LAZY RANDOM WALKS AND OPTIMAL TRANSPORT ON GRAPHS

CHRISTIAN LÉONARD

ABSTRACT. This paper is about the construction of displacement interpolations on a discrete metric graph. Our approach is based on the approximation of any optimal transport problem whose cost function is a distance on a discrete graph by a sequence of Schrödinger problems associated with random walks whose jump frequencies tend down to zero. Displacement interpolations are defined as the limit of the time-marginal flows of the solutions to the Schrödinger problems. This allows to work with these interpolations by doing stochastic calculus on the approximating random walks which are regular objects, and then to pass to the limit in a slowing down procedure. The main convergence results are based on Γ-convergence of entropy minimization problems.
As a by-product, we obtain new results about optimal transport on graphs.

CONTENTS

INTRODUCTION

Displacement interpolations on $\mathbb{R}^n$ were introduced by McCann in [McC94] and extended later to a geodesic space $(\mathcal{X}, d)$ where they are defined as minimizing geodesics on the space of all probability measures on $\mathcal{X}$ equipped with the Wasserstein pseudo-distance of order two. They appeared to be a basic and essential notion of the Lott-Sturm-Villani theory of lower bounded curvature of geodesic spaces, see [LV09, Stu06a, Stu06b, Vil09]. Indeed, as discovered by McCann [McC94, McC97], Otto, Villani [OV00], Cordero-Erausquin, McCann, Schmuckenschläger [CEMS01], Sturm and von Renesse [Sv05] in the Riemannian setting, lower bounded curvature is intimately linked to convexity properties

of the relative entropy with respect to the volume measure along displacement interpolations. It happens that these convexity properties admit natural analogues on a geodesic space.

It is tempting to try to implement a similar approach in a discrete setting. But, little is known in this case since a discrete space fails to be a length space. Indeed, any regular enough path on a discrete space is piecewise constant with instantaneous jumps, so that no speed and a fortiori no constant speed geodesic exist.

This paper is about the construction of displacement interpolations on a discrete[1] metric graph. It permits us to propose, in this discrete setting, natural substitutes for the constant speed geodesics. As a by-product of our approach, we also obtain new results about the optimal transport on a graph.

To recover some time regularity of the paths without allowing any mass transfer along the edges of a graph, one is enforced to do some averaging on ensembles of discontinuous sample paths. This means that, for defining instantaneous speeds and accelerations, one is obliged to consider expected values of random walks. So doing, one lifts the paths on the discrete state space $\mathcal{X}$ up to the continuous space $\mathrm{P}(\mathcal{X})$ of probability measures on $\mathcal{X}$. This lifting from $\mathcal{X}$ to $\mathrm{P}(\mathcal{X})$ has already been successfully used in the Lott-Sturm-Villani theory where minimizing geodesics on the length space $\mathcal{X}$ are embedded in the set of all minimizing geodesics on the Wasserstein space of order two $(\mathrm{P}(\mathcal{X}), W_2)$, i.e. W_2-displacement interpolations. As usual, the pseudo-distance W_2 is defined as the square root of the optimal value of a transport problem with cost function d^2, the square of a distance d on $\mathcal{X}$.

The author already proposed in [Léo12a] a construction of the W_2-displacement interpolations in $\mathbb{R}^n$ as limits of solutions of Schrödinger problems when the reference processes are slowed down to a no-motion process. In the present paper, we stay as close as possible to this strategy. It will lead us to a natural notion of W_1-displacement interpolation on a discrete metric graph. The main idea lies in the following thought experiment.

The cold gas experiment. Suppose you observe at time $t = 0$ a large collection of particles that are distributed with a profile close to the probability measure $\mu_0 \in \mathrm{P}(\mathcal{X})$ on the state space $\mathcal{X}$. As in the thought experiment proposed by Schrödinger in 1931 [Sch31, Sch32] or in its close variant described in Villani's textbook [Vil09, *Lazy gas experiment*, p. 445], ask them to rearrange into a new profile close to some $\mu_1 \in \mathrm{P}(\mathcal{X})$ at some later time $t = 1$.

Suppose that the particles are in contact with a heat bath. Since they are able to create mutual optimality (Gibbs conditioning principle), they find an optimal transference plan between the endpoint profiles μ_0 and μ_1. Now, suppose in addition that the typical speed of these particles is close to zero: the particles are lazy, or equivalently the heat bath is pretty cold. As each particle decides to travel at the lowest possible cost, it chooses an almost *geodesic* path. Indeed, with a very high probability each particle is very slow, so that it is typically expected that its final position is close to its initial one. But it is required by the optimal transference plan that it must reach a distant final position. Hence, *conditionally on the event that the target μ_1 is finally attained*, each particle follows an almost geodesic path with a high probability. At the limit where the heat bath vanishes (zero temperature), each particle follows a geodesic while the whole system obeys the performs some optimal transference plan. This absolutely cold gas experiment is called the

---

[1]The epithet *discrete* is important since the standard definition of a (non-discrete) metric graph allows for continuous mass transfer along the edges. In the present paper, only pure jumps occur.

lazy gas experiment in [Vil09] where its dynamics is related to displacement interpolations. For further detail with graphical illustrations, see [Léoa, §6]

With this thought experiment in mind, one can guess that a slowing down procedure enforces the appearance of individual geodesics. A displacement interpolation is a mixture of these geodesics which is specified by some optimal transport plan between the endpoint profiles μ_0 and μ_1. We will see that displacement interpolations on a discrete metric graph $(\mathcal{X}, \sim, d)$ are minimizing geodesics on the Wasserstein space of order one $(\mathrm{P}(\mathcal{X}), W_1)$ where the pseudo-distance W_1 on $\mathrm{P}(\mathcal{X})$ is defined as the optimal value of a transport problem with the distance d as its cost function.

Notation. Before going on we need some general notation. We denote by $\mathrm{P}(Y)$ and $\mathrm{M}_+(Y)$ the sets of all probability and positive measures on a measurable set Y. The push-forward of a measure $\alpha \in \mathrm{M}_+(Y_1)$ by the measurable mapping $f : Y_1 \to Y_2$ is $f_\#\alpha(\cdot) := \alpha(f^{-1}(\cdot)) \in \mathrm{M}_+(Y_2)$.

Let $\Omega \subset \mathcal{X}^{[0,1]}$ be a set of paths from the time interval $[0,1]$ to the measurable state space $\mathcal{X}$. The canonical process $X = (X_t)_{0 \leq t \leq 1}$ is defined for all $\omega = (\omega_s)_{0 \leq s \leq 1} \in \Omega$ by $X_t(\omega) = \omega_t \in \mathcal{X}$ for each $0 \leq t \leq 1$. The set Ω is endowed with the σ-field generated by $(X_t; t \in [0,1])$. For any $t \in [0,1]$ and any $Q \in \mathrm{M}_+(\Omega)$, the push-forward

$$Q_t := (X_t)_\# Q \in \mathrm{M}_+(\mathcal{X}) \tag{0.1}$$

of Q by the measurable mapping X_t is the law of the random position X_t at time t if Q describes the behaviour of the random path. More specifically Q_0, Q_1 are the initial and final time-marginal projections of Q. Also

$$Q_{01} := (X_0, X_1)_\# Q \in \mathrm{M}_+(\mathcal{X}^2) \tag{0.2}$$

is the joint law of the random endpoint position (X_0, X_1). The xy-bridge of Q is the conditional probability measure

$$Q^{xy} := Q(\cdot | X_0 = x, X_1 = y) \in \mathrm{P}(\Omega), \quad x, y \in \mathcal{X}. \tag{0.3}$$

As a general result, we have the disintegration formula

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

If P is a probability measure on Ω, we sometimes use the probabilistic convention: $E_P(u) := \int_\Omega u \, dP$.

Implementing the slowing down procedure. As a consequence of the large deviation theory, the mathematical translation of the lazy gas experiment is in terms of some entropy minimization problems: the Schrödinger problems. More precisely, we are going to investigate the limit as a slowing down parameter k tends to infinity (for instance, the average number of jumps of each individual particle is of order $1/k$) of the sequence of minimizing problems

$$H(P|R^k) \to \min; \qquad P \in \mathrm{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1, \tag{0.4}$$

where the relative entropy of a probability measure p with respect to a reference measure r is defined by $H(p|r) := \int \log(dp/dr) \, dp$, $R^k \in \mathrm{M}_+(\Omega)$ describes the random behaviour of a slow walker (a single typical particle) on the graph and P is the unknown path probability measure which is subject to have initial and final prescribed marginal measures: $P_0 = \mu_0$ and $P_1 = \mu_1 \in \mathrm{P}(\mathcal{X})$, see notation (0.1). For more detail about Schrö dinger's problem, see the author's survey paper [Léo14].

By means of Γ-convergence, we prove that the unique minimizer $\widehat{P}^k$ of the entropy problem (0.4) admits a limit

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

that is characterized as a specific solution of a new auxiliary variational problem. It is the law of a random walk whose sample paths are piecewise constant geodesics with respect to a given distance d on the graph $(\mathcal{X}, \sim)$. Moreover, the joint law $\widehat{P}_{01} \in \mathrm{P}(\mathcal{X}^2)$, see notation (0.2), of the couple of its endpoint positions under $\widehat{P}$ is a singled out solution of the Monge-Kantorovich optimal transport problem

$$\int_{\mathcal{X}^2} d(x, y) \, \pi(dxdy) \to \min; \qquad \pi \in \mathrm{P}(\mathcal{X}^2), \pi_0 = \mu_0, \pi_1 = \mu_1, \qquad (0.5)$$

where π_0 and $\pi_1 \in \mathrm{P}(\mathcal{X})$ are the first and second marginals of π. The optimal value of this problem is the Wasserstein distance of order one $W_1(\mu_0, \mu_1)$.

This specific random walk $\widehat{P}$ is a geodesic bridge between μ_0 and μ_1. We call it a *displacement random walk* and its time-marginal flow $(\widehat{P}_t)_{0 \leq t \leq 1} \in \mathrm{P}(\mathcal{X})^{[0,1]}$, see notation (0.1), defines a *displacement interpolation* between μ_0 and μ_1. This definition is justified because of several analogies with the standard displacement interpolations on a geodesic space.

Approximating (0.5) **by means of Γ-asymptotic expansions.** Let us comment a little on the approximation of (0.5) by (0.4) as k tends to infinity. Instead of the discrete set of vertices $\mathcal{X}$, let us first consider the analogue of (0.5) on $\mathcal{X} = \mathbb{R}^k$ which is related to Monge's problem:

$$\int_{\mathbb{R}^k} d(x, T(x)) \, \mu_0(dx) \to \min; \qquad T : \mathbb{R}^k \to \mathbb{R}^k, T_{\#}\mu_0 = \mu_1, \qquad (0.6)$$

where the transport map T is assumed to be measurable. The Monge-Kantorovich problem (0.5) is a convex relaxation of (0.6) in the sense that $\pi \mapsto \int_{\mathcal{X}^2} d(x, y) \, \pi(dxdy)$ is a convex function on the convex subset $\{\pi \in \mathrm{P}(\mathcal{X}^2); \pi_0 = \mu_0, \pi_1 = \mu_1\}$ and $\pi^T := (\mathrm{Id}, T)_{\#}\mu_0$ gives $\int_{\mathcal{X}^2} d(x, y) \, \pi^T(dxdy) = \int_{\mathcal{X}} d(x, T(x)) \, \mu_0(dx)$.

Many known solutions of (0.6) *rely on approximations and variational methods.* Monge's original problem corresponds to d the standard Euclidean distance. Sudakov proposed an efficient, but still incomplete, strategy in [Sud79]. The first complete solution was obtained by Evans and Gangbo in [EG99]. It states that when μ_0 is absolutely continuous and μ_0, μ_1 have finite first moments (plus some restrictions on μ_0, μ_1), (0.6) admits a unique solution. Its proof is based on PDE arguments and an approximation of the "affine" cost $d(x, y) = \|y - x\|$ by the "strictly convex" costs $d^\epsilon(x, y) := \|y - x\|^{1+\epsilon}$, with $\epsilon > 0$ tending to zero, which entails a convergence of the corresponding Monge-Kantorovich problems. A natural generalization of Monge's original problem is obtained by replacing the Euclidean norm by any norm $\| \cdot \|$ on $\mathbb{R}^k$. With alternate approaches, but still taking advantage of the approximation $d^\epsilon \to d$, Caffarelli , Feldman and McCann in [CFM02] and Ambrosio and Pratelli in [Amb03, AP03], removed [EG99]'s restrictions and extended this existence and uniqueness result to the case where the norm $\| \cdot \|$ is assumed to be strictly convex. Later, Ambrosio, Kirchheim and Pratelli [AKP04] succeeded in the more difficult case where the norm is crystalline. In the general case without any restriction on the norm, the solution has recently been obtained by Champion and De Pascale in [CDP11]. Again, both [AKP04] and [CDP11] rely on variational methods and Γ- convergence.

The main Γ-convergence technic used during the proofs of [AP03, AKP04, CDP11] is an asymptotic expansion which was introduced by Anzellotti and Baldo [AB93], see also [Att96]. In the present paper, we also make a crucial use of this technique at Lemma 5.10. Instead of considering the approximation $d^\epsilon \to d$, the not convex enough problem (0.5) is approximated by the sequence of *strictly* convex entropy minimization problems (0.4) up to some normalization, see (S_{dyn}^k) at the end of Section 1.

A discrete metric measure graph $(\mathcal{X}, \sim, d, m)$. Let us take a distance d on $\mathcal{X}$ which is compatible with the graph structure $\sim$, in the sense of Hypothesis 2.1-(d), and a possibly unbounded positive measure m on $\mathcal{X}$. It will be shown that the sequence of reference Markov random walks $(R^k)_{k\geq 1}$ can be chosen such that for any $0 \leq t \leq 1$ and any $k \geq 1$, the t-marginal R_t^k is equal to m, the Markov generator L_t^k is self-adjoint in $L^2(\mathcal{X}, m)$ and the limiting displacement interpolation

$$\mu_t := \lim_{k \to \infty} \widehat{P}_t^k \in P(\mathcal{X}), \quad 0 \leq t \leq 1$$

is a minimizing constant speed geodesic in $(P(\mathcal{X}), W_1)$ where the Wasserstein distance W_1 is built upon the distance d.

Although what follows is not treated in the present paper, it seems plausible that, in the perspective of deriving displacement convexity properties of the entropy, it is worthwhile studying the functions $t \in [0, 1] \mapsto H\big((X_t)_\# \widehat{P}^k | (X_t)_\# R^k\big) = H(\widehat{P}_t^k | m)$ since the dynamics of R^k and $\widehat{P}^k$ are well understood, see [Léo14], and $\lim_{k\to\infty} H(\widehat{P}_t^k | m) = H(\mu_t | m)$. This will be explored elsewhere.

Alternate approaches for deriving displacement convexity in a discrete setting. Let us write a few words about already existing strategies in view of deriving displacement convexity properties of the entropy on a discrete space.

Midpoint interpolations. In the special case of the hypercube $\mathcal{X} = \{0, 1\}^n$ equipped with the Hamming distance, Ollivier and Villani [OV12] have introduced the most natural interpolation $(\mu_t)_{0\leq t\leq 1}$ between δ_x and δ_y which is defined as follows. For any $0 \leq t \leq 1$, μ_t is the uniform probability measure on the set of all t-midpoints of x and y. In this paper, the authors announce that it seems to be difficult to obtain displacement convexity of the entropy along these interpolations. However, they prove a Brunn-Minkowski inequality for the counting measure under the restriction $t = 1/2$ which allows for explicit combinatoric computations.

Approximate midpoint interpolations. Bonciocat and Sturm [BS09] have introduced h-approximate midpoint interpolations and a natural extension of the lower bounded curvature of geodesic spaces, again in terms of displacement convexity of the relative entropy. They show that the h-discretized graph of a geodesic space with a curvature lower bound K has a lower h-curvature bound which converges to K as the discretization mesh h tends to zero. They also compute h-curvature lower bounds of some planar graphs.

Maas-Mielke gradient flow. Recently, Maas [Maa11] and Mielke [Mie11] designed a new distance $\mathcal{W}$ on $P(\mathcal{X})$ which, unlike W_2, allows for regarding evolution equations of reversible Markov chains on the discrete space $\mathcal{X}$ as gradient flows of the entropy $H(\cdot|m)$ on $(P(\mathcal{X}), \mathcal{W})$, where m is the reversing measure of the Markov chain. Soon after, following the Lott-Sturm-Villani strategy, Maas and Erbar [EM12] and Mielke [Mie] applied this gradient flow approach to obtain convexity properties of relative entropies in $(P(\mathcal{X}), \mathcal{W})$ leading to new interesting results about lower bounded curvature on discrete spaces. The distance $\mathcal{W}$ is a Riemannian distance that is obtained by plugging discrete objects,

such as a discrete gradient and a discrete divergence, into the standard Benamou-Brenier formula, see (1.4). It is not clear that $\mathcal{W}$ and the corresponding minimizing geodesics on $(\mathrm{P}(\mathcal{X}), \mathcal{W})$ are associated with an optimal transport problem related to some distance d on $\mathcal{X}$. Hence, although this approach is respectful of the measure space structure $(\mathcal{X}, m)$, it might not be linked to a *metric* structure $(\mathcal{X}, d)$. This still needs to be clarified if, starting from this Riemannian structure on $\mathrm{P}(\mathcal{X})$, one wishes to define a notion of Ricci type curvature on $\mathcal{X}$ (which should be related to the variation of the discrete volume m along "geodesics" on $(\mathcal{X}, d)$).

Entropic interpolations. Recently, the author [Léoa] studied convexity properties of the relative entropy along *entropic* interpolations, i.e. time marginal flow of the solution of the entropy minimization problem (0.4) without slowing down ($k = 1$).

Binomial interpolations. Consider the set $\{0, \ldots, n\}$ with the graph structure of $\mathbb{Z}$. The binomial interpolation between δ_0 and δ_n is defined for any $0 \leq t \leq 1$ and $x \in \{0, \ldots, n\}$ by $\mu_t(x) = \binom{n}{x} t^x (1-t)^{n-x}$. It means that $\mu_t = \mathcal{B}(n, t)$ is the binomial distribution with parameters n and t. This interpolation was introduced by Johnson [Joh07] on $\mathbb{N}$ to obtain displacement convexity of the entropy with motivations slightly different from the Lott-Sturm-Villani theory. Now, on a graph equipped with the standard graph distance, a binomial interpolation between δ_x and δ_y is a mixture of the binomial interpolations along the geodesic chains $x = x_0 \sim x_1 \sim \cdots \sim x_{d(x,y)} = y$ that connect x and y. These interpolations allow for displacement convexity of the entropy with respect to the counting measure m. It has been successfully used to prove displacement convexity of the entropy by Hillion [Hil, Hil10, Hil12] on some trees and Gozlan, Roberto, Samson and Tetali [GRST] which have worked out the examples of the complete graph, $\mathbb{Z}^n$ and the hypercube.

It happens that binomial interpolations are specific instances of the displacement interpolations which are built in the present paper. They correspond to the simple random walk R and the volume measure m^o, see (A.3) and (A.4) at the appendix, with the standard graph distance. It will be seen at Claim 3.17 that they are closely related to bridges of the Poisson process.

Outline of the paper. Section 1 is a continuation of this introduction where we briefly present the analogies between usual displacement interpolations on a Riemannian manifold and displacement interpolations on a graph. The results are stated at Sections 2 and 3. Their proofs are done in the last Sections 5, 6 and 7.

Section 2 is devoted to the displacement random walks: Theorem 2.7 gives the Γ-convergence results with in particular $\lim_{k \to \infty} \widehat{P}^k = \widehat{P}$, and Theorems 2.10 and 2.11 describe the Markov dynamics of $\widehat{P}$. In Section 3, the results about displacement interpolations are stated. This section also includes the proof of a Benamou-Brenier type formula at Theorem 3.3 and a discussion about constant speed interpolations and natural substitutes for the geodesics on a graph. Theorem 3.15 is a statement about the conservation of average rate of mass displacement. The Schrödinger problems are introduced at Section 4 where a set of assumptions for the existence of their solutions is also discussed. The Γ-convergence of the sequence of slowed down Schrödinger problems to the optimal transport problem of order one is studied at Section 5. The proofs rely on Girsanov's formula for the Radon-Nykodim density dR^k/dR. The dynamics of the limit $\widehat{P}$ is worked out at Section 6; some effort is needed to show that $\widehat{P}$ is Markov. Finally, the conservation of the average mass displacement along interpolations is proved at last Section 7.

Basic information about random walks and relative entropy with respect to an unbounded measure is provided at the appendix sections A and B.

1. Defining displacement interpolations on a graph by analogy

To stress the analogies between displacement interpolations in discrete and continuous settings, we first recall their main properties on a Riemannian manifold. Then, we briefly introduce the main properties of an object in a discrete setting, see (1.6), which will be defined as a displacement interpolation because of the strong analogies between its properties and the corresponding properties of the displacement interpolations in a continuous setting, see Definitions 1.12.

