GROMOV-HAUSDORFF CONVERGENCE OF DISCRETE TRANSPORTATION METRICS

NICOLA GIGLI AND JAN MAAS

ABSTRACT. This paper continues the investigation of 'Wasserstein-like' transportation distances for probability measures on discrete sets. We prove that the discrete transportation metrics on the d-dimensional discrete torus $\mathbf{T}_N^d$ with mesh size $\frac{1}{N}$ converge, when $N \to \infty$, to the standard 2-Wasserstein distance on the continuous torus in the sense of Gromov–Hausdorff. This is the first result on a passage to the limit from a discrete transportation problem to a continuous one, and it is a key ingredient needed to prove convergence of gradient flows. Furthermore, the result shows compatibility of the recently developed discrete metrics and the well-established 2-Wasserstein metric.

CONTENTS

1. INTRODUCTION

It is by now well known that large classes of PDEs can be formulated as gradient flows of entropy functionals, e.g., Fokker-Planck equations [12], porous medium equations [16], granular media equations [5] and other equations. In all of these equations, the metric used in the gradient flow formulation is the 2-Wasserstein metric.

Much more recently, it has been realised that an even larger class of natural models can be formulated as gradient flows of entropy functionals, provided that one replaces the Wasserstein metric by a different metric, which is obtained by adapting the Benamou-Brenier formulation of the Wasserstein metric [3]. Examples include finite Markov chains [13] (e.g. statistical

mechanical models), systems of reaction diffusion equations [14], as well as fermionic Fokker-Planck equations [4] and other dissipative quantum mechanical evolutions [15].

An important consequence of this development is that it opens the door to study discretisations of continuous models, in a framework which preserves the gradient flow structure at the discrete level. Typical examples are discretisations of Fokker-Planck equations, for which a new gradient flow structure has been found in [6, 13, 14]. In such a situation, involving functionals defined on discrete spaces, it is natural to ask whether the associated gradient flows converge to their continuous counterpart.

There are various recent papers dealing with convergence of gradient flows under Γ-convergence of functionals (see [7, 11, 17]). A crucial ingredient in this analysis is the convergence of the metric spaces on which the functionals are defined, often understood in the sense of Gromov-Hausdorff. To realise the power of this approach, note that in typical applications it allows to prove convergence of second order PDEs, while the Gromov-Hausdorff convergence of the spaces is, heuristically speaking, a zeroth order condition.

The present paper proves a Gromov-Hausdorff convergence result in an important special case and may serve as guide to prove Gromov-Hausdorff convergence results in other situations: we present a first example of passage to the limit from discrete to continuous, well suited to study convergence of gradient flows.

We consider the space $\mathscr{P}(\mathbf{T}^d)$ of probability measures on the torus $\mathbf{T}^d := \mathbf{R}^d/\mathbf{Z}^d$, endowed with the usual 2-Wasserstein metric W_2. We also consider the d-dimensional periodic lattice $\mathbf{T}_N^d := (\mathbf{Z}/N\mathbf{Z})^d$ with mesh size $\frac{1}{N}$, and endow the space of probability measures $\mathscr{P}(\mathbf{T}_N^d)$ with its renormalized discrete transportation metric $\mathcal{W}_N$ as defined in [6, 13, 14] (see Section 2 below). This metric relies on a modification of the Benamou-Brenier formula, and is defined by minimizing an action functional over curves in the spaces of probability measures, rather than a cost functional over measures on the product space.

The main result of this paper is the following theorem, which proves compatibility between the discrete and the continuous theory.

Theorem 1.1. *Let $d \geq 1$. Then the metric spaces $(\mathscr{P}(\mathbf{T}_N^d), \mathcal{W}_N)$ converge to $(\mathscr{P}(\mathbf{T}^d), W_2)$ in the sense of Gromov-Hausdorff as $N \to \infty$.*

We refer to Section 3.1 for an outline of the strategy of the proof. The crucial estimates needed in our argument are contained in Section 3.2. The proof is completed in Section 3.3.

To finish the introduction, let us discuss in some detail how Theorem 1.1 can be used to prove convergence of the associated gradient flows.

