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, when a fluctuation parameter tends down to zero, of a sequence of entropy minimization problems, the so-called Schrödinger problems. We prove the convergence of the entropic optimal values to the optimal transport cost as the fluctuations decrease to zero, and we also show that the cluster 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; these Γ-convergence results which are interesting in their own right are also proved.

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 optimal values tend to the optimal transport cost as the fluctuations decrease to zero, and also that the cluster 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 2, they are Theorems 2.4, 2.7 and 2.8.

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. We tried as much as possible to formulate the probabilistic notions in terms of

---

Date: Revised version, October 2011.

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

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

analysis and measure theory. A short reminder of the basic definitions and results of large deviation theory is given at the beginning of Appendix A.

The connection between large deviation and optimal transport was discovered by Mikami in the context of the quadratic transport [Mik04]. Although no relative entropy appears in [Mik04] where an optimal control approach is performed, the results of the present paper can be seen as extensions of Mikami's ones.

Some notation. Let us introduce briefly some notation and conventions before presenting our main results. For any topological space Z, we denote by $\mathrm{P}(Z)$ the set of all probability measures on the Borel σ-field of Z.
In the whole paper, $\mathcal{X}$ denotes a state space which is assumed to be Polish, i.e. complete metric and separable. As usual when working with stochastic processes, we are going to consider probability measures on the space

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

of all $\mathcal{X}$-valued paths on the time interval $[0,1]$ which are left continuous and right limited. For any $P \in \mathrm{P}(\Omega)$, i.e. P is the law of a random path $(X_t)_{0\le t\le 1}$, and any $0 \le t \le 1$, we denote by $P_t = (X_t)_\# P \in \mathrm{P}(\mathcal{X})$ the law of the random location X_t at time t, where we write $f_\# m$ for the push-forward of the measure m by the measurable mapping f. In particular, P_0 and P_1 are the laws of the initial and final random locations under P. Also useful is the joint law of the initial and final locations (X_0, X_1): $P_{01} = (X_0, X_1)_\# P \in \mathrm{P}(\mathcal{X}^2)$. The disintegration of P with respect to (X_0, X_1) is $P(d\omega) = \int_{\mathcal{X}^2} P^{xy}(d\omega)\, P_{01}(dxdy)$ where $P^{xy}(\cdot) := P(\cdot \mid X_0 = x, X_1 = y), x, y \in \mathcal{X}$ is the bridge of P between x and y, i.e. P conditioned by the event $(X_0, X_1) = (x, y)$.
When working with the product space

$$\mathcal{X}^2 = \mathcal{X}^{\{0,1\}},$$

one sees the first and second factors $\mathcal{X}$ as the sets of initial and final states respectively. Therefore, denoting the canonical projections $X_0(x,y) := x$ and $X_1(x,y) := y$, $(x,y) \in \mathcal{X}^2$, the first and second marginals of the probability measure $\pi \in \mathrm{P}(\mathcal{X}^2)$ are $\pi_0 := (X_0)_\#\pi \in \mathrm{P}(\mathcal{X})$ and $\pi_1 := (X_1)_\#\pi \in \mathrm{P}(\mathcal{X})$.

Monge-Kantorovich and Schrödinger problems. 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, see [Sch32].

Monge-Kantorovich problem. It is about finding the cheapest 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 $\mu_1 \in \mathrm{P}(\mathcal{X})$. 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

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

It is worth introducing a dynamic version of this static problem. For this purpose, let us take a cost function on the path space $C : \Omega \to [0,\infty]$ and consider the corresponding geodesic problem, for all $x, y \in \mathcal{X}$:

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

The dynamic Monge-Kantorovich problem is

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

The following results about (MK) and ($\mathrm{MK_{dyn}}$) are part of Theorem 2.8 below. We have $\inf\,(\mathrm{MK_{dyn}}) = \inf\,(\mathrm{MK})$, whenever c and C are related by

$$c(x,y) = \inf\,(\mathrm{G}^{xy}) := \inf\{C(\omega); \omega \in \Omega : \omega_0 = x, \omega_1 = y\}, \quad x, y \in \mathcal{X}. \tag{1}$$

This will be assumed once for all.
If $\widehat{P}$ is a minimizer of ($\mathrm{MK_{dyn}}$), then $\widehat{P}_{01}$ minimizes (MK). Moreover, the formula

$$\widehat{P}(d\omega) = \int_{\mathcal{X}^2} \delta_{\Gamma^{xy}}(d\omega)\,\widehat{\pi}(dxdy),$$

is a one-one relation between the minimizers $\widehat{P}$ of ($\mathrm{MK_{dyn}}$) and $\Big(\widehat{\pi}, (\delta_{\Gamma^{xy}})_{x,y\in\mathcal{X}}\Big)$ where $\widehat{\pi}$ solves (MK) and $(x,y) \mapsto \delta_{\Gamma^{xy}} \in \mathrm{P}(\Omega)$ is any measurable mapping such that for $\widehat{\pi}$-almost each (x,y), $\delta_{\Gamma^{xy}}$ concentrates on the set of *geodesic paths*

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

i.e. $\delta_{\Gamma^{xy}}(\Gamma^{xy}) = 1$.
Optimal transport is an active field of research. For a remarkable overview of this topic, see Villani's textbook [Vil09] and the references therein.

Schrödinger problem. In his Saint-Flour lecture notes [Föl88], Föllmer gave a modern translation of a statistical physics problem that Schrödinger addressed in 1932 in connection with the newly born quantum physics. The relative entropy of the probability p with respect to the *reference* probability r is defined by

$$H(p|r) := \begin{cases} \int \log\left(\frac{dp}{dr}\right)\,dp \in [0,\infty], & \text{if } p \ll r \\ \infty, & \text{otherwise.} \end{cases}$$

Take $R \in \mathrm{P}(\Omega)$ which is seen as the law of some reference[1] stochastic process and two probability measures $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$. Schrödinger's dynamic problem is

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

where as for the Monge-Kantorovich problem, μ_0 and $\mu_1 \in \mathrm{P}(\mathcal{X})$ are prescribed. In order to obtain a minimization problem on the same set $\mathrm{P}(\mathcal{X}^2)$ as (MK), let us project ($\mathrm{S_{dyn}}$) from $\mathrm{P}(\Omega)$ onto $\mathrm{P}(\mathcal{X}^2)$ by taking the push-forward $P_{01} = (X_0, X_1)_\# P \in \mathrm{P}(\mathcal{X}^2)$ of any $P \in \mathrm{P}(\Omega)$. The reference probability on $\mathcal{X}^2$ becomes

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

and we call

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

the *Schrödinger problem.*
One can prove that $\inf\,(\mathrm{S_{dyn}}) = \inf\,(\mathrm{S}) \in [0,\infty]$ and since the relative entropy is a strictly convex function, if $\inf\,(\mathrm{S_{dyn}}) < \infty$, then ($\mathrm{S_{dyn}}$) and (S) admit respectively a *unique* minimizer $\widehat{P} \in \mathrm{P}(\Omega)$ and $\widehat{\pi} \in \mathrm{P}(\mathcal{X}^2)$. Moreover, these solutions are related by $\widehat{\pi} = \widehat{P}_{01}$ and

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

[1]The unusual letter R is used to stress the fact that R is the *reference* process.

i.e. $\widehat{P}$ is the $\widehat{\pi}$-mixture of the bridges R^{xy} of the reference process R. Indeed, as noticed in [Föl88], these results follow from the disintegration $P(\cdot) = \int_{\mathcal{X}^2} P^{xy}(\cdot)\, P_{01}(dxdy)$, the tensorization formula $H(P|R) = H(P_{01}|R_{01}) + \int_{\mathcal{X}^2} H(P^{xy}|R^{xy})\, P_{01}(dxdy)$ and the fact that $H(P^{xy}|R^{xy})$ attains uniquely its minimal value zero at $P^{xy} = R^{xy}$.

A link between these problems. In order to reinforce the formal resemblance between the Monge-Kantorovich and the Schrödinger problems, it is necessary to understand how the reference process R encodes the dynamic cost C. This will be done by *replacing R with a sequence $(R^k)_{k\geq 1}$ of reference processes which satisfies a large deviation principle with scale k, as k tends to infinity, and rate function C.* It roughly means that

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

for a large class of measurable subsets $A \subset \Omega$.
Very informally, the most likely paths ω correspond to high values $R^k(d\omega)$ and therefore to low values of $C(\omega)$. As k tends to infinity, the support of R^k shrinks down to a subset of the minimizers of C. Under endpoint constraints, it shouldn't be surprising to meet the family (G^{xy}) of geodesic problems: *The large deviation behaviour of the sequence $(R^k)_{k\geq 1}$ brings us a family of geodesic paths.* More precisely, it can be proved that for R_{01}-almost all (x, y), the cluster points of the sequence $(R^{k,xy})_{k\geq 1}$ of bridges of $(R^k)_{k\geq 1}$ concentrate on the set Γ^{xy} of solutions to the geodesic problem (G^{xy}): $\lim_{k\to\infty} R^{k,xy} = \delta_{\Gamma^{xy}}$.

