(\hat{x}_{k,y}, y)$.

Lower bound. Let $(y_k)_{k\geq 1}$ be any converging sequence in Y such that $\lim_{k\to\infty} y_k = y^*$. we want to show that

$$\liminf_{k\to\infty} \psi_k(y_k) \geq \psi(y^*).$$

Suppose that $\liminf_{k\to\infty} \psi_k(y_k) < \infty$, since otherwise there is nothing to prove. We denote $x_k^* = \hat{x}_{k,y_k}$. Then,

$$\liminf_{k\to\infty} \psi_k(y_k) = \liminf_{k\to\infty} g_k(x_k^*, y_k) \overset{(a)}{=} \lim_{m\to\infty} g_m(x_m^*, y_m) \overset{(b)}{=} \lim_{n\to\infty} g_n(x_n^*, y_n)$$

where the index m at equality (a) means that we have extracted a subsequence such that $\liminf_{k\to\infty} = \lim_{n\to\infty}$. At equality (b), once again a new subsequence is extracted in order that $(x_n^*)_{n\geq 1}$ converges to some limit point x^* :

$$\lim_{n\to\infty} x_n^* = x^*.$$

The existence of a limit point x^* is insured by our assumptions that $\liminf_{k\to\infty} \psi_k(y_k) < \infty$ and $\bigcup_{k\geq 1}\{g_k \leq a\}$ is relatively compact for all $a \geq 0$. Now, by filling the holes in an approriate way one can construct a sequence $(\tilde{x}_k)_{k\geq 1}$ which admits $(x_n)_{n\geq 1}$ as a subsequence and such that $\lim_{k\to\infty} \tilde{x}_k = x^*$. It follows that

$$\liminf_{k\to\infty} \psi_k(y_k) = \lim_{n\to\infty} g_n(x_n^*, y_n) \geq \liminf_{k\to\infty} g_k(\tilde{x}_k, y_k) \overset{\checkmark}{\geq} g(x^*, y^*) \geq \psi(y^*)$$

which is the desired result. At the marked inequality, we have used our assumption that $\Gamma\text{-}\lim_{k\to\infty} g_k = f$.

Recovery sequence. Under our assumptions, the Γ-limit g is coercive on $X \times Y$, see [Mas93, Thm 7.8]. It follows that $g(\cdot, y^*)$ is also coercive and that there exists $\hat{x} \in \operatorname{argmin} g(\cdot, y^*)$. Let $(x_k, y_k)_{k\geq 1}$ be a recovery sequence of $(g_k)_{k\geq 1}$ at $(\hat{x}, y^*)$. This means that $\lim_{k\to\infty}(x_k, y_k) = (\hat{x}, y^*)$ and $\liminf_{k\to\infty} g_k(x_k, y_k) \leq g(\hat{x}, y^*) = \psi(y^*)$. We see eventually that

$$\liminf_{k\to\infty} \psi_k(y_k) \leq \liminf_{k\to\infty} g_k(x_k, y_k) \leq \psi(y^*),$$

which is the desired estimate. □

Let us fix $y^* \in Y$. By Lemma 7.2, there exists a sequence $(y^*_k)_{k\geq 1}$ such that

$$\lim_{k\to\infty} y^*_k = y^*, \quad \lim_{k\to\infty} \psi_k(y^*_k) = \psi(y^*). \tag{46}$$

Let us define

$$\varphi_k(x) := g_k(x, y^*_k), \quad \varphi(x) := g(x, y^*), \quad x \in X$$

for all $k \geq 1$. Since g_k is coercive, φ_k is also coercive. In particular, if $\psi(y^*) = \inf_X \varphi < \infty$, its minimum value $\psi_k(y^*_k) = \inf_X \varphi_k$ is finite and therefore attained at some $\hat{x}_k \in X$.