McCann displacement interpolations. The quadratic Monge-Kantorovich optimal transport problem on the metric space $(\mathcal{X}, d)$ associated with the pair of prescribed probability measures $\mu_0, \mu_1 \in P(\mathcal{X})$ is

$$\int_{\mathcal{X}^2} \frac{1}{2} d^2(x, y) \, \pi(dxdy) \to \min; \qquad \pi \in P(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1 \qquad (\mathrm{MK}_2)$$

where we denote $\pi_0(dx) := \pi(dx \times \mathcal{X})$ and $\pi_1(dy) := \pi(\mathcal{X} \times dy)$ the marginals of $\pi \in P(\mathcal{X}^2)$. The Wasserstein pseudo-distance of order 2 is defined by

$$W_2(\mu_0, \mu_1) := \sqrt{2 \inf (\mathrm{MK}_2)}$$

where $\inf (\mathrm{MK}_2)$ is the value of the minimization problem (MK_2). As usual, we denote

$$P_2(\mathcal{X}) := \left\{ \mu \in P(\mathcal{X}); \int_{\mathcal{X}} d^2(x_o, x) \, \mu(dx) < \infty \right\}$$

for some $x_o \in \mathcal{X}$, so that W_2 is a distance on $P_2(\mathcal{X})$. A displacement interpolation is a minimizing geodesic $[\mu_0, \mu_1] := (\mu_t)_{t \in [0,1]}$ on $(P_2(\mathcal{X}), W_2)$ joining μ_0 and μ_1 in $P_2(\mathcal{X})$, i.e.

$$W_2(\mu_s, \mu_t) = |t - s| \, W_2(\mu_0, \mu_1), \quad s, t \in [0, 1].$$

On a Riemannian manifold $\mathcal{X}$ equipped with its Riemannian distance d, it appears that it is also an action minimizing geodesic in the following sense. Let $\Omega_{\mathrm{ac},2}$ be the space of all absolutely continuous paths $\omega = (\omega_t)_{t \in [0,1]}$ from the time interval $[0, 1]$ to $\mathcal{X}$ such that $\int_{[0,1]} |\dot{\omega}_t|^2_{\omega_t} \, dt < \infty$ where $\dot{\omega}_t$ is the generalized derivative of ω at time t and let $P(\Omega_{\mathrm{ac},2})$ be the corresponding space of probability measures. It appears that $[\mu_0, \mu_1]$ is the time marginal flow (recall notation (0.1))

$$\mu_t = \widehat{P}_t, \quad t \in [0, 1] \tag{1.1}$$

of some solution $\widehat{P} \in P(\Omega_{\mathrm{ac},2})$ of the following dynamical version of (MK_2),

$$\int_{\Omega_{\mathrm{ac},2}} C_{\mathrm{kin}}(\omega) \, P(d\omega) \to \min; \quad P \in P(\Omega_{\mathrm{ac},2}) : P_0 = \mu_0, P_1 = \mu_1 \qquad (\mathrm{MK}_{\mathrm{dyn},2})$$

where the kinetic action C_{kin} is defined by

$$C_{\mathrm{kin}}(\omega) := \int_{[0,1]} \frac{1}{2} |\dot{\omega}_t|^2_{\omega_t} \, dt \in [0, \infty], \quad \omega \in \Omega_{\mathrm{ac},2}.$$

Suppose for simplicity that any solution $\pi^* \in P(\mathcal{X}^2)$ of (MK_2) gives a zero mass to the cut-locus so that there exists a unique minimizing geodesic $\gamma^{xy} \in \Omega_{\mathrm{ac},2}$ joining x and y

for π^*-almost every $x, y \in \mathcal{X}$. Then, any solution $\widehat{P} \in \mathrm{P}(\Omega_{\mathrm{ac},2})$ of $(\mathrm{MK_{dyn,2}})$ is in one-one correspondence with a solution $\widehat{\pi} \in \mathrm{P}(\mathcal{X}^2)$ of $(\mathrm{MK_2})$ via the relation

$$\widehat{P}(\cdot) = \int_{\mathcal{X}^2} \delta_{\gamma^{xy}}(\cdot)\,\widehat{\pi}(dxdy) \in \mathrm{P}(\Omega_{\mathrm{ac},2}) \tag{1.2}$$

where δ stands for a Dirac probability measure. With (1.1), we see that the displacement interpolation $[\mu_0, \mu_1]$ satisfies

$$\mu_t = \int_{\mathcal{X}^2} \delta_{\gamma_t^{xy}}(\cdot)\,\widehat{\pi}(dxdy) \in \mathrm{P}(\mathcal{X}), \quad t \in [0,1]. \tag{1.3}$$

In particular, with $\mu_0 = \delta_x$ and $\mu_1 = \delta_y$, we obtain $[\delta_x, \delta_y] = (\delta_{\gamma_t^{xy}})_{t\in[0,1]}$. This signifies that the notion of displacement interpolation lifts the notion of action minimizing geodesic from the manifold $\mathcal{X}$ onto the Wasserstein space $\mathrm{P}_2(\mathcal{X})$.

It follows from (1.1) and $(\mathrm{MK_{dyn,2}})$ that $W_2(\mu_0, \mu_1)$ admits the Benamou-Brenier representation [BB00]:

$$W_2^2(\mu_0, \mu_1) = \inf_{(\nu,v)} \left\{ \int_{[0,1]\times\mathcal{X}} |v_t(x)|_x^2\,\nu_t(dx)dt \right\} \tag{1.4}$$

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

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

Displacement interpolations on a discrete metric graph $(\mathcal{X}, \sim, d)$. The countable set $\mathcal{X}$ of vertices is endowed with a graph structure: $x \sim y$ means that $x \neq y$ and $\{x, y\}$ is an undirected edge. Let $\Omega \subset \mathcal{X}^{[0,1]}$ be the natural path space for a random walk on the graph $(\mathcal{X}, \sim) : \Omega$ is the space of all left-limited, right-continuous, piecewise constant paths $\omega = (\omega_t)_{0\leq t\leq 1}$ on $\mathcal{X}$ with finitely many jumps such that: $\forall t \in (0,1), \omega_{t^-} \neq \omega_t \Rightarrow \omega_{t^-} \sim \omega_t$. The distance d is in accordance with the graph structure, meaning that it is required to be intrinsic in the discrete sense, that is $d(x, y) = \inf\{\ell(\omega); \omega \in \Omega : \omega_0 = x, \omega_1 = y\}$, $x, y \in \mathcal{X}$ with

$$\ell(\omega) := \sum_{0<t<1} d(\omega_{t^-}, \omega_t), \quad \omega \in \Omega \tag{1.5}$$

the discrete length of the discontinuous path ω. As before, we are going to define the displacement interpolation $[\mu_0, \mu_1] = (\mu_t)_{t\in[0,1]}$ by formula (1.1):

$$\mu_t := \widehat{P}_t, \quad 0 \leq t \leq 1 \tag{1.6}$$

where $\widehat{P} \in \mathrm{P}(\Omega)$ is some *singled out solution* of the following order-one analogue of $(\mathrm{MK_{dyn,2}})$:

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

The random walk $\widehat{P}$ minimizes the average length while transporting the mass distribution μ_0 on $\mathcal{X}$ onto another mass distribution μ_1.

What is meant by "singled out solution" when talking about $\widehat{P}$ will be made precise in a while. For the moment, let us say that $\widehat{P}$ is selected among the infinitely many – see (1.7) below – solutions of $(\mathrm{MK_{dyn}})$, as being the limit of a sequence of solutions of entropy

minimizing problems associated with the slowing down procedure that was invoked when describing the lazy gas experiment. See (1.16) below.

Since d is assumed to be intrinsic, we shall see that the push-forward of (MK_{dyn}) onto $\mathcal{X}^2$ is the Monge-Kantorovich problem

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

It leads to the Wasserstein pseudo-distance of order one

$$W_1(\mu_0, \mu_1) := \inf (MK)$$

which is a distance on

$$P_1(\mathcal{X}) := \left\{ \mu \in P(\mathcal{X}); \int_{\mathcal{X}} d(x_o, x)\,\mu(dx) < \infty \right\}.$$

For each $x, y \in \mathcal{X}$, let us denote

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

the set of all *geodesics* joining x and y. Remark that when x and y are distinct, Γ^{xy} contains infinitely many paths since it is characterized by ordered sequences of visited states, regardless of the instants of jumps. On the other hand, it is easily seen that for any measurable *geodesic kernel* $(Q^{xy} \in P(\Gamma^{xy}); x, y \in \mathcal{X})$ and for any π^* solution of (MK),

$$P^*(\cdot) := \int_{\mathcal{X}^2} Q^{xy}(\cdot)\,\pi^*(dxdy) \in P(\Omega) \tag{1.7}$$

solves (MK_{dyn}). It follows that (MK_{dyn}) admits infinitely many solutions and also that the static and dynamical Monge-Kantorovich problems have the same optimal value:

$$\inf (MK) = \inf (MK_{dyn}). \tag{1.8}$$

The $(X_t)_{t \in [0,1]}$-push forward of the minimizer P^* given at (1.7) is

$$P_t^*(\cdot) = \int_{\mathcal{X}^2} Q_t^{xy}(\cdot)\,\pi^*(dxdy) \in P(\mathcal{X}), \quad 0 \le t \le 1. \tag{1.9}$$

Remark that (1.7)-(1.9) has the same structure as (1.2)-(1.3). The slowing down procedure selects one singled out geodesic kernel $(G^{xy} \in P(\Gamma^{xy}); x, y \in \mathcal{X})$ which does not depend on the specific choice of μ_0 and μ_1, see Theorem 2.7 below, and one singled out solution $\widehat{\pi} \in P(\mathcal{X}^2)$ of (MK) such that

$$\widehat{P}(\cdot) = \int_{\mathcal{X}^2} G^{xy}(\cdot)\,\widehat{\pi}(dxdy) \in P(\Omega) \tag{1.10}$$

and the displacement interpolation $[\mu_0, \mu_1]$ satisfies

$$\mu_t(\cdot) = \int_{\mathcal{X}^2} G_t^{xy}(\cdot)\,\widehat{\pi}(dxdy) \in P(\mathcal{X}), \quad t \in [0,1] \tag{1.11}$$

where for each $x, y \in \mathcal{X}$, $G_t^{xy} \in P(\mathcal{X})$ is the t-marginal of $G^{xy} \in P(\Gamma^{xy})$.

It will be proved that $\widehat{P}$ and every G^{xy} have the Markov property.

The defining identity $\mu_t := \widehat{P}_t$, $t \in [0,1]$, states that $\widehat{P}$ is a dynamical coupling of $[\mu_0, \mu_1]$.

Comparing (1.2) and (1.10) leads us to the following analogies:

- The optimal plan $\widehat{\pi}$ in (1.11) refers to (MK), while in (1.3) it refers to (MK_2).

- The Markov random walk $G^{xy} \in \mathrm{P}(\Gamma^{xy})$ in (1.11) corresponds to $\delta_{\gamma^{xy}}$ in (1.3). In particular, we see that the deterministic behaviour of $\delta_{\gamma^{xy}}$ must be replaced with a genuinely random walk G^{xy}.

The geodesic kernel $(G^{xy} \in \mathrm{P}(\Gamma^{xy}); x, y \in \mathcal{X})$ encodes some geodesic dynamics of the discrete metric graph $(\mathcal{X}, \sim, d)$.

The analogy between (1.2)-(1.3) and (1.10)-(1.11) entitles us to propose the following definitions.

Definitions 1.12. *We call $\mu = (\widehat{P}_t)_{0 \leq t \leq 1}$ the (R, d)-displacement interpolation and $\widehat{P}$ the (R, d)-displacement random walk between μ_0 and μ_1.*
We denote $\mu = [\mu_0, \mu_1]^{(R,d)}$ or more simply $[\mu_0, \mu_1]^R$ or $[\mu_0, \mu_1]$ when the context is clear.

It follows from $\mu_t := \widehat{P}_t, 0 \leq t \leq 1$, ($\mathrm{MK_{dyn}}$) and the Markov property of $\widehat{P}$ that $W_1(\mu_0, \mu_1)$ admits the Benamou-Brenier type representation

$$W_1(\mu_0, \mu_1) = \inf_{(\nu, j)} \int_{[0,1]} dt \int_{\mathcal{X}^2} d(z, w) \, \nu_t(dz) j_{t,z}(dw) < \infty \tag{1.13}$$

where the infimum is taken over all couples (ν, j) such that $\nu = (\nu_t)_{t \in [0,1]} \in \mathrm{P}(\mathcal{X})^{[0,1]}$ is a time-differentiable flow of probability measures on $\mathcal{X}$, $j = (j_{t,z})_{t \in [0,1], z \in \mathcal{X}} \in \mathrm{M}_+(\mathcal{X})^{[0,1] \times \mathcal{X}}$ is a measurable jump kernel, ν and j are linked by the current equation

$$\begin{cases} \partial_t \nu_t(z) + \int_{\mathcal{X}} [\nu_t(z) j_{t,z}(dw) - \nu_t(dw) j_{t,w}(z)] = 0, & 0 < t < 1, \ z \in \mathcal{X} \\ \nu_0 = \mu_0, \ \nu_1 = \mu_1 \end{cases}$$

with $\int_{\mathcal{X}} \nu_t(z) j_{t,z}(dw) < \infty$ for all $0 < t < 1, \ z \in \mathcal{X}$. The infimum $\inf_{(\nu, j)}$ is attained at $(\mu, \widehat{J})$ where μ is the displacement interpolation and $\widehat{J}$ is the jump kernel of the Markov displacement random walk $\widehat{P}$. Hence

$$W_1(\mu_0, \mu_1) = \int_{[0,1]} dt \int_{\mathcal{X}^2} d(z, w) \, \mu_t(dz) \widehat{J}_{t,z}(dw).$$

This is in complete analogy with the standard Benamou-Brenier formula (1.4). The detailed formulation of this Benamou-Brenier formula is given at Theorem 3.3.

The discrete metric measure graph $(\mathcal{X}, \sim, d, m)$. Let us equip $(\mathcal{X}, \sim, d)$ with a positive measure $m \in \mathrm{M}_+(\mathcal{X})$. We introduce a sequence $R_k \in \mathrm{M}_+(\Omega)$ of *slowed down* continuous-time Markov random walks on $\mathcal{X}$ which is respectful of the graph and the metric measure structures of $(\mathcal{X}, \sim, d, m)$.

For each $k \geq 1$, the Markov generator $L^k = (L_t^k)_{0 \leq t \leq 1}$ ot R^k is defined for any finitely supported function u by

$$L_t^k u(x) = \sum_{y : y \sim x} k^{-d(x,y)} J_{t,x}(y) [u(y) - u(x)], \quad t \in [0, 1], x \in \mathcal{X}, k \geq 1. \tag{1.14}$$

In this formula $J_{t,x}(y) > 0$ is the average rate of jumps of the random walk from x to y at time t when $k = 1$. For some detail about random walks, see the appendix Section A. As k increases, the sequence of generators $(L^k)_{k \geq 1}$ tends down to zero. It describes the dynamics of slowed down random walks which ultimately do not move anymore. We call them *lazy random walks*. For each k, the random walk with generator L^k and initial measure $m \in \mathrm{M}_+(\mathcal{X})$ is described by a positive measure

$$R^k \in \mathrm{M}_+(\Omega)$$

on the path space Ω (the letter R stands for *reference* measure).

We also assume that the graph $(\mathcal{X}, \sim)$ is irreducible and that the initial positive measure $m = (m_x)_{x \in \mathcal{X}}$ and the jump kernel J satisfy the detailed balance conditions

$$m_x J_{t,x}(y) = m_y J_{t,y}(x), \quad \forall x, y \in \mathcal{X}, t \in [0,1]. \tag{1.15}$$

Together with the irreducibility assumption, these conditions imply that $m_x > 0, \forall x \in \mathcal{X}$, whenever $m(\mathcal{X}) > 0$, which is always assumed. A large class of m-reversible random walks is described at the appendix Section A, see (A.5).

Let us comment on the connection between this sequence of lazy random walks and the metric graph measure structure of $(\mathcal{X}, \sim, d, m)$.

($\sim$) As for each k, t and x, the support of the jump measure $J_{t,x} = \sum_{y:y \sim x} J_{t,x}(y)\, \delta_y \in \mathrm{M}_+(\mathcal{X})$ is the set of neighbours of x, the random walker which is governed by R^k is allowed to jump from x to y if and only if x and y are neighbours.

(m) Because of (1.15), the Markov random walks R^k are assumed to be m-stationary, i.e. for all $t \in [0,1]$ the t-marginal R_t^k of R^k satisfies $R_t^k = m$. Furthermore, for all $0 \le t \le 1$ and $k \ge 1$, the generators L_t^k are self-adjoint on $L^2(\mathcal{X}, m)$.

(d) The connection with the distance d is less immediate. For any $x, y \in \mathcal{X}$, we shall prove at Theorem 2.9 that the sequence of bridges $R^{k,xy} \in \mathrm{P}(\Omega)$ from x to y (see notation (0.3)) converges to a probability measure which is concentrated on the set Γ^{xy} of all geodesics joining x and y.

Therefore, in some sense, the sequence $(R^k)_{k \ge 1}$ of lazy random walks is respectful of the discrete metric measure graph structure of $(\mathcal{X}, \sim, d, m)$.

Optimal transport appears at the limit of the slowing down procedure. Let us briefly sketch the connection between lazy random walks and optimal transport. The main idea is to consider, as in [Léo12a], a sequence of entropy minimizing problems which are called Schrödinger problems. This is the mathematical implementation of the lazy gas experiment that was described in the introduction. Details about the connection between the lazy gas experiment and the Schrödinger problem are given in the author's survey paper [Léo14, §6].

Recall that the relative entropy of a probability measure p on the measurable space Y with respect to some reference positive measure $r \in \mathrm{M}_+(Y)$ on Y is roughly defined by

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

if p is absolutely continuous with respect to r and $+\infty$ otherwise. For a rigorous definition, see the appendix Section B. For each $k \ge 2$, consider the entropy minimizing problem that is called a dynamical *Schrödinger problem*

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

See [Léo14] for more detail about Schrödinger's problem. It should be compared with ($\mathrm{MK}_{\mathrm{dyn}}$). Note that unlike the affine minimization problem ($\mathrm{MK}_{\mathrm{dyn}}$), ($\mathrm{S}_{\mathrm{dyn}}^k$) is a *strictly* convex minimization problem. Hence, ($\mathrm{S}_{\mathrm{dyn}}^k$) admits at most one solution while ($\mathrm{MK}_{\mathrm{dyn}}$) admits an infinite convex set of solutions.

It is proved at Lemma 5.5 that

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

where this Γ-convergence refers to the standard narrow topology $\sigma(\mathrm{P}(\Omega), C_b(\Omega))$ on $\mathrm{P}(\Omega)$ when Ω is endowed with the Skorokhod topology, see (5.2) for detail about Ω. For each k, let $\widehat{P}^k \in \mathrm{P}(\Omega)$ denote the *unique* solution of ($\mathrm{S}_{\mathrm{dyn}}^k$). As a consequence of this Γ-limit, one

expects that any limit point of $(\widehat{P}^k)_{k\geq 1}$ solves $(\mathrm{MK_{dyn}})$. We shall do better at Theorem 2.7 which states that $(\widehat{P}^k)_{k\geq 1}$ is a convergent sequence and that

$$\lim_{k\to\infty} \widehat{P}^k := \widehat{P} \in \mathrm{P}(\Omega) \qquad (1.16)$$

is represented by (1.10) where the geodesic kernel $(G^{xy})_{x,y}$ is expressed in terms of the geodesic sets $(\Gamma^{xy})_{x,y}$, J and $R := R^{k=1}$.

Pushing forward (S^k_{dyn}) from $\mathrm{P}(\Omega)$ onto $\mathrm{P}(\mathcal{X}^2)$ via the $(0,1)$-marginal projection (X_0, X_1) gives us

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

where $R^k_{01} \in \mathrm{P}(\mathcal{X}^2)$ is the joint law of the initial and final positions of the random walk R^k, see notation (0.2). A consequence of $\Gamma\text{-}\lim_{k\to\infty} (\mathrm{S}^k_{dyn}) = (\mathrm{MK_{dyn}})$ is

$$\Gamma\text{-}\lim_{k\to\infty} (\mathrm{S}^k) = (\mathrm{MK})$$

and it will proved that for each k, the unique solution $\widehat{\pi}^k \in \mathrm{P}(\mathcal{X}^2)$ of (S^k) is the joint law of the initial and final positions of the random walk $\widehat{P}^k$, i.e. $\widehat{\pi}^k := \widehat{P}^k_{01}$. Therefore, the slowing down procedure $\lim_{k\to\infty} \widehat{P}^k = \widehat{P}$ selects

(1) one solution $\widehat{\pi} := \lim_{k\to\infty} \widehat{\pi}^k \in \mathrm{P}(\mathcal{X}^2)$ of the static problem (MK) and
(2) one random dynamics encoded in $(G^{xy}; x, y \in \mathcal{X})$.

It is interesting to note that the convergence $\lim_{k\to\infty} \widehat{\pi}^k \in \mathrm{P}(\mathcal{X}^2) = \widehat{\pi}$ is a by-product of its dynamical analogue.

2. Main results about displacement random walks

We gather our assumptions before stating our main results about the displacement random walks.

The underlying hypotheses. The following set of hypotheses will prevail for the rest of the paper.

Hypotheses 2.1. *The vertex set $\mathcal{X}$ is countable.*

($\sim$) - *$(\mathcal{X}, \sim)$ is irreducible: for any $x, y \in \mathcal{X}$, there exists a finite chain $x_1, x_2, \ldots, x_n$ in $\mathcal{X}$ such that $x = x_1 \sim x_2 \sim \cdots \sim x_n = y$.*
- *$(\mathcal{X}, \sim)$ contains no loop: $x \sim x$ is forbidden.*
- *$(\mathcal{X}, \sim)$ is locally finite: any vertex $x \in \mathcal{X}$ admits finitely many neighbours*

$$n_x := \# \{y \in \mathcal{X}; y \sim x\} < \infty, \quad \forall x \in \mathcal{X}. \qquad (2.2)$$

(d) - *The distance d is positively lower bounded: for all $x \neq y \in \mathcal{X}$, $d(x,y) \geq 1$.*
- *The distance d is intrinsic in the discrete sense:*

$$d(x,y) = \inf \{\ell(\omega); \omega \in \Omega : \omega_0 = x, \omega_1 = y\}, \quad x, y \in \mathcal{X}$$

where the discrete length ℓ is defined at (1.5).