Instead of ($\mathrm{S}_{\mathrm{dyn}}$) and (S), we introduce the sequence of problems $(\tilde{\mathrm{S}}^k_{\mathrm{dyn}})_{k\geq 1}$ and $(\tilde{\mathrm{S}}^k)_{k\geq 1}$:

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

and

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

where

$$\rho^k := R^k_{01}$$

and $1/k$ is a normalization factor which prevents $H(P|R^k)$ from exploding as k tends to infinity. Note that, as a direct consequence of the contraction principle (Theorem A.1), $(\rho^k)_{k\geq 1}$ satisfies the large deviation principle with scale k and the rate function c on $\mathcal{X}^2$ which is defined by (1).
The aim of this paper is to prove the informal statement

$$\begin{cases} (a) & \lim_{k\to\infty} (\tilde{\mathrm{S}}^k) & = & (\mathrm{MK}), \\ (b) & \lim_{k\to\infty} (\tilde{\mathrm{S}}^k_{\mathrm{dyn}}) & = & (\mathrm{MK}_{\mathrm{dyn}}), \end{cases} \tag{4}$$

having in mind

- $\lim_{k\to\infty} \inf (\tilde{\mathrm{S}}^k) = \inf (\mathrm{MK})$
- and any cluster point of the sequence $(\widehat{\pi}^k)_{k\geq 1}$ of minimizers of $(\tilde{\mathrm{S}}^k)_{k\geq 1}$ is a minimizer of (MK).

And also similar statements for the dynamic problems. It will be seen later that it is necessary to replace $(\tilde{\mathrm{S}}^k)$ and $(\tilde{\mathrm{S}}^k_{\mathrm{dyn}})$ by some modified problems (S^k) and $(\mathrm{S}^k_{\mathrm{dyn}})$, see the statements of Theorems 2.4 and 2.7 respectively. The relevant notion of convergence in (4) is the Γ-convergence, see (6) below and Theorem 2.1 for more details.

Presentation of the results. In this introduction, we briefly present our main results by means of the specific important example which was Schrödinger's original motivation in [Sch32].

Schrödinger's heat bath. Our general results applied to Schrödinger's original example[2] lead us to a restatement in terms of relative entropy of the main result of Mikami's article [Mik04]. For each integer $k \geq 1$, let R^k be the law of a Brownian motion Y^k with diffusion coefficient $1/k$:

$$dY_t^k = \sqrt{1/k}\, dB_t \tag{5}$$

where B is a standard Brownian motion. With these dynamics, it is enough to consider the path space $C([0,1], \mathcal{X}) \subset \Omega$ of all continuous trajectories from $[0,1]$ to $\mathcal{X} = \mathbb{R}^d$. As k tends to infinity, R^k tends to some deterministic dynamics: R^∞ describes a deterministic flow since it only gives mass to constant paths.
The rate function C corresponding to (5) is given by Schilder's theorem (Theorem A.3), 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. With Jensen's inequality, one easily sees that the corresponding static cost defined by (1) is the standard quadratic cost

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

For each $k \geq 1$, the solution to ($\tilde{\mathrm{S}}^k_{\mathrm{dyn}}$) is

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

where $R^{k,xy}$ is the bridge between x and y of the Brownian motion with variance $1/k$, and $\widehat{\pi}^k$ is the minimizer of ($\tilde{\mathrm{S}}^k$) with

$$\rho^k(dy|X_0 = x) = R^k(X_1 \in dy|X_0 = x) = (2\pi/k)^{-d/2} \exp(-k|y-x|^2/2)\, dy, \quad x \in \mathbb{R}^d.$$

If the quadratic cost transport problem on $\mathcal{X} = \mathbb{R}^d$ 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 [Bre91], then $(\widehat{P}^k)_{k\geq 1}$ converges to

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

where for each $x,y \in \mathcal{X}$, γ^{xy} is the constant velocity geodesic path between x and y, $\delta_{\gamma^{xy}}$ is the Dirac measure at γ^{xy} and $\widehat{\pi} \in \mathrm{P}(\mathcal{X}^2)$ is the unique solution to the Monge-Kantorovich transport problem (MK) with a quadratic cost. Observing

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

and comparing the expressions of $\widehat{P}^k$ and $\widehat{P}$, it is worth remarking that

- "$-\lim_{k\to\infty} \frac{1}{k} \log \rho^k(dxdy) \simeq c(x,y)$", which reflects the fact that $(\rho^k(X_1 \in \cdot|X_0 = x))_{k\geq 1}$ obeys the large deviation principle with scale k and rate function $c(x,\cdot)$,
- $\lim_{k\to\infty} R^{k,xy} = \delta_{\gamma^{xy}}$, which is a consequence of (1) and (3),
- $\lim_{k\to\infty} \widehat{\pi}^k = \widehat{\pi}$, which is part of (4)-(a) when $\widehat{\pi}$ is unique.

[2]In [Sch32], the semi-classical limit $k \to \infty$ is not investigated.

Note also that $\widehat{P}$ is a process with a deterministic dynamics and a random initial condition. Moreover, it is the solution to (MK$_{\mathrm{dyn}}$) and its time-marginal flow is given by

$$\widehat{P}_t = \int_{\mathcal{X}^2} \delta_{\gamma_t^{xy}}(\cdot)\, \widehat{\pi}(dxdy) \in \mathrm{P}(\mathcal{X}), \quad 0 \le t \le 1.$$

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. Therefore,

$$\widehat{P}_t^k(\cdot) = \int_{\mathcal{X}^2} R_t^{k,xy}(\cdot)\, \widehat{\pi}^k(dxdy), \quad 0 \le t \le 1$$

is an entropic approximation of the displacement interpolation $(\widehat{P}_t)_{0\le t\le 1}$.

These results hold true with the unmodified minimization problems ($\tilde{\mathrm{S}}^k$) and ($\tilde{\mathrm{S}}^k_{\mathrm{dyn}}$) whenever both μ_0 and μ_1 are assumed to be absolutely continuous with respect to Lebesgue measure. But our general results allow us to remove this restriction, see Theorems 2.4 and 2.7. They are stated in terms of Γ-limits of some modifications (S^k) and ($\mathrm{S}^k_{\mathrm{dyn}}$) of ($\tilde{\mathrm{S}}^k$) and ($\tilde{\mathrm{S}}^k_{\mathrm{dyn}}$): the informal statement (4) must be replaced by

$$\begin{cases} (a) & \Gamma\text{-}\lim_{k\to\infty} (\mathrm{S}^k) & = & (\mathrm{MK}), \\ (b) & \Gamma\text{-}\lim_{k\to\infty} (\mathrm{S}^k_{\mathrm{dyn}}) & = & (\mathrm{MK}_{\mathrm{dyn}}). \end{cases} \tag{6}$$

The definition and some basic results about Γ-convergence are recalled at the beginning of Section 2. In particular, Theorem 2.1 tells us that Γ-convergence is well-suited for obtaining

- $\lim_{k\to\infty} \inf (\mathrm{S}^k) = \inf (\mathrm{MK})$;
- any cluster point of the minimizers $(\widehat{\pi}^k)_{k\ge1}$ of $(\mathrm{S}^k)_{k\ge1}$ is a minimizer of (MK);

and their dynamic analogues.

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

Results. Our main results are Theorems 2.4, 2.7 and 2.8.

- The quadratic cost is an important instance of transport cost, but our results are valid for any cost functions c and C satisfying (1) and (3), plus some coercivity properties. In addition, it is not even necessary that c is derived from a dynamic cost C via (1): the convergence (6)-(a) holds in a more general setting, this is the content of Theorem 2.4. Its dynamic analogue (6)-(b) is stated at Theorem 2.7 and the connection between the dynamic and static minimizers is described at Theorem 2.8.
- Examples of random dynamics $(R^k)_{k\ge1}$ are introduced, leading to several cost functions C and c. They are mainly based on random walks and the Brownian motion so that one can compute explicitly the corresponding cost functions. 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 3.5 for such dynamics based on the Brownian motion.
- The relevant tool for handling convergence of minimization problems is the Γ-convergence theory. We also prove technical results about Γ-convergence which we didn't find in the literature. A typical result about the Γ-convergence of a sequence of convex functions $(f_k)_{k\ge1}$ is: If the sequence of the convex conjugates $(f_k^*)_{k\ge1}$ converges pointwise, then $(f_k)_{k\ge1}$ 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 5.2 is such a result in this weak topology setting.

- Finally, we also prove Theorem 6.1 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.

Organization of the paper. Our main results are stated at Section 2. 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 3 where these costs functions are derived for a large family of random dynamics. The proofs of our main results are done at Section 4. They are partly based on two Γ-convergence results which are stated and proved at Sections 5 and 6.

