The Annals of Probability
2016, Vol. 44, No. 3, 1864–1915
DOI: 10.1214/15-AOP1012

LAZY RANDOM WALKS AND OPTIMAL TRANSPORT ON GRAPHS

BY CHRISTIAN LÉONARD

Modal-X. Université Paris Ouest

This paper is about the construction of displacement interpolations of probability distributions 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 entropy minimization problems under marginal constraints, called Schrödinger problems, which are associated with random walks. Displacement interpolations are defined as the limit of the time-marginal flows of the solutions to the Schrödinger problems as the jump frequencies of the random walks tend down to zero. 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.

Introduction.

Aim of the paper. Displacement interpolations on $\mathbb{R}^n$ were introduced by McCann (1994) and extended later to a geodesic space $(\mathcal{X}, d)$ where they are defined as constant speed 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 Lott and Villani (2009), Sturm (2006a, 2006b), Villani (2009). Indeed, as discovered by McCann (1994, 1997), Otto and Villani (2000), Cordero-Erausquin, McCann and Schmuckenschläger (2001), von Renesse and Sturm (2005) 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 displacement 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 between probability measures on a discrete metric graph. In this discrete setting, we pro-

Received November 2013; revised February 2015.
MSC2010 subject classifications. 60J27, 65K10.
Key words and phrases. Displacement interpolation, discrete metric graph, optimal transport, Schrödinger problem, random walks, entropy minimization, Γ-convergence.

pose natural substitutes for the constant speed geodesics. As a by-product of our approach, we also obtain new results about optimal transport on graphs.

Although these displacement interpolations are designed for obtaining displacement convexity of the entropy, this topic is not investigated in the present article.

Some notions to be made precise. Let us begin with some terminology and notation.

The graph $(\mathcal{X}, \sim)$. Let $\mathcal{X}$ be a countable set of vertices equipped with the symmetric relation $x \sim y$ which means that the vertices x and y are distinct neighbours and that $\{x, y\}$ is an undirected edge.

The path space Ω. The path space $\Omega \subset \mathcal{X}^{[0,1]}$ on $\mathcal{X}$ to be considered in the present paper is the space of all left-limited, right-continuous, piecewise constant paths $\omega = (\omega_t)_{0 \le t \le 1}$ on $\mathcal{X}$ with finitely many jumps such that for all $t \in (0, 1)$, $\omega_{t^-} \neq \omega_t$ implies that ω_{t^-} and ω_t are neighbours, so that Ω is respectful of the graph structure of $(\mathcal{X}, \sim)$. See (5.2), (5.3).

Discrete metric graph. The distance d is in accordance with the graph structure of $(\mathcal{X}, \sim)$ since it is required to be intrinsic in the discrete sense. This means that for all $x, y \in \mathcal{X}$,

$$d(x, y) = \inf\{\ell(\omega); \omega \in \Omega : \omega_0 = x, \omega_1 = y\}, \tag{0.1}$$

where

$$\ell(\omega) := \sum_{0<t<1} d(\omega_{t^-}, \omega_t), \qquad \omega \in \Omega \tag{0.2}$$

is the discrete length of the discontinuous path ω. The epithet *discrete* in the expression "discrete metric graph" is important. Indeed, the standard definition of a (nondiscrete) metric graph allows for continuous mass transfer along edges. In the present paper, the topology induced by the distance d on the set $\mathcal{X}$ of vertices is discrete and the edges are not part of the state space.

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

$$\Gamma^{xy} := \{\omega \in \Omega; \omega_0 = x, \omega_1 = y, \ell(\omega) = d(x, y)\} \tag{0.3}$$

the set of all *geodesics* joining x and y: they achieve the minimum value of (0.1). Since it will be assumed that the graph is irreducible, only finitely many jumps occur along any geodesic. This is the main reason why Ω as defined above is the relevant path space for our investigation. Remark that when x and y are distinct, Γ^{xy} contains infinitely many paths since geodesics are characterized by ordered sequences of visited states, regardless of the instants of jumps. *The cut-locus is quite large.*

Random walk. We give a specific meaning to the expression "random walk". A random walk is any positive measure on the path space Ω. This is not the customary usage, but it turns out to be convenient. As a measure, it specifies the behaviour of a continuous-time random process. This piecewise constant process may not be Markov. The Markov property of a measure on Ω is defined at Definition 6.1, in accordance with the usual notion. More detail about Markov random walks are given at the Appendix A.

Why random walks have to be considered. To recover some time regularity of the paths in order to define notions of speed and acceleration without allowing any mass transfer along edges, one is enforced to do some averaging on ensembles of discontinuous sample paths. This means that one is obliged to consider expected values of piecewise constant *random* paths on $\mathcal{X}$. Therefore, we are going to consider random walks in the wide sense that was described some lines above. So doing, one lifts the paths from the discrete state space $\mathcal{X}$ up to the continuum $\mathrm{P}(\mathcal{X})$ of all probability measures on $\mathcal{X}$. Instead of searching for geodesic paths on $\mathcal{X}$, we are going to build geodesics on $\mathrm{P}(\mathcal{X})$.

A problem to be solved. The main trouble in a discrete setting comes from a serious degeneracy problem. Let us call any random walk which is supported on the set of geodesic paths on $\mathcal{X}$, a geodesic random walk. We will see below at Result 0.1 that the interpolations on $\mathrm{P}(\mathcal{X})$ that we are on the way to build are time-marginal flows of geodesic random walks. Yet we have just seen that the cut-locus is huge. *What geodesic random walk should be selected among these infinitely many candidates*? *What should be the random behaviour of the instants of jumps to specify an interpolation which is well suited for the displacement convexity of the entropy*? This article proposes a solution to this problem.

Approximation by entropy minimizers. The author proposed in Léonard (2012a) a construction of the McCann displacement interpolations in $\mathrm{P}(\mathbb{R}^n)$ as limits of minimizers of some relative entropy under marginal constraints (see the Schrödinger problem at Section 4). In the present paper, we stay as close as possible to this strategy. In a discrete metric graph setting, it will lead us to a natural way of selecting one displacement interpolation among infinitely many candidates, which is suitable for the displacement convexity of the entropy.

The main idea about this selection procedure 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 (1931, 1932) or in its close variant described in Villani's textbook Villani [(2009), Lazy gas experiment, page 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 performs some optimal transference plan. This (absolutely) cold gas experiment is called the *lazy gas experiment* in Villani (2009) where its dynamics is related to displacement interpolations. For further detail with graphical illustrations, see Léonard (2012c), Section 6.

With this thought experiment in mind, one can guess that a slowing down procedure enforces the appearance of individual geodesics.

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 $m \in \mathrm{M}_+(Y_1)$ by the measurable mapping $f : Y_1 \to Y_2$ is $f_\# m(\cdot) := m(f^{-1}(\cdot)) \in \mathrm{M}_+(Y_2)$.

Let $\Omega \subset \mathcal{X}^{[0,1]}$ be a set of paths from the time interval $[0, 1]$ to the measurable state space $\mathcal{X}$ (mainly think of Ω as defined above, but sometimes for purposes of comparison we shall refer to Ω as a set of continuous paths on a Riemannian manifold $\mathcal{X}$). The canonical process $X = (X_t)_{0 \le t \le 1}$ is defined for all $\omega = (\omega_s)_{0 \le s \le 1} \in \Omega$ by $X_t(\omega) = \omega_t \in \mathcal{X}$ for each $0 \le t \le 1$. 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

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

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

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

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

$$Q^{xy} := Q(\cdot | X_0 = x, X_1 = y) \in \mathrm{P}(\Omega), \qquad x, y \in \mathcal{X}. \tag{0.4}$$

As a general result, we have the disintegration formula

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

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

Presentation of the main results. Let us present the main results of the article in the easy framework of a discrete metric graph $(\mathcal{X},\sim,d)$ equipped with the standard distance specified by (0.1) and

$$\forall x,y\in\mathcal{X},\qquad d(x,y)=1 \quad\Longleftrightarrow\quad x\sim y. \tag{0.5}$$

Let us take any Markov random walk $R\in\mathrm{M}_+(\Omega)$ such that jumps are allowed between any neighbours in both directions. Precise assumptions are stated at Hypotheses 2.1. For instance, one may choose the simple random walk; see (A.3) at the Appendix A for detail.

For any $k\geq 1$, the slowed down version of R is the random walk $R^k\in\mathrm{M}_+(\Omega)$ defined in such a way that, if R is the law of the process $(Y_t)_{0\leq t\leq 1}$, then R^k is the law of $(Y_{t/k})_{0\leq t\leq 1}$.

As a consequence of the large deviation theory, the mathematical translation of the cold gas experiment is in terms of some entropy minimization problems. We are going to investigate the limit as k tends to infinity of the following sequence of dynamical *Schrödinger problems*:

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

where the relative entropy of a probability measure p with respect to a reference measure r is defined by

$$H(p|r):=\int\log(dp/dr)\,dp.$$

Schrödinger problem and its relations with the cold gas experiment are investigated in the author's survey paper Léonard (2014a). Relative entropy is detailed at the Appendix B.

The Monge–Kantorovich optimal transport problem is defined for any $\mu_0,\mu_1\in\mathrm{P}(\mathcal{X})$ by

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

where π_0 and $\pi_1\in\mathrm{P}(\mathcal{X})$ are the first and second marginals of π. Its dynamical version is

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

where the length ℓ is defined at (0.2). Unlike the affine minimization problems (MK) and ($\mathrm{MK}_{\mathrm{dyn}}$), for each k the minimization problem ($\mathrm{S}^k_{\mathrm{dyn}}$) is *strictly* convex.

Hence, it admits a *unique* solution $\widehat{P}^k \in \mathrm{P}(\Omega)$ while ($\mathrm{MK}_{\mathrm{dyn}}$) admits an infinite convex set of solutions. We shall prove by means of Γ-convergence that $(\widehat{P}^k)_{k\geq 1}$ is a convergent sequence whose limit $\widehat{P} \in \mathrm{P}(\Omega)$ solves ($\mathrm{MK}_{\mathrm{dyn}}$). To give a more precise statement at Result 0.1, let us introduce the auxiliary entropy minimization problem

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

where the auxiliary reference random walk G is defined by

$$G := \mathbf{1}_\Gamma \exp\left(\int_{[0,1]} J_{t,X_t}(\mathcal{X})\,dt\right) R \in \mathrm{M}_+(\Omega) \tag{0.6}$$

with:

- $\Gamma := \bigcup_{x,y\in\mathcal{X}} \Gamma^{xy}$ the set of all geodesics;
- $J_{t,x}(\mathcal{X}) = \sum_{y:\, y\sim x} J_{t,x}(y)$ where $J_{t,x}(y)$ is the instantaneous rate of jump of the random walk R at time t from x to y; see the Appendix A;
- $\mathcal{S}_{\mathrm{MK}}(\mu_0, \mu_1) \subset \mathrm{P}(\mathcal{X}^2)$ the set of all solutions of (MK).