(R) *The reference path measure $R \in \mathrm{M}_+(\Omega)$ is assumed to be Markov with a forward jump kernel $(J_{t,x} \in \mathrm{M}_+(\mathcal{X}); t \in [0,1], x \in \mathcal{X})$ such that:*
- *For any $x, y \in \mathcal{X}$, we have $J_{t,x}(y) > 0$, $\forall t \in [0,1]$ if and only if $x \sim y$.*
- *J is uniformly bounded, i.e.*

$$\sup_{t\in[0,1], x\in\mathcal{X}} J_{t,x}(\mathcal{X}) < \infty. \qquad (2.3)$$

(R^k) *For each $k \geq 1$, the slowed down random walk $R^k \in M_+(\Omega)$ is the Markov measure with the forward jump kernel*

$$J_{t,x}^k := \sum_{y:y \sim x} k^{-d(x,y)} J_{t,x}(y) \, \delta_y, \qquad t \in [0,1], x \in \mathcal{X}, \tag{2.4}$$

and the initial measure is $R_0^k = m \in M_+(\mathcal{X})$ with $m_x > 0$ for all $x \in \mathcal{X}$.

(μ) *The prescribed probability measures μ_0 and $\mu_1 \in P(\mathcal{X})$ satisfy the following requirements. There exists some $\pi^o \in P(\mathcal{X}^2)$ such that $\pi_0^o = \mu_0$, $\pi_1^o = \mu_1$, $\int_{\mathcal{X}^2} E_{R^{xy}}(\ell) \, \pi^o(dxdy) < \infty$ and $H(\pi^o|R_{01}) < \infty$.*

We give a simple criterion for the Hypothesis (μ) to be verified. Remark that for the problems (S_{dyn}^k) and (S^k) to admit solutions, it is necessary that $H(\mu_0|R_0), H(\mu_1|R_1) < \infty$.

Proposition 2.5. *For the Hypothesis 2.1-(μ) to be satisfied, it is enough that in addition to $H(\mu_0|R_0), H(\mu_1|R_1) < \infty$, there exists a nonnegative function A on $\mathcal{X}$ such that*

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

(ii) $R_{01}(dxdy) \geq e^{-A(x)-A(y)} R_0(dx) R_1(dy)$

(iii) $E_{R^{xy}}(\ell) \leq A(x) + A(y)$, *for all $x, y \in \mathcal{X}$*

(iv) $\int_{\mathcal{X}} A \, d\mu_0, \int_{\mathcal{X}} A \, d\mu_1 < \infty$.

In particular, Hypothesis 2.1-(μ) holds when $\mathcal{X}$ is finite.

The proof of Proposition 2.5 is done at Section 4.

The minimization problems. Let us recall for convenience the Monge-Kantorovich problems

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

and

$$\int_{\Omega} \ell(\omega) \, P(d\omega) \to \min; \qquad P \in P(\Omega) : P_0 = \mu_0, P_1 = \mu_1 \tag{MK$_{dyn}$}$$

and the Schrödinger problems which are defined for each $k \geq 1$, by

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

and

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

We denote

$$\Gamma := \cup_{x,y \in \mathcal{X}} \Gamma^{xy}$$

the set of all geodesics. It is proved at Lemma 5.11 that it is measurable, so that one is allowed to define the path measure

$$G := \mathbf{1}_\Gamma \exp\left(\int_{[0,1]} J_{t,X_t}(\mathcal{X}) \, dt\right) R \in M_+(\Omega). \tag{2.6}$$

Let us denote $\mathcal{S}_{MK}(\mu_0, \mu_1) \subset P(\mathcal{X}^2)$ the set of all solutions to the Monge-Kantorovich problem (MK) and introduce the subsequent auxiliary entropic minimization problems

$$H(\pi|G_{01}) \to \min; \qquad \pi \in \mathcal{S}_{MK}(\mu_0, \mu_1). \tag{$\widetilde{S}$}$$

and

$$H(P|G) \to \min; \qquad P \in P(\Omega) : P_{01} \in \mathcal{S}_{MK}(\mu_0, \mu_1). \tag{$\widetilde{S}_{dyn}$}$$

Results about the displacement random walks. We are now ready to state the main results about the random walks. Their consequences in terms of interpolations will be made precise at next section.

Theorem 2.7. *The Hypotheses 2.1 are assumed to hold.*

(1) *For all $k \geq 2$, the problems (S^k) and (S^k_{dyn}) admit respectively a unique solution $\widehat{\pi}^k \in \mathrm{P}(\mathcal{X}^2)$ and $\widehat{P}^k \in \mathrm{P}(\Omega)$.*
    Moreover, $\widehat{P}^k$ is Markov and $\widehat{\pi}^k = \widehat{P}^k_{01}$.
(2) *$(\widetilde{S})$ has a unique solution $\widehat{\pi} \in \mathrm{P}(\mathcal{X}^2)$ and $\lim_{k \to \infty} \widehat{\pi}^k = \widehat{\pi}$.*
    As a definition, $\widehat{\pi}$ also solves (MK).
(3) *$(\widetilde{S}_{dyn})$ has a unique solution $\widehat{P} \in \mathrm{P}(\Omega)$ and $\lim_{k \to \infty} \widehat{P}^k = \widehat{P}$.*
    The limit $\widehat{P}$ also solves (MK_{dyn}).
(4) *$\widehat{P}$ is the following mixture of bridges of G:*

$$\widehat{P}(\cdot) = \int_{\mathcal{X}^2} G^{xy}(\cdot)\, \widehat{\pi}(dxdy) \in \mathrm{P}(\Omega), \tag{2.8}$$

meaning that it satisfies $\widehat{P}_{01} = \widehat{\pi}$ and that $\widehat{P}$ shares its bridges with the geodesic path measure G defined by (2.6): $\widehat{P}^{xy} = G^{xy}$ for $\widehat{\pi}$-almost every (x, y).

The proof of Theorem 2.7-(1) is done at the end of Section 4 and the proof of Theorem 2.7-(2-3-4) is done at the end of Section 5.

As a corollary, we obtain the following result.

Theorem 2.9. *For any $x, y \in \mathcal{X}$ such that $E_{R^{xy}}(\ell) < \infty$, the sequence $(R^{k,xy})_{k \geq 1}$ of bridges of $(R^k)_{k \geq 1}$ is convergent and $\lim_{k \to \infty} R^{k,xy} = G^{xy}$.*

Proof. Under the marginal constraints $\mu_0 = \delta_x$ and $\mu_1 = \delta_y$, we have for all $k \geq 2$, $\widehat{\pi}^k = \widehat{\pi} = \delta_{(x,y)}$ and $\widehat{P}^k = R^{k,xy}$ by (4.2). It remains to apply Theorem 2.7. □

We need some additional preliminary material to describe the dynamics of $\widehat{P}$ and of the bridge G^{xy}. Recall that a directed tree is a directed graph $(\mathcal{Z}, \to)$ that contains no circuit (directed loop). We denote $z \to z'$ when the directed edge $(z, z') \in \mathcal{Z}^2$ exists and we define the order relation $\preceq$ by: $z \preceq z'$ if $z = z'$ or if there exists a finite path $z = z_1 \to z_2 \to \cdots \to z_n = z'$.
Unlike the following configuration (a),
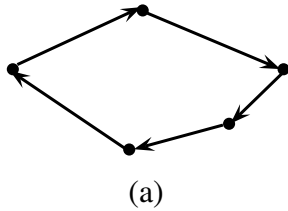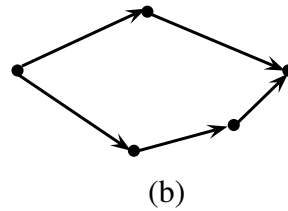
(a)   (b)

configuration (b) is not a circuit and it may enter a directed tree.

We have in mind the directed tree $(\Gamma^{xy}([0, 1]), \to)$ related to the set Γ^{xy} of all the geodesics from x to y on $(\mathcal{X}, \sim)$, where $z \to z' \in \Gamma^{xy}([0, 1])$ if $z \sim z' \in \mathcal{X}$ and there are some $\gamma \in \Gamma^{xy}$ and $0 \leq t < t' \leq 1$ such that $\gamma_t = z$ and $\gamma_{t'} = z'$. This tree describes the successive occurrence of the states which are visited by the geodesics from x to y. It keeps the information of the order of occurrence, but it is regardless of the instants of jump.

Theorem 2.10 (The dynamics of G^{xy}). *The Hypotheses 2.1 are assumed to hold.*

(1) *Although G is not Markov in general, for every $x, y \in \mathcal{X}$, its bridge G^{xy} is Markov.*

(2) *For every $x, y \in \mathcal{X}$, the jump kernel of the Markov measure G^{xy} is given by*

$$J_{t,z}^{G,y} = \sum_{w \in \{z \to \cdot\}^y} \frac{g_t^y(w)}{g_t^y(z)} J_{t,z}(w) \, \delta_w, \qquad 0 \le t < 1, \; z \in \Gamma^{xy}([0,1]),$$

where $\{z \to \cdot\}^y := \{w \in \Gamma^{zy}([0,1]); z \to w\}$ is the set of all successors of z in the directed tree $(\Gamma^{zy}([0,1]), \to)$ and

$$g_t^y(z) := E_R\left[\exp\left(\int_t^1 J_{s,X_s}(\mathcal{X}) \, ds \right) \mathbf{1}_{\Gamma(t,z;1,y)} \mid X_t = z \right]$$

with

$$\Gamma(t, z; 1, y) := \left\{ \omega \in \Omega; \omega_{|[t,1]} = \gamma_{|[t,1]} \text{ for some } \gamma \in \Gamma, \omega_t = z, \omega_1 = y \right\},$$

the set of all geodesics from z to y on the time interval $[t, 1]$.

Remark that since G^{xy} only visits $\Gamma^{xy}([0,1])$, one can put $J_{t,z}^{G,y} = 0$ for any $z \notin \Gamma^{xy}([0,1])$.

The proof of Theorem 2.10 is given at Section 6.

Theorem 2.11 (The dynamics of $\widehat{P}$)**.** *The Hypotheses 2.1 are assumed to hold.*

(1) *The limiting random walk $\widehat{P}$ is Markov.*

(2) *The jump kernel of the Markov measure $\widehat{P} \in \mathrm{P}(\Omega)$ is given by*

$$\widehat{J}_{t,z}(\cdot) = \int_{\mathcal{X}} J_{t,z}^{G,y}(\cdot) \, \widehat{P}(X_1 \in dy | X_t = z).$$

It is a mixture of the jump kernels $J^{G,y}$ of G^{xy}, see Theorem 2.10.

The statement of Theorem 2.11-(1) is the content of Proposition 6.7 which is proved at Section 6. The second statement is proved at the end of Section 6.

Gathering Theorems 2.10 and 2.11, one obtains for all t and z such that $\mu_t(z) > 0$,

$$\widehat{J}_{t,z}(\cdot) = \int_{\mathcal{X}} \left(\sum_{w \in \{z \to \cdot\}^y} \frac{E_R[\exp(\int_t^1 J_{s,X_s}(\mathcal{X}) \, ds) \mathbf{1}_{\Gamma(t,w;1,y)} \mid X_t = w]}{E_R\left[\exp(\int_t^1 J_{s,X_s}(\mathcal{X}) \, ds) \mathbf{1}_{\Gamma(t,z;1,y)} \mid X_t = z \right]} J_{t,z}(w) \, \delta_w(\cdot) \right) \tag{2.12}$$
$$\times \quad \widehat{P}(X_1 \in dy | X_t = z).$$

3. Main results about displacement interpolations

We have defined the displacement interpolation $[\mu_0, \mu_1]$ as the marginal flow $\mu_t := \widehat{P}_t$, $0 \le t \le 1$, of the displacement random walk $\widehat{P}$. A direct consequence of Theorem 2.11 is the

Corollary 3.1 (The dynamics of μ)**.** *The Hypotheses 2.1 are assumed to hold. The displacement interpolation $[\mu_0, \mu_1]$ solves the following evolution equation*

$$\begin{cases} \partial_t \mu_t(z) = \sum_w [\mu_t(w) \widehat{J}_{t,w}(z) - \mu_t(z) \widehat{J}_{t,z}(w)], & 0 \le t \le 1, z \in \mathcal{X}, \\ \mu_0, & t = 0. \end{cases}$$

A Benamou-Brenier type formula. Let us have a closer look at the Benamou-Brenier type formula (1.13).

Definition 3.2 (Change of time). *A change of time τ is an absolutely continuous function $\tau : [0,1] \to [0,1]$ such that $\tau(0) = 0$, $\tau(1) = 1$ and with a nonnegative generalized derivative $0 \le \dot{\tau} \in L^1([0,1])$.*

For any change of time τ and any measure $Q \in M_+(\Omega)$, we denote

$$Q^\tau := (X_\tau)_\# Q$$

where $X_\tau(t) := X_{\tau(t)}, t \in [0,1]$. For any flow ν of probability measures and any jump kernel $\mathcal{J}$, we denote $\nu \mathcal{J}_t(dxdy) := \nu_t(dx)\mathcal{J}_{t,x}(dy)$ and $\nu \mathcal{J}_t(\mathcal{X}^2) := \int_{\mathcal{X}^2} \nu_t(dx)\mathcal{J}_{t,x}(dy)$.

The following theorem is a consequence of the Markov property of $\widehat{P}$ which was stated at Theorem 2.11.

Theorem 3.3 (A Benamou-Brenier type formula). *Suppose that the Hypotheses 2.1 are satisfied.*

(1) We have

$$W_1(\mu_0, \mu_1) = \inf_{\nu,j} \int_{[0,1]} dt \int_{\mathcal{X}^2} d(z,w)\,\nu j_t(dzdw) < \infty \tag{3.4}$$

where the infimum is taken over all couples (ν, j) such that $\nu = (\nu_t)_{t\in[0,1]} \in P(\mathcal{X})^{[0,1]}$ is a time-differentiable flow of probability measures on $\mathcal{X}$, $j = (j_{t,z})_{t\in[0,1],z\in\mathcal{X}} \in M_+(\mathcal{X})^{[0,1]\times\mathcal{X}}$ is a measurable jump kernel, ν and j are linked by

$$\begin{cases} \partial_t \nu_t(z) + \int_{\mathcal{X}}[\nu_t(z)j_{t,z}(dw) - \nu_t(dw)j_{t,w}(z)] = 0, & 0 < t < 1, \ z \in \mathcal{X} \\ \nu_0 = \mu_0, \ \nu_1 = \mu_1 \end{cases} \tag{3.5}$$

with $\int_{\mathcal{X}} \nu_t(z)j_{t,z}(dw) < \infty$ for all $0 < t < 1, \ z \in \mathcal{X}$.

(2) The infimum $\inf_{\nu,j}$ in (3.4) is attained at $(\mu, \widehat{J})$ where μ is the displacement interpolation and $\widehat{J}$ is the jump kernel of $\widehat{P}$. Hence

$$W_1(\mu_0, \mu_1) = \int_{[0,1]} dt \int_{\mathcal{X}^2} d(z,w)\,\mu\widehat{J}_t(dzdw).$$

(3) For any change of time τ, the infimum $\inf_{\nu,j}$ in (3.4) is also attained at $(\mu^\tau, \widehat{J}^\tau)$ where μ^τ is the displacement interpolation and $\widehat{J}^\tau$ is the jump kernel that are associated with $\widehat{P}^\tau$. Moreover, $\widehat{P}^\tau$ is the displacement random walk associated with R^τ, i.e. the analogue of $\widehat{P}$ when R is replaced with R^τ and

$$W_1(\mu_0, \mu_1) = \int_{[0,1]} dt \int_{\mathcal{X}^2} d(z,w)\,\mu\widehat{J}_t^\tau(dzdw)$$

where for almost every $t \in [0,1]$, $\mu\widehat{J}_t^\tau := \dot{\tau}(t)\mu\widehat{J}_{\tau(t)}$ is the mass displacement distribution of $\widehat{P}^\tau$ at time t.

Proof. • Proof of (1) and (2). With (1.8) and Theorem 2.7-(3), we have

$$W_1(\mu_0, \mu_1) = \inf(MK) = \inf(MK_{dyn}) = E_{\widehat{P}}(\ell).$$

But, for any Markov random walk $P \in P(\Omega)$ on $(\mathcal{X}, \sim)$ with jump kernel j and such that $E_P(\ell) < \infty$, we have

$$E_P(\ell) = E_P \int_{[0,1]\times\mathcal{X}} d(X_t, y)\,j_{t,X_t}(dy)dt = \int_{[0,1]} dt \int_{\mathcal{X}^2} d(z,w)\,P_t(dz)j_{t,z}(dw).$$

This proves that $W_1(\mu_0, \mu_1) = \inf \{E_P(\ell); P \text{ Markov on } (\mathcal{X}, \sim) : P_0 = \mu_0, P_1 = \mu_1\} = E_{\widehat{P}}(\ell)$, which is the announced result since $\nu_t := P_t$ solves the Fokker-Planck equation in (3.5).

• Proof of (3). We denote P^* the displacement random walk associated to $(\mathrm{S}^k_{\mathrm{dyn}})_{k \geq 2}$ with R^τ instead of R. As X_τ is injective, we have $H(P|R^k) = H(P^\tau|R^{\tau,k})$ for all $P \in \mathrm{P}(\Omega)$ and $k \geq 1$. This implies that $P^* = \widehat{P}^\tau$. Hence (3) follows from (2). $\qquad\square$

Constant speed and minimizing displacement interpolations. Next proposition is another consequence of the Markov property of $\widehat{P}$.

Proposition 3.6. *For all $0 \leq s \leq t \leq 1$, let $\widehat{P}_{st} := (X_s, X_t)_{\#}\widehat{P} \in \mathrm{P}(\mathcal{X}^2)$ be the joint law of the positions at time s and t under $\widehat{P}$. Then,*

(1) $\widehat{P}_{st} \in \mathrm{P}(\mathcal{X}^2)$ is an optimal coupling of μ_s and μ_t, meaning that $\widehat{P}_{st}$ is a solution of (MK) with μ_s and μ_t as prescribed marginal constraints;

(2) $W_1(\mu_s, \mu_t) = \int_{[s,t]} dr \int_{\mathcal{X}^2} d(z,w) \, \mu\widehat{J}_r(dzdw).$

Proof. Both statements are consequences of

- the Markov property of $\widehat{P}$, see Theorem 2.11, which allows for surgery by gluing the bridges of $\widehat{P}_{[s,t]}$ together with the restrictions $\widehat{P}_{[0,s]}$ and $\widehat{P}_{[t,1]}$, where we denote $P_{[u,v]} := (X_t; u \leq t \leq v)_{\#}P$;
- the fact that $\ell_{st} := \sum_{s < r < t} d(X_{r^-}, X_r)$ is insensitive to changes of time: i.e. for any strictly increasing mapping $\theta : [s,t] \to [0,1]$ with $\theta(s) = 0$, $\theta(t) = 1$, we have $\ell_{st} = \ell_{01}(X_\theta)$.

A standard ad absurdum reasoning leads to (1). Statement (2) follows from (1), a change of variables formula based on any absolutely continuous change of time $\theta : [s,t] \to [0,1]$ and the general identity

$$\mathcal{J}^\theta_r = \dot{\theta}(r)\mathcal{J}_{\theta(r)}, \quad \text{for almost every } r \in (s,t) \tag{3.7}$$

where $(\mathcal{J}_u; u \in [0,1])$ is any jump kernel and $(\mathcal{J}^\theta_r; r \in [s,t])$ the jump kernel resulting from the mapping X_θ. $\qquad\square$

Proposition 3.6-(2) entitles us to define the speed of the displacement interpolation μ at time t by

$$\mathrm{speed}(\mu)_t := \int_{\mathcal{X}^2} d(z,w) \, \mu\widehat{J}_t(dzdw), \quad 0 \leq t \leq 1, \tag{3.8}$$

to obtain

$$W_1(\mu_0, \mu_1) = \int_{[0,1]} \mathrm{speed}(\mu)_t \, dt.$$

For any change of time $\tau : [0,1] \to [0,1]$, we see with this identity, (3.7) and the change of variable formula that

$$W_1(\mu_0^\tau, \mu_1^\tau) = W_1(\mu_0, \mu_1).$$

Hence, there are infinitely many μ^τ that minimize the action in formula (3.4). This is the content of Theorem 3.3-(3). On the other hand, next result states that only one of them has a constant speed.

Proposition 3.9. *Under the Hypotheses 2.1, there exists a unique change of time τ_o such that μ^{τ_o} has a constant speed, i.e.*

$$W_1(\mu_s^{\tau_o}, \mu_t^{\tau_o}) = (t - s)W_1(\mu_0, \mu_1), \quad \forall 0 \leq s \leq t \leq 1.$$

Proof. Indeed, this equation is equivalent to

$$\dot{\tau}(s)\psi(\tau(s)) = W_1(\mu_0, \mu_1), \quad \text{a.e.,} \tag{3.10}$$

where $\psi(t) := \int_{\mathcal{X}^2} d(z,w)\, \mu \widehat{J}_t(dzdw)$, a.e. Clearly, the assumption that d is uniformly lower bounded and (2.12) imply that $\psi > 0$. Hence, a solution of (3.10) is given by

$$\tau_o(s) = \Psi_{\mu_0,\mu_1}^{-1}(W_1(\mu_0,\mu_1)\, s), \quad s \in [0,1] \tag{3.11}$$

where for all $t \in [0,1]$,

$$0 \le \Psi_{\mu_0,\mu_1}(t) := \int_{[0,t]} \psi(r)\, dr = \int_{[0,t]} dr \int_{\mathcal{X}^2} d(z,w)\, \mu \widehat{J}_r(dzdw) \le W_1(\mu_0,\mu_1) < \infty.$$

Let us prove the uniqueness. Remark that, as a continuous strictly monotone function, τ_o is bijective. In addition, it is absolutely continuous. Hence, any change of time τ is equal to $\tau_o \circ \sigma$ for some change of time σ. Now, instead of starting from μ, let us do a change of time σ on μ^{τ_o}. Defining $\psi_o(u) := \int_{\mathcal{X}^2} d(z,w)\, \mu \widehat{J}_u^{\tau_o}(dzdw)$, a.e. instead of ψ, we arrive similarly at $\dot{\sigma}(u)\psi_o(\sigma(u)) = W_1(\mu_0,\mu_1)$, a.e. But, $\psi_o(u) = W_1(\mu_0,\mu_1)$ for all u. Hence, $\dot{\sigma} = 1$, from which the desired result follows. $\square$

Definition 3.12 (Constant speed displacement interpolation). *The time changed displacement interpolation μ^{τ_o} with τ_o given at (3.11) is called a constant speed displacement interpolation.*

As a definition, a constant speed displacement interpolation is a minimizing geodesic between μ_0 and μ_1 in the Wasserstein space $(\mathrm{P}_1(\mathcal{X}), W_1)$. But the general definition of a minimizing geodesic is misleading in the present non-strictly convex setting, since there are infinitely many *action minimizing* W_1-geodesics which do not have a constant speed.