- Let π_N be the uniform probability measure on $\mathbf{T}_N^d$ and let π be the Lebesgue measure on $\mathbf{T}^d$. It is not difficult to see that the relative entropy functionals Ent_{π_N} on $\mathscr{P}(\mathbf{T}_N^d)$ Γ-converge, as $N \to \infty$, to the relative entropy functional Ent_{π} on $\mathscr{P}(\mathbf{T}^d)$.
- In [10] it has been proved that the functionals Ent_{π_N} are all geodesically convex on $(\mathscr{P}(\mathbf{T}_N^d), \mathcal{W}_N)$.
- From [13] we know that the gradient flow of Ent_{π_N} with respect to $\mathcal{W}_N$ produces solutions to the heat flow.
- These results can be combined to obtain convergence of gradient flows, since it has been proved in [11] that gradient flows of λ-geodesically convex functionals on Gromov-Hausdorff convergent spaces are stable with respect to Γ-convergence.

Of course, the convergence of the discrete heat flow to the continuous one is not a new result, and could also be proved directly, for instance using the explicit formulas for the heat kernels.

Yet the argument pointed out here has the advantage of being based on discrete Ricci bounds and Gromov-Hausdorff convergence only, and as such this strategy can be useful also in other situations. For instance, in [2] the stability of the heat flow in a continuous and non-smooth context has been used to show the stability of Ricci curvature bounds in conjunction with the linearity of the heat flow.

2. Preliminaries

2.1. **The 2-Wasserstein metric.** Let $\mathcal{M}$ be a compact smooth Riemannian manifold and $\mathscr{P}(M)$ the set of Borel probability measures on it. The Wasserstein distance W_2 on $\mathscr{P}(M)$ is usually defined by minimizing the transport cost with respect to the cost function distance-squared. It has been emphasized by Benamou and Brenier [3] that a completely different introduction to the subject can be given in terms of solutions to the continuity equation. The following result has been proved for $M = \mathbf{R}^d$ in [1] (see also [16]), the case of general manifolds being a consequence of Nash's embedding theorem (see also [9, Proposition 2.5] for a direct proof on manifolds).

Proposition/Definition 2.1. *Let $\mathcal{M}$ be a compact smooth Riemannian manifold and $\mu, \nu \in \mathscr{P}(M)$. Then we have*

$$W_2^2(\mu, \nu) = \min \int_0^1 \int_M |v_t|^2(x) \, \mathrm{d}\mu_t(x) \, \mathrm{d}t \,, \tag{2.1}$$

the minimum being taken among all distributional solutions (μ_t, v_t) of the continuity equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t + \nabla \cdot (v_t \mu_t) = 0 \,, \tag{2.2}$$

such that $t \mapsto \mu_t$ is weakly continuous in duality with $C(M)$ and $\mu_0 = \mu$, $\mu_1 = \nu$.

In the sequel, when considering the continuous setting we will work with M being the d-dimensional torus $\mathbf{T}^d := \mathbf{R}^d/\mathbf{Z}^d$ and we will consider solutions to the continuity equation in terms of probability densities and momentum vector fields. To fix the ideas, we give the following definition.

Definition 2.2 (Solutions to the continuity equation in the continuous torus)**.** *Consider the mappings $[0,1] \times \mathbf{T}^d \ni (t,x) \mapsto \rho_t(x) \in \mathbf{R}$ and $[0,1] \times \mathbf{T}^d \mapsto V_t(x) \in \mathbf{R}^d$. We say that (ρ_t, V_t) solves the continuity equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_t + \nabla \cdot V_t = 0 \,, \tag{2.3}$$

provided both $(t,x) \mapsto \rho_t(x)$ and $(t,x) \mapsto V_t(x)$ are in $L^1([0,1] \times \mathbf{T}^d)$, $t \mapsto \rho_t$ is continuous with respect to convergence in duality with $C(\mathbf{T}^d)$, and (2.3) is satisfied in the sense of distributions.