Since (MK) is an affine minimization problem, its convex set $\mathcal{S}_{\mathrm{MK}}(\mu_0, \mu_1)$ of solutions may be infinite.

We are now ready to give the following partial statement of Theorem 2.1.

RESULT 0.1. The limit $\lim_{k\to\infty} \widehat{P}^k =: \widehat{P} \in \mathrm{P}(\Omega)$ exists. It solves ($\mathrm{MK}_{\mathrm{dyn}}$) and is the unique solution of ($\tilde{\mathrm{S}}_{\mathrm{dyn}}$). Moreover, $\widehat{P}(\Gamma) = 1$, meaning that the sample paths of $\widehat{P}$ are piecewise constant geodesics.

Last statement follows from the absolute continuity of $\widehat{P}$ with respect to G and $G(\Omega \setminus \Gamma) = 0$.

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. It is selected among the infinitely many [see (1.5) below] solutions of ($\mathrm{MK}_{\mathrm{dyn}}$).

DEFINITIONS. We call $\widehat{P}$ a *displacement random walk* and its time-marginal flow $[\mu_0, \mu_1] = (\mu_t)_{0\leq t\leq 1} := (\widehat{P}_t)_{0\leq t\leq 1} \in \mathrm{P}(\mathcal{X})^{[0,1]}$ defines a *displacement interpolation* between μ_0 and μ_1.

This last definition is justified because of several analogies with the standard displacement interpolations on a geodesic space; see Section 1. The defining identity $\mu_t := \widehat{P}_t$, $t \in [0, 1]$ means that $\widehat{P}$ is a coupling of the interpolation $[\mu_0, \mu_1]$. Be aware that, unlike the interpolation $[\mu_0, \mu_1]$, the random walk $\widehat{P}$ encodes the information of all the marginal laws of $(X_{t_1}, \dots, X_{t_k})$.

Clearly, these notions depend on both choices of d and R via Γ and J entering the expression (0.6) of G and also via ($\mathrm{MK}_{\mathrm{dyn}}$). This will be made explicit at Definitions 1.1.

Pushing forward ($\mathrm{MK_{dyn}}$) from $\mathrm{P}(\Omega)$ onto $\mathrm{P}(\mathcal{X}^2)$ via the $(0,1)$-marginal projection (X_0, X_1) gives (MK). Similarly, the push-forward of ($\tilde{\mathrm{S}}_{\mathrm{dyn}}$) is the auxiliary entropic minimization problem

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

We shall obtain at Theorem 2.1 the following natural corollary of Result 0.1.

RESULT 0.2. The law $\widehat{P}_{01} \in \mathrm{P}(\mathcal{X}^2)$ of the couple of endpoint positions (X_0, X_1) under $\widehat{P}$ is a singled out solution of the Monge–Kantorovich problem (MK). It is the unique solution of ($\tilde{\mathrm{S}}$).

We have selected one solution of (MK).

Optimal transport appears at the limit of the slowing down procedure. Again, pushing forward ($\mathrm{S}^k_{\mathrm{dyn}}$) from $\mathrm{P}(\Omega)$ onto $\mathrm{P}(\mathcal{X}^2)$ gives

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

where $R^k_{01} \in \mathrm{P}(\mathcal{X}^2)$ is the joint law of the initial and final positions of R^k.

RESULT 0.3. 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 $\widehat{P}^k$, that is, $\widehat{\pi}^k := \widehat{P}^k_{01}$. Therefore, the slowing down procedure $\lim_{k\to\infty} \widehat{P}^k = \widehat{P}$ selects:

1. one solution $\widehat{\pi} := \lim_{k\to\infty} \widehat{\pi}^k = \widehat{P}_{01} \in \mathrm{P}(\mathcal{X}^2)$ 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 $\lim_{k\to\infty} \widehat{\pi}^k \in \mathrm{P}(\mathcal{X}^2) = \widehat{\pi}$ to the singled out solution $\widehat{\pi}$ of (MK) is a by-product of its dynamical analogue.

A Benamou–Brenier formula. Although G is not Markov, it will be proved at Theorem 2.4 that the displacement random walk $\widehat{P}$ is Markov. It follows from the definition $\mu_t := \widehat{P}_t, 0 \le t \le 1$, ($\mathrm{MK_{dyn}}$) and the Markov property of $\widehat{P}$ that the Wasserstein distance $W_1(\mu_0, \mu_1) := \inf(\mathrm{MK})$ admits a Benamou–Brenier representation. This result is stated at Theorem 3.1. It is in complete analogy with the standard Benamou–Brenier formula (3.1).

Constant speed geodesics. This notion is basic for the Lott–Sturm–Villani theory where the displacement interpolations are constant speed W_2-geodesics [W_2 is defined at (3.1) and constant speed geodesics are defined at page 1880]. As a corollary of the Benamou–Brenier formula, we obtain at Theorem 3.2 the following analogous result.

RESULT 0.4 (d is the standard graph distance). Any displacement interpolation is a constant speed W_1-geodesic.

A discussion about the links between the notions of constant speed geodesics, minimizing geodesics and displacement interpolations is provided at Section 3.

Dynamics of the displacement random walk. The Markov random walk $\widehat{P}$ disintegrates as $\widehat{P}(\cdot) = \int_{\mathcal{X}^2} G^{xy}(\cdot)\widehat{\pi}(dx\,dy)$ where G^{xy} is the xy-bridge of G and $\widehat{\pi}$ is the unique solution of ($\tilde{\mathrm{S}}$); see (1.8), (0.4), (0.6) and Result 0.2. Its dynamics is specified at Theorem 2.4. Let us describe G^{xy} in a special simple case.

Binomial interpolations. Let $\mathcal{X} = \mathbb{Z}$ be equipped with its natural graph structure and let the reference random walk R be a simple walk with jump kernel $J_z = (\delta_{z-1} + \delta_{z+1})/2$, $z \in \mathbb{Z}$ and some unspecified initial marginal with a full support. The content of Proposition 3.4 is the following. Let $(N_t)_{0\le t\le 1}$ be a standard Poisson process, then:

(i) For any $x < y \in \mathbb{Z}$, $G^{xy}(\cdot) = \mathrm{Proba}(x + N \in \cdot | N_1 = d(x, y))$.
(ii) For any $x > y \in \mathbb{Z}$, $G^{xy}(\cdot) = \mathrm{Proba}(y - N \in \cdot | N_1 = d(x, y))$.

Since for each $t \in [0, 1]$, the law of N_t conditionally on $N_1 = d(x, y)$ is the binomial distribution $\mathcal{B}(d(x, y), t)$, the displacement interpolation $[\delta_x, \delta_y] = (G_t^{xy})_{0\le t\le 1}$ is sometimes called a *binomial interpolation*; see (3.9).

What happens when d is not the standard distance. The case of a general distance d is treated in this article. The main difference with the standard distance is that the slowing down procedure takes d into account explicitly as follows. Let J be the jump kernel of R. Instead of taking $J^k = J/k$ as above, we choose

$$J_{t,x}^k := \sum_{y:\, y\sim x} k^{-d(x,y)} J_{t,x}(y)\delta_y, \qquad t \in [0, 1], x \in \mathcal{X}. \tag{0.11}$$

Of course, when d is the standard distance one recovers $J^k = J/k$. The larger is the distance $d(x, y)$ between two neighbours x and y, the lower is the frequency of jumps from x to y as k is large. Therefore, having the cold gas experiment in mind, one sees that the larger $d(x, y)$ is, the more it costs to a lazy particle to jump from x to y. The no-motion limit $k \to \infty$ simulates the metric structure of the graph.

Alternate notions of interpolations. Several approaches for deriving displacement convexity in a discrete setting have lead to alternate notions of interpolations.

In the special case of the hypercube $\mathcal{X} = \{0, 1\}^n$ equipped with the Hamming distance, Ollivier and Villani (2012) introduced the most natural interpolation $(\mu_t)_{0\le t\le 1}$ between δ_x and δ_y which is called *midpoint interpolation* and is defined as follows. For any $0 \le t \le 1$, μ_t is the uniform probability measure on the set of all t-midpoints of x and y. Bonciocat and Sturm (2009) introduced *h-approximate midpoint interpolations*.

Maas (2011) and Mielke (2011) designed a new distance $\mathcal{W}$ on $\mathrm{P}(\mathcal{X})$ which, unlike the Wasserstein distance W_2, allows for regarding evolution equations of reversible Markov chains on the discrete space $\mathcal{X}$ as *gradient flows* of some entropy on $(\mathrm{P}(\mathcal{X}), \mathcal{W})$.

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 $x = x_0 \sim x_1 \sim \cdots \sim x_{d(x,y)} = y$ connecting x and y that have been constructed above. It has been successfully used to prove displacement convexity of the entropy by Johnson (2007), Gozlan et al. (2014) and Hillion (2012, 2014a, 2014b, 2014c). We have seen that binomial interpolations are specific instances of the displacement interpolations which are built in the present paper.

Recently, the author Léonard (2012c) studied convexity properties of the relative entropy along an *entropic* interpolation, that is, the time-marginal flow of the minimizer of $(\mathrm{S}_{\mathrm{dyn}}^{k=1})$ without slowing down, that is, with $k = 1$.

Some remarks and open questions about the links between displacement interpolations on graphs and previous works. It is natural to seek for relations between these different notions of interpolations. However, not very much is known by now.

We have already noticed that binomial interpolations are special instances of displacement interpolations. It seems also that Hillion's (2014a) $W_{1,+}$-interpolations are very close to displacement interpolations. Are they the same?

The distance $\mathcal{W}$ on $\mathrm{P}(\mathcal{X})$ which is introduced in Maas (2011) and Mielke (2011) is a successful object to derive lower curvature results about the dynamics of the random walk with consequences in terms of the rate of convergence to equilibrium; see Erbar and Maas (2012). No $\mathcal{W}$-interpolation is built except heat flows which interpolate between μ_0 at time $t = 0$ and the equilibrium measure at time $t = \infty$. They are interpreted as gradient flows on $(\mathrm{P}(\mathcal{X}), \mathcal{W})$ of the relative entropy with respect to the equilibrium measure. It not clear that $\mathcal{W}$ is related to some transport cost built upon a distance on the graph. Nevertheless, $\mathcal{W}$ is related to some carré du champ operator; it is a Riemannian object. This in contrast with the displacement interpolations of the present article which are related to a Wasserstein distance of order 1.