Lemma 7.3. *In addition to the assumptions of Lemma 7.2, suppose that* $\inf_X \varphi < \infty$*. For each* k*, let* $\hat{x}_k$ *be a minimizer of* φ_k*. Then the sequence* $(\hat{x}_k)_{k\geq 1}$ *admits limit points in* X *and any limit point is a minimizer of* φ*.*

Proof. We have already noticed that for each k, φ_k is coercive so that it admits one or several minimizers. Since $\lim_{k\to\infty} \inf_X \varphi_k = \inf_X \varphi < \infty$, we see that $\sup_k \inf_X \varphi_k < \infty$. It follows from the assumed relative compactness of $\bigcup_{k\geq 1}\{g_k \leq a\}$ for all $a \geq 0$, that $\bigcup_{k\geq 1} \operatorname{argmin} \varphi_k$ is also relatively compact. Therefore any sequence $(\hat{x}_k)_{k\geq 1}$ of minimizers $\hat{x}_k \in \operatorname{argmin} \varphi_k$ admits at least one limit point.
As $\varphi_k(\hat{x}_k) = \psi_k(y^*_k)$, we see with (46) that

$$\lim_{k\to\infty} \varphi_k(\hat{x}_k) = \inf \varphi.$$

On the other hand, let $\hat{x}$ be any limit point of $(\hat{x}_k)_{k\geq 1}$. There exists a subsequence (indexed by m with an abuse of notation) such that $\lim_{m\to\infty} \hat{x}_m = \hat{x}$. Because of the assumed Γ-limit: $\Gamma\text{-}\lim_{k\to\infty} g_k = g$, we obtain

$$\varphi(\hat{x}) := g(\hat{x}, y^*) \leq \liminf_{m\to\infty} g_m(\hat{x}_m, y^*_m) := \liminf_{m\to\infty} \varphi_m(\hat{x}_m) = \lim_{k\to\infty} \varphi_k(\hat{x}_k) = \inf \varphi.$$

It follows that $\hat{x}$ is a minimizer of φ. □

Proof of Theorem 7.1. Consider the functions

$$g_k(x, y) := f_k(x) + \iota_{\{y=\theta(x)\}}, \quad (x, y) \in X \times Y,$$

for each $k \geq 1$ and

$$g(x, y) := f(x) + \iota_{\{y=\theta(x)\}}, \quad (x, y) \in X \times Y.$$

Because of Lemmas 7.2, 7.3 and (46), to complete the proof it is enough to show that

$$\Gamma\text{-}\lim_{k\to\infty} g_k = g \tag{47}$$

together with the coerciveness assumptions of these lemmas.

Let us begin with the coerciveness. Since for each $k \geq 1$, f_k is coercive and θ is continuous, we see that for any large enough a, $\{g_k \leq a\} = \{(x, y) \in X \times Y; x \in \{f_k \leq a\}, y = \theta(x)\}$ is compact, i.e. for each $k \geq 1$, g_k is coercive. As $(f_k)_{k\geq 1}$ is assumed to be equi-coercive, its Γ-limit f is coercive and it follows by the same argument that g is also coercive. We also see that $\bigcup_{k\geq 1}\{g_k \leq a\} = \{(x, y) \in X \times Y; x \in \bigcup_{k\geq 1}\{f_k \leq a\}, y = \theta(x)\}$ is relatively compact, i.e. $(g_k)_{k\geq 1}$ is equi-coercive.

Let us prove that (47) holds true. Let $(x, y) \in X \times Y$ be fixed. We have to prove that:

(i) For any sequence $(x_k, y_k)_{k\geq 1}$ such that $\lim_{k\to\infty}(x_k, y_k) = (x, y)$, $\liminf_{k\to\infty} f_k(x_k) + \iota_{\{y_k=\theta(x_k)\}} \geq f(x) + \iota_{\{y=\theta(x)\}}$.

(ii) There exists a sequence $(\tilde{x}_k, \tilde{y}_k)_{k\geq 1}$ such that $\lim_{k\to\infty}(\tilde{x}_k, \tilde{y}_k) = (x, y)$, and $\liminf_{k\to\infty} f_k(\tilde{x}_k) + \iota_{\{\tilde{y}_k=\theta(\tilde{x}_k)\}} \leq f(x) + \iota_{\{y=\theta(x)\}}$.