Literature. We already mentioned Mikami's important contribution [Mik04] which connects large deviation and optimal quadratic transport. Let us note also that an interesting aspect of [Mik04], which is uncovered by the present paper, is that the stochastic control approach together with c-cyclical monotonicity arguments, provide a stochastic proof to the existence of the solution to Monge problem with a quadratic cost without relying on Brenier and McCann's results [Bre91, McC95]. Still using optimal control, Mikami and Thieullen [MT06, MT08] obtained Kantorovich type duality results.

Recently, Adams, Dirr, Peletier and Zimmer [ADPZ11] 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 (MK) is penalized by a relative entropy, leading to the minimization problem

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

where $\rho \in \mathrm{P}(\mathcal{X}^2)$ is a fixed reference probability measure on $\mathcal{X}^2$, for instance $\rho = \mu_0 \otimes \mu_1$. Putting $\rho^k = Z_k^{-1} e^{-kc}\, \rho$ 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 ($\tilde{\mathrm{S}}^k$). 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.

2. Statements of the main results

The statements of our results are in terms of Γ-convergence and large deviation principle. We start introducing their definitions, together with general notation.

Some more notation and conventions. For any topological space Z, $\mathrm{P}(Z)$ is furnished with the usual *narrow topology* $\sigma(\mathrm{P}(Z), \mathrm{C}_b(Z))$ weakened by the space $\mathrm{C}_b(Z)$ of all continuous bounded functions on Z and the corresponding Borel σ-field.

We denote $X = (X_t)_{t\in[0,1]}$ the canonical process which is defined for all $t \in [0, 1]$ by

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

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].

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 its sublevel sets are compact in the Hausdorff space X. A sequence $(f_k)_{k\ge1}$ is said to be *equi-coercive*, if for all $a \ge 0$, $\bigcup_{k\ge1}\{f_k \le a\}$ is relatively compact in X. The convex analysis indicator of a set $A \subset X$ is defined by

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

Large deviation principle. Let X be a Polish space furnished with its Borel σ-field. One says that the sequence $(p_k)_{k\ge1}$ of probability measures on X satisfies the large deviation principle (LDP for short) with scale k and rate function I, if for each Borel measurable subset A of X we have

$$-\inf_{x\in \operatorname{int} A} I(x) \overset{\text{(i)}}{\le} \liminf_{k\to\infty} \frac{1}{k}\log p_k(A) \le \limsup_{k\to\infty} \frac{1}{k}\log p_k(A) \overset{\text{(ii)}}{\le} -\inf_{x\in \operatorname{cl} A} I(x) \tag{7}$$

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.
Important standard LD results are recalled at Appendix A.

Γ-convergence. Recall that if it exists, the Γ-limit of the sequence $(f_k)_{k\ge1}$ 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 $(x_k)_{k\ge1}$ such that $\lim_{k\to\infty} x_k = x$,
$$\liminf_{k\to\infty} f_k(x_k) \ge f(x)$$
(ii) and there exits a sequence $(\tilde{x}_k)_{k\ge1}$ such that $\lim_{k\to\infty} \tilde{x}_k = x$ and
$$\liminf_{k\to\infty} f_k(\tilde{x}_k) \le f(x).$$

Item (i) is called the *lower bound* and the sequence $(\tilde{x}_k)_{k\ge1}$ in item (ii) is the *recovery sequence*. An important standard Γ-convergence result is the following

Theorem 2.1. *Let $(f_k)_{k\ge1}$ be an equi-coercive sequence of $(-\infty,\infty]$-valued functions which Γ-converges to some function f which is not identically equal to $+\infty$.*

1. *Then, f is coercive and $\min_{x\in X} f(x) = \lim_{k\to\infty} \inf_{x\in X} f_k(x)$.*
2. *For each $k \ge 1$, let $\hat{x}_k$ be a minimizer of f_k. If $(\hat{x}_k)_{k\ge1}$ converges to some $\hat{x} \in X$, then $\hat{x}$ is a minimizer of f.*

For a proof, see [DM93, Thm 7.8, Cor 7.20].

The main results. The state space $\mathcal{X}$ is assumed to be Polish.

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

Static results. For each integer $k \ge 1$, 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 $[0,\infty]$-valued functions on $\mathrm{P}(\mathcal{X}^2)$:

$$\left\{ \begin{array}{lcl} \mathcal{C}_{01}^{k,\mu_0}(\pi) & := & \frac{1}{k}H(\pi|\rho^{k,\mu_0}) + \iota_{\{\pi_0=\mu_0\}} \\ \mathcal{C}_{01}^{\mu_0}(\pi) & := & \int_{\mathcal{X}^2} c\, d\pi + \iota_{\{\pi_0=\mu_0\}} \end{array} \right. , \quad \pi \in \mathrm{P}(\mathcal{X}^2).$$

Proposition 2.2. *We assume that for each $x \in \mathcal{X}$, the sequence $((X_1)_\# \rho^{k,x})_{k\ge 1}$ 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

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

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 will follow from Proposition 2.2 and Theorem 6.1.

Corollary 2.3. *Under the assumptions of Proposition 2.2, 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 $(\mu_1^k)_{k\ge 1}$ such that $\lim_{k\to\infty} \mu_1^k = \mu_1$ in $\mathrm{P}(\mathcal{X})$ and $\lim_{k\to\infty} T_{01}^k(\mu_0,\mu_1^k) = T_{01}(\mu_0,\mu_1) \in [0,\infty]$.

The main result of the paper is the following theorem.

Theorem 2.4. *Let us consider the sequence $(\mathrm{S}^k)_{k\ge 1}$ of minimization problems which is given for each $k \ge 1$, by*

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

where $(\mu_1^k)_{k\ge 1}$ is a sequence in $\mathrm{P}(\mathcal{X})$ as in Corollary 2.3.
Under the assumptions of Proposition 2.2, for any $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$ we have $\lim_{k\to\infty} \inf(\mathrm{S}^k) = \inf(\mathrm{MK}) \in [0,\infty]$.
Suppose that in addition $\inf(\mathrm{MK}) < \infty$, *then for each large enough k, (S^k) admits a* unique *solution $\widehat{\pi}^k \in \mathrm{P}(\mathcal{X}^2)$. Moreover, any cluster point of the sequence $(\widehat{\pi}^k)_{k\ge 1}$ 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 $\lim_{k\to\infty} \widehat{\pi}^k = \widehat{\pi}$.

Remark that $\lim_{k\to\infty} \inf(\mathrm{S}^k) = \inf(\mathrm{MK})$ is a restatement of $\lim_{k\to\infty} T_{01}^k(\mu_0,\mu_1^k) = T_{01}(\mu_0,\mu_1)$ in Corollary 2.3.

Dynamic results. Proposition 2.2 and Theorem 2.4 admit a dynamic 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 a stochastic process and $R^{k,x} = R^k(\cdot \mid X_0 = x)$ is its conditional law knowing that $X_0 = x$. For any $\mu_0 \in \mathrm{P}(\mathcal{X})$, denote

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

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 $R_{01}^{k,\mu_0} = (X_0, X_1)_\# R^{k,\mu_0}$ is the joint law of the initial and final positions under R^{k,μ_0}. Let us define the functions

$$\left\{ \begin{array}{lcl} \mathcal{C}^{k,\mu_0}(P) & := & \frac{1}{k} H(P|R^{k,\mu_0}) + \iota_{\{P_0=\mu_0\}} \\ \mathcal{C}^{\mu_0}(P) & := & \int_\Omega C\,dP + \iota_{\{P_0=\mu_0\}} \end{array} \right. , \quad P \in \mathrm{P}(\Omega)$$

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

Proposition 2.5. *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 $[0,\infty]$-valued* coercive *rate function on Ω*

$$C^x = C + \iota_{\{X_0=x\}}$$

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

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

and

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

Corollary 2.6. *Under the assumptions of Proposition 2.5, 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 $\lim_{k\to\infty} T^k(\mu_0, \mu_1^k) = T(\mu_0, \mu_1) \in [0, \infty]$.

Theorem 2.7. *Let us consider the sequence $(\mathrm{S}^k_{\mathrm{dyn}})_{k\geq 1}$ of minimization problems which is given for each $k \geq 1$, by*

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

where $(\mu_1^k)_{k\geq 1}$ *is a sequence in* $\mathrm{P}(\mathcal{X})$ *as in Corollary 2.6.*
Under the assumptions of Proposition 2.5, for any $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$ *we have* $\lim_{k\to\infty} \inf(\mathrm{S}^k_{\mathrm{dyn}}) = \inf(\mathrm{MK}_{\mathrm{dyn}}) \in [0,\infty]$.
Suppose that in addition $\inf(\mathrm{MK}_{\mathrm{dyn}}) < \infty$, *then for each large enough* k, $(\mathrm{S}^k_{\mathrm{dyn}})$ *admits a* unique *solution* $\widehat{P}^k \in \mathrm{P}(\Omega)$. *Moreover, any cluster point of the sequence* $(\widehat{P}^k)_{k\geq 1}$ *in* $\mathrm{P}(\Omega)$ *is a solution to* $(\mathrm{MK}_{\mathrm{dyn}})$.
In particular, if $(\mathrm{MK}_{\mathrm{dyn}})$ *admits a unique solution* $\widehat{P} \in \mathrm{P}(\Omega)$, *then* $\lim_{k\to\infty} \widehat{P}^k = \widehat{P}$.