The connection between Ollivier's (2009) coarse curvature and displacement interpolations on graphs remains to be explored. Although Ollivier's approach to graph curvature does not necessitate interpolations (discrete-time Markov chains are enough), it is defined in terms of optimal transport of order one, a common feature with displacement interpolations. The power and beauty of Ollivier's theory are both its simplicity and wide applicability. One can expect that a more sophisticated approach based on the displacement convexity of the relative entropy should bring sharper results than the consequences of Ollivier's coarse curvature regarding concentration of measure and rate of convergence to equilibrium.

Outline of the paper. Section 1 briefly presents 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 2.1 gives the Γ-convergence results: in particular $\lim_{k\to\infty} \widehat{P}^k = \widehat{P}$, and Theorems 2.3 and 2.4 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 3.1 and a discussion about constant speed interpolations and natural substitutes for the geodesics on a graph. Theorem 3.2 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 Appendices 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 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 1.1.

McCann displacement interpolations. On a Riemannian manifold $\mathcal{X}$ equipped with its Riemannian distance d, any displacement interpolation 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 $\mathrm{P}(\Omega_{\mathrm{ac},2})$ be the corresponding space of probability measures. Let (MK_2) be the quadratic Monge–Kantorovich problem where d in (MK) is replaced by its square d^2. It appears that $[\mu_0, \mu_1]$ is the time-marginal flow

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

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

$$\int_{\Omega_{\mathrm{ac},2}} C_{\mathrm{kin}}(\omega) P(d\omega) \to \min; \qquad P \in \mathrm{P}(\Omega_{\mathrm{ac},2}) : P_0 = \mu_0, P_1 = \mu_1, \tag{1.2}$$

where the kinetic action C_{kin} is defined by $C_{\mathrm{kin}}(\omega) := \int_{[0,1]} \frac{1}{2}|\dot\omega_t|^2_{\omega_t}\, dt \in [0,\infty]$, $\omega \in \Omega_{\mathrm{ac},2}$. Suppose for simplicity that any solution $\pi^* \in \mathrm{P}(\mathcal{X}^2)$ 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 (1.2) is in one–one correspondence with a solution $\widehat{\pi} \in \mathrm{P}(\mathcal{X}^2)$ of (MK_2) via the relation

$$\widehat{P}(\cdot) = \int_{\mathcal{X}^2} \delta_{\gamma^{xy}}(\cdot)\widehat{\pi}(dx\, dy) \in \mathrm{P}(\Omega_{\mathrm{ac},2}), \tag{1.3}$$

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

$$\mu_t = \int_{\mathcal{X}^2} \delta_{\gamma_t^{xy}}(\cdot)\widehat{\pi}(dx\, dy) \in \mathrm{P}(\mathcal{X}), \qquad t \in [0,1]. \tag{1.4}$$

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 $\mathrm{P}(\mathcal{X})$. This lifting from $\mathcal{X}$ to $\mathrm{P}(\mathcal{X})$ was used successfully in the Lott–Sturm–Villani theory on geodesic spaces.

Displacement interpolations on a discrete metric graph $(\mathcal{X}, \sim, d)$. As before, we are going to define the displacement interpolation $[\mu_0, \mu_1] = (\mu_t)_{t\in[0,1]}$ by formula (1.1): $\mu_t := \widehat{P}_t$, $0 \le t \le 1$, where $\widehat{P} \in \mathrm{P}(\Omega)$ is some *singled out solution* of $(\mathrm{MK}_{\mathrm{dyn}})$ which is an order-one analogue of (1.2); see Result 0.1.

Recall that for each distinct $x, y \in \mathcal{X}$, the set of geodesics Γ^{xy} contains infinitely many paths; see (0.3). On the other hand, it is easily seen that for any measurable kernel $(Q^{xy} \in \mathrm{P}(\Omega); x, y \in \mathcal{X})$ which is geodesic in the sense that: $Q^{xy}(\Gamma^{xy}) = 1$ for all x, y (such a kernel is called a *geodesic kernel*) and for any π^* solution of (MK),

$$P^*(\cdot) := \int_{\mathcal{X}^2} Q^{xy}(\cdot)\pi^*(dx\, dy) \in \mathrm{P}(\Omega) \tag{1.5}$$

solves $(\mathrm{MK}_{\mathrm{dyn}})$. It follows that $(\mathrm{MK}_{\mathrm{dyn}})$ admits infinitely many solutions and also that the static and dynamical Monge–Kantorovich problems have the same optimal value:

$$\inf(\mathrm{MK}) = \inf(\mathrm{MK}_{\mathrm{dyn}}). \tag{1.6}$$

The $(X_t)_{t\in[0,1]}$-push forward of the minimizer P^* given at (1.5) is

$$P_t^*(\cdot) = \int_{\mathcal{X}^2} Q_t^{xy}(\cdot)\pi^*(dx\, dy) \in \mathrm{P}(\mathcal{X}), \qquad 0 \le t \le 1. \tag{1.7}$$

Remark that (1.5)–(1.7) has the same structure as (1.3)–(1.4).

The slowing down procedure selects one singled out geodesic kernel ($G^{xy} \in \mathrm{P}(\Omega); x, y \in \mathcal{X}$) which does not depend on the specific choice of μ_0 and μ_1 (see Theorem 2.1 below) and one singled out solution $\widehat{\pi} \in \mathrm{P}(\mathcal{X}^2)$ of (MK) such that

$$\widehat{P}(\cdot) = \int_{\mathcal{X}^2} G^{xy}(\cdot)\widehat{\pi}(dx\,dy) \in \mathrm{P}(\Omega) \tag{1.8}$$

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

$$\mu_t(\cdot) = \int_{\mathcal{X}^2} G_t^{xy}(\cdot)\widehat{\pi}(dx\,dy) \in \mathrm{P}(\mathcal{X}), \qquad t \in [0, 1], \tag{1.9}$$

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

We observe several analogies between the continuous and discrete settings.

- There is an analogy between (1.3)–(1.4) and (1.8)–(1.9):
  - The optimal plan $\widehat{\pi}$ in (1.9) refers to (MK), while in (1.4) it refers to (MK_2).
  - The geodesic Markov random walk G^{xy} in (1.9) corresponds to $\delta_{\gamma^{xy}}$ in (1.4). 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 ($G^{xy} \in \mathrm{P}(\Omega); x, y \in \mathcal{X}$) *encodes some geodesic dynamics of the discrete metric graph* ($\mathcal{X}, \sim, d$).
- While previous item is related to the notion of minimizing geodesic, there is also an analogy in terms of constant speed geodesics. McCann displacement interpolations are constant speed W_2-geodesics, while Result 0.4 states that displacement interpolations on a discrete graph equipped with the standard distance are constant speed W_1-geodesics. See also Proposition 3.3 for a similar result when the distance is general.
- Similarly, as in the present article, it is shown in Léonard (2012a) by means of Γ-convergence technics that McCann displacement interpolations are limits of solutions to entropic problems analogous to ($\mathrm{S}^k_{\mathrm{dyn}}$) where R^k is the law of a slowed down diffusion process and the normalization is $1/k$ instead of $1/\log k$.

These strong analogies entitle us to propose the following.

DEFINITIONS 1.1. Let $R \in \mathrm{M}_+(\Omega)$ be a random walk on the discrete metric graph $(\mathcal{X}, \sim, d)$. We call $\mu = (\widehat{P}_t)_{0\le t\le 1}$ the (R, d)-displacement interpolation and $\widehat{P}$ the (R, d)-displacement random walk between μ_0 and μ_1 specified by (1.8) and (1.9).

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.

2. Main results about displacement random walks. We gather our assumptions before stating rigorously 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 $x_1, x_2, \dots, x_n$ 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

$$n_x := \#\{y \in \mathcal{X}; y \sim x\} < \infty \qquad \forall x \in \mathcal{X}. \tag{2.1}$$

(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: it satisfies (0.1).

(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, that is,

$$\sup_{t \in [0,1], x \in \mathcal{X}} J_{t,x}(\mathcal{X}) < \infty. \tag{2.2}$$

(R^k) For each $k \geq 1$, the slowed down random walk $R^k \in \mathrm{M}_+(\Omega)$ is the Markov measure with the forward jump kernel (0.11) and the initial measure is $R_0^k =: m \in \mathrm{M}_+(\mathcal{X})$ with $m_x > 0$ for all $x \in \mathcal{X}$.

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

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

PROPOSITION 2.1. *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) $\int_{\mathcal{X}^2} e^{-A(x) - A(y)} R_{01}(dx\, dy) < \infty$,
(ii) $\int_E R_{01}(dx\, dy) \geq \int_E e^{-A(x) - A(y)} R_0(dx) R_1(dy)$, *for any* $E \subset \mathcal{X}^2$,
(iii) $E_{R^{xy}}(\ell) \leq A(x) + A(y)$, *for all* $x, y \in \mathcal{X}$,
(iv) $\int_{\mathcal{X}} A\, d\mu_0, \int_{\mathcal{X}} A\, d\mu_1 < \infty$.

As a corollary of the proposition, Hypothesis 2.1(μ) holds when $\mathcal{X}$ is finite. The proof of Proposition 2.1 is done at Section 4.

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 2.1. *Hypotheses 2.1 are assumed to hold.*

1. *For all $k \geq 2$, the problems* (S^k) *and* ($\mathrm{S}^k_{\mathrm{dyn}}$) *admit, respectively, a unique solution* $\widehat{\pi}^k \in \mathrm{P}(\mathcal{X}^2)$ *and* $\widehat{P}^k \in \mathrm{P}(\Omega)$.
Moreover, $\widehat{P}^k$ *is Markov and* $\widehat{\pi}^k = \widehat{P}^k_{01}$.
2. ($\tilde{\mathrm{S}}$) *has a unique solution* $\widehat{\pi} \in \mathrm{P}(\mathcal{X}^2)$ *and* $\lim_{k\to\infty} \widehat{\pi}^k = \widehat{\pi}$.
By definition of ($\tilde{\mathrm{S}}$), $\widehat{\pi}$ *also solves* (MK).
3. ($\tilde{\mathrm{S}}_{\mathrm{dyn}}$) *has a unique solution* $\widehat{P} \in \mathrm{P}(\Omega)$ *and* $\lim_{k\to\infty} \widehat{P}^k = \widehat{P}$.
The random walk $\widehat{P}$ *also solves* ($\mathrm{MK}_{\mathrm{dyn}}$).
4. *The random walk* $\widehat{P}$ *disintegrates as* (1.8). *This means that it satisfies* $\widehat{P}_{01} = \widehat{\pi}$ *and that* $\widehat{P}$ *shares its bridges with the geodesic path measure* G *defined by* (0.6): $\widehat{P}^{xy} = G^{xy}$ *for* $\widehat{\pi}$*-almost every* (x, y).

The proof of Theorem 2.1(1) is done at the end of Section 4 and the proof of Theorem 2.1(2–3–4) is done at the end of Section 5.