Suppose first that $y \neq \theta(x)$. Then (ii) is obvious and due to the continuity of θ, for any sequence $(x_k, y_k)_{k\geq 1}$ such that $\lim_{k\to\infty}(x_k, y_k) = (x, y)$ we have that for all large enough k, $\theta(x_k) \neq y_k$. This proves (i).
Now, suppose that $y = \theta(x)$. Then (i) follows from $\liminf_{k\to\infty} f_k(x_k) + \iota_{\{y_k=\theta(x_k)\}} \geq \liminf_{k\to\infty} f_k(x_k) \geq f(x) = f(x) + \iota_{\{y=\theta(x)\}}$, whenever $\lim_{k\to\infty} x_k = x$. To prove (ii), take a recovering sequence $(\tilde{x}_k)_{k\geq 1}$ for $(f_k)_{k\geq 1}$ at x, i.e. $\liminf_{k\to\infty} f_k(\tilde{x}_k) \leq f(x)$ and put $\tilde{y}_k = \theta(\tilde{x}_k)$, for each $k \geq 1$. By the continuity of θ, $\lim_{k\to\infty} \tilde{y}_k = y$, so that $\lim_{k\to\infty}(\tilde{x}_k, \tilde{y}_k) = (x, y)$. We also have $\liminf_{k\to\infty} f_k(\tilde{x}_k) + \iota_{\{\tilde{y}_k=\theta(\tilde{x}_k)\}} = \liminf_{k\to\infty} f_k(\tilde{x}_k) \leq f(x) = f(x) + \iota_{\{y=\theta(x)\}}$, which proves (ii) and completes the proof of the theorem. □

Appendix A. Large deviations

Large deviation principle. We refer to the monograph by Dembo and Zeitouni [DZ98] for a clear exposition of the subject. Let X be a polish space furnished with its Borel σ-field. One says that the sequence $(\gamma_n)_{n\geq 1}$ of probability measures on X satisfies the large deviation principle (LDP for short) with scale n and rate function I, if for each Borel measurable subset A of X we have

$$-\inf_{x\in \operatorname{int} A} I(x) \overset{\text{(i)}}{\leq} \liminf_{n\to\infty} \frac{1}{n} \log \gamma_n(A) \leq \limsup_{n\to\infty} \frac{1}{n} \log \gamma_n(A) \overset{\text{(ii)}}{\leq} -\inf_{x\in \operatorname{cl} A} I(x) \tag{48}$$

where $\operatorname{int} A$ and $\operatorname{cl} A$ are respectively the topological interior and closure of A in X and the rate function $I : X \to [0, \infty]$ is lower semicontinuous. The inequalities (i) and (ii) are called respectively the *LD lower bound* and *LD upper bound*, where LD is an abbreviation for large deviation. The LDP is the exact statement of what was meant in previous section when writing

$$\gamma_n(A) \underset{n\to\infty}{\asymp} \exp\left(-n \inf_{x\in A} I(x)\right)$$

for "all" $A \subset X$.
It is sometimes too much demanding to have the upper bound $\limsup_{n\to\infty} \frac{1}{n} \log \gamma_n(C) \leq -\inf_{x\in C} I(x)$ for all *closed* sets C. One says that we have the *weak LD upper bound* if

$$\limsup_{n\to\infty} \frac{1}{n} \log \gamma_n(K) \leq -\inf_{x\in K} I(x)$$

for every *compact* subset K of X. In case $(\gamma_n)_{n\geq 1}$ satisfies the LD lower bound (for all open subsets) and the weak LD upper bound (for all compact subsets), one says that $(\gamma_n)_{n\geq 1}$ satisfies the *weak LDP*.

An important instance of large deviation principle is given by the Sanov theorem. Consider a probability measure $R \in \mathrm{P}(\mathcal{X})$ on the polish space $\mathcal{X}$ and furnish $\mathrm{P}(\mathcal{X})$ with the narrow topology $\sigma(\mathrm{P}(\mathcal{X}), \mathrm{C}_b(\mathcal{X}))$ and the corresponding Borel σ-field. Let $Z_1, Z_2, \dots$ be a sequence of independent $\mathcal{X}$-valued random variables with common law R, i.e. $\mathbb{P}(Z_i \in B) = R(B)$ for any Borel measurable subset $B \subset \mathcal{X}$ and any $i \geq 1$. In other words $(Z_1, \dots, Z_n)_\#\mathbb{P} = R^{\otimes n}$ for all $n \geq 1$.