One must be aware that, in general, the change of time τ_o depends on μ_0 and μ_1. Nevertheless, we shall see below at Theorem 3.15 that in the special important case where the distance d is the standard graph distance $d_\sim$ specified by

$$\forall x, y \in \mathcal{X}, \quad d_\sim(x,y) = 1 \iff x \sim y, \tag{3.13}$$

for any μ_0, μ_1, the displacement interpolation $[\mu_0, \mu_1]$ has a constant speed.

Let us go back to a general discrete metric measure graph $(\mathcal{X}, \sim, d, m)$ where d might differ from the standard graph distance $d_\sim$. When transferring analogically the Lott-Sturm-Villani theory to this graph setting, it is likely that one should consider a reference random walk R that satisfies the detailed balance conditions (1.15):

$$m_x J_{t,x}(y) = m_y J_{t,y}(x), \quad \forall x, y \in \mathcal{X}, t \in [0,1],$$

for being respectful of the metric measure graph structure. Indeed, this implies that for all $0 \le t \le 1$ and $k \ge 1$, $R_t^k = m$ and also that the generators L_t^k, see (1.14), are self-adjoint on $L^2(\mathcal{X}, m)$.

Pick μ_0 and μ_1 and consider the associated constant speed interpolation μ^{τ_o}. Since, for any change of time τ, the mapping $Q \in \mathrm{M}_+(\Omega) \mapsto Q^\tau \in \mathrm{M}_+(\Omega)$ is injective, we have $H(P^\tau|R^\tau) = H(P|R)$ for any $P \in \mathrm{P}(\Omega)$. This implies that μ^{τ_o} is the displacement interpolation associated with R^{τ_o} :

$$\mu^{\tau_o} = [\mu_0, \mu_1]^R \circ \tau_o = [\mu_0, \mu_1]^{R^{\tau_o}}.$$

But, since $J_t^\tau = \dot{\tau}(t) J_{\tau(t)}$, R^{τ_o} inherits of the detailed balance condition satisfied by R :

$$m_x J_{t,x}^{\tau_o}(y) = m_y J_{t,y}^{\tau_o}(x), \quad \forall x, y \in \mathcal{X}, t \in [0,1].$$

Therefore, the constant speed displacement interpolation μ^{τ_o} is really a displacement interpolation in the sense of Definition 1.12 with respect to the reference random walk R^{τ_o} which is respectful of the discrete metric measure graph $(\mathcal{X}, \sim, d, m)$.

Conservation of the average rate of mass displacement. Next result tells us that along any displacement interpolation $[\mu_0, \mu_1]$, the average rate of mass displacement, as defined below, doesn't depend on time.

Definitions 3.14 (Rate of mass displacement)**.** *For any Markov random walk $P \in \mathrm{P}(\Omega)$ with jump kernel $(j_{t,x}; t \in [0,1], x \in \mathcal{X})$, we denote $\nu_t = P_t \in \mathrm{P}(\mathcal{X})$ and call*

$$\nu j_t(dxdy) := \nu_t(dx) j_{t,x}(dy), \qquad 0 \le t \le 1$$

the distribution of the rate of mass displacement *of P at time t.*
We also call

$$\nu j_t(\mathcal{X}^2) := \int_{\mathcal{X}^2} \nu_t(dx) j_{t,x}(dy), \qquad 0 \le t \le 1$$

the average rate of mass displacement *of P at time t.*

Theorem 3.15 (Conservation of the average rate of mass displacement)**.** *Suppose that the Hypotheses 2.1 are satisfied. Let $\widehat{J}$ be the jump kernel of the displacement random walk $\widehat{P}$ and μ the corresponding displacement interpolation. There exists some $K > 0$ such that*

$$\mu \widehat{J}_t(\mathcal{X}^2) = K, \quad \forall t \in [0,1].$$

In particular, when the distance d is the standard discrete distance $d_\sim$, see (3.13), the displacement interpolation μ has a constant speed.

Last statement simply relies on the remark that when $d = d_\sim$, the speed of μ coincides with its average rate of mass displacement.

Theorem 3.15 is a restatement of Theorem 7.4 which is proved at Section 7.

Corollary 3.16. *The constant speed displacement interpolation μ^{τ^o} defined at Definition 3.12 has also a constant average rate of mass displacement.*

Proof. Since $\mu^{\tau^o} = [\mu_0, \mu_1]^{R^{\tau^o}}$ is the R^{τ^o}-displacement interpolation, one can apply Theorem 3.15. $\qquad\square$

Natural substitutes for the constant speed geodesics on a discrete metric graph. Let R be given. When specifying $\mu_0 = \delta_x$ and $\mu_1 = \delta_y$, the displacement random walk $\widehat{P}$ is simply G^{xy}. Moreover, there exists a unique change of time τ^{xy} such that

$$\mu^{xy} := [\delta_x, \delta_y] \circ \tau^{xy} = G^{xy}_{\tau^{xy}}$$

has a constant speed. Its dynamics is given by the current equation

$$\begin{cases} \partial_t \mu_t^{xy}(z) - \dot{\tau}_t^{xy} \sum_w [\mu_t^{xy}(w) J^{G,y}_{\tau_t^{xy},w}(z) - \mu_t^{xy}(z) J^{G,y}_{\tau_t^{xy},z}(w)] = 0, & 0 \le t \le 1, z \in \mathcal{X}, \\ \mu_0^{xy} = \delta_x, & t = 0. \end{cases}$$

The (R,d)-displacement interpolation μ^{xy} is a natural time-continuous averaging of the piecewise constant paths $t \mapsto \delta_{\gamma(t)}$ with γ in the set Γ^{xy} of all d-geodesics joining x and y. It depends on the choice of the reference random walk R.

Binomial interpolations. We present some easy examples of constant speed interpolations μ^{xy}. Let us consider the simplest and important setting where R is the reversible simple random walk and the distance $d = d_\sim$ is the standard graph distance. The jump kernel is described at (A.3): $J_x(y) = 1/n_x$, $\forall x \sim y$, and with the initial "volume measure" given at (A.4): $m_x = n_x$, $\forall x \in \mathcal{X}$.

Let x and y be fixed . We know by Theorem 3.15 that the displacement interpolation $\mu^{xy} = [\delta_x, \delta_y] = (G_t^{xy})_{0 \le t \le 1}$ has a constant speed. The dynamics of G^{xy} is specified at Theorem 2.10, for any $t \in [0, 1)$, $z \in \Gamma^{xy}([0, 1])$ and $w \in \{z \to \cdot\}^y$, by

$$J_{t,z}^{G,y}(w) = \mathbf{1}_{\{z \ne y\}} n_z^{-1} \frac{g_t^y(w)}{g_t^y(z)} = \mathbf{1}_{\{z \ne y\}} n_z^{-1} \frac{R(\Gamma(t, w; 1, y) \mid X_t = w)}{R(\Gamma(t, z; 1, y) \mid X_t = z)}$$

The complete graph. Let $\mathcal{X} = \{1, \ldots, n\}$ with $x \sim y$ for all $x \ne y \in \mathcal{X}$. Then, for all $x \ne y$ and all $0 \le t < 1$, we have $J_{t,x}^{G,y}(y) = 1/(1-t)$ and the probability that no jump occurred before time t is $\mathrm{Proba}(N_{\lambda(t)} = 0)$ where N_λ denotes a random variable distributed according to $\mathrm{Poisson}(\lambda)$ and $\lambda(t) = \int_0^t \frac{1}{1-s} ds = -\log(1-t)$. Therefore, $\mathrm{Proba}(N_{\lambda(t)} = 0) = \exp(-\lambda(t)) = 1 - t$ and

$$G_t^{xy} = (1-t)\delta_x + t\delta_y = \sum_{z \in \{x,y\}} t^{d(x,z)}(1-t)^{d(z,y)} \delta_z,$$

for any $0 \le t \le 1$. This law is in one-one correspondence with the Bernoulli law $\mathcal{B}(t)$ which is the specific binomial law $\mathcal{B}(2, t)$.

The graph $\mathbb{Z}$. We consider the simple situation where $(\mathcal{X}, \sim)$ is the set of integers $\mathbb{Z}$ with its natural graph structure. The reference random walk R is the simple walk with $J_z = (\delta_{z-1} + \delta_{z+1})/2$, $z \in \mathbb{Z}$, and the counting measure as its initial measure. Take $x < y \in \mathbb{Z}$. Then, for any $0 \le t < 1$ and $x \le z < y$, denoting N_{1-t} a random variable distributed according to the $\mathrm{Poisson}(1-t)$ law and $\delta = d(z, y) = y - z$, we obtain

$$\begin{aligned} J_{t,z}^{G,y}(z+1) &= \frac{1}{2} \frac{\mathrm{Proba}(N_{1-t} = d(z+1, y))(1/2)^{d(z+1,y)}}{\mathrm{Proba}(N_{1-t} = d(z, y))(1/2)^{d(z,y)}} \\ &= \frac{1}{2} \frac{e^{1-t} 2^{-(\delta-1)}(1-t)^{\delta-1}/(\delta-1)!}{e^{1-t} 2^{-\delta}(1-t)^\delta/\delta!} = d(z, y)/(1-t). \end{aligned}$$

Claim 3.17. *This proves that G^{xy} has the same law as $x + \widetilde{N}$ where $\widetilde{N}$ is the bridge of a Poisson process $(N_t)_{0 \le t \le 1}$, that is $\mathrm{Proba}(G^{xy} \in \cdot) = \mathrm{Proba}(N \in \{\cdot\} - x \mid N_1 = d(x, y))$.*

Since for each $t \in [0, 1]$, the law of $\widetilde{N}_t$ is the binomial $\mathcal{B}(d(x, y), t)$, $[\delta_x, \delta_y]$ is sometimes called a binomial interpolation, see [GRST, Hil12, Hil]. For each $0 < t < 1$, the support of G_t^{xy} is $\{x, x+1, \ldots, y\}$ and we have

$$G_t^{xy} = \sum_{z: x \le z \le y} \binom{d(x, y)}{d(x, z)} t^{d(x,z)}(1-t)^{d(z,y)} \delta_z.$$

The hypercube. Consider the hypercube $\mathcal{X} = \{0, 1\}^n$ with its natural graph structure so that the graph distance is the Hamming distance $d(x, y) = \sum_{1 \le i \le n} \mathbf{1}_{\{x_i \ne y_i\}}$ where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. The reference path measure R is the simple random walk with the uniform measure as its initial law. The directed tree that describes the geodesic dynamics between x and y has exactly $d(x, y)!$ directed chains with length $d(x, y)$ and endpoints x and y. The law of G^{xy} is the uniform mixture of the $d(x, y)!$ corresponding Poisson bridges.

In order to describe for any $0 \leq t \leq 1$, the law G_t^{xy}, let us encode each intermediate state by an ordered sequence in $\{\mathsf{d}, \mathsf{s}\}^{d(x,y)}$ where d and s stand respectively for "different" and "same". With this encoding, $(\mathsf{d}, \ldots, \mathsf{d})$ is x since x has $d(x,y)$ components that are different from y, of course $(\mathsf{s}, \ldots, \mathsf{s})$ is y and we see that the support S^{xy} of G_t^{xy} consists of $2^{d(x,y)}$ intermediate states at any time $0 < t < 1$. A short computation shows that for each $0 < t < 1$, we have

$$G_t^{xy} = \sum_{z \in S^{xy}} t^{d(x,z)}(1-t)^{d(z,y)} \, \delta_z.$$

4. The Schrödinger problem

The reversing measure m of the simple random walk on an infinite graph is unbounded, see (A.4). Since it is the analogue of the volume measure of a Riemannian manifold, it is likely that the relative entropy with respect to m should play an important role when trying to develop a Lott-Sturm-Villani theory on infinite graphs. Consequently, in this case the reference path measure R is unbounded.

In order to state the Schrödinger problem, it will be necessary to have in mind some basic facts about relative entropy with respect to a possibly *unbounded* reference measure. They are collected at Section B.

Schrödinger problem. We briefly introduce the main features of the Schrödinger problem. For more detail, see for instance the survey paper [Léo14].

The dynamic Schrödinger problem associated with the random walk $R \in \mathrm{M}_+(\Omega)$ is the following entropic minimization problem

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

where $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$ are prescribed initial and final marginals. As a *strictly* convex problem, it admits at most one solution.

Let us particularize the consequences of the additivity formula (B.4) to $r = R$, $p = P$ and $\phi = (X_0, X_1)$. Denoting $Q_{01} = (X_0, X_1)_\# Q \in \mathrm{M}_+(\mathcal{X}^2)$ and $Q^{xy} = Q(\cdot \mid X_0 = x, X_1 = y) \in \mathrm{P}(\Omega)$, we have for all $P \in \mathrm{P}(\Omega)$,

$$H(P|R) = H(P_{01}|R_{01}) + \int_{\mathcal{X}^2} H(P^{xy}|R^{xy}) \, P_{01}(dxdy) \tag{4.1}$$

which implies that $H(P_{01}|R_{01}) \leq H(P|R)$ with equality (when $H(P|R) < \infty$) if and only if

$$P^{xy} = R^{xy} \tag{4.2}$$

for P_{01}-almost every $(x,y) \in \mathcal{X}^2$, see (B.5). Therefore $\widehat{P}$ is the (unique) solution of ($\mathrm{S_{dyn}}$) if and only if it disintegrates as

$$\widehat{P}(\cdot) = \int_{\mathcal{X}^2} R^{xy}(\cdot) \, \widehat{\pi}(dxdy) \in \mathrm{P}(\Omega) \tag{4.3}$$

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

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

where $\pi_0, \pi_1 \in \mathrm{P}(\mathcal{X})$ are respectively the first and second marginals of $\pi \in \mathrm{P}(\mathcal{X}^2)$. Identity (4.3) means that

- $\widehat{P}$ shares its bridges with the reference path measure R, i.e. (4.2);

- these bridges are mixed according to

$$\widehat{\pi} = \widehat{P}_{01},$$

  the unique solution of (S).

The entropic minimization problem (S) is called the (static) *Schrödinger problem*. With (4.3), we see that

$$\inf(S_{\text{dyn}}) = \inf(S) \in (-\infty, \infty]. \tag{4.4}$$

Proofs of Theorem 2.7-(1) and Proposition 2.5. We begin with a key technical statement.

Girsanov's formula. We shall take advantage, several times in the remainder of the article, of the absolute continuity of R^k with respect to R. Girsanov's formula gives the expression of the Radon-Nykodim derivative of R^k with respect to R:

$$Z^k := \frac{dR^k}{dR} = \exp\left(-(\log k)\ell + U^k\right) \tag{4.5}$$

where $U^k := \int_{[0,1]\times\mathcal{X}}(1 - k^{-d(X_t, y)})\, J_{t, X_t}(dy)dt$ and ℓ is the length defined at (1.5).

Proof of Theorem 2.7-(1). The uniqueness follows from the strict convexity of the Schrödinger problem and we have just seen that $\widehat{\pi} = \widehat{P}_{01}$. The Markov property of $\widehat{P}$ which is inherited from the Markov property of R is proved at [Léo14, Prop. 2.10].
It remains to show the existence. For any $P \in P(\Omega)$ and any $k \geq 1$, with (4.5) we see that

$$
\begin{aligned}
H(P|R^k) &= H(P|R) + \log k \; E_P(\ell) - E_P U^k \\
&\leq H(P|R) + \log k \; E_P(\ell) \\
&= H(P_{01}|R_{01}) + \int_{\mathcal{X}^2} H(P^{xy}|R^{xy})\, P_{01}(dxdy) + \log k \; E_P(\ell).
\end{aligned}
$$

Choosing

$$P^o(\cdot) := \int_{\mathcal{X}^2} R^{xy}(\cdot)\, \pi^o(dxdy), \tag{4.6}$$

we have

$$H(P^o|R^k) \leq H(\pi^o|R_{01}) + \log k \; \int_{\mathcal{X}^2} E_{R^{xy}}(\ell)\, \pi^o(dxdy). \tag{4.7}$$

With Hypothesis 2.1-(μ) we obtain $\inf(S^k_{\text{dyn}}) < \infty$ and it follows that (S^k_{dyn}) and (S^k) admit a solution, see [Léo14, Lemma 2.4]. $\qquad\square$

Proof of Proposition 2.5. Taking $\pi^o = \mu_0 \otimes \mu_1$ in (4.7) gives

$$H(P^o|R^k) \leq H(\mu_0 \otimes \mu_1|R_{01}) + \log k \; \int_{\mathcal{X}^2} E_{R^{xy}}(\ell)\, \mu_0(dx)\mu_1(dy).$$

It is proved at [Léo14, Prop. 2.5] that the assumptions (i), (ii) and (iv) together with $H(\mu_0|R_0), H(\mu_1|R_1) < \infty$ imply $H(\mu_0 \otimes \mu_1|R_{01}) < \infty$. As regards the last term, it is clear that (i) and (iii) imply $\int_{\mathcal{X}^2} E_{R^{xy}}(\ell)\, \mu_0(dx)\mu_1(dy) < \infty$. $\qquad\square$

5. Lazy random walks converge to displacement random walks

The aim of this section is to make precise the convergence of $(S_{dyn}^k)_{k\geq 2}$. It is proved at Theorem 2.7 that the sequence of minimizers of $(S_{dyn}^k)_{k\geq 2}$ has a limit in $P(\Omega)$ which is singled out among the infinitely solutions of the dynamic Monge-Kantorovich problem (MK_{dyn}). As a corollary, we describe at Theorem 2.9 the convergence of the sequence of bridges $(R^{k,xy})_{k\geq 1}$.

The topological path space Ω. The countable set $\mathcal{X}$ is equipped with its discrete topology. The set $D([0,1], \mathcal{X})$ of all left-limited right-continuous paths on $[0,1)$ and left-continuous at the terminal time $t = 1$, be equipped with the Skorokhod topology. Note that, although for any $0 < t < 1$, the mapping $X_t : D([0,1], \mathcal{X}) \to \mathcal{X}$ is discontinuous, both the endpoint positions X_0 and X_1 are continuous. This will be used later several times. Let us denote the total number of jumps, defined on $D([0,1], \mathcal{X})$, by

$$N := \sum_{0<t<1} \mathbf{1}_{\{X_{t^-} \neq X_t\}} \in \mathbb{N} \cup \{\infty\}. \tag{5.1}$$

We consider $\widetilde{\Omega} = \{\omega \in D([0,1], \mathcal{X}); \forall t \in (0,1), \omega_{t^-} \neq \omega_t \implies \omega_{t^-} \sim \omega_t\}$ the subset of all paths compatible with the graph structure and introduce

$$\Omega := \{N < \infty\} \cap \widetilde{\Omega}, \tag{5.2}$$

the set of all càdlàg paths from $[0,1]$ to $\mathcal{X}$ which are compatible with the graph structure and piecewise constant with finitely many jumps.

Our Hypotheses 2.1, and in particular (2.3), imply that the support of each R^k is included in Ω. Since, $\widehat{P}^k$ is absolutely continous with respect to R^k, it also lives on Ω which appears to be the relevant path space.

As $\mathcal{X}$ is a discrete space, for each $n \in \mathbb{N}$, $\{N = n\} \cap \widetilde{\Omega}$ is a closed and open (clopen) set. In particular Ω is closed in $D([0,1], \mathcal{X})$ and it inherits its (trace) Polish topological structure and the corresponding (trace) Borel σ-field which is generated by the canonical process (restricted to Ω). The path space $\Omega = \bigsqcup_{n\in\mathbb{N}} \{N = n\} \cap \widetilde{\Omega}$ is partitioned by the disjoint clopen sets $\{N = n\} \cap \widetilde{\Omega}$. A small neighbourhood of $\omega \in \Omega$ consists of paths visiting exactly the same states as ω in the same order of occurrence and with jump times close to ω's ones. From now on, any topological statement on Ω refers to this topology and the canonical process $(X_t)_{0\leq t\leq 1}$ lives on Ω.

Γ-convergence. The right notion of convergence for the sequences of minimization problems $(S_{dyn}^k)_{k\geq 2}$ and $(S^k)_{k\geq 2}$ is the Γ-convergence which is briefly over-viewed now. Recall that $\Gamma\text{-}\lim_{k\to\infty} f^k = f$ on the metric space Y if and only if for any $y \in Y$,

(a) $\liminf_{k\to\infty} f^k(y_k) \geq f(y)$ for any convergent sequence $y_k \to y$,

(b) $\lim_{k\to\infty} f^k(y_k^o) = f(y)$ for some sequence $y_k^o \to y$.

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

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

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

For more detail about Γ-convergence, see [DM93] for instance.

The convergences of $(\mathrm{S}^k_{\mathrm{dyn}})_{k\geq 2}$ **and** $(\mathrm{S}^k)_{k\geq 2}$. As $(\mathrm{S}^k_{\mathrm{dyn}})$ and (S^k) are deeply linked to each other via the relations (4.3) and (4.4), the convergence of the static problems will follow from the convergence of the dynamic problems $(\mathrm{S}^k_{\mathrm{dyn}})$.

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

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

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

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

Let us rewrite $(I \to \min)$ as the following dynamic Monge-Kantorovich problem:

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

Otherwise stated, the topologies on $\mathrm{P}(\Omega)$ and $\mathrm{P}(\mathcal{X}^2)$ are respectively the topologies of narrow convergence: $\sigma(\mathrm{P}(\Omega), C_b(\Omega))$ and $\sigma(\mathrm{P}(\mathcal{X}^2), C_b(\mathcal{X}^2))$ which are weakened by the spaces $C_b(\Omega)$ and $C_b(\mathcal{X}^2)$ of all numerical continuous and bounded functions. The Γ-convergences are related to these topologies.