As a corollary, we obtain the following result.

THEOREM 2.2. *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 2.1. □

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 configuration (a) in Figure 1, configuration (b) is not a circuit. Hence, it may enter a directed tree.

(a) (b)

FIG. 1. *Circuit and loop.*

We have in mind the directed tree $(\Gamma^{xy}([0,1]), \to)$. Its set of vertices is $\Gamma^{xy}([0,1]) := \{\gamma_t \in \mathcal{X}; 0 \le t \le 1, \gamma \in \Gamma^{xy}\}$ and $z \to z' \in \Gamma^{xy}([0,1])$ if $z \sim z' \in \mathcal{X}$ and there are some $\gamma \in \Gamma^{xy}$ and $0 \le t < t' \le 1$ 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, regardless of the instants of jump.

THEOREM 2.3 (The dynamics of G^{xy}). *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*

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

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*

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

with

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

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 $J^{G,y}_{t,z} = 0$ for any $z \notin \Gamma^{xy}([0,1])$. The proof of Theorem 2.3 is given at Section 6.

THEOREM 2.4 (The dynamics of $\widehat{P}$). *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*

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

It is a mixture of the jump kernels $J^{G,y}$ *of* G^{xy}, *see Theorem* 2.3.

The statement of Theorem 2.4(1) is the content of Proposition 6.2 which is proved at Section 6. The second statement is proved at the end of Section 6.

Gathering Theorems 2.3 and 2.4, one obtains for all t and z such that $\widehat{P}_t(z) > 0$,

$$\begin{aligned} &\widehat{J}_{t,z}(\cdot) \\ &= \int_{\mathcal{X}} \left(\sum_{w \in \{z \to \cdot\}^y} \frac{E_R[\exp(\int_t^1 J_{s,X_s}(\mathcal{X})\,ds)\mathbf{1}_{\Gamma(t,w;1,y)} | X_t = w]}{E_R[\exp(\int_t^1 J_{s,X_s}(\mathcal{X})\,ds)\mathbf{1}_{\Gamma(t,z;1,y)} | X_t = z]} J_{t,z}(w)\delta_w(\cdot)\right) \\ &\qquad \times \widehat{P}(X_1 \in dy | X_t = z). \end{aligned} \tag{2.3}$$

3. Main results about displacement interpolations. A direct consequence of Theorem 2.4 is the following.

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

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

Benamou–Brenier formula. The identity $W_1(\mu_0, \mu_1) := \inf(\text{MK})$ defines the Wasserstein distance on $\mathrm{P}_1(\mathcal{X}) := \{\mu \in \mathrm{P}(\mathcal{X}) : \int_{\mathcal{X}} d(x_o, y)\mu(dy) < \infty\}$ (for some $x_o \in \mathcal{X}$). More generally, on any Polish space $(\mathcal{X}, d)$, replacing the cost function d in (MK) by d^p with $1 \le p < \infty$ gives rise to the Monge–Kantorovich problem (MK_p) of order p and the corresponding *Wasserstein distance* defined by $W_p(\mu_0, \mu_1) := \inf(\text{MK}_p)^{1/p}$ for any μ_0, μ_1 in $\mathrm{P}_p(\mathcal{X}) := \{\mu \in \mathrm{P}(\mathcal{X}) : \int_{\mathcal{X}} d^p(x_o, y)\mu(dy) < \infty\}$.

In the usual Riemannian setting, it follows from (1.1) and (1.2) that $W_2(\mu_0, \mu_1)$ admits the Benamou and Brenier (2000) representation:

$$W_2^2(\mu_0, \mu_1) = \inf_{(\nu, v)} \left\{ \int_{[0,1]\times\mathcal{X}} |v_t(x)|_x^2 \nu_t(dx)\, dt \right\}, \tag{3.1}$$

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

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

We are on the way to state a similar representation in the discrete graph setting. The following theorem is a consequence of the Markov property of $\widehat{P}$ which was stated at Theorem 2.4.

THEOREM 3.1 (Benamou–Brenier formula). *Suppose that the Hypotheses 2.1 are satisfied.*

(1) *We have*

$$W_1(\mu_0, \mu_1) = \inf_{\nu, j} \int_{[0,1]} dt \int_{\mathcal{X}^2} d(z, w)\nu j_t(dz\, dw) < \infty, \tag{3.2}$$

where the infimum is taken over all couples (ν, j) such that:

(i) $\nu = (\nu_t)_{t \in [0,1]} \in \mathrm{P}(\mathcal{X})^{[0,1]}$ *is a time-differentiable flow of probability measures on $\mathcal{X}$*,

(ii) $j = (j_{t,z})_{t \in [0,1], z \in \mathcal{X}} \in \mathrm{M}_+(\mathcal{X})^{[0,1] \times \mathcal{X}}$ *is a measurable jump kernel*,

(iii) *ν and j are linked by the current equation*

$$\begin{cases} \partial_t \nu_t(z) + \int_{\mathcal{X}} [\nu_t(z) j_{t,z}(dw) - \nu_t(dw) j_{t,w}(z)] = 0, \\ \qquad\qquad 0 < t < 1, z \in \mathcal{X}, \\ \nu_0 = \mu_0, \qquad \nu_1 = \mu_1 \end{cases} \tag{3.3}$$

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

(2) *The infimum* $\inf_{\nu,j}$ *in* (3.2) *is attained at* $(\mu, \widehat{J})$ *where* μ *is the displacement interpolation and* $\widehat{J}$ *is the jump kernel of* $\widehat{P}$. *Hence*,

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

PROOF. With (1.6) and Theorem 2.1(3), we have

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

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

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

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 (3.3). $\square$

Constant speed geodesics, *minimizing geodesics and displacement interpolations*. Let us start recalling some basic definitions.

Constant speed geodesics. Recall that a constant speed geodesic between a_0 and a_1 in the metric space (A, d) is a path $(a_t)_{0 \le t \le 1}$ in A such that for all $0 \le s \le t \le 1$, $\mathsf{d}(a_s, a_t) = (t - s)\mathsf{d}(a_0, a_1)$. The metric space (A, d) is said to be geodesic if there exists at least a constant speed geodesic joining any pair of points a and b in A.

In the special case where $\mathcal{X}$ is a Riemannian manifold with its Riemannian distance d, $(\mathrm{P}_2(\mathcal{X}), W_2)$ is a geodesic space and the McCann displacement interpolations are exactly the constant speed W_2-geodesics; see Villani (2009), for instance. Hence, it is a natural idea to look also for some constant speed W_p-geodesics in a discrete setting. But, it is argued in Maas [(2011), Remark 2.1] that when $\mathcal{X}$ is discrete, for any $p > 1$, the only constant speed W_p-geodesics are the constant paths on $\mathrm{P}_p(\mathcal{X})$. Therefore, only W_1-geodesics have to be retained. And indeed, the displacement interpolations that are built in this article are constant speed W_1-geodesics.

In the present context:

- a constant speed geodesic refers to a constant speed W_1-geodesic in the length space $(\mathrm{P}_1(\mathcal{X}), W_1)$;
- a minimizing geodesic is any solution of the action minimizing problem $(\mathrm{MK}_{\mathrm{dyn}})$;
- displacement interpolations are defined at Definition 1.1.

Only the knowledge of the distance d is required for defining constant speed geodesics, while both the distance d and the reference random walk R are required for defining minimizing geodesics and displacement interpolations. In the standard Riemannian setting, because of the strict convexity of the action, the minimizing geodesic is unique and has a constant speed. In the present setting things differ significantly. Indeed, the action in $(\mathrm{MK}_{\mathrm{dyn}})$ is affine so that there are infinitely many minimizing geodesics. However, we shall see in Proposition 3.3 that thanks to the positive 1-homogeneity of the action, any minimizing geodesic can be transformed via a change of time into a constant speed minimizing geodesic.

Let us explore a little further some connections between the notions of minimizing/constant speed geodesics and displacement interpolations.

The next proposition is another consequence of the Markov property of $\widehat{P}$.

PROPOSITION 3.1. *For all $0 \le s \le t \le 1$, 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. *$W_1(\mu_s, \mu_t) = \int_{[s,t]} dr \int_{\mathcal{X}^2} d(z, w)\mu\widehat{J}_r(dz\,dw)$.*

PROOF. Both statements are consequences of:

- the Markov property of $\widehat{P}$, see Theorem 2.4, 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 $P_{[u,v]} := (X_t; u \le t \le v)_\# P$;
- the fact that $\ell_{st} := \sum_{s<r<t} d(X_{r^-}, X_r)$ is insensitive to changes of time: that is, 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

$$\mathcal{J}_r^\theta = \dot{\theta}(r)\mathcal{J}_{\theta(r)} \qquad \text{for almost every } r \in (s,t), \tag{3.4}$$

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

Proposition 3.1(2) entitles us to define the speed of the displacement interpolation μ at time t by

$$\text{speed}(\mu)_t := \int_{\mathcal{X}^2} d(z,w)\mu\widehat{J}_t(dz\,dw), \qquad 0 \le t \le 1, \tag{3.5}$$

to obtain

$$W_1(\mu_0,\mu_1) = \int_{[0,1]} \text{speed}(\mu)_t\,dt. \tag{3.6}$$

DEFINITION 3.1 (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 $Q \in \mathrm{M}_+(\Omega)$, 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 $\nu\mathcal{J}_t(dx\,dy) := \nu_t(dx)\mathcal{J}_{t,x}(dy)$ and $\nu\mathcal{J}_t(\mathcal{X}^2) := \int_{\mathcal{X}^2} \nu_t(dx)\mathcal{J}_{t,x}(dy)$.

For any change of time $\tau : [0,1] \to [0,1]$, we see with (3.4), (3.6) 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 (3.2). This implies that, for fixed d and R, there are infinitely many minimizing geodesics.

PROPOSITION 3.2. *For any change of time τ, the infimum $\inf_{\nu,j}$ in (3.2) is also attained at $(\mu^\tau, \widehat{J}^\tau)$ where μ^τ is the (R^τ, d)-displacement interpolation and $\widehat{J}^\tau$ is the jump kernel of $\widehat{P}^\tau$. Moreover, $\widehat{P}^\tau$ is the displacement random walk associated with R^τ, that is, the analogue of $\widehat{P}$ when R is replaced with R^τ and*

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

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. We denote P^* the (R^τ, d)-displacement random walk. 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 $k \ge 1$. This implies that $P^* = \widehat{P}^\tau$. Hence, the result follows from Theorem 3.1(2). $\square$

The next result states that among all the displacement interpolations μ^τ as τ varies, only one has a constant speed.