Theorem A.1 (Sanov's theorem)**.** *Under the above assumptions, the empirical measure*

$$L^n := \frac{1}{n}\sum_{i=1}^n \delta_{Z_i} \in \mathrm{P}(\mathcal{X})$$

satisfies the LDP [7] *in* $\mathrm{P}(\mathcal{X})$ *with scale* n. *Its rate function is* $H(\cdot|R) : \mathrm{P}(\mathcal{X}) \to [0,\infty]$, *the relative entropy with respect to the reference probability measure* R.

Here, the LDP stands in $X = \mathrm{P}(\mathcal{X})$ and for each n, $\gamma_n = (L_n)_\#\mathbb{P} \in \mathrm{P}(\mathrm{P}(\mathcal{X}))$. For a proof of this result, see [DZ98, Thm 6.2.10].

Next theorem states that the continuous image of a LDP is still a LDP with the same scale.

Theorem A.2 (Contraction principle)**.** *Let* $(\gamma_n)_{n\geq 1}$ *satisfy the LDP in* X *with scale* n *and rate function* I. *Suppose in addition that* I *is not only lower semicontinuous, but that it is coercive. For any continuous function* $f : X \to Y$ *from* X *to another polish space* Y *furnished with its Borel* σ*-field,*

$$(f_\#\gamma_n)_{n\geq 1}$$

satisfies the LDP in Y *with scale* n *and the rate function*

$$J(y) = \inf\{I(x); x : f(x) = y\}, \quad y \in Y.$$

Moreover, J *is also coercive.*

For a proof, see [DZ98, Thm 4.2.1].

Let us look at an example of application of the contraction principle which is in the mood of this article. Consider an independent sequence of identically distributed random paths, i.e. $(Z_1, \dots, Z_n)_\#\mathbb{P} = R^{\otimes n}$ where the reference probability measure R belongs to $\mathrm{P}(\Omega)$. The empirical measure L^n is a $\mathrm{P}(\Omega)$-valued random variable. Now let f be the marginal projection

$$f(P) = (X_0, X_1)_\# P = (P_0, P_1) \in \mathrm{P}(\mathcal{X}) \times \mathrm{P}(\mathcal{X}), \quad P \in \mathrm{P}(\Omega).$$

It is a continuous function. This is clear when $\Omega = C([0,1], \mathcal{X})$ and it remains true when $\Omega = D([0,1], \mathcal{X})$ ($t = 0, 1$ being the initial and final times, X_0 and X_1 turns out to be Skorokhod-continuous). Using the notation of the previous section, we see that

$$f(L^n) = (L^n_0, L^n_1).$$

By Sanov's theorem, the sequence of empirical measures L^n satisfies the LDP in $\mathrm{P}(\Omega)$ with scale n and rate function $H(\cdot|R)$. Applying the contraction principle with f as above, we see that $(L^n_0, L^n_1)_{n\geq 1}$ satisfies the LDP in $\mathrm{P}(\mathcal{X}) \times \mathrm{P}(\mathcal{X})$ with scale n and rate function

$$J(\mu_0, \mu_1) = \inf\{H(P|R); P \in \mathrm{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\} \in [0,\infty], \quad \mu_0, \mu_1 \in \mathrm{P}(\mathcal{X}),$$

compare (S).

Theorem A.3 (Laplace-Varadhan principle)**.** *Suppose that* $(\gamma_n)_{n\geq 1}$ *satisfy the LDP in* X *with a coercive rate function* $I : X \to [0,\infty]$, *and let* f *be a continuous function on* X. *Assume further that*

$$\lim_{M\to\infty} \liminf_{n\to\infty} \frac{1}{n} \log \int_X e^{nf(x)} \mathbf{1}_{\{f\geq M\}}\, \gamma_n(dx) = -\infty.$$

[7] This is an abuse of definition. The correct statement should be: the sequence $((L^n)_\#\mathbb{P})_{n\geq 1}$ satisfies the LDP.