From the dynamic to a static version. Once we have the dynamic results, the static ones can be derived by means of the continuous projection $P \in \mathrm{P}(\Omega) \mapsto (X_0, X_1)_\# P = P_{01} \in \mathrm{P}(\mathcal{X}^2)$. The LD tool which is behind this transfer is the contraction principle which is recalled at Theorem A.1 below. The connection between the dynamic cost C and the static cost c is Eq. (1).
Since C^x is coercive for all $x \in \mathcal{X}$, there exists at least one solution to the geodesic problem (G^{xy}) which we call 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 2.2 even if it is not derived from a dynamic rate function C via the identity (1). 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. Theorem 2.8 below tells us that it is jointly lower semicontinuous.
The coerciveness of C^x also guarantees that the set Γ^{xy} of all geodesic paths from x to y is a nonempty compact subset of Ω as soon as $c(x,y) < \infty$. In particular, it is a Borel measurable subset.

Theorem 2.8. *Suppose that the assumptions of Proposition 2.5 are satisfied.*

(1) *Then, not only the dynamic results Corollary 2.6 and Theorem 2.7 are satisfied with the cost function* C, *but also the static results Proposition 2.2, Corollary 2.3 and Theorem 2.4 hold with the cost function* c *which is derived from* C *by means of* (1). *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* $k \geq 1$, (S^k) *and* $(\mathrm{S}^k_{\mathrm{dyn}})$ *admit respectively a unique solution* $\widehat{\pi}^k \in \mathrm{P}(\mathcal{X}^2)$ *and* $\widehat{P}^k \in \mathrm{P}(\Omega)$. *Furthermore,* $\widehat{\pi}^k = \widehat{P}^k_{01}$ *and more precisely*

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

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*

$$\widehat{P}^{xy}(\Gamma^{xy}) = 1, \quad \forall (x,y) \in \mathcal{X},\ \widehat{P}_{01}\text{-a.e.} \tag{8}$$

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*

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

Several comments and remarks. • The Schrödinger dynamic problem (S$^k_{\mathrm{dyn}}$) looks like the Monge-Kantorovich dynamic problem (MK$_{\mathrm{dyn}}$) not only because of μ_0 and $\mu_1^k \underset{k\to\infty}{\to} \mu_1$, but also because of some cost of transportation. Indeed, if the random dynamics creates a trend to move in some direction rather than in another one, it costs less to a dynamic particle system whose empirical measure is close to P to end up at some configurations close to μ_1 than other ones. For the particle system picture associated with Schrödinger's problem, see [Sch32, Föl88]. Even if no direction is favoured by a drift vector field, we see that the structure of the random fluctuations which is described by the sequence $(R^k)_{k\ge1}$ encodes some zero-fluctuation cost functions C on Ω (via the LD assumption of Proposition 2.5) and c on $\mathcal{X}^2$ (via (1)).

• Let us comment on the necessity for replacing the entropy minimization problems $(\tilde{\mathrm{S}}^k)_{k\ge1}$ with $(\mathrm{S}^k)_{k\ge1}$ and consequently $(\tilde{\mathrm{S}}^k_{\mathrm{dyn}})_{k\ge1}$ by $(\mathrm{S}^k_{\mathrm{dyn}})_{k\ge1}$. Recall that we introduced

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

For inf $(\tilde{\mathrm{S}}^k)$ to be finite, it is necessary that μ_0 and μ_1 are respectively absolutely continuous with respect to ρ_0^k and ρ_1^k. Considering ρ^{k,μ_0} instead of ρ^k guarantees $\rho_0^{k,\mu_0} = \mu_0$. To see why μ_1 must be approximated by some sequence $(\mu_1^k)_{k\ge1}$, let us consider two examples.

- Take $\mathcal{X} = \mathbb{R}$, $\mu_0 = \mu_1 = \delta_0$ and choose ρ^{k,μ_0} to be the Gaussian distribution Gauss(0,1/k) with mean zero and variance $1/k$. Since μ_1 and ρ^{k,μ_0} are mutually singular, we have for each k, inf $(\tilde{\mathrm{S}}^k) = \inf \emptyset = \infty$. But this sequence $(\rho^k)_{k\ge1}$ corresponds to the quadratic cost (see Schrödinger's heat bath at the introductory section or Example 3.4-(1) below). Therefore (MK) admits the unique solution $\widehat{\pi} = \delta_{(0,0)}$ and it is necessary to approximate $\mu_1 = \delta_0$ by a sequence of absolutely continuous probability measures $(\mu_1^k)_{k\ge1}$ to obtain $\lim_{k\to\infty} \inf\,(\mathrm{S}^k) = \inf\,(\mathrm{MK}) = 0$.
- Take $\mathcal{X} = \mathbb{R}$, $\mu_0 = \delta_0$ and choose ρ^{k,μ_0} as the law of the random variable $Y^k = 2S_k/k - 1$ where S_k has a binomial law $\mathrm{Bin}(k, 1/2)$, i.e. S_k is the number of successes after tossing a fair coin k times. Clearly, Y^k lives in $\Sigma_k := \{2n/k - 1; 0 \le n \le k\} \subset [-1, 1]$ and if μ_1 is absolutely continuous with respect to the Lebesgue measure on $[0, 1]$, $(\tilde{\mathrm{S}}^k)$ has no solution at all and inf $(\tilde{\mathrm{S}}^k) = \infty$ for all $k \ge 1$. It is necessary that μ_1 be approximated by μ_1^k whose support is included in Σ_k. The cost function c corresponding to this sequence $(\rho^k)_{k\ge1}$ is given below at Example 3.4-(2) and it is immediate to see that taking $\mu_0 = \delta_0$ and any μ_1 with a support included in $[-1, 1]$ leads to inf (MK) $\le \log 2$ which clearly implies that (MK) admits a (unique) solution.

• Let us comment on the necessity for replacing the pointwise convergence in (4) with the Γ-convergence in (6). Considering the Brownian dynamics described by (5), one obtains a sequence $(R^k)_{k\ge1}$ of mutually singular reference probability measures on Ω. Indeed, for each k, R^k concentrates of the set of all continuous paths with quadratic variation t/k,

$0 \le t \le 1$. Hence, for any $P \in \mathrm{P}(\Omega)$, we have $H(P|R^k) = \infty$ for all but at most one $k \ge 1$. This rules out the pointwise convergence in (4)-(b) which has to be replaced with $\Gamma\text{-}\lim_{k\to\infty} H(\cdot|R^k) = \langle C, \cdot\rangle_\Omega$ where the recovery sequences $(P^k)_{k\ge1}$ satisfy $\lim_{k\to\infty} P^k = P$ and $H(P^k|R^k) < \infty$ for all large enough k, whenever $\int_\Omega C\, dP < \infty$. The supports of the P^k's must follow the drifting supports of the R^k's. Of course, this is the case in particular for the sequence $(\widehat{P}^k)_{k\ge1}$ of the minimizers of $(\mathrm{S}^k_{\mathrm{dyn}})_{k\ge1}$.

Remarks 2.9 (about related literature). • Proposition 2.5 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 2.5 in the present paper and in [DG94] differ significantly. Dawson-Gärtner's proof is essentially probabilistic while ours is analytic. The strategies 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 approaches and of the weakening of the hypotheses in the present paper, we provide a complete analytic proof of Proposition 2.5 at Section 4.
• Although one may interpret $1/k$ as something 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 $k \to \infty$.
• Schrödinger's paper is the starting point of the Euclidean quantum mechanics which was developed by Zambrini [CZ08].

Remarks 2.10 (about Theorem 2.8). • Formula (8) 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.
• In case of uniqueness as in the last statement of Theorem 2.8, 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 associated with the cost c.
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.
• In case when no uniqueness holds for $(\mathrm{MK}_{\mathrm{dyn}})$, any flow $(X_t)_\#\widehat{P} \in \mathrm{P}(\mathcal{X})$, $t \in [0,1]$ is still a good candidate for being called a displacement interpolation between μ_0 and μ_1. It would be interesting to know if $(\widehat{P}^k)_{k\ge1}$ converges to some "entropic" $\widehat{P}$, selecting a privileged displacement interpolation.

3. From stochastic processes to transport cost functions

We have seen in the introductory section 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 $(R^k)_{k\ge1}$ in $\mathrm{P}(\Omega)$ which give rise to various cost functions c on $\mathcal{X}^2$.