PROPOSITION 3.3. *Under the Hypotheses 2.1, there exists a unique change of time τ_o such that the (R^{τ_o}, d)-displacement interpolation μ^{τ_o} has a constant speed, that is,*

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

PROOF. Indeed, this equation is equivalent to

$$\dot{\tau}(s)\psi\big(\tau(s)\big) = W_1(\mu_0, \mu_1) \qquad \text{a.e.,} \tag{3.7}$$

where $\psi(t) := \int_{\mathcal{X}^2} d(z,w)\mu\widehat{J}_t(dz\,dw)$, a.e. Clearly, the assumption that d is uniformly lower bounded and (2.3) imply that $\psi > 0$. Hence, a solution of (3.7) is given by

$$\tau_o(s) = \Psi^{-1}_{\mu_0,\mu_1}\big(W_1(\mu_0,\mu_1)s\big), \qquad s \in [0,1], \tag{3.8}$$

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

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

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 $\psi_o(u) := \int_{\mathcal{X}^2} d(z,w)\mu\widehat{J_u^{\tau_o}}(dz\,dw)$, 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 3.2 (Constant speed displacement interpolation). The time changed displacement interpolation μ^{τ_o} with τ_o given at (3.8) is called the constant speed (R,d)-displacement interpolation.

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 3.2 that in the special important case where the distance d is the standard graph distance specified by (0.5), for any μ_0, μ_1, without changing time the displacement interpolation $[\mu_0, \mu_1]$ has a constant speed.

Conservation of the average rate of mass displacement. The next result tells us that along any displacement interpolation $[\mu_0, \mu_1]$, the average rate of mass displacement, as defined below, does not depend on time.

DEFINITIONS 3.1 (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

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

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 3.2 (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*

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

In particular, when the distance d is the standard discrete distance specified by (0.5), *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 3.2 is a restatement of Theorem 7.1 which is proved at Section 7.

COROLLARY 3.2. *The constant speed displacement interpolation μ^{τ^o} defined at Definition 3.2 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 3.2. $\square$

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

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

The constant speed $(R^{\tau^{xy}}, d)$-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.

Special interpolations. We present some easy examples of constant speed interpolations μ^{xy} in the simplest and important setting where:

(i) R is the reversible simple random walk;
(ii) d 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 3.2 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 2.3, for any $t \in [0, 1)$, $z \in \Gamma^{xy}([0, 1])$ and $w \in \{z \to \cdot\}^y$, by

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

The complete graph. Let $\mathcal{X} = \{1, \dots, n\}$ with $x \sim y$ for all $x \neq y \in \mathcal{X}$. Then, for all $x \neq y$ 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 Poisson(λ) 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 Poisson$(1-t)$ law and $\delta = d(z, y) = y - z$, we obtain

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

This proves the following.

PROPOSITION 3.4. *For any $x < y$, $\mathrm{Proba}(G^{xy} \in \cdot) = \mathrm{Proba}(x + N \in \cdot | N_1 = d(x, y))$ where $(N_t)_{0\le t\le 1}$ is a standard Poisson process.*

For any $x < y$ and each $0 < t < 1$, the support of G_t^{xy} is $\{x, x+1, \dots, y\}$ and

$$G_t^{xy} = \sum_{z : x\le z\le y} \binom{d(x, y)}{d(x, z)} t^{d(x,z)}(1-t)^{d(z,y)}\delta_z. \tag{3.9}$$

The hypercube. Consider the hypercube $\mathcal{X} = \{0, 1\}^n$ with its natural graph structure so that the graph distance is the Hamming distance defined by $d(x, y) = \sum_{1\le i\le n} \mathbf{1}_{\{x_i \neq y_i\}}$ where $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$. 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 $0 \le t \le 1$, the law G^{xy}_t, 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, $(\mathsf{d}, \dots, \mathsf{d})$ is x since x has $d(x, y)$ components that are different from y, of course $(\mathsf{s}, \dots, \mathsf{s})$ is y and we see that the support S^{xy} of G^{xy}_t 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

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

4. The Schrödinger problem.

The reversing measure $m \in \mathrm{M}_+(\mathcal{X})$ 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 the Appendix B.

Schrödinger problem. We briefly introduce the main features of the Schrödinger problem. For more detail, see, for instance, the survey paper Léonard (2014a). The dynamical Schrödinger problem associated with the random walk $R \in \mathrm{M}_+(\Omega)$ is the following entropic minimization problem

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

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 $\phi = (X_0, X_1)$. We have for all $P \in \mathrm{P}(\Omega)$

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

which implies that $H(P_{01}|R_{01}) \le H(P|R)$ 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 ($\mathrm{S}_{\mathrm{dyn}}$) if and only if it disintegrates as

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

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, that is, (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

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

Proofs of Theorem 2.1(1) *and Proposition* 2.1. 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:

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

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 (0.2).

Proof of Theorem 2.1(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 Léonard (2014a), Proposition 2.10.

It remains to show the existence. For any $P \in \mathrm{P}(\Omega)$ and any $k \geq 1$, with (4.5), we see that

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

Choosing

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

we have

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

With Hypothesis 2.1-(μ), we obtain $\inf(\mathrm{S}^k_{\mathrm{dyn}}) < \infty$ and it follows that $(\mathrm{S}^k_{\mathrm{dyn}})$ and (S^k) admit a solution; see Léonard (2014a), Lemma 2.4.

Proof of Proposition 2.1. Taking $\pi^o = \mu_0 \otimes \mu_1$ in (4.7) gives

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

It is proved at Léonard [(2014a), Proposition 2.5] that 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 $\int_{\mathcal{X}^2} E_{R^{xy}}(\ell)\mu_0(dx) \times \mu_1(dy) < \infty$.

5. Lazy random walks converge to displacement random walks. The aim of this section is to make precise the convergence of $(\mathrm{S}^k_{\mathrm{dyn}})_{k\ge 2}$. It is proved at Theorem 2.1 that the sequence of minimizers of $(\mathrm{S}^k_{\mathrm{dyn}})_{k\ge 2}$ has a limit in $\mathrm{P}(\Omega)$ which is singled out among the infinitely solutions of the dynamic Monge–Kantorovich problem ($\mathrm{MK}_{\mathrm{dyn}}$). As a corollary, we describe at Theorem 2.2 the convergence of the sequence of bridges $(R^{k,xy})_{k\ge 1}$.

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 $\widetilde{\Omega} = \{\omega \in D([0,1],\mathcal{X}); \forall t \in (0,1), \omega_{t^-} \neq \omega_t \Longrightarrow \omega_{t^-} \sim \omega_t\}$ 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 with finitely many jumps. A typical path ω in Ω is either constant or writes as

$$\omega = \sum_{0\le i\le n-1} x_i \mathbf{1}_{[t_i,t_{i+1})} + x_n \mathbf{1}_{[t_n,1]} \tag{5.3}$$

with $n \ge 1$, $0 = t_0 < t_1 < \cdots < t_n < 1$ and $x_0 \sim x_1 \sim \cdots \sim x_n \in \mathcal{X}$.

Our Hypotheses 2.1, and in particular (2.2), imply that the support of each R^k is included in Ω. Since, $\widehat{P}^k$ is absolutely continuous 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 $(X_t)_{0\le t\le 1}$ lives on Ω.

Γ-convergence. The right notion of convergence for the sequences of minimization problems $(\mathrm{S}^k_{\mathrm{dyn}})_{k\ge 2}$ and $(\mathrm{S}^k)_{k\ge 2}$ 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) $\liminf_{k\to\infty} f^k(y_k) \ge f(y)$ for any convergent sequence $y_k \to y$,
(b) $\lim_{k\to\infty} f^k(y^o_k) = f(y)$ for some sequence $y^o_k \to y$.

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

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

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

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

For more detail about Γ-convergence, see Dal Maso (1993), for instance.

The convergences of $(\mathrm{S}^k_{\mathrm{dyn}})_{k\ge 2}$ and $(\mathrm{S}^k)_{k\ge 2}$. 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 $k \ge 2$ [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 dynamical Monge–Kantorovich problem:

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

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 5.3 that $\Gamma\text{-}\lim_{k\to\infty} I^k = I$, meaning that ($\mathrm{MK}_{\mathrm{dyn}}$) is the limit of $(\mathrm{S}^k_{\mathrm{dyn}})_{k\ge 2}$.

LEMMA 5.1. *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 $\ell \ge 0$ is continuous, $P \mapsto E_P(\ell) = \sup_{n\ge 1} E_P(\ell \wedge n)$ 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 $a \ge 0$, $\{P : P_0 = \mu_0, E_P(\ell) \le a\}$ is uniformly tight in $\mathrm{P}(\Omega)$.

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

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

As it is assumed that the graph $(\mathcal{X}, \sim)$ is locally finite [see (2.1)], $\{X_0 \in K_n, N \le n\}$ 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. $\square$

LEMMA 5.2. *For any $P \in \mathrm{P}(\Omega)$, there exists a sequence $(P_n)_{n\ge 1}$ 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 $n \ge 1$.*

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 (2.2) 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 5.2.

PROOF OF LEMMA 5.2. 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: $P = \sum_n P(\ell = c_n)P(\cdot|\ell = c_n)$.

(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 $\mathrm{P}(\Omega)$ 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 $\mathrm{P}(\Omega)$ 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 can choose $Q_n^\omega \in \mathrm{P}(\Omega)$ 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 nonnegative continuous functions φ_i^n have, for each fixed $n \geq 1$, nonoverlapping 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 $(\omega_0, \omega_{t_1}, \ldots, \omega_{t_k})$, does not 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 assumption (2.2), 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 $\mathrm{P}(\Omega)$. This completes the proofs of (c) and the lemma. $\square$

LEMMA 5.3. *The sequence $(I^k)_{k\geq 2}$ is equi-coercive and $\Gamma\text{-}\lim_{k\to\infty} I^k = I$.*

PROOF. Let us denote

$$H^k(P) := H\big(P|R^k\big)/\log k, \qquad P \in \mathrm{P}(\Omega).$$

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

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

Because of assumption (2.2), we have the uniform bounds

$$\begin{aligned} &0 \leq E_P\int_0^1 J_{t,X_t}(\mathcal{X})\,dt, \\ &E_P\int_{[0,1]\times\mathcal{X}} k^{-d(X_t,y)} J_{t,X_t}(dy)\,dt \leq \sup_{t,x} J_{t,x}(\mathcal{X}) < \infty \qquad \forall P \in \mathrm{P}(\Omega), k \geq 2. \end{aligned} \tag{5.4}$$

Hence, we obtain with $I - [(-\inf(\mathrm{S}) \vee 0) + \sup_{t,x} J_{t,x}(\mathcal{X})]/\log 2 \leq I^k$ for all $k \geq 2$, and Lemma 5.1 that $(I^k)_{k\geq 2}$ is equi-coercive.