2.2. **Discrete transportation metrics.** In several recent works [6, 13, 14] discrete analogues of W_2 have been considered, which are well suited to study evolution equations in a discrete setting. The definition of the Wasserstein distance requires a metric on the underlying space. In [13], instead, the starting point is a Markov kernel K on the finite set $\mathcal{X}$, i.e., we assume that $K : \mathcal{X} \times \mathcal{X} \to \mathbf{R}_+$ satisfies $\sum_{y \in \mathcal{X}} K(x,y) = 1$ for all $x \in \mathcal{X}$. We assume that

K is irreducible and denote the unique steady state by π. Thus π is the unique probability measure on $\mathcal{X}$ satisfying

$$\pi(y) = \sum_{x \in \mathcal{X}} \pi(x) K(x, y)$$

for all $y \in \mathcal{X}$. We shall assume that K is reversible, i.e., the detailed balance equations

$$K(x, y)\pi(x) = K(y, x)\pi(y)$$

hold for all $x, y \in \mathcal{X}$. Since basic Markov chain theory implies that π is strictly positive, we can – and will – identify probability measures on $\mathcal{X}$ with their densities with respect to π, i.e., we set

$$\mathscr{P}(\mathcal{X}) := \left\{ \rho : \mathcal{X} \to \mathbf{R}_+ \mid \sum_{x \in \mathcal{X}} \pi(x)\rho(x) = 1 \right\} .$$

In order to define the metric $\mathcal{W}$ on $\mathscr{P}(\mathcal{X})$, we let $\theta : \mathbf{R}_+ \times \mathbf{R}_+ \to \mathbf{R}_+$ denote the logarithmic mean, which is defined by

$$\theta(s, t) = \int_0^1 s^{1-p} t^p \, \mathrm{d}p .$$

For $\rho \in \mathscr{P}(\mathcal{X})$ and $x, y \in \mathcal{X}$ we set

$$\hat{\rho}(x, y) = \theta(\rho(x), \rho(y)) ,$$

which can be regarded informally as being "the density ρ at the edge (x, y)". According to [10, Lemma 2.6], the following definition can be taken as one of the equivalent definitions of the transportation metric $\mathcal{W}$ on $\mathscr{P}(\mathcal{X})$.

Definition 2.3. *Let K be an irreducible and reversible Markov kernel on a finite set $\mathcal{X}$, and let $\bar{\rho}_0, \bar{\rho}_1 \in \mathscr{P}(\mathcal{X})$. The distance $\mathcal{W}(\bar{\rho}_0, \bar{\rho}_1)$ is defined by*

$$\mathcal{W}(\bar{\rho}_0, \bar{\rho}_1)^2 = \inf \left\{ \frac{1}{2} \int_0^1 \sum_{x, y \in \mathcal{X}} \frac{V_t(x, y)^2}{\hat{\rho}_t(x, y)} K(x, y)\pi(x) \, \mathrm{d}t \right\} , \tag{2.4}$$

where the infimum runs over all curves $[0, 1] \ni t \mapsto (\rho_t, V_t)$ such that:

   (i) $\rho_t \in \mathscr{P}(\mathcal{X})$ *for any $t \in [0, 1]$, the function $t \mapsto \rho_t(x)$ is continuous for any $x \in \mathcal{X}$, and $\rho_0 = \bar{\rho}_0$, $\rho_1 = \bar{\rho}_1$;*
  (ii) $V_t : \mathcal{X} \times \mathcal{X} \to \mathbf{R}$ *for any $t \in [0, 1]$, and the function $t \mapsto V_t(x, y)$ belongs to $L^1(0, 1)$ for any $x, y \in \mathcal{X}$;*
 (iii) the "discrete continuity equation"

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_t(x) + \frac{1}{2} \sum_{y \in \mathcal{X}} \big(V_t(x, y) - V_t(y, x)\big) K(x, y) = 0 \tag{2.5}$$

   holds for all $x \in \mathcal{X}$ in the sense of distributions.