Simple random walks on $\mathbb{R}^d$. Instead of (5), let us consider

$$Y_t^{k,x} = x + W_t^k, \quad 0 \le t \le 1, \tag{9}$$

where for each $k \ge 1$, 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 $(Z_m)_{m\ge1}$ of a random variable Z in $\mathbb{R}^d$. For each integer $k \ge 1$, W^k is the rescaled random walk defined for all $0 \le t \le 1$, by

$$W_t^k = \frac{1}{k} \sum_{j=1}^{\lfloor kt \rfloor} Z_j \tag{10}$$

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

$$c_Z(v) := \sup_{\zeta \in \mathbb{R}^d} \left\{ \zeta \cdot v - \log \int_{\mathbb{R}^d} e^{\zeta \cdot z}\, m_Z(dz) \right\}, \quad v \in \mathbb{R}^d. \tag{11}$$

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 $\operatorname{supp} m_Z$ of the probability measure m_Z. Under the assumption (12) 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 3.1 (Mogulskii's theorem)**.** *Under the assumption*

$$\int_{\mathbb{R}^d} e^{\zeta \cdot z}\, m_Z(dz) < +\infty, \quad \forall \zeta \in \mathbb{R}^d, \tag{12}$$

for each $x \in \mathbb{R}^d$ the sequence $(R^{k,x})_{k\ge1}$ of the laws of $(Y^{k,x})_{k\ge1}$ specified by (9) 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

$$C(\omega) = C_Z(\omega) := \begin{cases} \int_{[0,1]} c_Z(\dot\omega_t)\, dt & \text{if } \omega \in \Omega_{\mathrm{ac}} \\ +\infty & \text{otherwise} \end{cases}, \quad \omega \in \Omega. \tag{13}$$

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

$$\sigma^{xy} : t \in [0,1] \mapsto (1-t)x + ty \in \mathbb{R}^d. \tag{14}$$

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.1, one deduces immediately from Mogulskii's theorem the simplest result of LD theory which is the Cramér theorem.

Corollary 3.2 (A complicated version of Cramér's theorem)**.** *Under the assumption* (12)*, for each $x \in \mathbb{R}^d$ the sequence $(\rho^{k,x})_{k\geq 1}$ of the laws of $(Y_1^{k,x})_{k\geq 1}$ 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 (11)*.*
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 (14) a few lines earlier. Cramér's theorem corresponds to the case when $x = 0$ and only the deviations of $Y_1^{k,x=0} = \frac{1}{k}\sum_{j=1}^k Z_j$ in $\mathbb{R}^d$ are considered.

Theorem 3.3 (Cramér's theorem)**.** *Under the assumption* (12)*, 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* (11)*.*

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

We have just described a general procedure which converts the law $m_Z \in \mathrm{P}(\mathbb{R}^{\mathrm{d}})$ into the cost functions $C = C_Z$ and

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

Here are some examples with explicit computations.

Examples 3.4. 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 $\mathrm{Proba}(Z=+1) = \mathrm{Proba}(Z=-1) = 1/2$, i.e. $m_Z = (\delta_{-1} + \delta_{+1})/2$ leads to
$$c_Z(v) = \begin{cases} [(1+v)\log(1+v) + (1-v)\log(1-v)]/2, & \text{if } -1 < v < +1 \\ \log 2, & \text{if } v \in \{-1,+1\} \\ +\infty, & \text{if } v \notin [-1,+1]. \end{cases}$$
(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 $v < 0$.
(5) 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 $b \in \mathbb{R}^d$.

Although Example 3.4-(3) does not satisfy Assumption (12), Mogulskii's theorem still holds true and c_Z is strictly convex since the log-Laplace transform of m_Z is a steep function.

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 $v = \mathbb{E}Z$.

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 3.4-(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 $|z|_p^p = \sum_{i\le d}|z_i|^p$. This gives $c_Z(v) = |v|_p^p/p(1+o_{|v|\to\infty}(1))$.

Nonlinear transformations. By means of the contraction principle (Theorem A.1), 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 $(V^k)_{k\ge1}$ satisfies a LDP which is not given by Cramér's theorem. Let $(Z_j)_{j\ge1}$ be as above and let α be any continuous mapping on $\mathbb{R}^d$. Consider

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

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

$$c = c_Z \circ \alpha^{-1}. \tag{15}$$

For instance, if Z is a standard normal vector as in Example 3.4-(1), we know that the empirical mean of independent copies of Z : $\frac{1}{k}\sum_{1\le j\le k} Z_j$, 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

$$V^k \overset{\text{Law}}{=} (2k)^{-1/p}|Z|^{2/p-1}Z, \tag{16}$$

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 (15):

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

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

$$Y_t^{k,x} = x + V_t^k, \quad 0\le t\le 1 \tag{17}$$

and $(V^k)_{k\ge1}$ 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 $B = (B_t)_{0\le t\le 1}$ 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) below is already known to us, we recall it for the comfort of the reader.

Examples 3.5.

(1) An important example is given by

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

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 3.4-(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 (16). 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

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

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 (16). But, this time the associated dynamic cost is given for all $\omega \in \Omega$ by

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

Recall that as a definition, a geodesic path from x to y is some $\omega \in \Omega_{\mathrm{ac}}$ which solves the minimization problem (G^{xy}) associated with the cost function C. It is well known that the geodesic paths for Item (1) are the constant velocity paths σ^{xy}, see (14). The geodesic paths for Item (2) are still straight lines but with a time dependent speed (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 (17) with $V^k = W^k$ given by (10). We introduce a generalization which is defined by (17) with

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

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 $0 < t \leq 1$.
For all $x \in \mathbb{R}^d$ and all $k \geq 1$, 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 $\Phi(\omega) = (\Phi_t(\omega))_{0\leq t\leq 1}$ where

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

As for (15), the LD rate function of $(Y^{k,x})_{k\geq 1}$ is $C^x = C + \iota_{\{X_0 = x\}}$ where

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

and C_Z is given at (13). 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 $x, y \in \mathbb{R}^d$, (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 (14). 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 (1), i.e. $c(x,y) = C(\gamma^{xy})$. 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 (15), 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.

4. Proofs of the main results

We give the proofs of the results which were stated at Section 2. The main technical result is Proposition 2.5. See Remark 2.9 about the contribution of Dawson and Gärtner [DG94] to this result.

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 when $\Omega = D([0,1],\mathcal{X})$ is furnished with the Skorokhod topology, it is known that X_t is not continuous in general [Bil68]. But, it remains true that X_0 and X_1 are continuous, due to the specific form of the metric at the endpoints.

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.

Proof of Proposition 2.5. It is organized as follows:

$$\left.\begin{array}{ccccc} & & \text{Lemma } 4.1 & \to & \text{Lemma } 4.2 \quad \text{(a)} \\ \text{Lemma } 4.3 & \to & \text{Lemma } 4.4 & \to & \text{Lemma } 4.5 \quad \text{(b)} \\ & & \text{Theorem A.2} & \to & \text{Lemma } 4.6 \quad \text{(c)} \\ & & & & \text{Corollary } 5.4 \quad \text{(d)} \end{array}\right\} \to \text{Proposition } 2.5$$

where Theorem A.2 is the Laplace-Varadhan principle and Corollary 5.4 is about Γ-convergence with respect to a weak topology and is the main result of Section 5.

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.

(a) *Proof of Lemma 4.2.* We start proving the preliminary Lemma 4.1. 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 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 $\Theta^*(Q) := \sup_{f\in \mathrm{C}_b(\Omega)} \{\langle f, Q\rangle - \Theta(f)\}$.

Lemma 4.1. $\{\Theta^* < \infty\} \subset \mathrm{M}_b^+(\Omega)$.

Proof. 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.

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 $\lim_{n\to\infty}\langle f_n, Q\rangle = 0$. 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. $\square$

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 4.2. *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

$$H(Q|R) + \iota_{\mathrm{P}(\Omega)}(Q) = \sup_{f\in \mathrm{C}_b(\Omega)} \left\{\langle f, Q\rangle - \log\langle e^f, R\rangle\right\}, \quad Q \in \mathrm{M}_b(\Omega) \tag{18}$$

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

Proof. It has been proved at Lemma 4.1 that any $Q \in \mathrm{C}_b(\Omega)'$ such that $\Theta^*(Q) < \infty$ is in $\mathrm{M}_b^+(\Omega)$. Let us take such a Q. Choosing $f = \phi(X_0)$ with $\phi \in \mathrm{C}_b(\mathcal{X})$, we see that $\sup_{\phi \in \mathrm{C}_b(\mathcal{X})} \int_{\mathcal{X}} \phi \, d(Q_0 - \mu_0) \leq \Theta^*(Q)$. Hence, $\Theta^*(Q) < \infty$ implies that $Q_0 = \mu_0$. As Q is positive, we also see that 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

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

and 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 (18) is used at the marked equality and last equality follows from the tensorization property of the entropy. 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 (18). 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 (18) again. This completes the proof of the lemma. □

(b) *Proof of Lemma 4.5.* We start proving the preliminary Lemmas 4.3 and 4.4.