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

$$E_P(\ell) < \infty \quad \text{and} \quad H(P|R) < \infty. \tag{5.5}$$

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

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

Since $P \mapsto E_P(\ell)$ is lower semi-continuous and $H(\cdot|R) \geq 0$, with (5.4), we obtain for any convergent sequence $P_k \underset{k\to\infty}{\to} P$ that $\liminf_{k\to\infty} H^k(P_k) \geq E_P(\ell)$. Lemma 5.2 tells us that from any recovery sequence $(P_n)_{n\geq 1}$ for the lower semi-continuity of $P \mapsto E_P(\ell)$, that is, such that $\lim_{n\to\infty} E_{P_n}(\ell) = E_P(\ell)$, one can build a recovery sequence for $(H^k)_{k\geq 1}$, that is, $\lim_{k\to\infty} H^k(P^k) = E_P(\ell)$. 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. □

PROPOSITION 5.1. *For any $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$, we have*

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

PROOF. It is a direct corollary of (4.4) and Lemma 5.3. □

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

$$H(P|R_J) \to \min; \qquad P \in \mathcal{M}_{\mathrm{dyn}}(\mu_0, \mu_1), \tag{5.6}$$

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

$$R_J := \exp\biggl(\int_0^1 J_{t,X_t}(\mathcal{X})\, dt\biggr) R \in \mathrm{M}_+(\Omega).$$

Remark that (2.2) ensures the finiteness of the integral in the above exponential.

LEMMA 5.4. (a) *For each* $k \geq 2$, ($\mathrm{S}^k_{\text{dyn}}$) *has a unique solution* $\widehat{P}^k$.

(b) $\mathcal{M}_{\text{dyn}}(\mu_0, \mu_1)$ *is a nonempty convex compact subset of* $\mathrm{P}(\Omega)$.

(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* (5.6).

Under the assumptions of Lemma 5.4, Lemma 5.3 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 5.4 asserts that there is indeed a unique limit point.

PROOF OF LEMMA 5.4. This proof relies on Anzellotti and Baldo's (1993) Γ-asymptotic expansion technic. For a clear exposition of this technique, see Ambrosio and Pratelli (2003), Section 4.

We have seen at (5.5), 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 nonemptiness 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 5.1.

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

$$J^k(P) := \log(k)\bigl(I^k(P) - i\bigr), \qquad P \in \mathrm{P}(\Omega).$$

We have

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

and, using the coerciveness of $H(\cdot|R_J)$ and (5.4), it is easily seen that:

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

As $H(\cdot|R_J)$ is strictly convex, so is J and (5.6) admits a unique minimizer $\widehat{P}$ on the convex set $\mathcal{M}_{\mathrm{dyn}}(\mu_0,\mu_1)=\{P: P_0=\mu_0, P_1=\mu_1, E_P(\ell)=i\}$. One completes the proof of the lemma, noticing that $\operatorname{argmin} J^k=\operatorname{argmin} I^k=\{\widehat{P}^k\}$, for each $k\geq 2$. $\square$

For the definition of G at (0.6) to be mathematically consistent, it is necessary that Γ is a measurable set.

LEMMA 5.5. *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 $\Gamma:=\bigcup_{x,y\in\mathcal{X}}\Gamma^{xy}$ of all geodesics, is also measurable. $\square$

LEMMA 5.6. *The set $\mathcal{M}_{\mathrm{dyn}}(\mu_0,\mu_1)$ consists of all $P\in\mathrm{P}(\Omega)$ concentrated on Γ, that is, $P(\Gamma)=1$, 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: $P(\cdot)=\int_{\mathcal{X}^2}P^{xy}(\cdot)P_{01}(dx\,dy)$. Thus, $E_P(\ell)=\int_{\mathcal{X}^2}E_{P^{xy}}(\ell)P_{01}(dx\,dy)$. As $\ell\geq d(x,y)$ on Ω^{xy} and $\Gamma^{xy}=\{\ell=d(x,y)\}$, we have $E_P(\ell)\geq\int_{\mathcal{X}^2}d(x,y)P_{01}(dx\,dy)$ 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 $E_P(\ell)=\int_{\mathcal{X}^2}d(x,y)P_{01}(dx\,dy)$ and the conclusion about P_{01} follows immediately. $\square$

Proof of Theorem 2.1(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 ($\tilde{\mathrm{S}}_{\mathrm{dyn}}$) and ($\tilde{\mathrm{S}}$).

We start proving the statements about the dynamical problems ($\mathrm{S}^k_{\mathrm{dyn}}$) and ($\mathrm{S}_{\mathrm{dyn}}$). 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 5.4 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

$$P\in\mathcal{M}_{\mathrm{dyn}}(\mu_0,\mu_1)\quad\Longleftrightarrow\quad\begin{cases}P_{01}\in\mathcal{S}_{\mathrm{MK}}(\mu_0,\mu_1),\\ P(\Gamma)=1\end{cases}$$

(see Lemma 5.6), 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 (1.8) with $\widehat{\pi}$ the unique solution of the strictly convex problem ($\tilde{\mathrm{S}}$), 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 ($\mathrm{S}^k_{\mathrm{dyn}}$) and ($\mathrm{S}_{\mathrm{dyn}}$) 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.

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

$$\int_{\mathbb{R}^k} d\big(x, T(x)\big)\mu_0(dx) \to \min; \qquad T : \mathbb{R}^k \to \mathbb{R}^k, T_\# \mu_0 = \mu_1, \tag{5.7}$$

where the transport map T is assumed to be measurable. The Monge–Kantorovich problem (MK) is a convex relaxation of (5.7) in the sense that $\pi \mapsto \int_{\mathcal{X}^2} d(x, y) \times \pi(dx\,dy)$ 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 $\int_{\mathcal{X}^2} d(x, y)\pi^T(dx\,dy) = \int_{\mathcal{X}} d(x, T(x))\mu_0(dx)$.

Many known solutions of (5.7) *rely on approximations and variational methods*. Monge's original problem corresponds to d the standard Euclidean distance. Sudakov (1979) proposed an efficient, but still incomplete, strategy. The first complete solution was obtained by Evans and Gangbo (1999). 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), (5.7) 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 $d^\varepsilon(x, y) := \|y - x\|^{1+\varepsilon}$, with $\varepsilon > 0$ 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 $d^\varepsilon \to d$, Caffarelli, Feldman and McCann (2002) and Ambrosio (2003), Ambrosio and Pratelli (2003), removed Evans and Gangbo's (1999) 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 (2004) 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 (2011). Again, both Ambrosio, Kirchheim and Pratelli (2004) and Champion and De Pascale (2011) rely on variational methods and Γ-convergence.

The main Γ-convergence technic used during the proofs of Ambrosio, Kirchheim and Pratelli (2004), Ambrosio and Pratelli (2003), Champion and De Pascale (2011) is an asymptotic expansion which was introduced by Anzellotti and Baldo (1993); see also Attouch (1996). In the present paper, we also have made a crucial use of this technique at Lemma 5.4. Instead of considering the approximation $d^\varepsilon \to d$, the not convex enough problem (MK) is approximated by the sequence of *strictly* convex entropy minimization problems (S^k).

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 2.3 and 2.4. 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 Léonard, Rœlly and Zambrini (2014) and the references therein.

DEFINITIONS 6.1. (a) A measure $Q \in \mathrm{M}_+(\Omega)$ is said to be *Markov* if for any $0 \le t \le 1$, $Q(X_{[t,1]} \in \cdot | X_{[0,t]}) = Q(X_{[t,1]} \in \cdot | X_t)$.

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

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

LEMMA 6.1. (a) *Any Markov measure is reciprocal* (*but the converse is false*).
(b) *Almost every bridge of a reciprocal measure is Markov.*

PROOF. See Léonard, Rœlly and Zambrini (2014). □

LEMMA 6.2. *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

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

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', we have

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

which is the desired result. □

Basic properties of G. We apply Lemmas 6.1 and 6.2 to G defined at (0.6).

PROPOSITION 6.1. *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 $E_{R_J}(ab|X_t) = E_{R_J}(a|X_t)E_{R_J}(b|X_t)$. 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

$$\begin{aligned} E_{R_J}(ab|X_t) &= \frac{E_R(a\alpha b\beta|X_t)}{E_R(\alpha\beta|X_t)} = \frac{E_R(a\alpha|X_t)}{E_R(\alpha|X_t)}\frac{E_R(b\beta|X_t)}{E_R(\beta|X_t)} \\ &= E_{R_J}(a|X_t)E_{R_J}(b|X_t), \end{aligned}$$

where last equality is obtained by plugging successively $b = 1$ and $a = 1$ in $E_{R_J}(ab|X_t) = \frac{E_R(a\alpha|X_t)}{E_R(\alpha|X_t)}\frac{E_R(b\beta|X_t)}{E_R(\beta|X_t)}$. This shows that R_J is Markov.

We conclude with Lemma 6.2 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 6.1. *For every $(x, y) \in \mathcal{X}^2$, the bridge G^{xy} is Markov.*

PROOF. This follows from Lemma 6.1 and Proposition 6.1, 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 $(\Gamma^{xy}([0,1]), \to)$, between the statements of Theorems 2.1 and 2.3, 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 2.3. The Markov property is already proved at Corollary 6.1.

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

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

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 $g_t^y(z) = E_R(G(t, z; 1, y)|X_t = z)$.

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

$$G(t_1, z_1; t_2, z_2; t_3, z_3) = G(t_1, z_1; t_2, z_2)G(t_2, z_2; t_3, z_3) \qquad \text{on } \Gamma^{xy}. \tag{6.1}$$

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

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

For any $0 \le t < 1$, $z \in \Gamma^{xy}([0, 1])$, with (0.6) we see that

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

where last equality follows from (6.1) and the Markov property of R.

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

$$\begin{aligned}&E_{G^{xy}}(U_{t+h} - U_t|X_t = z)\\&\qquad = \frac{E_R[(U_{t+h} - U_t)G(0, x; t, z)G(t, z; 1, y)|X_t = z]}{E_R[G(0, x; t, z)|X_t = z]g_t^y(z)}\end{aligned}$$

$$\overset{(6.1)}{=} \frac{1}{g_t^y(z)} E_R\big[(U_{t+h} - U_t)\mathbf{1}_{\{z \le X_{t+h} \le y\}} G(t, z; t+h, X_{t+h}) \\ \times G(t+h, X_{t+h}; 1, y) | X_t = z\big]$$

$$= \frac{1}{g_t^y(z)} E_R\big[(U_{t+h} - U_t)\mathbf{1}_{\{z \le X_{t+h} \le y\}} G(t, z; t+h, X_{t+h}) \\ \times g_{t+h}^y(X_{t+h}) | X_t = z\big],$$

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 (2.2). Hence,

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

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.

$\widehat{P}$ is Markov. It follows from (1.8), Proposition 6.1 and Léonard, Rœlly and Zambrini [(2014), Proposition 2.8] that the limiting path measure $\widehat{P}$ is reciprocal. We can do better, but it requires some effort.