It is shown below at Lemma 5.5 that $\Gamma\text{-}\lim_{k\to\infty} I^k = I$, meaning that $(\mathrm{MK}_{\mathrm{dyn}})$ is the limit of $(\mathrm{S}^k_{\mathrm{dyn}})_{k\geq 2}$.

Lemma 5.3. *The function I is coercive.*

Proof. As $\{P : P_1 = \mu_1\}$ is closed, it is enough to show that the function $P \mapsto E_P(\ell) + \iota_{\{P:P_0=\mu_0\}}$ is coercive. Since $\ell \geq 0$ is continuous, $P \mapsto E_P(\ell) = \sup_{n\geq 1} E_P(\ell \wedge n)$ is lower semi-continuous. As in addition $\{P : P_0 = \mu_0\}$ is closed, the function $P \mapsto E_P(\ell) + \iota_{\{P:P_0=\mu_0\}}$ is also lower semi-continuous. It remains to show that for every $a \geq 0$, $\{P : P_0 = \mu_0, E_P(\ell) \leq a\}$ is uniformly tight in $\mathrm{P}(\Omega)$.

For any $n \geq 1$, there is some compact (finite) subset K_n of $\mathcal{X}$ such that $\mu_0(K_n) \geq 1 - 1/n$. We have $\ell \geq N$ where N in the number of jumps, see (5.1). Hence, any P such that $P_0 = \mu_0$ and $E_P(\ell) \leq a$ satisfies:

$$P(X_0 \in K_n, N \leq n) \geq 1 - P(X_0 \notin K_n) - P(\ell > n) \geq 1 - 1/n - a/n.$$

As it is assumed that the graph $(\mathcal{X}, \sim)$ is locally finite, see (2.2), $\{X_0 \in K_n, N \leq n\}$ is a compact subset of Ω (recall that Ω is compatible with the graph structure, see (5.2)). This proves the desired uniform tightness and completes the proof of the lemma. $\qquad\square$

Lemma 5.4. *For any $P \in \mathrm{P}(\Omega)$, there exists a sequence $(P_n)_{n\geq 1}$ in $\mathrm{P}(\Omega)$ such that $\lim_{n\to\infty} P_n = P$, $\lim_{n\to\infty} E_{P_n}(\ell) = E_P(\ell)$, and $H(P_n|R) < \infty$ for all $n \geq 1$.*

A similar result would fail in a diffusion setting with for instance $\Omega = C([0,1], \mathbb{R})$ and R the reversible Wiener measure (with Lebesgue measure as initial marginal). Here, we are going to take advantage of the countability of the discrete space $\mathcal{X}$ and of assumption (2.3) which allowed us to introduce at (5.2) a made-to-measure definition of the path space Ω. Indeed, this definition is entirely motivated by Lemma 5.4.

Proof of Lemma 5.4. Let us pick $P \in \mathrm{P}(\Omega)$.
(a) The set $\ell(\Omega)$ of all possible values of ℓ is countable; let us enumerate it: $\ell(\Omega) = \{c_n; n \geq 1\}$ and expand P along these values: $P = \sum_n P(\ell = c_n) P(\cdot \mid \ell = c_n)$.
(b) Suppose that $Q \in \mathrm{P}(\Omega)$ is concentrated on the set $\{\ell = c\}$. As $\{\ell = c\}$ is metric separable, there exists a sequence of convex combinations of Dirac masses: $Q_n = \sum_{i=1}^n a^{in} \delta_{\omega^{in}}$ with $\omega^{in} \in \{\ell = c\}$ such that $\lim_{n\to\infty} Q_n = Q$.
(c) Let $\omega \in \Omega$ be a fixed path. We shall prove below that there is a sequence $(Q_n^\omega)_{n\geq 1}$

in $P(\Omega)$ such that $\lim_{n\to\infty} Q_n^\omega = \delta_\omega$ and for each n, Q_n^ω is concentrated on $\{\ell = \ell(\omega)\}$ and $H(Q_n^\omega|R) < \infty$.

Putting (a), (b) and (c) together, it is not hard to check with the aid of Jensen's inequality applied to the convex function $H(\cdot|R)$, that there exists a sequence $(P_n)_{n\geq 1}$ in $P(\Omega)$ such that $\lim_{n\to\infty} P_n = P$ and for each n, $E_{P_n}(\ell) = E_P(\ell)$ and $H(P_n|R) < \infty$, which is the desired result.

It remains to prove (c), taking advantage of the specificity of the path space Ω. Let $\omega \in \Omega$ be fixed. It is completely described by its jump times $0 < t_1 < \cdots < t_k < 1$ and the corresponding states $(\omega_0, \omega_{t_1}, \ldots, \omega_{t_k})$. One choose $Q_n^\omega \in P(\Omega)$ as a Markov probability measure with initial marginal δ_{ω_0} and jump kernel $J_n^\omega = \sum_{i=1}^k \varphi_i^n(t)\, dt \delta_{\omega_{t_i}}$ where the non-negative continuous functions φ_i^n have, for each fixed $n \geq 1$, non-overlapping compact supports as $1 \leq i \leq k$ varies and are such that for each $1 \leq i \leq k$, $(\varphi_i(t)\, dt)_{n\geq 1}$ is an approximation of δ_{t_i}. Changing the jump times but keeping the order of $(\omega_0, \omega_{t_1}, \ldots, \omega_{t_k})$, doesn't change the value $\ell(\omega)$. Therefore , Q_n^ω is concentrated on $\{\ell = \ell(\omega)\}$. We have $H(Q_n^\omega|R) = E_{Q_n^\omega} \int_{[0,1]\times\mathcal{X}} h(\frac{dJ_n^\omega(t,X_t)}{dJ(t,X_t)}(y))\, J_{t,X_t}(dy)dt$ with $h(a) = a\log a - a + 1$ if $a > 0$ and $h(0) = 1$. One easily sees that $H(Q_n^\omega|R) < \infty$, using the assumption (2.3) , the fact that ω is compatible with the graph structure (by the very definition of Ω) and also that $J_{t_i,\omega_{t_i}}(\omega_{t_{i+1}}) > 0$ for all i (since by Hypothesis 2.1-(R), $J_{t,x}(y) > 0$, for all $t, x \sim y$). On the other hand, the compactness of the common initial law δ_{ω_0} and the weak convergence of the jump kernels $(J_n^\omega)_{n\geq 1}$ to $\sum_{i=1}^k \delta_{\omega_{t_i}}\delta_{t_i}$ which is the jump kernel of δ_ω implies that $\lim_{n\to\infty} Q_n^\omega = \delta_\omega$ in $P(\Omega)$. This completes the proofs of (c) and the lemma. $\qquad\square$

Lemma 5.5. *The sequence $(I^k)_{k\geq 2}$ is equi-coercive and $\Gamma\text{-}\lim_{k\to\infty} I^k = I$.*

Proof. Let us denote

$$H^k(P) := H(P|R^k)/\log k, \quad P \in P(\Omega).$$

We first prove the equi-coercivity of $(I^k)_{k\geq 2}$. Using (4.5), we obtain

$$
\begin{aligned}
H^k(P) &= E_P\left[\log(dP/dR) - \log Z^k\right]/\log k \\
&= E_P(\ell) + \left[H(P|R) - E_P\int_0^1 J_{t,X_t}(\mathcal{X})\, dt\right]/\log k + E_P\int_{[0,1]\times\mathcal{X}} k^{-d(X_t,y)}\, J_{t,X_t}(dy)dt/\log k
\end{aligned}
$$

Because of assumption (2.3), we have the uniform bounds

$$0 \leq E_P\int_0^1 J_{t,X_t}(\mathcal{X})\, dt, \ E_P\int_{[0,1]\times\mathcal{X}} k^{-d(X_t,y)}\, J_{t,X_t}(dy)dt \leq \sup_{t,x} J_{t,x}(\mathcal{X}) < \infty, \forall P \in P(\Omega), k \geq 2. \tag{5.6}$$

Hence, we obtain with $I - [(-\inf(S) \vee 0) + \sup_{t,x} J_{t,x}(\mathcal{X})]/\log 2 \leq I^k$ for all $k \geq 2$, and Lemma 5.3 that $(I^k)_{k\geq 2}$ is equi-coercive.

For future use, remark that $H^k(P) < \infty$ if and only if

$$E_P(\ell) < \infty \quad \text{and} \quad H(P|R) < \infty. \tag{5.7}$$

Now, we prove that $\Gamma\text{-}\lim_{k\to\infty} I^k = I$. As the constraint set $\{P \in P(\Omega); P_0 = \mu_0, P_1 = \mu_1\}$ is closed, it is enough to show that

$$\Gamma\text{-}\lim_{k\to\infty} H^k(P) = E_P(\ell), \quad \forall P \in P(\Omega).$$

Since $P \mapsto E_P(\ell)$ is lower semi-continuous and $H(\cdot|R) \geq 0$, with (5.6), we obtain for any convergent sequence $P_k \xrightarrow[k\to\infty]{} P$ that $\liminf_{k\to\infty} H^k(P_k) \geq E_P(\ell)$. Lemma 5.4 tells us that from any recovery sequence $(P_n)_{n\geq 1}$ for the lower semi-continuity of $P \mapsto E_P(\ell)$,

i.e. such that $\lim_{n\to\infty} E_{P_n}(\ell) = E_P(\ell)$, one can build a recovery sequence for $(H^k)_{k\geq 1}$, i.e. $\lim_{k\to\infty} H^k(P^k) = E_P(\ell)$. Namely, take $P^k = P_{n(k)}$ with $k \mapsto n(k)$ increasing to infinity slowly enough for $\lim_{k\to\infty} H(P_{n(k)}|R)/\log k = 0$. This completes the proof the proposition. $\square$

Proposition 5.8. *For any μ_0, $\mu_1 \in P(\mathcal{X})$, we have*

$$\lim_{k\to\infty} \inf (S^k) = \lim_{k\to\infty} \inf (S^k_{\text{dyn}}) = \inf (MK_{\text{dyn}}) = \inf (MK) \in (-\infty, \infty].$$

Proof. It is a direct corollary of (4.4) and Lemma 5.5. $\square$

The following auxiliary entropic minimization problem will be needed for identifying the limit of $\widehat{P}^k$ as k tends to infinity:

$$H(P|R_J) \to \min; \qquad P \in \mathcal{M}_{\text{dyn}}(\mu_0, \mu_1) \tag{5.9}$$

where $\mathcal{M}_{\text{dyn}}(\mu_0, \mu_1) \subset P(\Omega)$ denotes the set of all minimizers of (MK_{dyn}) and

$$R_J := \exp\left(\int_0^1 J_{t,X_t}(\mathcal{X})\, dt\right) R \in M_+(\Omega).$$

Remark that (2.3) ensures the finiteness of the integral in the above exponential.

Lemma 5.10.
(a) *For each $k \geq 2$, (S^k_{dyn}) has a unique solution $\widehat{P}^k$;*
(b) *$\mathcal{M}_{\text{dyn}}(\mu_0, \mu_1)$ is a non-empty convex compact subset of $P(\Omega)$;*
(c) *The sequence $(\widehat{P}^k)_{k\geq 2}$ is convergent and its limit $\lim_{k\to\infty} \widehat{P}^k = \widehat{P} \in \mathcal{M}_{\text{dyn}}(\mu_0, \mu_1)$ is the unique minimizer of (5.9).*

Under the assumptions of Lemma 5.10, Lemma 5.5 ensures that the limit points of $(\widehat{P}^k)_{k\geq 2}$ belong to $\mathcal{M}_{\text{dyn}}(\mu_0, \mu_1)$. But statement (c) of Lemma 5.10 asserts that there is indeed a unique limit point.

Proof of Lemma 5.10. This proof relies on Anzellotti and Baldo's Γ-asymptotic expansion technic [AB93]. For a clear exposition of this technique, see [AP03, §4].

We have seen at (5.7), that $I^k(P) < \infty$ if and only if $E_P(\ell) < \infty$, $H(P|R) < \infty$ and $P_0 = \mu_0, P_1 = \mu_1$. Therefore, taking P^o as in (4.6), we see that $I^k(P^o) < \infty$, for all $k \geq 2$. Together with the considerations of the preceding section, this proves statement (a).

The non-emptiness and convexity parts of statement (b) are immediate. The compactness is a standard consequence of the lower semi-continuity of $H(\cdot|R)$, the continuity of $P \mapsto P_1$ and the coerciveness of $P \mapsto E_P(\ell) + \iota_{\{P:P_0=\mu_0\}}$, see Lemma 5.3.

Let us prove (c). Denote $i := \inf(MK_{\text{dyn}}) < \infty$ and consider the subsequent renormalization of I^k :

$$J^k(P) := \log(k)\big(I^k(P) - i\big), \qquad P \in P(\Omega).$$

We have

$$J^k(P) = \iota_{\{P:P_0=\mu_0,P_1=\mu_1\}} + \log(k)\big(E_P(\ell) - i\big) + H(P|R_J) + E_P \int_{[0,1]\times\mathcal{X}} k^{-d(X_t,y)} J_{t,X_t}(dy)\, dt$$

and, using the coerciveness of $H(\cdot|R_J)$ and (5.6), it is easily seen that:

- $(J^k)_{k\geq 2}$ is equi-coercive;
- $\Gamma\text{-}\lim_{k\to\infty} J^k = J$ with $J(P) = \iota_{\{P:P_0=\mu_0,P_1=\mu_1,E_P(\ell)=i\}} + H(P|R_J), \ \ P \in P(\Omega)$.

As $H(\cdot|R_J)$ is strictly convex, so is J and (5.9) admits a unique minimizer $\widehat{P}$ on the convex set $\mathcal{M}_{\text{dyn}}(\mu_0, \mu_1) = \{P : P_0 = \mu_0, P_1 = \mu_1, E_P(\ell) = i\}$. One completes the proof of the lemma, noticing that $\operatorname{argmin} J^k = \operatorname{argmin} I^k = \{\widehat{P}^k\}$, for each $k \geq 2$. $\square$

Lemma 5.11. *For each $x, y \in \mathcal{X}$, Γ^{xy} is measurable. So is Γ.*

Proof. For each $x, y \in \mathcal{X}$, denote $\Omega_x := \{X_0 = x\}$ and $\Omega^{xy} := \{X_0 = x, X_1 = y\}$. The set of d-geodesics from x to y is $\Gamma^{xy} := \{\omega \in \Omega^{xy}; \ell(\omega) = d(x, y)\}$. Since ℓ is continuous and it controls the total number of jumps, the restriction $\ell_x = \ell_{|\Omega_x}$ of ℓ to the closed set Ω_x is coercive. Hence $\Gamma^{xy} = \{\omega \in \Omega_x; \ell_x = d(x, y)\} \cap \{X_1 = y\}$ is a compact subset of Ω (in particular, it is measurable). As a countable union of measurable sets, the set $\Gamma := \cup_{x, y \in \mathcal{X}} \Gamma^{xy}$ of all geodesics, is also measurable. $\qquad\square$

Lemma 5.12. *The set $\mathcal{M}_{\mathrm{dyn}}(\mu_0, \mu_1)$ consists of all $P \in \mathrm{P}(\Omega)$ concentrated on Γ, i.e. $P(\Gamma) = 1$, and such that the endpoint marginal $P_{01} \in \mathrm{P}(\mathcal{X}^2)$ solves* (MK).

Proof. Any $P \in \mathrm{P}(\Omega)$ disintegrates as: $P(\cdot) = \int_{\mathcal{X}^2} P^{xy}(\cdot) \, P_{01}(dxdy)$. Thus, $E_P(\ell) = \int_{\mathcal{X}^2} E_{P^{xy}}(\ell) \, P_{01}(dxdy)$. As $\ell \geq d(x, y)$ on Ω^{xy} and $\Gamma^{xy} = \{\ell = d(x, y)\}$, we have $E_P(\ell) \geq \int_{\mathcal{X}^2} d(x, y) \, P_{01}(dxdy)$ with equality if and only if $P^{xy}(\Gamma^{xy}) = 1$, for P_{01}-almost every (x, y). This means that $P(\Gamma) = 1$, in which case $E_P(\ell) = \int_{\mathcal{X}^2} d(x, y) \, P_{01}(dxdy)$ and the conclusion about P_{01} follows immediately. $\qquad\square$

Proof of Theorem 2.7-(2-3-4). Denote $\widehat{P} \in \mathrm{P}(\Omega)$ and $\widehat{\pi} \in \mathrm{P}(\mathcal{X}^2)$ the unique solutions (if they exist) of $(\widetilde{S}_{\mathrm{dyn}})$ and $(\widetilde{S})$.

We start proving the statements about the dynamical problems (S^k_{dyn}) and (S_{dyn}). Let P^o be defined by (4.6). Then our assumptions on μ_0 and μ_1 are equivalent to: $P^o_0 = \mu_0$, $P^o_1 = \mu_1$, $E_{P^o}(\ell) < \infty$ and $H(P^o|R) < \infty$, which are the hypotheses of Lemma 5.10 which tells us that $\lim_{k \to \infty} \widehat{P}^k = P^*$ with P^* the unique solution of

$$H(P|R_J) \to \min; \qquad P \in \mathcal{M}_{\mathrm{dyn}}(\mu_0, \mu_1).$$

Together with $P \in \mathcal{M}_{\mathrm{dyn}}(\mu_0, \mu_1) \iff \begin{cases} P_{01} \in \mathcal{S}_{\mathrm{MK}}(\mu_0, \mu_1) \\ P(\Gamma) = 1 \end{cases}$ (see Lemma 5.12), and the

identity $H(P|G) = \begin{cases} H(P|R_J), & \text{if } P(\Gamma) = 1 \\ \infty, & \text{otherwise} \end{cases}$, this yields the identity $P^* = \widehat{P}$.

Formula (2.8) with $\widehat{\pi}$ the unique solution of the strictly convex problem $(\widetilde{S})$, follows from a reasoning similar to the one leading to (4.3) and based on the additive disintegration formula (4.1).

We prove the statements about the static problems (S^k) and (S) by pushing forward (S^k_{dyn}) and (S_{dyn}) with the mapping (X_0, X_1) to obtain the measures $\widehat{\pi}^k = \widehat{P}^k_{01}$, $\widehat{\pi} = \widehat{P}_{01}$ and considering (4.1) again. $\qquad\square$

6. Dynamics of the displacement random walk

To give some detail about the dynamics of the displacement interpolation μ, it is necessary to study the dynamics of the displacement random walk $\widehat{P}$. It is shown below that for any $x, y \in \mathcal{X}$, G^{xy} and $\widehat{P}$ are Markov and we compute their jump kernels at Theorems 2.10 and 2.11. To achieve this goal, we need some preliminary material involving the reciprocal and Markov properties.

Reciprocal path measure. The reciprocal property extends the notion of Markov property. For more detail, see [LRZ] and the references therein.

Definition 6.1. *A measure $Q \in \mathrm{M}_+(\Omega)$ is said to be* reciprocal *if for any $0 \leq u \leq v \leq 1$,*
$$Q(X_{[u,v]} \in \cdot \mid X_{[0,u]}, X_{[v,1]}) = Q(X_{[u,v]} \in \cdot \mid X_u, X_v).$$

The following lemma is standard. We state it for the comfort of the reader.

Lemma 6.2.
(a) Any Markov measure is reciprocal (but the converse is false).
(b) Almost every bridge of a reciprocal measure is Markov.

Proof. Let us prove (a). Take $Q \in \mathrm{M}_+(\Omega)$ any Markov measure. We have for any measurable subsets $A \in \Omega_{[u,v]}$ and $B \in \Omega_{[v,1]}$ with $Q(X_{[v,1]} \in B) > 0$,

$$
\begin{aligned}
Q(X_{[u,v]} \in A \mid X_{[0,u]}, X_{[v,1]} \in B) &= \frac{Q(X_{[u,v]} \in A, X_{[v,1]} \in B \mid X_{[0,u]})}{Q(X_{[v,1]} \in B \mid X_{[0,u]})} \\
&= \frac{Q(X_{[u,v]} \in A, X_{[v,1]} \in B \mid X_u)}{Q(X_{[v,1]} \in B \mid X_u)} \\
&= Q(X_{[u,v]} \in A \mid X_u, X_{[v,1]} \in B),
\end{aligned}
$$

meaning that $Q(X_{[u,v]} \in \cdot \mid X_{[0,u]}, X_{[v,1]}) = Q(X_{[u,v]} \in \cdot \mid X_u, X_{[v,1]})$. And one concludes using the time-symmetry of the Markov property.