Lemma 4.3. *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 from [DM93, Prop 5.4] that it is a Γ-convergent sequence and

$$\Gamma\text{-}\lim_{n \to \infty} g_n = \lim_{n \to \infty} g_n = g. \tag{19}$$

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

$$\inf_\Omega g_n = \inf_K g_n \tag{20}$$

for all n. This and the convergence (19) allow to apply [DM93, Thm 7.4] 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 (20) is true. Let $\omega_* \in \Omega$ be such that $J(\omega_*) < \infty$ (if $J \equiv +\infty$, there is nothing to prove). Then, $\inf_\Omega g_n \le g_n(\omega_*) = J(\omega_*) - f_n(\omega_*) \le J(\omega_*) - f(\omega_*) \le J(\omega_*) - \inf_\Omega f$. On the other hand, for all n, $f_n \le f_1 \le A := \sup f_1$. Let $B := A + 1 + J(\omega_*) - \inf_\Omega f$. For all ω such that $J(\omega) > B$, we have $g_n(\omega) > B - \sup_\Omega f_n \ge B - A \ge J(\omega_*) - \inf_\Omega f + 1$. We have just seen that for all n,

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

This proves (20) with the compact level set $K = \{J \le B\}$ and completes the proof of the lemma. $\square$

Let us denote

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

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 4.4. $\{\Lambda^* < \infty\} \subset \mathrm{M}_b^+(\Omega)$.

Proof. Let us show that $Q \ge 0$ if $\Lambda^*(Q) < \infty$. Let $f \in \mathrm{C}_b(\Omega)$ be such that $f \ge 0$. As $\inf C = 0$, $\Lambda(af) \le 0$ for all $a \le 0$, and we conclude as in Lemma 4.1.

As in the proof of Lemma 4.1, a positive element $Q \in \mathrm{C}_b(\Omega)'$ is in $\mathrm{M}_b(\Omega)$ if and only if for all decreasing sequence $(f_n)_{n\ge1}$ in $\mathrm{C}_b(\Omega)$ such that $\lim_{n\to\infty} f_n = 0$ pointwise, we have $\lim_{n\to\infty}\langle f_n, Q\rangle = 0$. Let us take a decreasing sequence $(f_n)_{n\ge1}$ in $\mathrm{C}_b(\Omega)$ which converges pointwise to zero. By Lemma 4.3, for all $x \in \mathcal{X}$, $\big(\sup_\Omega\{f_n - C^x\}\big)_{n\ge1}$ is a decreasing sequence and $\lim_{n\to\infty} \sup_\Omega\{f_n - C^x\} = 0$. As $|\sup_\Omega\{f_n - C^x\}| \le \sup_\Omega |f_1| < \infty$ 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 $a \ge 0$ and we conclude as in Lemma 4.1.

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, Lem. 3.7,3.8] which tells us that for each $n \ge 1$ and each Borel probability measure μ on $\mathcal{X}$, there exists a Borel measurable function $\tilde{u}_n$ on $\mathcal{X}$ such that $\tilde{u}_n \le u_n$ and $\tilde{u}_n(x) = u_n(x)$ for μ-a.e. $x \in \mathcal{X}$. $\square$

Lemma 4.5. *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 the 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

$$\begin{aligned}\Lambda^*(Q) := \sup_{f\in \mathrm{C}_b(\Omega)} \left\{\langle f, Q\rangle - \int_{\mathcal{X}} \sup_\Omega\{f - C^x\}\, \mu(dx)\right\} \\ = \int_\Omega C\, dQ + \iota_{\{Q\in \mathrm{P}(\Omega): \theta_\# Q = \mu\}}, \quad Q \in \mathrm{M}_b(\Omega). \qquad (21)\end{aligned}$$

Note that since C is measurable and nonnegative, the integral $\int_\Omega C\, dP$ makes sense in $[0,\infty]$ for any $P \in \mathrm{P}(\Omega)$.

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 4.4 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, $\sup_{\phi \in \mathrm{C}_b(\mathcal{X})} \int_{\mathcal{X}} \phi\, d(\theta_\# Q - \mu) \le \Lambda^*(Q) < \infty$ 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 $f \in \mathrm{C}_b(\Omega)$,

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

Optimizing, we obtain

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

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 (21) 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 $(C_n)_{n\ge1}$ of functions in $\mathrm{C}_b(\Omega)$ which converges pointwise to C. Therefore,

$$\begin{aligned}
&\sup_{f \in \mathrm{C}_b(\Omega)} \left\{ \langle f, P\rangle - \int_{\mathcal{X}} \sup_\Omega\{f - C^x\}\, \mu(dx) \right\} \\
\le\ & \int_\Omega C\, dP \\
\overset{\text{(i)}}{=}\ & \sup_{n\ge1} \int_\Omega C_n\, dP \\
\overset{\text{(ii)}}{=}\ & \sup_{n\ge1} \sup_{f \in \mathrm{C}_b(\Omega)} \left\{ \langle f, P\rangle - \int_{\mathcal{X}} \sup_\Omega\{f - C_n^x\}\, \mu(dx) \right\} \\
=\ & \sup_{f \in \mathrm{C}_b(\Omega)} \left\{ \langle f, P\rangle + \sup_{n\ge1} \int_{\mathcal{X}} \inf_\Omega\{C_n^x - f\}\, \mu(dx) \right\} \\
\overset{\text{(iii)}}{\le}\ & \sup_{f \in \mathrm{C}_b(\Omega)} \left\{ \langle f, P\rangle + \int_{\mathcal{X}} \inf_\Omega\{C^x - f\}\, \mu(dx) \right\} \\
=\ & \sup_{f \in \mathrm{C}_b(\Omega)} \left\{ \langle f, P\rangle - \int_{\mathcal{X}} \sup_\Omega\{f - C^x\}\, \mu(dx) \right\},
\end{aligned}$$

which proves the desired identity (21).

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. □

(c) *Proof of Lemma 4.6.* Let us introduce for each $k \ge 1$,

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

The keystone of the proof of Proposition 2.5 is the following consequence of the Laplace-Varadhan principle.

Lemma 4.6. *Under the assumptions of Proposition 2.5, for all $f \in \mathrm{C}_b(\Omega)$, we have*
(1) $\lim_{k\to\infty} \Lambda_k(f) = \Lambda(f)$;
(2) $\sup_{k\ge1} |\Lambda_k(f)| \le \|f\|, \quad |\Lambda(f)| \le \|f\| := \sup_\Omega |f|$.
The functions Λ_k and Λ are convex.

Proof. Our assumptions allow us to apply the Laplace-Varadhan principle, see Theorem A.2. 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 $k \ge 1$, $|\frac{1}{k}\log\langle e^{kf}, R^{k,x}\rangle| \le \|f\|$. Passing to the limit, we also get $|\sup_\Omega\{f - C^x\}| \le \|f\|$. 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. □

(d) *Completion of the proof of Proposition 2.5.* With Lemma 4.2, we see that

$$\mathcal{C}^{k,\mu_0}(Q) = \Lambda_k^*(Q), \quad Q \in \mathrm{C}_b(\Omega)' \tag{22}$$

where Λ_k^* is the convex conjugate of Λ_k with respect to the duality $\langle \mathrm{C}_b(\Omega), \mathrm{C}_b(\Omega)'\rangle$.
Let us equip $\mathrm{C}_b(\Omega)'$ with the $*$-weak topology $\sigma(\mathrm{C}_b(\Omega)', \mathrm{C}_b(\Omega))$. By Lemma 4.6 and Corollary 5.4, we have $\Gamma\text{-}\lim_{k\to\infty} \Lambda_k^* = \Lambda^*$ in $\mathrm{C}_b(\Omega)'$. By Lemma 4.4, this limit still holds in $\mathrm{M}_b(\Omega) \subset \mathrm{C}_b(\Omega)'$:

$$\Gamma\text{-}\lim_{k\to\infty} \Lambda_k^* = \Lambda^* \quad \text{in } \mathrm{M}_b(\Omega). \tag{23}$$

As the function C of Proposition 2.5 is such that C^x is a LD rate function for all $x \in \mathcal{X}$, it satisfies the assumption of Lemma 4.5 which is $\inf_\Omega C^x = 0$ for μ-almost every $x \in \mathcal{X}$. Therefore, we have $\Lambda^*(Q) = \int_\Omega C\, dQ + \iota_{\{Q\in \mathrm{P}(\Omega): \theta_\# Q = \mu\}}$, $Q \in \mathrm{M}_b(\Omega)$.
Together with (22) and (23), this completes the proof of Proposition 2.5. □

Proofs of the remaining results. The main ingredients of the proofs of the remaining results are Proposition 2.5 and Theorem 6.1 which is the main result of Section 6.