PROPOSITION 6.2. *The limiting path measure $\widehat{P}$ is Markov.*

PROOF. By Theorem 2.1(1), for each $k \ge 2$, $\widehat{P}^k$ inherits the Markov property of R. We show below at Lemma 6.6 that, as k tends to infinity, $\widehat{P}^k$ converges in variation norm to $\widehat{P}$ and we conclude with Lemma 6.3 below that $\widehat{P}$ is Markov. □

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

PROOF. See Nagasawa (1993), Lemma 5.3. □

There are counter-examples of sequences $(P_k)_{k\geq 1}$ of Markov measures converging narrowly to a non-Markov P. The following standard lemmas are preliminary results for Lemma's 6.6 proof.

LEMMA 6.4 (Scheffé's theorem). *Let r be a positive measure and $p_k = z_k r$, $k \geq 1$, $p = zr$ be probability measures which are absolutely continuous with respect to r. If* $\lim_{k\to\infty} z_k = z$, r, *then*

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

PROOF. See Billingsley (1968), Appendix. □

LEMMA 6.5 (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*

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

PROOF. See Dembo and Zeitouni (1998), Section 4.3. □

LEMMA 6.6. *We have* $\lim_{k\to\infty} \|\widehat{P}^k - \widehat{P}\|_{TV} = 0$.

PROOF. By Lemma 6.4, it is enough to prove that $\lim_{k\to\infty} d\widehat{P}^k/dR = d\widehat{P}/dR$, R. For any $P \in \mathrm{P}(\Omega)$ 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}. As, $\widehat{P}^{xy} = G^{xy}$ and $\widehat{P}^{k,xy} = R^{k,xy}$, this amounts to prove that

$$\lim_{k\to\infty} \frac{dG^{xy}}{dR^{k,xy}} = 1, \qquad R^{xy} \tag{6.2}$$

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

$$\begin{aligned}
&J_{t,z}^{k,xy}(w) \\
&\quad = \frac{R^k(X_1 = y|X_t = w)}{R^k(X_1 = y|X_t = z)} k^{-d(z,w)} J_{t,z}(w) \\
&\quad = E_R\Bigl[k^{-\{d(z,w)+\ell_{[t,1]}\}} \exp\Bigl(\int_{[t,1]\times\mathcal{X}} (1 - k^{-d(X_s,a)}) J_{s,X_s}(da)\, ds\Bigr) \\
&\qquad\qquad \times \mathbf{1}_{(X_1=y)} \Big| X_t = w\Bigr] J_{t,z}(w) \\
&\qquad \Big/ \Bigl(E_R\Bigl[k^{-\ell_{[t,1]}} \exp\Bigl(\int_{[t,1]\times\mathcal{X}} (1 - k^{-d(X_s,a)}) J_{s,X_s}(da)\, ds\Bigr) \\
&\qquad\qquad \times \mathbf{1}_{(X_1=y)} \Big| X_t = z\Bigr]\Bigr).
\end{aligned}$$

With Lemma 6.5, we obtain that

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

meaning, with Theorem 2.3, 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 (6.2) is true, completing the proof of the lemma. □

REMARK 6.1. Theorem 2.3 and Lemma 6.6 together, give an alternate proof of the convergence $\lim_{k\to\infty} R^{k,xy} = G^{xy}$ stated in Theorem 2.2, 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 Villani (2009), Chapter 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 $\operatorname{supp}\widehat{\pi}$ 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 6.2, we obtain the following result.

PROPOSITION 6.3. *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 6.2, 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: that is, 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 2.4(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 $\widehat{P}_{t1}(dz\,dy) = \mu_t(dz)\widehat{p}(t,z;dy)$.

As in Theorem 2.3, we compute a stochastic derivative. A general disintegration result tells us that

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

since, for μ_0-almost any x_o,

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

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 $0 < t < t+h < 1$,

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

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 $\widehat{P}$.

7. Conservation of the average rate of mass displacement. The main result of this section is Theorem 7.1. It states that the constant average rate of mass displacement (recall Definitions 3.1) is conserved along the displacement interpolation. It is a consequence of the Corollary 7.1 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 ($\mathrm{S}_{\mathrm{dyn}}$).

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 time-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

$$\nu j_t^{\widehat{\tau}}(\mathcal{X}^2) - \nu J_{\widehat{\tau}(t)}(\mathcal{X}^2) = K \qquad \textit{for almost every } t \in [0, 1], \tag{7.1}$$

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 7.1. *Let $\widetilde{P}$ be the solution of ($\mathrm{S}_{\mathrm{dyn}}$) and denote ν its time-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 ($\mathrm{S}_{\mathrm{dyn}}$), we have $H(\widetilde{P}|R) < \infty$ and also $\nu \widetilde{J}_t(\mathcal{X}^2) > 0$ for all $t \in [0, 1]$; see Léonard (2014a), Proposition 4.2. Hence, we are allowed to apply

Proposition 7.1. But as $\widetilde{P}$ solves ($\mathrm{S}_{\mathrm{dyn}}$), the only time change $\widetilde{\tau}$ which minimizes $\tau \mapsto H(\widetilde{P}^{\tau}|R)$ is the identity; recall that ($\mathrm{S}_{\mathrm{dyn}}$) admits a unique minimizer. Therefore, $\widetilde{\tau}(t) = t$ for all $t \in [0, 1]$ and (7.1) 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 (7.8) for taking *absolutely continuous* changes of time into account and in the present case $\tau = \mathrm{Id}$ is differentiable everywhere. $\square$

With Corollary 7.1 at hand, we can prove the main result of the section.

THEOREM 7.1. *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 $d_\sim$: that is, 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 (3.5).

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

$$\lim_{k\to\infty} \int_{[0,t]} \mu^k \widehat{J}^k_s(\mathcal{X}^2)\, ds = \int_{[0,t]} \mu \widehat{J}_s(\mathcal{X}^2)\, ds. \tag{7.2}$$

As $\int_{[0,t]} \mu^k \widehat{J}^k_s(\mathcal{X}^2)\, ds = E_{\widehat{P}^k} N_t$ where $N_t := \sum_{0<s<t} \mathbf{1}_{\{X_{s^-} \neq X_s\}}$ is the number of jumps during $[0, t]$, (7.2) does not directly follow from $\lim_{k\to\infty} \widehat{P}^k = \widehat{P}$ (see Theorem 2.1), because N_t is unbounded. We must strengthen Theorem 2.1 to authorize the test functions $0 \le N_t \le N_1 := N$, $0 \le t \le 1$. To do so, it is enough to show that Lemma 5.3 can be improved as follows: the sequence $(I^k)_{k\ge 2}$ is equi-coercive and Γ-$\lim_{k\to\infty} I^k = I$ in $\mathrm{P}_N(\Omega) := \{P \in \mathrm{P}(\Omega); E_P N < \infty\}$ with respect to the topology weakened by the functions $C_b(\Omega) \cup \{N_t; 0 \le t \le 1\}$. For this strengthening to hold, it is sufficient that $(I^k)_{k\ge 2}$ is equi-coercive in $\mathrm{P}_N(\Omega)$. Inspecting the proof of Lemma 5.3, we see that $I^k \ge I + c$ 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 5.1, noticing that $0 \le N_t \le N \le \ell$. This completes the proof of (7.2).

We are going to apply Corollary 7.1 to the solution $\widehat{P}^k$ of ($\mathrm{S}^k_{\mathrm{dyn}}$), for any large enough $k \ge 1$. Hence, we have to check that $\int_{[0,1]} \mu^k \widehat{J}^k_t(\mathcal{X}^2)/\mu^k J^k_t(\mathcal{X}^2)\, dt > 1$. By (2.2) and since it is assumed that $d(x, y) \ge 1$ for all distinct $x, y \in \mathcal{X}$,

$$\sup_{0\le t\le 1} \mu^k J^k_t(\mathcal{X}^2) \le \bar{J}/k, \tag{7.3}$$

where $\bar{J} := \sup_{t,x} J_{t,x}(\mathcal{X}) < \infty$ and with (7.2),

$$\lim_{k\to\infty} \int_{[0,1]} \mu^k \widehat{J}_t^k(\mathcal{X}^2)\,dt = \int_{[0,1]} \mu \widehat{J}_t(\mathcal{X}^2)\,dt =: K > 0. \tag{7.4}$$

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 7.1 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 (7.3) and (7.4), integrating and taking the limit leads us to

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

We conclude with (7.2) that $\int_{[0,t]} \mu \widehat{J}_s(\mathcal{X}^2)\,ds = Kt$ for all $0 \le t \le 1$, which is the announced result. $\square$

Proof of Proposition 7.1. A consequence of Girsanov's formula [see Léonard (2012b), Theorem 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

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

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

$$\tau \mapsto \int_{[0,1]} L(\mathbf{q}_{\tau(t)}, \dot{\tau}(t)\mathbf{v}_{\tau(t)})\,dt \tag{7.5}$$

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 7.1, 7.3 and 7.5. With Lemma 7.1, we see that we are in position to apply Lemma 7.5. We conclude with (7.5), Lemmas 7.3 and 7.5.

It remains to prove Lemmas 7.1, 7.3 and 7.5.

Statement and proof of Lemma 7.1. We show that the finite entropy condition $H(P|R)<\infty$ implies that the average number of jumps under P is finite.

LEMMA 7.1. *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 $N=\sum_{0<t<1}\mathbf{1}_{\{X_{t^-}\neq X_t\}}$ 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 (2.2) 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 $H(P|R)<\infty$. $\square$

Statement and proof of Lemma 7.3. To prove Lemma 7.3, we need the preliminary Lemma 7.2 which is stated and proved below. We consider the set

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

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

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

Let us fix $K\ge 0$ 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

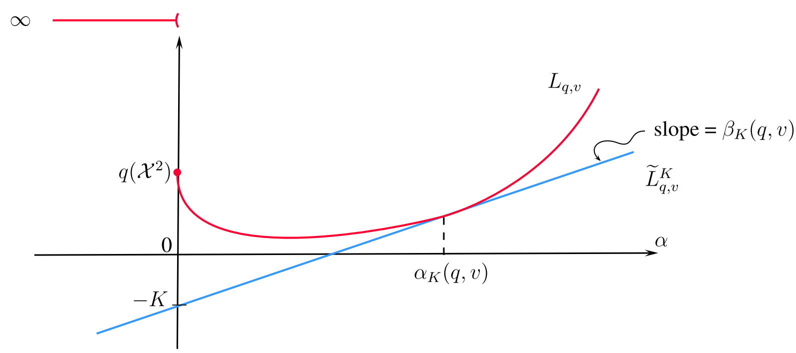

FIG. 2. *$L_{q,v}$ and its tangent line $\widetilde{L}^K_{q,v}$.*

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); see Figure 2. Define

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

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 7.2. *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

$$H_{q,v}(\beta_K(q,v)) = K. \tag{7.6}$$

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

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

where

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

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,

$$L(q, v) = \langle p^v, v\rangle_q - H\bigl(q, p^v\bigr). \tag{7.7}$$

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 (7.7) 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 (7.6) 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. $\square$

As a consequence of Lemma 7.2, we obtain the following result.