Let us prove (b). Let $Q \in \mathrm{M}_+(\Omega)$ be reciprocal. We have for any $x, y \in \mathcal{X}$ and $0 \le t \le 1$,

$$
\begin{aligned}
Q^{xy}(X_{[t,1]} \in \cdot \mid X_{[0,t]}) &= Q(X_{[t,1]} \in \cdot \mid X_0 = x, X_{[0,t]}, X_1 = y) \\
&= Q(X_{[t,1]} \in \cdot \mid X_t, X_1 = y) = Q^{xy}(X_{[t,1]} \in \cdot \mid X_t)
\end{aligned}
$$

where last equality is obtained repeating the argument with X_t instead of $X_{[0,t]}$. This is the announced result.   □

Lemma 6.3. *Let $Q \in \mathrm{M}_+(\Omega)$ be a reciprocal measure and $G \subset \Gamma$ a measurable subset of Ω consisting of geodesics. Then, the measure $Q' := \mathbf{1}_G Q \in \mathrm{M}_+(\Omega)$ is still reciprocal.*

Proof. We use the following property of a geodesic: the restriction $\gamma_{[u,v]}$ of any geodesic $\gamma \in \Gamma$, is still a geodesic of $\Omega_{[u,v]}$. Therefore, $X \in \Gamma_{[0,1]}^{X_0, X_1}$ implies that $X_{[u,v]} \in \Gamma_{[u,v]}^{X_u, X_v}$ which implies that

$$
\begin{aligned}
&Q(X \in G, X_{[u,v]} \in A \mid X_{[0,u]}, X_{[v,1]}) \\
&= \mathbf{1}_{\{X_{[0,u]} \in G_{[0,u]}, X_{[v,1]} \in G_{[v,1]}\}} Q(X_{[u,v]} \in G_{[u,v]} \cap A \mid X_{[0,u]}, X_{[v,1]}) \\
&= \mathbf{1}_{\{X_{[0,u]} \in G_{[0,u]}, X_{[v,1]} \in G_{[v,1]}\}} Q(X_{[u,v]} \in G_{[u,v]} \cap A \mid X_u, X_v)
\end{aligned}
$$

for any measurable set $A \subset \Omega_{[u,v]}$, where the last equality follows from the reciprocal property. Therefore, since $X_{[0,u]} \in G_{[0,u]}$ and $X_{[v,1]} \in G_{[v,1]}$, Q'-a.e., we have

$$
\begin{aligned}
Q'(X_{[u,v]} \in A \mid X_{[0,u]}, X_{[v,1]}) &= Q(X_{[u,v]} \in G_{[u,v]} \cap A \mid X_u, X_v) \\
&= Q'(X_{[u,v]} \in A \mid X_u, X_v), \qquad Q'\text{-a.e.}
\end{aligned}
$$

which is the desired result.   □

Basic properties of G. We apply Lemmas 6.2 and 6.3 to G defined at (2.6).

Proposition 6.4. *If R is reversible, then G is also reversible.*
The measure G is reciprocal (but not Markov in general) and it concentrates on the set Γ of all geodesics.

Proof. If R is reversible, the time-reversal invariances of $\int_{[0,1]} J_{X_t}(\mathcal{X}) \, dt$ and Γ together with the symmetry of the distance d immediately imply the reversibility of G.

Let us show that $R_J := \exp(\int_{[0,1]} J_{t,X_t}(\mathcal{X}) \, dt) R$ is Markov by proving that for each $t \in [0,1]$ and all bounded measurable functions $a \in \sigma(X_{[0,t]})$ and $b \in \sigma(X_{[t,1]})$, we have

$E_{R_J}(ab \mid X_t) = E_{R_J}(a \mid X_t)E_{R_J}(b \mid X_t)$. Denoting $\alpha := \exp(\int_{[0,t]} J_{s,X_s}(\mathcal{X})\,ds) \in \sigma(X_{[0,t]})$ and $\beta := \exp(\int_{[t,1]} J_{s,X_s}(\mathcal{X})\,ds) \in \sigma(X_{[t,1]})$, by the Markov property of R, we have

$$E_{R_J}(ab \mid X_t) = \frac{E_R(a\alpha b\beta \mid X_t)}{E_R(\alpha\beta \mid X_t)} = \frac{E_R(a\alpha \mid X_t)}{E_R(\alpha \mid X_t)}\frac{E_R(b\beta \mid X_t)}{E_R(\beta \mid X_t)} = E_{R_J}(a \mid X_t)E_{R_J}(b \mid X_t)$$

where last equality is obtained by plugging successively $b = 1$ and $a = 1$ in $E_{R_J}(ab \mid X_t) = \frac{E_R(a\alpha|X_t)}{E_R(\alpha|X_t)}\frac{E_R(b\beta|X_t)}{E_R(\beta|X_t)}$. This shows that R_J is Markov.
We conclude with Lemma 6.3 that $G = \mathbf{1}_\Gamma R_J$ is reciprocal. $\square$

Although G is reciprocal, it is not Markov. To see this, remark that the time reversed of a geodesic is also geodesic. If the geodesic walker only knows that he stands at z at time t, having forgotten his past history and in particular that his previous state before jumping was z', he cannot decide to forbid z' to be his next state. Nevertheless, the bridges G^{xy} are Markov.

Corollary 6.5. *For every $(x,y) \in \mathcal{X}^2$, the bridge G^{xy} is Markov.*

Proof. This follows from Lemma 6.2 and Proposition 6.4, remarking that under our irreducibility assumption, R_{01}-almost everywhere is equivalent to everywhere on $\mathcal{X}^2$. $\square$

As G^{xy} is Markov, it is sufficient to compute its jump kernel to characterize its dynamics. Recall the definition of the directed tree $(\Gamma^{xy}([0,1]), \rightarrow)$, between the statements of Theorems 2.7 and 2.10, that describes the successive occurrence of the states which are visited by the geodesics from x to y, regardless of the instants of jump.

Proof of Theorem 2.10. The Markov property is already proved at Corollary 6.5. Let us begin with some notation. For all $0 \le t_1 \le t_2 \le t_3 \le 1$, $z_1, z_2, z_3 \in \mathcal{X}$, we denote

$$\Gamma(t_1, z_1; t_2, z_2) := \left\{\omega \in \Omega; \omega_{|[t_1,t_2]} = \gamma_{|[t_1,t_2]} \text{ for some } \gamma \in \Gamma, \omega_{t_1} = z_1, \omega_{t_2} = z_2\right\}$$
$$\Gamma(t_1, z_1; t_2, z_2; t_3, z_3) := \Gamma(t_1, z_1; t_3, z_3) \cap \{X_{t_2} = z_2\}.$$

In particular, we have $\Gamma^{xy} = \Gamma(0, x; 1, y)$. We also introduce the functions on Ω :

$$G(t_1, z_1; t_2, z_2) := \exp\left(\int_{t_1}^{t_2} J_{t,X_t}(\mathcal{X})\,dt\right) \mathbf{1}_{\Gamma(t_1,z_1;t_2,z_2)}$$

$$G(t_1, z_1; t_2, z_2; t_3, z_3) := \exp\left(\int_{t_1}^{t_3} J_{t,X_t}(\mathcal{X})\,dt\right) \mathbf{1}_{\Gamma(t_1,z_1;t_2,z_2;t_3,z_3)}.$$

We see that $g_t^y(z) = E_R\left(G(t, z; 1, y) \mid X_t = z\right)$.
As a direct consequence of the definition of a geodesic, for all $0 \le t_1 \le t_2 \le t_3 \le 1$, $z_1 \preceq z_2 \preceq z_3 \in \Gamma^{xy}([0,1])$, we have

$$\Gamma(t_1, z_1; t_2, z_2; t_3, z_3) = \Gamma(t_1, z_1; t_2, z_2) \cap \Gamma(t_2, z_2; t_3, z_3)$$

which implies that

$$G(t_1, z_1; t_2, z_2; t_3, z_3) = G(t_1, z_1; t_2, z_2)G(t_2, z_2; t_3, z_3) \quad \text{on } \Gamma^{xy}. \qquad (6.6)$$

As G^{xy} is Markov, to derive the infinitesimal generator of its Markov semi-group, it is enough to compute its forward stochastic derivative

$$L_t^{G,xy}u(z) := \lim_{h\downarrow 0} E_{G^{xy}}\left[u(X_{t+h}) - u(X_t)|X_t = z\right], \quad 0 \le t < 1, z \in \Gamma^{xy}([0,1]).$$

For any $0 \le t < 1, z \in \Gamma^{xy}([0,1])$, with (2.6) we see that

$$
\begin{aligned}
G^{xy}(\cdot \mid X_t = z) &= \frac{G(0,x;t,z;1,y)}{E_R\left[G(0,x;t,z;1,y)|X_t = z\right]}R(\cdot \mid X_t = z) \\
&= \frac{G(0,x;t,z)G(t,z;1,y)}{E_R\left[G(0,x;t,z)|X_t = z\right]g_t^y(z)}R(\cdot \mid X_t = z)
\end{aligned}
$$

where last equality follows from (6.6) and the Markov property of R.

We set $U_t = u(X_t)$ for short. For any finitely supported function u and any $0 \le t < t + h \le 1$,

$$
\begin{aligned}
&E_{G^{xy}}(U_{t+h} - U_t|X_t = z) \\
&= \frac{E_R\left[(U_{t+h} - U_t)G(0,x;t,z)G(t,z;1,y)|X_t = z\right]}{E_R[G(0,x;t,z)|X_t = z]g_t^y(z)} \\
&\overset{(6.6)}{=} \frac{1}{g_t^y(z)}E_R\left[(U_{t+h} - U_t)\mathbf{1}_{\{z \le X_{t+h} \le y\}}G(t,z;t+h,X_{t+h})G(t+h,X_{t+h};1,y)|X_t = z\right] \\
&= \frac{1}{g_t^y(z)}E_R\left[(U_{t+h} - U_t)\mathbf{1}_{\{z \le X_{t+h} \le y\}}G(t,z;t+h,X_{t+h})g_{t+h}^y(X_{t+h})|X_t = z\right]
\end{aligned}
$$

where the Markov property of R is used at last equality.

When $X_t = z$ and h tends down to 0 we have:

$$
(U_{t+h}-U_t)\mathbf{1}_{\{z \le X_{t+h} \le y\}} = \begin{cases} 0, & \text{if } X_{t+h} = X_t = z \text{ with probability } 1 - J_{t,z}(\mathcal{X})h + o(h) \\ u(w) - u(z), & \text{if } X_{t+h} = w \leftarrow z \text{ with probability } J_{t,z}(w)h + o(h) \\ *, & \text{otherwise with probability } o(h) \end{cases}
$$

where $*$ is something bounded by $2\sup|u|$, and

$$
G(t,z;t+h,X_{t+h}) = \begin{cases} 1 + O(h), & \text{if } X_{t+h} = X_t = z \text{ with probability } 1 - J_{t,z}(\mathcal{X})h + o(h) \\ 1 + O(h), & \text{if } X_{t+h} = w \leftarrow z \text{ with probability } J_{t,z}(w)h + o(h) \\ *, & \text{otherwise with probability } o(h). \end{cases}
$$

where $*$ is something bounded because of the assumption (2.3). Hence,

$$
h^{-1}E_{G^{xy}}[U_{t+h} - U_t|X_t = z] = \sum_{w:z \to w}[u(w) - u(z)]\frac{g_t^y(w)}{g_t^y(z)}J_{t,z}(w) + o_{h\downarrow 0}(1)
$$

which shows that $L^{G,xy}u(z) = \sum_{w:z\to w}[u(w) - u(z)]\frac{g_t^y(w)}{g_t^y(z)}J_{t,z}(w)$ and completes the proof of the theorem. $\qquad \square$

$\widehat{P}$ **is Markov.** It follows from (2.8), Proposition 6.4 and [LRZ, Prop. 2.8] that the limiting path measure $\widehat{P}$ is reciprocal. We can do better, but it requires some effort.

Proposition 6.7. *The limiting path measure $\widehat{P}$ is Markov.*

Proof. By Theorem 2.7-(1), for each $k \ge 2$, $\widehat{P}^k$ inherits the Markov property of R. We show below at Lemma 6.12 that, as k tends to infinity, $\widehat{P}^k$ converges in variation norm to $\widehat{P}$ and we conclude with Lemma 6.8 that $\widehat{P}$ is Markov. $\qquad \square$

Recall that the total variation norm of the signed bounded measure q on Y is $\|q\|_{\mathrm{TV}} := |q|(Y) = q^+(Y) + q^-(Y) = \sup_{f:\sup|f|\le 1}\int_Y f\,dq = \sup_{A\subset Y}(|q(A)| + |q(A^c)|)$

Lemma 6.8 (Nagasawa, [Nag93]). *Let $(P_k)_{k\ge 1}$ be a sequence in $\mathrm{P}(\Omega)$ of Markov probability measures which converges in variation norm to P, then P is also Markov.*

Proof. We have to show that for all $0 \leq t \leq 1$ and any bounded functions $a \in \sigma(X_{[0,t]})$ and $c \in \sigma(X_{[t,1]})$, we have $E_P(ac) = E_P(aE_P(c|X_t))$.

For any bounded functions $b \in \sigma(X_t)$ and $c \in \sigma(X_{[t,1]})$, we have

$$
\begin{aligned}
& \left| E_P[bE_P(c|X_t)] - E_P[bE_{P^k}(c|X_t)] \right| \\
\leq{} & \left| E_P(bc) - E_{P^k}(bc) \right| + \left| E_{P^k}[bE_{P^k}(c|X_t)] - E_P[bE_{P^k}(c|X_t)] \right| \\
\leq{} & 2\|b\|\|c\|\|P^k - P\|_{\mathrm{TV}} \underset{k\to\infty}{\to} 0
\end{aligned}
$$

Hence,

$$
\lim_{k\to\infty} E_{P^k}(c|X_t) = E_P(c|X_t) \quad \text{in } L^1(P). \tag{6.9}
$$

Let $a \in \sigma(X_{[0,t]})$,

$$
\begin{aligned}
& \left| E_P(ac) - E_P[aE_P(c|X_t)] \right| \\
\leq{} & \left| E_P(ac) - E_{P^k}(ac) \right| + \left| E_{P^k}[aE_{P^k}(c|X_t)] - E_P[aE_{P^k}(c|X_t)] \right| \\
& + \left| E_P[aE_{P^k}(c|X_t)] - E_P[aE_P(c|X_t)] \right|
\end{aligned}
$$

But,

$$
\left| E_P(ac) - E_{P^k}(ac) \right| + \left| E_{P^k}[aE_{P^k}(c|X_t)] - E_P[aE_{P^k}(c|X_t)] \right| \leq 2\|a\|\|c\|\|P^k - P\|_{\mathrm{TV}} \underset{k\to\infty}{\to} 0
$$

and $\left| E_P[aE_{P^k}(c|X_t)] - E_P[aE_P(c|X_t)] \right| \underset{k\to\infty}{\to} 0$ because of (6.9). This completes the proof of the lemma. $\qquad\square$

There are counter-examples of sequences $(P_k)_{k\geq 1}$ of Markov measures converging narrowly to a non-Markov P.

The following standard lemma (Scheffé's theorem) is a preliminary result for Lemma 6.12's proof.

Lemma 6.10. *Let r be a positive measure and $p_k = z_k\, r$, $k \geq 1$, $p = z\, r$ be probability measures which are absolutely continuous with respect to r. If $\lim_{k\to\infty} z_k = z$, r-a.e., then*

$$
\|z\, r - z_k\, r\|_{\mathrm{TV}} = \int |z - z_k|\, dr \underset{k\to\infty}{\to} 0.
$$

Proof. Set $d_k = z - z_k$. Then, $\int d_k\, dr = 0$. For any measurable subset A,

$$
\left| \int_A d_k\, dr \right| + \left| \int_{A^c} d_k\, dr \right| \leq \int |d_k|\, dr
$$

with equality when $A = \{d_k \geq 0\}$. Therefore, $\|z\, r - z_k\, r\|_{\mathrm{TV}} = \int |d_k|\, dr$. As $\lim_{k\to\infty} d_k^+ = 0$, r-a.e. and $0 \leq d_k^+ \leq z$, we obtain $\lim_{k\to\infty} \int |d_k|\, dr = 2\lim_{k\to\infty} \int d_k^+\, dr = 0$. $\qquad\square$

The following standard lemma will also be used during the proof of Lemma 6.12.

Lemma 6.11 (Laplace principle)**.** *Let r be a positive measure on the measurable set Y. For any measurable function $F : Y \to [-\infty, \infty]$ which is not identically equal to $-\infty$ and any measurable subset $Y' \subset Y$ such that $r(Y') < \infty$, we have*

$$
\lim_{\epsilon \to 0} \epsilon \log \int_{Y'} e^{F/\epsilon}\, dr = r\text{-ess sup}_{Y'}\, F \in (-\infty, \infty].
$$

Proof. Considering the restriction of r to Y', one can assume without loss of generality that $Y' = Y$ and $r(Y) < \infty$. To simplify notation, let us denote $b := r\text{-ess sup}\, F \in (-\infty, \infty]$

and for any $\delta > 0$, $b_\delta := \min[b - \delta, 1/\delta]$. Since $F \leq b$, r-a.e., we have $\epsilon \log \int e^{F/\epsilon} \, dr \leq b + \epsilon \log r(Y)$ and a fortiori

$$\limsup_{\epsilon \to 0} \epsilon \log \int e^{F/\epsilon} \, dr \leq b.$$

On the other hand, for any $\epsilon, \delta > 0$,

$$\epsilon \log \int e^{F/\epsilon} \, dr \geq \epsilon \log \int \mathbf{1}_{\{F \geq b_\delta\}} e^{F/\epsilon} \, dr \geq \epsilon \log r(F \geq b_\delta) + b_\delta.$$

By the definition of r-ess sup F, for any $\delta > 0$ we have $0 < r(F \geq b_\delta) \leq r(Y) < \infty$. Therefore, taking the $\liminf_{\epsilon \to 0}$ and then letting δ tend down to zero, we obtain

$$\liminf_{\epsilon \to 0} \epsilon \log \int e^{F/a_\epsilon} \, dr \geq b,$$

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

Lemma 6.12. *We have:* $\lim_{k \to \infty} \|\widehat{P}^k - \widehat{P}\|_{TV} = 0$.

Proof. By Lemma 6.10, it is enough to prove that $\lim_{k \to \infty} d\widehat{P}^k/dR = d\widehat{P}/dR$, R-a.e. For any $P \in P(\Omega)$ such that $P \ll R$, we have $\frac{dP}{dR} = \sum_{x,y \in \mathcal{X}} \mathbf{1}_{\{X_0 = x, X_1 = y\}} \frac{P_{01}(x,y)}{R_{01}(x,y)} \frac{dP^{xy}}{dR^{xy}}$. We also have $\lim_{k \to \infty} \widehat{P}^k_{01}(x, y) = \widehat{P}_{01}(x, y)$ and $\lim_{k \to \infty} \widehat{P}^{k,xy} = \widehat{P}^{xy}$ for all $x, y \in \mathcal{X}$. Therefore, it remains to show that for each $(x, y) \in \mathcal{X}$, $\lim_{k \to \infty} d\widehat{P}^{xy}/d\widehat{P}^{k,xy} = 1$, R^{xy}-a.e. As, $\widehat{P}^{xy} = G^{xy}$ and $\widehat{P}^{k,xy} = R^{k,xy}$, this amounts to prove that

$$\lim_{k \to \infty} \frac{dG^{xy}}{dR^{k,xy}} = 1, \quad R^{xy}\text{-a.e.} \tag{6.13}$$

for all x, y. By Girsanov's formula (4.5), for any $0 \leq t \leq 1$,

$$R^k_{[t,1]}(\cdot|X_t) = k^{-\ell_{[t,1]}} \exp\left(\int_{[t,1] \times \mathcal{X}} (1 - k^{-d(X_s, y)}) J_{s,X_s}(dy) \, ds\right) R_{[t,1]}(\cdot|X_t)$$

where $\ell_{[t,1]} := \sum_{t \leq s \leq 1} d(X_{s^-}, X_s)$. Hence, the jump measure of $R^{k,xy}$ is given for any $t \in [0,1]$, $z, w \in \mathcal{X}$ by

$$J^{k,xy}_{t,z}(w)$$
$$= \frac{R^k(X_1 = y|X_t = w)}{R^k(X_1 = y|X_t = z)} k^{-d(z,w)} J_{t,z}(w)$$
$$= \frac{E_R\left[k^{-\{d(z,w)+\ell_{[t,1]}\}} \exp\left(\int_{[t,1] \times \mathcal{X}} (1 - k^{-d(X_s, a)}) J_{s,X_s}(da) ds\right) \mathbf{1}_{(X_1 = y)} \mid X_t = w\right]}{E_R\left[k^{-\ell_{[t,1]}} \exp\left(\int_{[t,1] \times \mathcal{X}} (1 - k^{-d(X_s, a)}) J_{s,X_s}(da) ds\right) \mathbf{1}_{(X_1 = y)} \mid X_t = z\right]} J_{t,z}(w).$$

With Lemma 6.11, we obtain that

$$\lim_{k \to \infty} J^{k,xy}_{t,z}(w) = \frac{E_R\left[\exp\left(\int_t^1 J_{s,X_s}(\mathcal{X}) \, ds\right) \mathbf{1}_{\Gamma(t,w;1,y)}|X_t = w\right]}{E_R\left[\exp\left(\int_t^1 J_{s,X_s}(\mathcal{X}) \, ds\right) \mathbf{1}_{\Gamma(t,z;1,y)}|X_t = z\right]} J_{t,z}(w) =: J^{G,y}_{t,z}(w)$$

meaning, with Theorem 2.10, that the pointwise limit of $J^{k,xy}$ as k tends to infinity is the jump measure $J^{G,y}$ of G^{xy}. With Girsanov's formula, this implies that (6.13) is true, completing the proof of the lemma. $\qquad \square$

Remark 6.14. Theorem 2.10 and Lemma 6.12 together, give an alternate proof of the convergence $\lim_{k \to \infty} R^{k,xy} = G^{xy}$ stated in Theorem 2.9, with no reference to an entropy minimization problem.

A comparison with the usual continuous setting. For comparison, it is worthwhile recalling an analogous result in the usual setting of McCann's displacement interpolations on $\mathcal{X} = \mathbb{R}^n$. The Monge-Mather shortening principle, see [Vil09, Ch. 8], tells us that the optimal plan $\widehat{\pi}$ for the quadratic cost between μ_0 and μ_1 is such that two distinct geodesics interpolating between couples of endpoints in its support supp $\widehat{\pi}$, do not intersect.

In our graph setting, this would correspond to the nice situation where for any z and $t < 1$, $\widehat{P}(X_1 \in \cdot | X_t = z)$ reduces to a Dirac measure. But in our discrete setting, $\widehat{P}(\cdot | X_t = z)$ has sample paths which live on a directed geodesic tree with top leaves (at time 1) distributed according to a probability measure $\widehat{P}(X_1 \in \cdot | X_t = z)$ which might not reduce to a Dirac mass in the general case.

This directed tree structure is a consequence of the *Markov property* of $\widehat{P}$. It can also be seen as a consequence of the *shortening principle* which, in the present metric cost setting, is an elementary consequence of the triangle inequality.

As a corollary of Proposition 6.7, we obtain the following result.