2.3. **The transportation metric on the discrete torus.** In this paper we shall only be concerned with simple random walk on the d-dimensional discrete torus $\mathbf{T}_N^d := (\mathbf{Z}/N\mathbf{Z})^d = \{0, \ldots, N-1\}^d$, in which case the kernel $K_N : \mathbf{T}_N^d \times \mathbf{T}_N^d \to [0, 1]$ is given by

$$K_N(\mathbf{a}, \mathbf{b}) = \begin{cases} \frac{1}{2d}, & \mathbf{b} = \mathbf{a} \pm \mathbf{e}_i \mod N \text{ for some } i \in \{1, \ldots, d\} , \\ 0, & \text{otherwise} . \end{cases}$$

Here, $\mathbf{e}_i$ denotes the i-th unit vector. All computations in $\mathbf{T}_N^d$ will be performed modulo N without further mentioning.

In this case the stationary probability measure π_N is the uniform measure given by $\pi_N(\mathbf{a}) = N^{-d}$ for all $\mathbf{a} \in \mathbf{T}_N^d$. Therefore, the collection of probability densities with respect to π_N is given by

$$\mathscr{P}(\mathbf{T}_N^d) = \left\{ \rho_N : \mathbf{T}_N^d \to \mathbf{R}_+ \ \Big| \ \sum_{\mathbf{a} \in \mathbf{T}_N^d} \rho_N(\mathbf{a}) = N^d \right\} .$$

For functions $f, g : \mathbf{T}_N^d \to \mathbf{R}$ we consider the normalized L^2-inner product

$$\langle f, g \rangle_{L_N^2} = \frac{1}{N^d} \sum_{\mathbf{a} \in \mathbf{T}_N^d} f(\mathbf{a}) g(\mathbf{a})$$

and the Dirichlet form

$$\mathcal{E}_N(f, g) = \frac{1}{N^{d-2}} \sum_{\mathbf{a} \in \mathbf{T}_N^d} \sum_{i=i}^{d} \big(f(\mathbf{a} + \mathbf{e}_i) - f(\mathbf{a}) \big) \big(g(\mathbf{a} + \mathbf{e}_i) - g(\mathbf{a}) \big) .$$

Furthermore we set

$$\|f\|_{L_N^2} = \sqrt{\langle f, f \rangle_{L_N^2}} , \qquad \mathcal{E}_N(f) = \mathcal{E}_N(f, f) .$$

Let Δ_N be the discrete Laplacian, defined by

$$\Delta_N f(\mathbf{a}) = 2dN^2(K_N - I)f(\mathbf{a}) = N^2 \sum_{i=1}^{d} \Big(f(\mathbf{a} + \mathbf{e}_i) - 2f(\mathbf{a}) + f(\mathbf{a} - \mathbf{e}_i) \Big)$$

for $\mathbf{a} \in \mathbf{T}_N^d$. Notice that following integration by parts formula holds:

$$\mathcal{E}_N(f, g) = -\langle \Delta_N f, g \rangle_{L_N^2} . \tag{2.6}$$

Moreover, given $g : \mathbf{T}_N^d \to \mathbf{R}$, the equation $\Delta_N f = g$ can be solved if and only if $\sum_{\mathbf{a} \in \mathbf{T}_N^d} g(\mathbf{a}) = 0$, in which case the solution is unique. We shall use the well-known Poincaré inequality on $\mathbf{T}_N^d$, which we now recall.

Proposition 2.4 (Poincaré inequality on $\mathbf{T}_N^d$). *Let $d \geq 1$ and $N \geq 4$. For all $f : \mathbf{T}_N^d \to \mathbf{R}$ with $\sum_{\mathbf{a} \in \mathbf{T}_N^d} f(\mathbf{a}) = 0$ we have*

$$\|f\|_{L_N^2}^2 \leq \frac{1}{2N^2(1 - \cos(2\pi/N))} \mathcal{E}_N(f) ,$$

$$\mathcal{E}_N(\Delta_N^{-1} f) \leq \frac{1}{2N^2(1 - \cos(2\pi/N))} \|f\|_{L_N^2}^2 .$$

Proof. One way to prove the first inequality is as follows. If $d = 1$, then the spectrum of the operator $I - K_N$ on $L^2(\mathbf{T}_N^d, \pi_N)$ consists of the eigenvalues

$$1 - \cos(2\pi n/N) , \qquad 0 \leq n \leq N - 1 ,$$

(see, e.g., [8, Section 4.2]), which yields the result if $d = 1$. The result in dimension $d > 1$ follows by tensorization (see, e.g., [8, Lemma 3.2]).