• *Proof of Proposition 2.2.* Proposition 2.2 is a particular case of Proposition 2.5. 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 $\mathcal{X}^2$.

• *Proofs of Corollary 2.3 and Theorem 2.4.* With Proposition 2.2 in hand, Corollary 2.3 and Theorem 2.4 are immediate consequences of Theorem 6.1 and of the equi-coerciveness

with respect to the $*$-weak topology $\sigma(\mathrm{P}(\mathcal{X}), \mathrm{C}_b(\mathcal{X}))$ of $\{\mathcal{C}_{01}, \mathcal{C}_{01}^k; k \geq 1\}$. 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 $\pi_1 \in \{\mu_1, \mu_1^k; k \geq 1\}$ is relatively compact since $\lim_{k\to\infty} \mu_1^k = \mu_1$; a consequence of Prokhorov's theorem in a Polish space.
The uniqueness of the solution to (S^k) follows from the strict convexity of the relative entropy.
Note that, when C and c are linked by (1), one can also derive the equi-coerciveness of $\{\mathcal{C}_{01}, \mathcal{C}_{01}^k; k \geq 1\}$ from the equi-coerciveness of $\{\mathcal{C}, \mathcal{C}^k; k \geq 1\}$ (see below), as in the proof of Theorem 6.1.

• *Proofs of Corollary 2.6 and Theorem 2.7.* Similarly, once we have Proposition 2.5 in hand, Corollary 2.6 and Theorem 2.7 are immediate consequences of Theorem 6.1 and of the equi-coerciveness with respect to the $*$-weak topology $\sigma(\mathrm{P}(\Omega), \mathrm{C}_b(\Omega))$ of $\{\mathcal{C}, \mathcal{C}^k; k \geq 1\}$. This equi-coerciveness follows from Corollary 5.4 and Lemma 4.6. Again, the uniqueness of the solution to ($\mathrm{S}^k_{\mathrm{dyn}}$) follows from the strict convexity of the relative entropy.

• *Proof of Theorem 2.8.* It relies upon the subsequent lemma.

Lemma 4.7. *Under the assumptions of Proposition 2.5, the function c defined by (1) 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:

$$\Psi(\pi) = \sup_{f\in\mathcal{F}} \int_{\mathcal{X}^2} f\,d\pi, \quad \pi \in \mathrm{P}(\mathcal{X}^2) \tag{24}$$

where $\mathcal{F} := \{f \in \mathrm{C}_b(\mathcal{X}^2); f(X_0, X_1) \leq C\}$. For a proof of (24), one can rewrite mutatis mutandis the proof of the Kantorovich dual equality. See for instance [Léo11, 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

$$\Psi(\pi) = \int_{\mathcal{X}^2} \psi\,d\pi, \quad \pi \in \mathrm{P}(\mathcal{X}^2). \tag{25}$$

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

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

Let us show the converse inequality. With (24), we see that for each $f\in\mathcal{F}$ and all $(x,y)\in\mathcal{X}^2$, $\psi(x,y)=\Psi(\delta_{(x,y)})\geq\int_{\mathcal{X}^2}f\,d\delta_{(x,y)}=f(x,y)$. That is $f\leq\psi$, for all $f\in\mathcal{F}$. Therefore, $\Psi(\pi)=\sup_{f\in\mathcal{F}}\int_{\mathcal{X}^2}f\,d\pi\leq\int_{\mathcal{X}^2}\psi\,d\pi$, completing the proof of (25).

It remains to establish that $\psi=c$. With (24), we get $\psi=\sup\mathcal{F}$. But it is clear that $f\in\mathcal{F}$ if and only if for all $x,y\in\mathcal{X}$, $f(x,y)\leq\inf\{C(\omega);\omega\in\Omega:\omega_0=x,\omega_1=y\}:=c(x,y)$. Hence, ψ is the upper envelope of the set of all functions $f\in\mathrm{C}_b(\mathcal{X}^2)$ such that $f\leq c$. In other words ψ is the lower semicontinuous envelope $\mathrm{ls}\,c$ of c. Finally, for all $x,y\in\mathcal{X}$, $\mathrm{ls}\,c(x,y)=\psi(x,y)=\inf\left\{\int_\Omega C\,dP^{xy};P\in\mathrm{P}(\Omega)\right\}\geq c(x,y)\geq\mathrm{ls}\,c(x,y)$. This implies the desired result: $\psi=\mathrm{ls}\,c=c$. □

Proof of Theorem 2.8. It is assumed that for any $x\in\mathcal{X}$, $(R^{k,x})_{k\geq1}$ 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.1, we obtain that for any $x\in\mathcal{X}$, $(\rho^{k,x})_{k\geq1}$ 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 2.8 follows from the lower semicontinuity of c which was obtained at Lemma 4.7. Indeed, this shows that the assumptions of Proposition 2.2 are fulfilled. The identity $\inf\,(\mathrm{MK_{dyn}})=\inf\,(\mathrm{MK})$ is a direct consequence of Lemma 4.7.

• 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 Eq. (2).

• 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 (21) and the proof of Corollary 5.4, 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 4.7, 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), $\widehat{P}^{xy}(\Gamma^{xy})=1$. 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. □

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

During the proof of Proposition 2.5, we used Corollary 5.4 to derive the key identity (23). This section is devoted to the proof of Corollary 5.4 which is an easy consequence of Theorem 5.2 below.

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 5.1. *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 5.1 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 $C:\iota_C^*(y)=\sup_{x\in C}\langle x,y\rangle$, $y\in Y$.
The effective domain of an extended-real valued function f is defined as $\mathrm{dom}\, f:=\{f<\infty\}$.

Theorem 5.2. *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 5.5 and 5.6.

Remark 5.3. By [DM93, Prop. 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 5.2 is

Corollary 5.4. *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 $\operatorname{dom} f_k \subset K$ for all $k \geq 1$ and $\operatorname{dom} f \subset K$.

Proof. Under (b), (d") implies (c). As (d") implies (d'), we have (d) by Remark 5.3. 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 5.2. □

Before proving Theorem 5.2, let us show the preliminary Lemmas 5.5 and 5.6.

Lemma 5.5. *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

$$V \cap \operatorname{dom} f \neq \emptyset \quad \text{or} \quad V \cap \operatorname{cl}\operatorname{dom} f = \emptyset, \tag{26}$$

we have

$$\inf_{x\in V} f(x) = -\inf_{y\in Y}(f^*(y) + \iota_V^*(-y)) \in (-\infty, \infty] \tag{27}$$

and if V doesn't satisfy (26), we have

$$\inf_{x\in W} f(x) = -\inf_{y\in Y}(f^*(y) + \iota_W^*(-y)) = +\infty \tag{28}$$

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

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

$$\inf_{x\in X} f(x) > -\infty \tag{29}$$

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 when* $V\cap \operatorname{cl}\operatorname{dom} f = \emptyset$. 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 $\iota_V^*(y_o) = \sup_{x\in V}\langle x,y_o\rangle < \inf_{\operatorname{cl}\operatorname{dom} f}\langle x,y_o\rangle \le \inf_{x\in\operatorname{dom} f}\langle x,y_o\rangle$. Hence,

$$\inf_{x\in\operatorname{dom} f}\{\langle x,y_o\rangle - \iota_V^*(y_o)\} > 0 \tag{30}$$

and

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

where the last equality follows from (29) and (30). This proves that (28) holds with $W=V$.

• Finally, if (26) 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 5.6. *Let the $\sigma(X,Y)$-closed convex neighbourhood V of the origin be defined by*

$$V = \{x\in X; \langle y_i,x\rangle \le 1, 1\le i\le n\} \tag{31}$$

with $n\ge 1$ 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 $\{\iota_V^*\le b\} = b\operatorname{cv}\{y_1,\dots,y_n\}$ for each $b\ge 0$ 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 $x_o\in E := \cap_{1\le i\le n}\ker y_i$. Then, $\langle y_i, x_1+x_o\rangle = \langle y_i,x_1\rangle \le 1$. 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 $\gamma_C(y) := \inf\{\lambda\ge 0; y\in\lambda C\}, y\in Y$ is the support function of its polar set $C^\circ = \{x\in X; \langle x,y\rangle\le 1, \forall y\in C\}$. 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. $N^{\circ\circ}=\mathrm{cv}(N)$: the convex hull of N, we get $V=\mathrm{cv}(N)^{\circ}$ and

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

In particular, for all real b, $\iota_V^*(y)\le b\Leftrightarrow\gamma_{\mathrm{cv}(N)}(y)\le b\Leftrightarrow y\in b\ \mathrm{cv}(N)$. 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 5.2. 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

$$\lim_{k\to\infty}\inf_{x\in V}f_k(x)=\inf_{x\in V}f(x). \tag{32}$$

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

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

as a base where $(y_1,\dots,y_n),n\ge 1$ describes the collection of all the finite families of vectors in Y. By Lemma 5.5, there exists such a $V\subset U$ which satisfies

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

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

$$\iota_V^*(-y)=-\langle x_o,y\rangle+\iota_{V-x_o}^*(-y) \tag{33}$$

and by Lemma 5.6, the effective domain of ι_V^* is Z. Therefore, to prove (32) it remains to show that

$$\lim_{k\to\infty}\inf_{y\in Y}h_k^Z(y)=\inf_{y\in Y}h^Z(y). \tag{34}$$

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 [DM93, Prop. 6.25].