Proposition 6.15. *For all $0 \le s \le t \le 1$, $\widehat{P}_{st} \in \mathrm{P}(\mathcal{X}^2)$ is an optimal coupling of μ_s and μ_t, meaning that $\widehat{P}_{st}$ is a solution of* (MK) *with μ_s and μ_t as prescribed marginal constraints.*

Proof. It is a consequence of

(1) the Markov property of $\widehat{P}$, see Proposition 6.7, which allows surgery by gluing the bridges of $\widehat{P}_{[s,t]}$ together with the restrictions $\widehat{P}_{[0,s]}$ and $\widehat{P}_{[t,1]}$;

(2) the fact that $\ell_{st} := \sum_{s<r<t} d(X_{r^-}, X_r)$ is insensitive to changes of time: i.e. for any strictly increasing mapping $\theta : [s,t] \to [0,1]$ with $\theta(s) = 0$, $\theta(t) = 1$, we have $\ell_{st} = \ell_{01}(X \circ \theta)$.

A standard ad absurdum reasoning leads to the announced property. $\square$

Proof of Theorem 2.11-(2). Now, let us investigate the dynamics of $\widehat{P}$ in the general case where it interpolates between marginal constraints μ_0 and μ_1 which are not necessarily Dirac measures. Let us denote

$$\widehat{p}(t, z; dy) := \widehat{P}(X_1 \in dy | X_t = z)$$

so that the optimal coupling $\widehat{P}_{t1}$ of μ_t and μ_1 disintegrates as $\widehat{P}_{t1}(dzdy) = \mu_t(dz)\widehat{p}(t, z; dy)$. As in Theorem 2.10, we compute a stochastic derivative. A general disintegration result tells us that

$$\begin{aligned}
\widehat{P}_{[t,1]}(\cdot | X_t = z) &= \int_{\mathcal{X}} \widehat{P}_{[t,1]}(\cdot | X_t = z, X_1 = y) \, \widehat{P}(X_1 \in dy | X_t = z) \\
&= \int_{\mathcal{X}} G_{[t,1]}^{x_o,y}(\cdot | X_t = z,) \, \widehat{p}(t, z; dy)
\end{aligned}$$

since, for μ_0-almost any x_o,

$$\begin{aligned}
\widehat{P}_{[t,1]}(\cdot | X_t = z, X_1 = y) &= \widehat{P}_{[t,1]}(\cdot | X_0 = x_o, X_t = z, X_1 = y) \\
&= G_{[t,1]}^{x_o,y}(\cdot | X_t = z)
\end{aligned}$$

where we used the Markov property of $\widehat{P}$ at the first equality and the identity $\widehat{P}^{x_o,y} = G^{x_o,y}$ at the second equality. Therefore, for all $0 < t < t + h < 1$,

$$E_{\widehat{P}}\Big([u(X_{t+h}) - u(X_t)]/h | X_t = z \Big) = \int_{\mathcal{X}} E_{G^{x_o,y}}\Big([u(X_{t+h}) - u(X_t)]/h | X_t = z \Big) \widehat{p}(t, z; dy)$$

and letting h tend down to zero, we see that the stochastic derivative of $\widehat{P}$ is given by

$$\widehat{L}_t u(z) = \int_{\mathcal{X}} L_t^{G,y} u(z)\, \widehat{p}(t,z;dy)$$

which gives the desired expression for the jump kernel $\widehat{J}$.

The evolution equation for $[\mu_0, \mu_1]$ is the usual forward Fokker-Planck equation for the time-marginal flow $t \mapsto \widehat{P}_t$ of the Markov measure $\widehat{P}$. $\qquad\square$

7. Conservation of the average rate of mass displacement

The main result of this section is Theorem 7.4. It states that the constant average rate of mass displacement, recall Definition 3.14, is conserved along the displacement interpolation. It is a consequence of the Corollary 7.3 of Proposition 7.1 which also asserts that some similar quantity is conserved along the time marginal flow of the solution of the dynamical Schrödinger problem (S_{dyn}).

Main results of the section. The main technical result of this section is the following Proposition 7.1 whose proof is postponed to the next subsection.

Proposition 7.1. *Let $P \in \mathrm{P}(\Omega)$ be a random walk. We denote $(\nu_t)_{t\in[0,1]}$ its marginal flow and $(\nu j_t(\mathcal{X}^2))_{0 \le t \le 1}$ its average rate of mass displacement. Let us assume that*

(i) $H(P|R) < \infty$;
(ii) $\nu j_t(\mathcal{X}^2) > 0$ for all $t \in [0,1]$;
(iii) $1 < \int_{[0,1]} \nu j_t(\mathcal{X}^2)/\nu J_t(\mathcal{X}^2)\, dt \le \infty$.

Then, there exists a change of time $\widehat{\tau}$ which minimizes the function $\tau \mapsto H(P^\tau|R)$ among all the changes of time τ and verifies

$$\nu j_t^{\widehat{\tau}}(\mathcal{X}^2) - \nu J_{\widehat{\tau}(t)}(\mathcal{X}^2) = K, \quad \text{for almost every } t \in [0,1], \tag{7.2}$$

for some constant $K > 0$. We have denoted by $\nu j_t^{\widehat{\tau}}(\mathcal{X}^2) = \dot{\widehat{\tau}}(t)\nu j_{\widehat{\tau}(t)}(\mathcal{X}^2)$ the average rate of mass displacement of $P^{\widehat{\tau}}$.

Let us admit this proposition for a while and investigate its consequences.

Corollary 7.3. *Let $\widetilde{P}$ be the solution of (S_{dyn}) and denote ν its marginal flow and $\widetilde{J}$ its jump kernel. Suppose that $1 < \int_{[0,1]} \nu \widetilde{J}_t(\mathcal{X}^2)/\nu J_t(\mathcal{X}^2)\, dt \le \infty$. Then there exists some $K > 0$ such that $\nu \widetilde{J}_t(\mathcal{X}^2) - \nu J_t(\mathcal{X}^2) = K$, for all $t \in [0,1]$.*

Proof. As $\widetilde{P}$ solves (S_{dyn}), we have $H(\widetilde{P}|R) < \infty$ and also $\nu \widetilde{J}_t(\mathcal{X}^2) > 0$ for all $t \in [0,1]$, see [Léo14, Prop. 4.2]. Hence, we are allowed to apply Proposition 7.1. But as $\widetilde{P}$ solves (S_{dyn}), the only time change $\widetilde{\tau}$ which minimizes $\tau \mapsto H(\widetilde{P}^\tau|R)$ is the identity; recall that (S_{dyn}) admits a unique minimizer. Therefore $\widetilde{\tau}(t) = t$ for all $t \in [0,1]$ and (7.2) becomes $\nu \widetilde{J}_t(\mathcal{X}^2) - \nu J_t(\mathcal{X}^2) = K$, almost everywhere. Finally we can remove this "almost" since this limitation comes from (7.14) for taking *absolutely continuous* changes of time into account and in the present case $\tau = \mathrm{Id}$ is differentiable everywhere. $\qquad\square$

With Corollary 7.3 at hand, we can prove the main result of the section.

Theorem 7.4. *There exists some $K > 0$ such that $\mu \widehat{J}_t(\mathcal{X}^2) = K$, for all $t \in [0,1]$.*

In particular, when the distance d is the standard discrete distance $d_\sim$: i.e. for all $x, y \in \mathcal{X}$, $x \sim y \iff d_\sim(x,y) = 1$, then the displacement interpolation μ has a constant speed.

Proof. The second statement is a direct consequence of the first one with (3.8).

Let us prove the first statement. For each k, let $\widehat{P}^k$ be the solution of $(\mathrm{S}_{\mathrm{dyn}}^k)$. We denote μ^k and $\widehat{J}^k$ its marginal flow and jump kernel. Let us first show that for all $t \in [0,1]$,

$$\lim_{k\to\infty} \int_{[0,t]} \mu^k \widehat{J}_s^k(\mathcal{X}^2)\, ds = \int_{[0,t]} \mu \widehat{J}_s(\mathcal{X}^2)\, ds. \tag{7.5}$$

As $\int_{[0,t]} \mu^k \widehat{J}_s^k(\mathcal{X}^2)\, ds = E_{\widehat{P}^k} N_t$ where $N_t := \sum_{0<s<t} \mathbf{1}_{\{X_{s-}\neq X_s\}}$ is the number of jumps during $[0,t]$, (7.5) doesn't directly follow from $\lim_{k\to\infty} \widehat{P}^k = \widehat{P}$ (see Theorem 2.7), because N_t is unbounded. We must strengthen Theorem 2.7 to authorize the test functions $0 \leq N_t \leq N_1 := N$, $0 \leq t \leq 1$. To do so, it is enough to show that Lemma 5.5 can be improved as follows: the sequence $(I^k)_{k\geq 2}$ is equi-coercive and $\Gamma\text{-}\lim_{k\to\infty} I^k = I$ in $\mathrm{P}_N(\Omega) := \{P \in \mathrm{P}(\Omega); E_P N < \infty\}$ with respect to the topology weakened by the functions $C_b(\Omega) \cup \{N_t; 0 \leq t \leq 1\}$. For this strengthening to hold, it is sufficient that $(I^k)_{k\geq 2}$ is equi-coercive in $\mathrm{P}_N(\Omega)$. Inspecting the proof of Lemma 5.5, we see that $I^k \geq I + c$ for all k and some constant c. Consequently, it remains to show that the function I is coercive in $\mathrm{P}_N(\Omega)$. But this follows from the proof of Lemma 5.3, noticing that $0 \leq N_t \leq N \leq \ell$. This completes the proof of (7.5).

We are going to apply Corollary 7.3 to the solution $\widehat{P}^k$ of $(\mathrm{S}_{\mathrm{dyn}}^k)$, for any large enough $k \geq 1$. Hence, we have to check that $\int_{[0,1]} \mu^k \widehat{J}_t^k(\mathcal{X}^2)/\mu^k J_t^k(\mathcal{X}^2)\, dt > 1$. By (2.3) and since it is assumed that $d(x,y) \geq 1$ for all distinct $x,y \in \mathcal{X}$,

$$\sup_{0\leq t\leq 1} \mu^k J_t^k(\mathcal{X}^2) \leq \bar{J}/k \tag{7.6}$$

where $\bar{J} := \sup_{t,x} J_{t,x}(\mathcal{X}) < \infty$ and with (7.5),

$$\lim_{k\to\infty} \int_{[0,1]} \mu^k \widehat{J}_t^k(\mathcal{X}^2)\, dt = \int_{[0,1]} \mu \widehat{J}_t(\mathcal{X}^2)\, dt =: K > 0. \tag{7.7}$$

It follows that $\int_{[0,1]} \mu^k \widehat{J}_t^k(\mathcal{X}^2)/\mu^k J_t^k(\mathcal{X}^2)\, dt \geq Kk/(2\bar{J}) > 1$, for any large enough k.

Corollary 7.3 tells us that there exists $K_k > 0$ such that $\mu^k \widehat{J}_t^k(\mathcal{X}^2) - \mu^k J_t^k(\mathcal{X}^2) = K_k$, for all $t \in [0,1]$. With (7.6) and (7.7), integrating and taking the limit leads us to

$$\lim_{k\to\infty} \int_{[0,t]} \mu^k \widehat{J}_s^k(\mathcal{X}^2)\, ds = Kt, \quad t \in [0,1].$$

We conclude with (7.5) that $\int_{[0,t]} \mu \widehat{J}_s(\mathcal{X}^2)\, ds = Kt$ for all $0 \leq t \leq 1$, which is the announced result. $\qquad\square$

Proof of Proposition 7.1. A consequence of Girsanov's formula, see [Léo12b, Thm. 2.11] for a related result, is

$$H(P|R) = H(P_0|m) + \int_{[0,1]} dt \int_{\mathcal{X}^2} \rho^*\left(\frac{dj_{t,x}}{dJ_{t,x}}(y)\right) P_t(dx) J_{t,x}(dy),$$

where

$$\rho^*(a) := \begin{cases} a\log a - a + 1 & \text{if } a > 0, \\ 1 & \text{if } a = 0, \\ \infty & \text{if } a < 0. \end{cases}$$

Let us denote

$$
\begin{aligned}
\nu_t(dx) &:= P_t(dx) \\
\mathbf{q}_t(dxdy) &:= \nu_t(dx)J_{t,x}(dy) \\
\mathbf{v}_t(x,y) &:= \frac{dj_{t,x}}{dJ_{t,x}}(y)
\end{aligned}
$$

Remark that $\nu = (\nu_t)_{0 \le t \le 1}, \mathbf{q} = (\mathbf{q}_t)_{0 \le t \le 1}$ and $\mathbf{v} = (\mathbf{v}_t)_{0 \le t \le 1}$ are fixed quantities. For any positive bounded measure $q \in \mathcal{M}_{b,+}(\mathcal{X}^2)$ and any measurable function $v : \mathcal{X}^2 \to \mathbb{R}$ on $\mathcal{X}^2$, we define

$$
L(q,v) := \int_{\mathcal{X}^2} \rho^*(v)\, dq \in [0, \infty]
$$

For any change of time τ we have $P_0^\tau = P_0$ and

$$
H(P^\tau | R) = H(P_0|m) + \int_{[0,1]} L(\mathbf{q}_{\tau(t)}, \dot\tau(t)\mathbf{v}_{\tau(t)})\, dt.
$$

Therefore, we wish to minimize

$$
\tau \mapsto \int_{[0,1]} L(\mathbf{q}_{\tau(t)}, \dot\tau(t)\mathbf{v}_{\tau(t)})\, dt \tag{7.8}
$$

among all changes of time $\tau : [0,1] \to [0,1]$. To achieve this goal, some preliminary results are necessary. They are stated and proved below at Lemmas 7.9, 7.13 and 7.16. With Lemma 7.9, we see that we are in position to apply Lemma 7.16. We conclude with (7.8), Lemma 7.13 and Lemma 7.16.

It remains to prove Lemmas 7.9, 7.13 and 7.16.

Statement and proof of Lemma 7.9. We show that the finite entropy condition $H(P|R) < \infty$ implies that the average number of jumps under P is finite.

Lemma 7.9. *Let $P \in \mathrm{P}(\Omega)$ be a random walk. We denote $(\nu j_t(\mathcal{X}^2))_{0 \le t \le 1}$ its average rate of mass displacement. Let us assume that $H(P|R) < \infty$. Then, $\int_{[0,1]} \nu j_t(\mathcal{X}^2)\, dt < \infty$.*

Proof. We see that

$$
\int_{[0,1]} \nu j_t(\mathcal{X}^2)\, dt = E_P N
$$

with $N = \sum_{0 < t < 1} \mathbf{1}_{\{X_{t^-} \ne X_t\}}$ the number of jumps. Let us denote $R_* := \frac{dP_0}{dm}(X_0)\, R \in \mathrm{P}(\Omega)$ the Markov random walk with initial marginal P_0 and forward jump kernel J. We have $H(P|R) = H(P_0|m) + H(P|R_*)$ which implies that $H(P|R_*) < \infty$. Taking advantage of the Fenchel inequality $ab \le a \log a + e^{b-1}$, we obtain

$$
E_P N = E_{R_*}\left(\frac{dP}{dR_*}N\right) \le H(P|R_*) + E_{R_*}e^N.
$$

But our assumption (2.3) implies that $N_{\#}R_*$ is stochastically dominated by the Poisson law with parameter $\sup_{t,x} J_{t,x}(\mathcal{X})$. Therefore, $E_{R_*}e^N < \infty$ and we have proved that $\int_{[0,1]} \nu j_t(\mathcal{X}^2)\, dt < \infty$ when $H(P|R) < \infty$.   □

Statement and proof of Lemma 7.13. To prove Lemma 7.13, we need the preliminary Lemma 7.10 which is stated and proved below. We consider the set

$$\mathcal{K} := \left\{ (q, v); q \in \mathcal{M}_{b,+}(\mathcal{X}^2), v : \mathcal{X}^2 \to [0, \infty) \text{ measurable } : L(q, v) < \infty, \ q(v > 0) > 0 \right\}.$$

For any $(q, v) \in \mathcal{K}$, as q is a bounded measure, v belongs to the Orlicz space $L \log L(\mathcal{X}^2, q)$ and its corresponding norm $|v|_q := \|v\|_{L \log L(\mathcal{X}^2, q)}$ is finite. As in addition $q(v > 0) > 0$, we have $|v|_q > 0$ and we are allowed to define

$$L_{q,v}(\alpha) := L(q, \alpha v / |v|_q), \quad \alpha \in \mathbb{R}, (q, v) \in \mathcal{K}.$$

Let us fix $K \geq 0$ and define $\beta_K(q, v) \in (0, \infty)$ to be the slope of the affine function $\widetilde{L}^K_{q,v} : \mathbb{R} \to \mathbb{R}$ which is tangent to the convex function $L_{q,v}$ and satisfies $\widetilde{L}^K_{q,v}(0) = -K$: $\widetilde{L}^K_{q,v}(\alpha) = \beta_K(q, v)\alpha - K$, $\alpha \in \mathbb{R}$. Such a below tangent exists since $L_{q,v}(0) = L(q, 0) = q(\mathcal{X}^2) > 0$. Furthermore, since $L_{q,v} \geq 0$, we also have $\beta_K(q, v) > 0$.
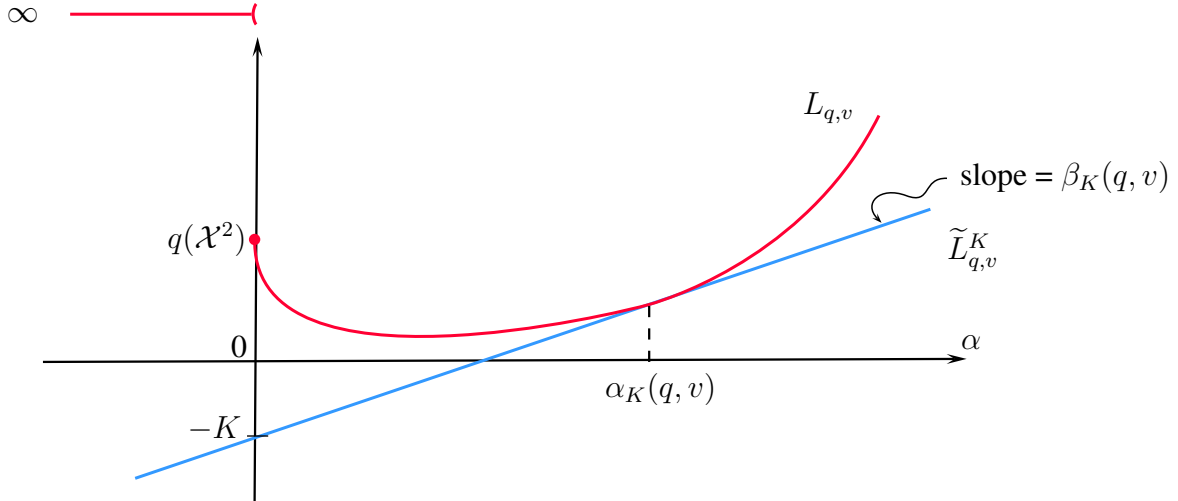
We denote $\alpha_K(q, v) > 0$ the solution of $L_{q,v}(\alpha) = \widetilde{L}^K_{q,v}(\alpha)$, which happens to be positive and unique (since $L_{q,v}$ is strictly convex). Define

$$\begin{cases} \lambda_K(q, v) &:= \beta_K(q, v)|v|_q, \\ a_K(q, v) &:= \alpha_K(q, v)/|v|_q, \\ v_K(q, v) &:= a_K(q, v)v, \end{cases} \quad (q, v) \in \mathcal{K}.$$

The functions $\lambda_K(q, \cdot)$, $a_K(q, \cdot)$ and $v_K(q, \cdot)$ are respectively positively homogeneous of degree 1, -1 and 0 on the convex cone $\mathcal{K}_q := \{v : (q, v) \in \mathcal{K}\}$. We also define

$$L_K(q, v) := \begin{cases} \lambda_K(q, v) - K, & \text{if } (q, v) \in \mathcal{K}, \\ -K, & \text{if } q(v \neq 0) = 0, \\ +\infty, & \text{if } q(v < 0) > 0. \end{cases}$$

Remark that $L_K(q, \cdot) + K$ is the largest positively 1-homogeneous function below the convex function $L(q, \cdot) + K$.

Lemma 7.10. *For any $K \geq 0$, we have $L_K \leq L$ and for any $(q, v) \in \mathcal{K}$ the equality $L(q, v) = L_K(q, v)$ is achieved if and only if $v = v_K(q, v)$ or equivalently $a_K(q, v) = 1$. Furthermore, for any $(q, v) \in \mathcal{K}$ such that $v > 0$, q-a.e., $\int_{\mathcal{X}^2}(v_K(q, v) - 1) \, dq = K$.*

Proof. Since $L_K(q, v) = \widetilde{L}^K_{q,v}(|v|_q)$, for any $(v, q) \in \mathcal{K}$, the inequality $\widetilde{L}^K_{q,v} \leq L_{q,v}$ implies that $L_K \leq L$ with equality on $\mathcal{K}$ if and only if $|v|_q = \alpha_K(q, v)$, that is $v = v_K(q, v)$.