LEMMA 7.3. *Suppose that $\widehat{\tau}$ is a change of time which solves the differential equation*

$$\dot{\tau}(t) = a_K(\mathbf{q}_{\tau(t)}, \mathbf{v}_{\tau(t)}), \qquad t \in [0, 1], \text{ a.e.} \tag{7.8}$$

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

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

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 (7.8), we have $a_K(\widehat{\mathbf{q}}_t, \widehat{\mathbf{v}}_t) = 1$, for a.e. $t \in [0, 1]$. With Lemma 7.2, 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,

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

is a quantity which does not depend on τ. Hence, for any change of time τ,

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

where the last inequality follows from $L_K \le L$; see Lemma 7.2. This proves the minimal attainment at $\widehat\tau$.

The last statement follows from (7.8) $\iff$ $\widehat{\mathbf{v}}_t = v_K(\widehat{\mathbf{q}}_t, \widehat{\mathbf{v}}_t)$ and Lemma 7.2. $\square$

Statement and proof of Lemma 7.5. Now, we prove Lemma 7.5. To do so, we need the following preliminary Lemma 7.4.

LEMMA 7.4. *For each $K \ge 0$,*

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

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 (7.6) that $K = H_{q,v}(\beta_K(q,v)) = \alpha_K(q,v)\int_{\mathcal{X}^2} u\,dq - q(\mathcal{X}^2)$. Hence, $a_K(q,v) = \alpha_K(q,v)/|v|_q = (K + q(\mathcal{X}^2))/\int_{\mathcal{X}^2} v\,dq$. $\square$

As a consequence of Lemma 7.4, we obtain the following result.

LEMMA 7.5. *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) $1 < \int_{[0,1]} \nu j_t(\mathcal{X}^2)/\nu J_t(\mathcal{X}^2)\,dt \le \infty$.

Then there exist a constant $\widehat K > 0$ and a change of time $\widehat\tau$ such that (7.8) *holds*:

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

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 $0 \le s \le 1$. Hence, with Lemma 7.4 we obtain that (7.8) writes as

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

where

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

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 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 $\lim_{K \to \infty} \Phi_K(1) = 0$. □

APPENDIX A: MARKOV RANDOM WALK ON A GRAPH

We give some basic information about Markov random walks on a graph.

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

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

where $j_{t,x}(y) \ge 0$ 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

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

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 other words, $Q \in \mathrm{M}_+(\Omega)$ is the unique solution to the martingale problem on Ω associated with the generator L defined on the space of finitely supported functions u on $\mathcal{X}$ with the prescribed initial law $Q_0 \in \mathrm{M}_+(\mathcal{X})$. It is a Markov measure in the sense of Definition 6.1.

In case L does not 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 reversible Markov random walks R. One may require that R is reversible. This means that R is invariant with respect to time reversal, that is, for any subinterval $[u, v] \subset [0, 1]$,

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

This implies that R is stationary, that is, there is some $m \in \mathrm{M}_+(\mathcal{X})$ such that $R_t = m$, $\forall 0 \le t \le 1$. If R is a time-homogeneous Markov random walk, this happens if and only if the following detailed balance condition

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

is satisfied. 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

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

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 $n_x \ge 1$ 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 $1 \le c < \infty$ such that

$$m_y/n_y \le c m_x/n_x \qquad \forall x \sim y$$

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

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

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, that is, 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 Léonard [(2014b), Section 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 $h(a) = a\log a - a + 1 \ge 0$ for all $a \ge 0$ [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 $p \in \mathrm{P}(Y)$,

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

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

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

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

$$H(p|r) = H(\phi_\# p|\phi_\# r) + \int_Z H\bigl(p(\cdot|\phi = z)|r(\cdot|\phi = z)\bigr)\phi_\# p(dz), \tag{B.4}$$

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

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

with equality if and only if

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

REFERENCES

AMBROSIO, L. (2003). Lecture notes on optimal transport problems. In *Mathematical Aspects of Evolving Interfaces* (*Funchal*, 2000). *Lecture Notes in Math.* **1812** 1–52. Springer, Berlin. MR2011032

AMBROSIO, L., KIRCHHEIM, B. and PRATELLI, A. (2004). Existence of optimal transport maps for crystalline norms. *Duke Math. J.* **125** 207–241. MR2096672

AMBROSIO, L. and PRATELLI, A. (2003). Existence and stability results in the L^1 theory of optimal transportation. In *Optimal Transportation and Applications* (*Martina Franca*, 2001). *Lecture Notes in Math.* **1813** 123–160. Springer, Berlin. MR2006307

ANZELLOTTI, G. and BALDO, S. (1993). Asymptotic development by Γ-convergence. *Appl. Math. Optim.* **27** 105–123. MR1202527

ATTOUCH, H. (1996). Viscosity solutions of minimization problems. *SIAM J. Optim.* **6** 769–806. MR1402205

BENAMOU, J.-D. and BRENIER, Y. (2000). A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math.* **84** 375–393. MR1738163

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

BONCIOCAT, A.-I. and STURM, K.-T. (2009). Mass transportation and rough curvature bounds for discrete spaces. *J. Funct. Anal.* **256** 2944–2966. MR2502429

CAFFARELLI, L. A., FELDMAN, M. and MCCANN, R. J. (2002). Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs. *J. Amer. Math. Soc.* **15** 1–26 (electronic). MR1862796

CHAMPION, T. and DE PASCALE, L. (2011). The Monge problem in $\mathbb{R}^d$. *Duke Math. J.* **157** 551–572. MR2785830

CORDERO-ERAUSQUIN, D., MCCANN, R. J. and SCHMUCKENSCHLÄGER, M. (2001). A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.* **146** 219–257. MR1865396

DAL MASO, G. (1993). *An Introduction to Γ-Convergence*. *Progress in Nonlinear Differential Equations and Their Applications* **8**. Birkhäuser, Boston, MA. MR1201152

DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **38**. Springer, New York. MR1619036

ERBAR, M. and MAAS, J. (2012). Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.* **206** 997–1038. MR2989449

EVANS, L. C. and GANGBO, W. (1999). Differential equations methods for the Monge–Kantorovich mass transfer problem. *Mem. Amer. Math. Soc.* **137** viii+66. MR1464149

GOZLAN, N., ROBERTO, C., SAMSON, P. M. and TETALI, P. (2014). Displacement convexity of entropy and related inequalities on graphs. *Probab. Theory Related Fields* **160** 47–94.

HILLION, E. (2012). Concavity of entropy along binomial convolutions. *Electron. Commun. Probab.* **17** no. 4, 9. MR2872573

HILLION, E. (2014a). $W_{1,+}$-interpolation of probability measures on graphs. *Electron. J. Probab.* **41** 679–698.

HILLION, E. (2014b). Entropy along $W_{1,+}$-geodesics on graphs. *Electron. J. Probab.* **19** 1–29.

HILLION, E. (2014c). Contraction of measures on graphs. *Potential Anal.* **41** 679–698. MR3264815

JOHNSON, O. (2007). Log-concavity and the maximum entropy property of the Poisson distribution. *Stochastic Process. Appl.* **117** 791–802. MR2327839

LÉONARD, C. (2012a). From the Schrödinger problem to the Monge–Kantorovich problem. *J. Funct. Anal.* **262** 1879–1920. MR2873864

LÉONARD, C. (2012b). Girsanov theory under a finite entropy condition. In *Séminaire de Probabilités XLIV*. *Lecture Notes in Math.* **2046** 429–465. Springer, Heidelberg. MR2953359

LÉONARD, C. (2012c). On the convexity of the entropy along entropic interpolations. Preprint. Available at arXiv:1310.1274.

LÉONARD, C. (2014a). A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.* **34** 1533–1574. MR3121631

LÉONARD, C. (2014b). Some properties of path measures. In *Séminaire de Probabilités de Strasbourg*, *Vol.* 46. 207–230. Springer, Berlin.

LÉONARD, C., RŒLLY, S. and ZAMBRINI, J.-C. (2014). Reciprocal processes. A measure-theoretical point of view. *Probab. Surv.* **11** 237–269. MR3269228

LOTT, J. and VILLANI, C. (2009). Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math.* (2) **169** 903–991. MR2480619

MAAS, J. (2011). Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.* **261** 2250–2292. MR2824578

MCCANN, R. J. (1994). A convexity theory for interacting gases and equilibrium crystals. Ph.D. thesis, Princeton Univ., Princeton, NJ.

MCCANN, R. J. (1997). A convexity principle for interacting gases. *Adv. Math.* **128** 153–179. MR1451422

MIELKE, A. (2011). A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Nonlinearity* **24** 1329–1346. MR2776123

NAGASAWA, M. (1993). *Schrödinger Equations and Diffusion Theory*. *Monographs in Mathematics* **86**. Birkhäuser, Basel. MR1227100

OLLIVIER, Y. (2009). Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.* **256** 810–864. MR2484937

OLLIVIER, Y. and VILLANI, C. (2012). A curved Brunn–Minkowski inequality on the discrete hypercube, or: What is the Ricci curvature of the discrete hypercube? *SIAM J. Discrete Math.* **26** 983–996. MR3022118

OTTO, F. and VILLANI, C. (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** 361–400. MR1760620

SCHRÖDINGER, E. (1931). Über die Umkehrung der Naturgesetze. *Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math.* **144** 144–153.

SCHRÖDINGER, E. (1932). Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Ann. Inst. H. Poincaré* **2** 269–310. MR1508000

STURM, K.-T. (2006a). On the geometry of metric measure spaces. I. *Acta Math.* **196** 65–131. MR2237206

STURM, K.-T. (2006b). On the geometry of metric measure spaces. II. *Acta Math.* **196** 133–177. MR2237207

SUDAKOV, V. N. (1979). Geometric problems in the theory of infinite-dimensional probability distributions. *Proc. Steklov Inst. Math.* **2** i–v, 1–178. Cover to cover translation of Trudy Mat. Inst. Steklov 141 (1976). MR0530375

VILLANI, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454

VON RENESSE, M.-K. and STURM, K.-T. (2005). Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.* **58** 923–940. MR2142879

MODAL-X. UNIVERSITÉ PARIS OUEST
BÂT. G, 200 AV. DE LA RÉPUBLIQUE
92001 NANTERRE
FRANCE
E-MAIL: christian.leonard@u-paris10.fr