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, Thm. 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 when $x_o\in\mathrm{dom}\,f$. We already know that (h_k^Z) Γ-converges to h^Z. To prove (34), it remains to check that the sequence (h_k^Z) is equicoercive, see Theorem 2.1.

For all $y\in Y$, $g(y)-\langle x_o,y\rangle\ge -f(x_o)$ and (33) imply $h^Z(y)\ge -f(x_o)+\iota_{V-x_o}^*(-y)$. Since, $-f(x_o)>-\infty$ and $\iota_{V-x_o}^*$ is coercive (Lemma 5.6), 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 (34).

The case when $x_o\in\mathrm{cl}\,\mathrm{dom}\,f$. In this case, there exists $x_o'\in\mathrm{dom}\,f$ such that $V'=x_o'+(V-x_o)/2=\{x\in X;|\langle 2y_i,x-x_o'\rangle|\le 1,i\le k\}\in\mathcal{N}(x_o')$ satisfies $x_o\in V'\subset V\subset U$. One deduces from the previous case, that (34) holds true with V' instead of V.

The case when $x_o\notin\mathrm{cl}\,\mathrm{dom}\,f$. As (h_k^Z) Γ-converges to h^Z, by [Bee93, Prop. 1.3.5], we have $\limsup_{n\to\infty}\inf_{y\in Y}h_k^Z(y)\le\inf_{y\in Y}h^Z(y)$. As $x_o\notin\mathrm{cl}\,\mathrm{dom}\,f$, 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 (34).

This completes the proof of Theorem 5.2. □

6. Γ-convergence of minimization problems under constraints

At Section 4, we derived the proofs of Corollary 2.3, Theorem 2.4, Corollary 2.6 and Theorem 2.7 from Propositions 2.2 and 2.5 by means of Theorem 6.1. The aim of this section is to prove this theorem. As this 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 6.1. *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 cluster point and any such cluster 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 6.2 and 6.3 below, is postponed after the proofs of these preliminary 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 6.2. *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 cluster point x^* :

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

The existence of a cluster 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 Theorem 2.1. 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 6.2, there exists a sequence $(y_k^*)_{k\geq 1}$ such that

$$\lim_{k\to\infty} y_k^* = y^*, \quad \lim_{k\to\infty} \psi_k(y_k^*) = \psi(y^*). \tag{35}$$

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 6.3. *In addition to the assumptions of Lemma 6.2, 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 cluster points in* X *and any cluster point is a minimizer of* φ.

Remark that this lemma doesn't assert that $(\varphi_k)_{k\geq 1}$ Γ-converges to 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 cluster point.
As $\varphi_k(\hat{x}_k) = \psi_k(y_k^*)$, we see with (35) that

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

On the other hand, let $\hat{x}$ be any cluster point of $(\hat{x}_k)_{k\ge1}$. 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

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

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

Proof of Theorem 6.1. Consider the functions

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

for each $k\ge1$ and

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

Because of Lemmas 6.2, 6.3 and (35), to complete the proof it is enough to show that

$$\Gamma\text{-}\lim_{k\to\infty} g_k=g \tag{36}$$

together with the coerciveness assumptions of these lemmas.

Let us begin with the coerciveness. Since for each $k\ge1$, f_k is coercive and θ is continuous, we see that for any large enough a, $\{g_k\le a\}=\{(x,y)\in X\times Y; x\in\{f_k\le a\}, y=\theta(x)\}$ is compact, i.e. for each $k\ge1$, g_k is coercive. As $(f_k)_{k\ge1}$ 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 $\bigcup_{k\ge1}\{g_k\le a\}=\{(x,y)\in X\times Y; x\in\bigcup_{k\ge1}\{f_k\le a\}, y=\theta(x)\}$ is relatively compact, i.e. $(g_k)_{k\ge1}$ is equi-coercive.

Let us prove that (36) holds true. Let $(x,y)\in X\times Y$ be fixed. We have to prove that:

(i) For any sequence $(x_k,y_k)_{k\ge1}$ such that $\lim_{k\to\infty}(x_k,y_k)=(x,y)$, $\liminf_{k\to\infty} f_k(x_k)+\iota_{\{y_k=\theta(x_k)\}}\ge f(x)+\iota_{\{y=\theta(x)\}}$.

(ii) There exists a sequence $(\tilde{x}_k,\tilde{y}_k)_{k\ge1}$ such that $\lim_{k\to\infty}(\tilde{x}_k,\tilde{y}_k)=(x,y)$, and $\liminf_{k\to\infty} f_k(\tilde{x}_k)+\iota_{\{\tilde{y}_k=\theta(\tilde{x}_k)\}}\le f(x)+\iota_{\{y=\theta(x)\}}$.

Suppose first that $y\ne\theta(x)$. Then (ii) is obvious and due to the continuity of θ, for any sequence $(x_k,y_k)_{k\ge1}$ such that $\lim_{k\to\infty}(x_k,y_k)=(x,y)$ we have that for all large enough k, $\theta(x_k)\ne y_k$. This proves (i).

Now, suppose that $y=\theta(x)$. Then (i) follows from $\liminf_{k\to\infty} f_k(x_k)+\iota_{\{y_k=\theta(x_k)\}}\ge\liminf_{k\to\infty} f_k(x_k)\ge f(x)=f(x)+\iota_{\{y=\theta(x)\}}$, whenever $\lim_{k\to\infty}x_k=x$. To prove (ii), take a recovering sequence $(\tilde{x}_k)_{k\ge1}$ for $(f_k)_{k\ge1}$ at x, i.e. $\liminf_{k\to\infty} f_k(\tilde{x}_k)\le f(x)$ and put $\tilde{y}_k=\theta(\tilde{x}_k)$, for each $k\ge1$. 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 $\liminf_{k\to\infty} f_k(\tilde{x}_k)+\iota_{\{\tilde{y}_k=\theta(\tilde{x}_k)\}}=\liminf_{k\to\infty} f_k(\tilde{x}_k)\le f(x)=f(x)+\iota_{\{y=\theta(x)\}}$, which proves (ii) and completes the proof of the theorem. □

Appendix A. Large deviations

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 $(p_k)_{k\ge1}$ of probability measures on X satisfies the large deviation principle (LDP for short) with scale k and rate function I, if for each Borel measurable subset A of X we have

$$-\inf_{x\in\operatorname{int}A} I(x)\overset{\text{(i)}}{\le}\liminf_{k\to\infty}\frac{1}{k}\log p_k(A)\le\limsup_{k\to\infty}\frac{1}{k}\log p_k(A)\overset{\text{(ii)}}{\le}-\inf_{x\in\operatorname{cl}A} I(x) \tag{37}$$

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.

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

Theorem A.1 (Contraction principle)**.** *Let $(p_k)_{k\geq 1}$ satisfy the LDP in X with scale k 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,*

$$(f_{\#}p_k)_{k\geq 1}$$

satisfies the LDP in Y with scale k 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].

Theorem A.2 (Laplace-Varadhan principle)**.** *Suppose that $(p_k)_{k\geq 1}$ 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*

$$\lim_{M\to\infty} \liminf_{k\to\infty} \frac{1}{k} \log \int_X e^{kf(x)} \boldsymbol{1}_{\{f\geq M\}}\, p_k(dx) = -\infty.$$

Then,

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

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

A well-known LD result is about the large deviations of the $\mathbb{R}^d$-valued process which we have already met at (5) and is defined by

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

where the initial condition $Y_0^{k,x} = x$ is deterministic, $B = (B_t)_{0\leq t\leq 1}$ is the Wiener process on $\mathbb{R}^d$.

Theorem A.3 (Schilder's theorem)**.** *The sequence of random processes $(Y^{k,x})_{k\geq 1}$ 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].

References

[ADPZ11] S. Adams, N. Dirr, M. Peletier, and J. Zimmer. From a large-deviation principle to the Wasserstein gradient flow: a new micro-macro passage. *Communications in Mathematical Physics*, 307:791–815, 2011.

[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. 98:423–487, 1994.

[DM93] G. Dal Maso. *An Introduction to* Γ*-Convergence*. Progress in Nonlinear Differential Equations and Their Applications 8. Birkhäuser, 1993.

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

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

[Léo11] C. Léonard. A saddle-point approach to the Monge-Kantorovich transport problem. *ESAIM-COCV (Control, Optimization and Calculus of Variations)*, 17(03):682–704, 2011.

[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. 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`