Let us prove last statement. Denoting $H_{q,v}(\beta) := \sup_{\alpha \in \mathbb{R}} \{\alpha\beta - L_{q,v}(\alpha)\} \in \mathbb{R}$, $\beta \in \mathbb{R}$, the convex conjugate of $L_{q,v}$, we have

$$H_{q,v}(\beta_K(q,v)) = K. \tag{7.11}$$

Let us define, for any measurable function $p : \mathcal{X}^2 \to \mathbb{R}$,

$$H(q,p) := \int_{\mathcal{X}^2} \rho(p)\,dq = \int_{\mathcal{X}^2} (e^p - 1)\,dq \in (-\infty, \infty]$$

where

$$\rho(b) := e^b - 1, \quad b \in \mathbb{R}$$

is the convex conjugate of ρ^*. For any real numbers $a, b : ab \le \rho^*(a) + \rho(b)$ with equality if and only if $a = e^b$. Hence, for any $(q,v) \in \mathcal{K}$ and any measurable function p on $\mathcal{X}^2$ such that $\int_{\mathcal{X}^2} e^p\,dq < \infty$, we have $-\infty \le \langle p, v\rangle_q := \int_{\mathcal{X}^2} pv\,dq \le L(q,v) + H(q,p) < \infty$ with equality if and only if $v = e^p$, q-a.e. As $v > 0$, q-a.e., the equality is realized with $p = p^v := \log v$, that is

$$L(q,v) = \langle p^v, v\rangle_q - H(q, p^v). \tag{7.12}$$

Now, let $\alpha := |v|_q > 0$ be given. For any $b \in \mathbb{R}$, $\alpha b \le L_{q,v}(\alpha) + H_{q,v}(b)$ and the equality is realized at $\beta = L'_{q,v}(\alpha) = \langle \log v, v/\alpha\rangle_q = \langle p^v, v\rangle_q/\alpha$. Therefore,

$$L(q,v) = L_{q,v}(\alpha) = \alpha\beta - H_{q,v}(\beta) = \langle p^v, v\rangle_q - H_{q,v}(\beta).$$

Comparing with (7.12) leads us to $H(q, p^v) = H_{q,v}(\beta)$. In particular, with $v = v_K(q,v)$, we see that $\alpha = |v_K(q,v)|_q = \alpha_K(q,v)$ and the corresponding conjugate parameter is $\beta = \beta_K(q,v)$. It follows with (7.11) that $H(q, \log v_K(q,v)) = H(q, p^{v_K(q,v)}) = H_{q,v}(\beta_K(q,v)) = K$, which is the desired result. $\qquad\square$

As a consequence of Lemma 7.10, we obtain the following result.

Lemma 7.13. *Suppose that $\widehat{\tau}$ is a change of time which solves the differential equation*

$$\dot{\tau}(t) = a_K(\mathbf{q}_{\tau(t)}, \mathbf{v}_{\tau(t)}), \quad t \in [0,1], \ a.e. \tag{7.14}$$

for some $K \ge 0$. Then, $\tau \mapsto \int_{[0,1]} L(\mathbf{q}_{\tau(t)}, \dot{\tau}(t)\mathbf{v}_{\tau(t)})\,dt$ attains its minimum value among all changes of time at $\tau = \widehat{\tau}$. Moreover, if for almost all $t \in [0,1]$, $\mathbf{v}_t > 0$, then

$$\int_{\mathcal{X}^2} (\widehat{\mathbf{v}}_t - 1)\,d\widehat{\mathbf{q}}_t = K, \quad \text{for a.e. } t \in [0,1]$$

where we denote $\widehat{\mathbf{q}}_t := \mathbf{q}_{\widehat{\tau}(t)}$ and $\widehat{\mathbf{v}}_t := \dot{\widehat{\tau}}(t)\mathbf{v}_{\widehat{\tau}(t)}$.

Proof. As $\widehat{\tau}$ solves (7.14), we have $a_K(\widehat{\mathbf{q}}_t, \widehat{\mathbf{v}}_t) = 1$, for a.e. $t \in [0,1]$. With Lemma 7.10, this implies that

$$\int_{[0,1]} L(\widehat{\mathbf{q}}_t, \widehat{\mathbf{v}}_t)\,dt = \int_{[0,1]} L_K(\widehat{\mathbf{q}}_t, \widehat{\mathbf{v}}_t)\,dt.$$

As $\lambda_K(q, \cdot)$ is 1-homogeneous, we see that for any change of time τ,

$$\int_{[0,1]} L_K(\mathbf{q}_{\tau(t)}, \dot{\tau}(t)\mathbf{v}_{\tau(t)})\,dt = \int_{[0,1]} \lambda_K(\mathbf{q}_{\tau(t)}, \dot{\tau}(t)\mathbf{v}_{\tau(t)})\,dt - K$$

$$= \int_{[0,1]} \lambda_K(\mathbf{q}_t, \mathbf{v}_t)\,dt - K = \int_{[0,1]} L_K(\mathbf{q}_t, \mathbf{v}_t)\,dt$$

is a quantity which doesn't depend on τ. Hence, for any change of time τ,

$$\int_{[0,1]} L(\widehat{\mathbf{q}}_t, \widehat{\mathbf{v}}_t)\, dt = \int_{[0,1]} L_K(\widehat{\mathbf{q}}_t, \widehat{\mathbf{v}}_t)\, dt = \int_{[0,1]} L_K(\mathbf{q}_{\tau(t)}, \dot{\tau}(t)\mathbf{v}_{\tau(t)})\, dt$$

$$\leq \int_{[0,1]} L(\mathbf{q}_{\tau(t)}, \dot{\tau}(t)\mathbf{v}_{\tau(t)})\, dt$$

where the last inequality follows from $L_K \leq L$, see Lemma 7.10. This proves the minimal attainment at $\widehat{\tau}$.

The last statement follows from (7.14) $\iff \widehat{\mathbf{v}}_t = v_K(\widehat{\mathbf{q}}_t, \widehat{\mathbf{v}}_t)$ and Lemma 7.10. $\qquad\square$

Statement and proof of Lemma 7.16. Now, we prove Lemma 7.16. To do so, we need the following preliminary Lemma 7.15.

Lemma 7.15. *For each $K \geq 0$,*

$$a_K(q, v) = \frac{K + q(\mathcal{X}^2)}{\int_{\mathcal{X}^2} v\, dq}, \quad (q, v) \in \mathcal{K}.$$

Proof. Let us pick $(q, v) \in \mathcal{K}$ and $\beta \in \mathbb{R}$. We have $H_{q,v}(\beta) = \sup_{a \in \mathbb{R}} \{a\beta - L_{q,v}(a)\}$ and the supremum is attained at α, solution of $\beta = \int_{\mathcal{X}^2} \rho^{*\prime}(\alpha u) u\, dq = \int_{\mathcal{X}^2} \log(\alpha u) u\, dq$ where $u := v/|v|_q$. Therefore, $H_{q,v}(\beta) = \alpha\beta - L_{q,v}(\alpha) = \int_{\mathcal{X}^2} [\alpha u \log(\alpha u) - \rho^*(\alpha u)]\, dq = \int_{\mathcal{X}^2} (\alpha u - 1)\, dq$. Choosing $\beta = \beta_K(q, v)$ corresponds to $\alpha_K(q, v)$ and we obtain with (7.11) that $K = H_{q,v}(\beta_K(q, v)) = \alpha_K(q, v) \int_{\mathcal{X}^2} u\, dq - q(\mathcal{X}^2)$. Hence, $a_K(q, v) = \alpha_K(q, v)/|v|_q = (K + q(\mathcal{X}^2))/\int_{\mathcal{X}^2} v\, dq$. $\qquad\square$

As a consequence of Lemma 7.15, we obtain the following result.

Lemma 7.16. *Let ν and j be such that*

(i) $\int_{[0,1]} \nu j_t(\mathcal{X}^2)\, dt < \infty$;

(ii) $\nu j_t(\mathcal{X}^2) > 0$, for all $t \in [0, 1]$;

(iii) $1 < \int_{[0,1]} \nu j_t(\mathcal{X}^2)/\nu J_t(\mathcal{X}^2)\, dt \leq \infty$.

Then, there exist a constant $\widehat{K} > 0$ and a change of time $\widehat{\tau}$ such that (7.14) holds:

$$\dot{\widehat{\tau}}(t) = a_{\widehat{K}}(\mathbf{q}_{\widehat{\tau}(t)}, \mathbf{v}_{\widehat{\tau}(t)}), \quad t \in [0, 1], a.e.$$

Proof. Plugging $\mathbf{q}_s := \nu J_s$ and $\mathbf{v}_s(x, y) := dj_{s,x}/dJ_{s,x}(y)$ into $a_K(\mathbf{q}_s, \mathbf{v}_s)$, we see that (ii) implies that $(\mathbf{q}_s, \mathbf{v}_s) \in \mathcal{K}$ for all $0 \leq s \leq 1$. Hence, with Lemma 7.15 we obtain that (7.14) writes as

$$\dot{\tau}(t)\varphi_K(\tau(t)) = 1$$

where

$$\varphi_K(s) := a_K(\mathbf{q}_s, \mathbf{v}_s)^{-1} = \frac{\nu j_s(\mathcal{X}^2)}{K + \nu J_s(\mathcal{X}^2)}, \quad 0 \leq s \leq 1.$$

Let $\Phi_K(t) := \int_0^t \varphi_K(s)\, ds \in [0, \infty]$. By assumption (i), for any $K > 0$, $\Phi_K(1) < \infty$. It follows that Φ_K is absolutely continuous. Assumption (ii) implies that it is strictly increasing so that one can define $\tau_K(t) := \Phi_K^{-1}(t)$ which solves $\dot{\tau}(t) = a_K(\mathbf{q}_{\tau(t)}, \mathbf{v}_{\tau(t)})$ almost everywhere on $[0, \Phi_K(1))$.

It remains to show that there exists some $\widehat{K} > 0$ such that $\Phi_{\widehat{K}}(1) = 1$. But this is insured by the assumption (iii) which states that $\Phi_0(1) > 1$, since with (i) we see that $K \in (0, \infty) \mapsto \Phi_K(1) \in (0, \infty)$ is a continuous decreasing function from $\Phi_0(1)$ to $\lim_{K \to \infty} \Phi_K(1) = 0$. $\qquad\square$

APPENDIX A. RANDOM WALK ON A GRAPH

We give some basic informations about random walks on a graph.

The jump kernel. A Markov random walk on the graph $(\mathcal{X}, \sim)$ is a time-continuous Markov process which is specified by its infinitesimal generator $L = (L_t)_{0 \leq t < 1}$ acting on any real function $u \in \mathbb{R}^{\mathcal{X}}$ with a finite support via the formula

$$L_t u(x) = \sum_{y:y \sim x} [u(y) - u(x)] \, j_{t,x}(y), \qquad x \in \mathcal{X}, \ 0 \leq t < 1$$

where $j_{t,x}(y) \geq 0$ is the average frequency of jumps from x to y at time t. As a convention, we set $j_{t,x}(y) = 0$ as soon as x and y are not neighbours. The corresponding jump kernel is

$$j_{t,x} := \sum_{y:y \sim x} j_{t,x}(y) \delta_y \in \mathrm{M}_+(\mathcal{X}), \qquad x \in \mathcal{X}, \ 0 \leq t < 1.$$

The operator L is well-defined for any bounded function $u \in \mathbb{R}^{\mathcal{X}}$ provided that

$$\sup_{x \in \mathcal{X}} j_{t,x}(\mathcal{X}) < \infty, \tag{A.1}$$

where $j_{t,x}(\mathcal{X}) := \sum_{y:y \sim x} j_{t,x}(y)$ denotes the global jump intensity at $x \in \mathcal{X}$. The bound (A.1) ensures that the random walk performs almost surely finitely many jumps during the unit time interval $[0, 1]$. Therefore, Ω as defined in this article is the relevant path space to be considered. When j is assumed to satisfy (A.1), it uniquely specifies $Q \in \mathrm{M}_+(\Omega)$ up to its initial law $Q_0 \in \mathrm{M}_+(\mathcal{X})$.

In case L doesn't depend on t, the random walk is said to be time-homogeneous. In this special case, its dynamics is described as follows. Once at site x, the walker waits during a random time with exponential law $\mathcal{E}(j_x(\mathcal{X}))$, and then decides to jump at y according to the probability $j_x(\mathcal{X})^{-1} \sum_{y:y \sim x} j_x(y) \delta_y$, and so on; all these random events being mutually independent.

Some interesting examples of reference random walks R. One may require that R is reversible. This means that it is time-homogeneous and that there is a (possibly unbounded) positive measure $m \in \mathrm{M}_+(\mathcal{X})$ on $\mathcal{X}$ such that, not only R is m-stationary i.e.: $R_t = m, \forall 0 \leq t \leq 1$, but also that R is invariant with respect to time reversal i.e.: for any subinterval $[u, v] \subset [0, 1]$,

$$(X_{(u+v-t)^-}; u \leq t \leq v)_{\#} R = (X_t; u \leq t \leq v)_{\#} R.$$

This happens if and only if the following time-homogeneous detailed balance condition

$$m_x J_x(y) = m_y J_y(x), \quad \forall x \sim y \in \mathcal{X} \tag{A.2}$$

is satisfied; compare (1.15). As it is assumed that $J_x(y) > 0, \forall x \sim y$ and the graph is irreducible, this implies that $m_x > 0$ for all $x \in \mathcal{X}$.

Simple random walk. An important example of such a walk is the simple random walk $R^o \in \mathrm{P}(\Omega)$ on $\mathcal{X}$. The dynamics of R^o is specified by the jump kernel

$$J_x^o := n_x^{-1} \sum_{y:y \sim x} \delta_y, \qquad x \in \mathcal{X}. \tag{A.3}$$

The successive waiting times are independent and identically distributed with the exponential law $\mathcal{E}(1)$ and the walker jumps from any site choosing a neighbour uniformly at

random. Solving (A.2), one sees that the corresponding reversing measures are multiples of

$$m^o := \sum_{x \in \mathcal{X}} n_x \delta_x. \tag{A.4}$$

Note that m^o is unbounded whenever $\mathcal{X}$ is an infinite set, since the irreducibility assumption implies that $n_x \geq 1$ for all x.

As the simple random walk is analogous to the Brownian motion on a Riemannian manifold, the measure m^o plays the role of the volume measure on the graph.

Counting random walk. It corresponds to $J_x = \sum_{y:y \sim x} \delta_y, x \in \mathcal{X}$ whose reversing measure is the counting measure $m = \sum_{x \in \mathcal{X}} \delta_x$.

A generic class of m-reversible random walks. Take some measure $m = \sum_{x \in \mathcal{X}} m_x \delta_x$ on $\mathcal{X}$ with $m_x > 0, \forall x \in \mathcal{X}$ and consider the jump kernel

$$J_x := \sum_{y:y \sim x} \frac{s(x,y)}{\sqrt{n_x n_y}} \sqrt{\frac{m_y}{m_x}} \, \delta_y, \qquad x \in \mathcal{X}, \tag{A.5}$$

where s is a symmetric function. Assume that there exists some constant $1 \leq c < \infty$ such that

$$m_y/n_y \leq c \, m_x/n_x, \quad \forall x \sim y$$

and some $0 < \sigma < \infty$ such that the symmetric function s satisfies

$$0 < s(x,y) = s(y,x) \leq \sigma, \qquad \forall x \sim y \in \mathcal{X}.$$

Then, J verifies (A.1), so that R is a measure on Ω. As J clearly verifies (A.2), R is m-reversible. Moreover, R is equivalent to the simple random walk R^o, i.e. for any measurable $A \subset \Omega$, $R(A) > 0 \iff R^o(A) > 0$.

APPENDIX B. RELATIVE ENTROPY

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

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

if $p \ll r$ and $H(p|r) = \infty$ otherwise.

More precisely, when r is a probability measure, we have

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

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

$$H(p|r) = \inf H(\cdot|r) = 0 \iff p = r. \tag{B.1}$$

If r is unbounded, one must restrict the definition of $H(\cdot|r)$ to some subset of $\mathrm{P}(Y)$ as follows. As r is assumed to be σ-finite, there exists a measurable function $W : Y \to [1, \infty)$ such that

$$z_W := \int_Y e^{-W} \, dr < \infty. \tag{B.2}$$

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

$$H(p|r) := H(p|r_W) - \int_Y W \, dp - \log z_W \in (-\infty, \infty] \qquad (\mathrm{B}.3)$$

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

Therefore, $H(p|r)$ is well-defined for any $p \in \mathrm{P}(Y)$ such that $\int_Y W \, dp < \infty$ for some measurable nonnegative function W verifying (B.2).

It follows from the strict convexity of $H(\cdot|r_W)$ and (B.3) that $H(\cdot|r)$ is also strictly convex.

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

$$q(dy) = \int_Z q(dy|\phi = z) \, \phi_\# q(dz)$$

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

$$H(p|r) = H(\phi_\# p|\phi_\# r) + \int_Z H\Big(p(\cdot \mid \phi = z)|r(\cdot \mid \phi = z)\Big) \, \phi_\# p(dz), \qquad (\mathrm{B}.4)$$

is valid for any $p \in \mathrm{P}(Y)$ and any σ-finite $r \in \mathrm{M}_+(Y)$. In particular, as $r(\cdot \mid \phi = z)$ is a probability measure for each z, with (B.1) we see that

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

with equality if and only if

$$p(\cdot \mid \phi = z) = r(\cdot \mid \phi = z), \quad \forall z, \phi_\# p\text{-a.s.} \qquad (\mathrm{B}.5)$$

REFERENCES

[AB93]   G. Anzellotti and S. Baldo. Asymptotic development by Γ-convergence. *Appl. Math. Optim.*, 27(2):105–123, 1993.

[AKP04]   L. Ambrosio, B. Kirchheim, and A. Pratelli. Existence of optimal transport maps for crystalline norms. *Duke Math. J.*, 125(2):207–241, 2004.

[Amb03]   L. Ambrosio. Lecture notes on optimal transport problems. In *Mathematical aspects of evolving interfaces (Funchal, 2000)*, volume 1812 of *Lecture Notes in Math.* Springer, Berlin, 2003.

[AP03]   L. Ambrosio and A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation. In L.A. Caffarelli and S. Salsa, editors, *Optimal transportation and applications (Martina Franca, 2001)*, volume 1813 of *Lecture Notes in Math.*, pages 123–160. Springer, Berlin, 2003.

[Att96]   H. Attouch. Viscosity solutions of minimization problems. *SIAM J. Optim.*, 3:769–806, 1996.

[BB00]   J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[BS09]   A.-I. Bonciocat and K.T. Sturm. Mass transportation and rough curvature bounds for discrete spaces. *J. Funct. Anal.*, 256:2944–2966, 2009.

[CDP11]   T. Champion and L. De Pascale. The Monge problem in $\mathbb{R}^d$. *Duke Math. J.*, 157(3):551–572, 2011.

[CEMS01] D. Cordero-Erausquin, R.J. McCann, and M. Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.*, 146(2):219–257, 2001.

[CFM02]   L. Caffarelli, M. Feldman, and R. McCann. Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs. *J. Amer. Math. Soc.*, 15(1):1–26, 2002.

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

[EG99]   L. C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem. *Mem. Amer. Math. Soc.*, 137(653):viii+66, 1999.

[EM12]   M. Erbar and J. Maas. Ricci curvature of finite markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.*, 206(3):997–1038, 2012.

[GRST]   N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Displacement convexity of entropy and related inequalities on graphs. Published online August 2013, Probab. Theory Related Fields.

[Hil]   E. Hillion. Contraction of measures on graphs. Preprint.

[Hil10]   E. Hillion. *Analysis and geometry in metric-measure spaces : Borell-Brascamp-Lieb inequalities and Olkin-Shepp conjecture.* PhD thesis, Toulouse U., 2010.

[Hil12]   E. Hillion. Concavity of entropy along binomial convolutions. *Electron. Commun. Probab.*, 17:no. 4, 9, 2012.

[Joh07]   O. Johnson. Log-concavity and the maximum entropy property of the Poisson distribution. *Stochastic Process. Appl.*, 117(6):791–802, 2007.

[Léoa]   C. Léonard. On the convexity of the entropy along entropic interpolations. Preprint, arXiv:1310.1274.

[Léob]   C. Léonard. Some properties of path measures. Preprint, arXiv:1308.0217.

[Léo12a] C. Léonard. From the Schrödinger problem to the Monge-Kantorovich problem. *J. Funct. Anal.*, 262(1879-1920), 2012.

[Léo12b] C. Léonard. Girsanov theory under a finite entropy condition. In *Séminaire de probabilités de Strasbourg, vol. XLIV.*, pages 429–465. Lecture Notes in Mathematics 2046. Springer, 2012.

[Léo14]   C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst. A*, 34(4):1533–1574, 2014.

[LRZ]   C. Léonard, S. Rœlly, and J.-C. Zambrini. On the time symmetry of some stochastic processes. Preprint, arXiv:1308.0576.

[LV09]   J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math.*, 169(3):903–991, 2009.

[Maa11]   J. Maas. Gradient flows of the entropy for finite markov chains. *J. Funct. Anal*, 261:2250–2292, 2011.

[McC94]   R.J. McCann. *A convexity theory for interacting gases and equilibrium crystals.* PhD thesis, Princeton Univ., 1994.

[McC97]   R.J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128:153–179, 1997.

[Mie]   A. Mielke. Geodesic convexity of the relative entropy in reversible markov chains. To appear. DOI: 10.1007/s00526-012-0538-8.

[Mie11]   A. Mielke. A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Nonlinearity*, 24(4):1329–1346, 2011.

[Nag93]   M. Nagasawa. *Schrödinger equations and diffusion theory*, volume 86 of *Monographs in Mathematics.* Birkhäuser Verlag, Basel, 1993.

[OV00]   F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.

[OV12]   Y. Ollivier and C. Villani. A curved Brunn–Minkowski inequality on the discrete hypercube, Or: What is the Ricci curvature of the discrete hypercube? *SIAM J. Discrete Math.*, 26(3):983–996, 2012.

[Sch31]   E. Schrödinger. Über die Umkehrung der Naturgesetze. *Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math.*, 144:144–153, 1931.

[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. Available at: `http://archive.numdam.org/ARCHIVE/AIHP/`.

[Stu06a] K-T. Sturm. On the geometry of metric measure spaces, I. *Acta Math*, 196:65–131, 2006.

[Stu06b] K-T. Sturm. On the geometry of metric measure spaces, II. *Acta Math*, 196:133–177, 2006.

[Sud79]   V. N. Sudakov. Geometric problems in the theory of infinite-dimensional probability distributions. *Proc. Steklov Inst. Math.*, (2):i–v, 1–178, 1979. Cover to cover translation of Trudy Mat. Inst. Steklov **1**41 (1976).

[Sv05]   K-T. Sturm and M-K. von Renesse. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005.

[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