Then,

$$\lim_{n\to\infty} \frac{1}{n} \log \int_X e^{nf(x)}\,\gamma_n(dx) = \sup_{x\in X}\{f(x) - I(x)\}.$$

For a proof, see [DZ98, Thm 4.3.1].

References

[ADPZ] S. Adams, N. Dirr, M. Peletier, and J. Zimmer. From a large-deviation proinciple to the Wasserstein gradient flow: a new micro-macro passage. Preprint. arXiv:1004.4076.

[Att84] H. Attouch. *Variational convergence for functions and operators*. Pitman Advanced Publishing Program. Pitman, 1984.

[Bee93] G. Beer. *Topologies on closed and closed convex sets*, volume 268 of *Mathematics and Its Applications*. Kluwer Academic Publishers, 1993.

[Bil68] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1968.

[Bre91] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44:375–417, 1991.

[BS09] M. Beiglböck and W. Schachermayer. Duality for Borel measurable cost functions. *Trans. Amer. Math. Soc.*, to appear, 2009.

[CZ08] K.L. Chung and J.C. Zambrini. *Introduction to Random Time and Quantum Randomness*. World Scientific, 2008.

[DG94] D.A. Dawson and J. Gärtner. Multilevel large deviations and interacting diffusions. *Probab. Theory Related Fields*, 98:423–487, 1994.

[DZ98] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications. Second edition.* Applications of Mathematics 38. Springer Verlag, 1998.

[Eke74] I. Ekeland. *La théorie des jeux et ses applications à l'économie mathématique*. Presses Universitaires de France, 1974.

[Föl88] H. Föllmer. *Random fields and diffusion processes, in Ecole d'Eté de Probabilités de Saint-Flour XV-XVII-1985-87*, volume 1362 of *Lecture Notes in Mathematics*. Springer, Berlin, 1988.

[GS] A. Galichon and B. Salanie. Matching with trade-offs: revealed preferences over competing characteristics. Preprint. `http://hal.archives-ouvertes.fr/hal-00473173/en/`.

[Kan42] L.V. Kantorovich. On the translocation of masses. *C. R. (Dokl.) Acad. Sci. URSS*, 37:199–201, 1942.

[Kan48] L.V. Kantorovich. On a problem of Monge (in Russian). *Uspekhi Mat. Nauk.*, 3:225–226, 1948.

[Léo] C. Léonard. A saddle-point approach to the Monge-Kantorovich transport problem. To appear in *ESAIM-COCV*. Published online: DOI: 10.1051/cocv/2010013.

[Mas93] G. Dal Maso. *An Introduction to* Γ*-Convergence*. Progress in Nonlinear Differential Equations and Their Applications 8. Birkhäuser, 1993.

[McC95] R. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80:309–323, 1995.

[Mik04] T. Mikami. Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probab. Theory Relat. Fields*, 129:245–260, 2004.

[MT06] T. Mikami and M. Thieullen. Duality theorem for the stochastic optimal control problem. *Stochastic Process. Appl.*, 116:1815–1835, 2006.

[MT08] T. Mikami and M. Thieullen. Optimal transportation problem by stochastic optimal control. *SIAM J. Control Optim.*, 47(3):1127–1139, 2008.

[Roc97] R.T. Rockafellar. *Convex Analysis*. Princeton landmarks in mathematics. Princeton University Press, Princeton, N.J., 1997. First published in the Princeton Mathematical Series in 1970.

[RT93] L. Rüschendorf and W. Thomsen. Note on the Schrödinger equation and I-projections. *Statist. Probab. Lett.*, 17:369–375, 1993.

[RT98] L. Rüschendorf and W. Thomsen. Closedness of sum spaces and the generalized Schrödinger problem. *Theory Probab. Appl.*, 42(3):483–494, 1998.

[RW98] R.T. Rockafellar and R. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften*. Springer, 1998.

[Sch32] E. Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Ann. Inst. H. Poincaré*, 2:269–310, 1932.

[Vil09] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.

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

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