The second inequality follows from the first one, using the integration by parts formula (2.6). $\qquad\square$

Remark 2.5. In the limit $N \to \infty$, one recovers the classical Poincaré inequality on the torus $\mathbf{T}^d$:

$$\|f\|^2_{L^2(\mathbf{T}^d)} \leq \frac{1}{2\pi^2} \|\nabla f\|^2_{L^2(\mathbf{T}^d)} \, ,$$

valid for any f with zero mean.

It will be useful to introduce some more notation. For $\mathbf{a} = (a_1, \ldots, a_d) \in \mathbf{T}^d_N$ we define the cube $Q^N_{\mathbf{a}}$ by

$$Q^N_{\mathbf{a}} := \Big[\frac{a_1}{N}, \frac{a_1+1}{N}\Big) \times \cdots \times \Big[\frac{a_d}{N}, \frac{a_d+1}{N}\Big) \subseteq \mathbf{T}^d \, ,$$

so that the torus $\mathbf{T}^d = \mathbf{R}^d / \mathbf{Z}^d$ can be written as the disjoint union

$$\mathbf{T}^d = \bigcup_{\mathbf{a} \in \mathbf{T}^d_N} Q^N_{\mathbf{a}} \, .$$

For $i = 1, \ldots, d$, the facets of $Q^N_{\mathbf{a}}$ will be denoted by

$$R^N_{\mathbf{a},i-} = \Big[\frac{a_1}{N}, \frac{a_1+1}{N}\Big] \times \; \cdots \; \Big\{\frac{a_i}{N}\Big\} \; \cdots \; \times \Big[\frac{a_d}{N}, \frac{a_d+1}{N}\Big] \, ,$$
$$R^N_{\mathbf{a},i+} = \Big[\frac{a_1}{N}, \frac{a_1+1}{N}\Big] \times \cdots \Big\{\frac{a_i+1}{N}\Big\} \cdots \times \Big[\frac{a_d}{N}, \frac{a_d+1}{N}\Big] \, .$$

The collection of all these facets $R^N_{\mathbf{a},i\pm}$ will be denoted by $\mathscr{R}^N$. For $R = R^N_{\mathbf{a},i\pm} \in \mathscr{R}^N$ we shall write

$$\hat{\rho}(R^N_{\mathbf{a},i\pm}) := \theta(Q^N_{\mathbf{a}}, Q^N_{\mathbf{a}\pm\mathbf{e}_i}) \, .$$

Notice that $K_N(\mathbf{a}, \mathbf{b})$ is non-zero only for $\mathbf{a}, \mathbf{b}$ such that $\mathbf{a} - \mathbf{b} = \pm\mathbf{e}_i$ for some $i = 1, \ldots, d$. Therefore we can think about the vector fields $V : \mathbf{T}^d_N \times \mathbf{T}^d_N \to \mathbf{R}$ appearing in Definition 2.3(ii) as being defined on facets $R \in \mathscr{R}^N$, rather than on generic couples $\mathbf{a}, \mathbf{b}$. This will be our convention from now on.

Let $\mathcal{W}_{K_N}$ denote the metric on $\mathscr{P}(\mathbf{T}^d_N)$ associated with the kernel K_N according to Definition 2.3. It will be convenient to work with the normalised metric

$$\mathcal{W}_N := \frac{\mathcal{W}_{K_N}}{N\sqrt{2d}} \, ,$$

which is a quantity of order 1.

Given a probability density $\rho_N \in \mathscr{P}(\mathbf{T}^d_N)$ and a 'momentum vector field' $V_N : \mathscr{R}^N \to \mathbf{R}$, the *action* $\mathcal{A}_N$ of (ρ_N, V_N) is defined by

$$\mathcal{A}_N(\rho_N, V_N) := \frac{1}{4d^2 N^{d+2}} \sum_{R \in \mathscr{R}^N} \frac{V_N(R)^2}{\hat{\rho}_N(R)} \, . \tag{2.7}$$

With this notation and taking Definition 2.3 into account, it is immediate to obtain the following expression for the metric $\mathcal{W}_N$.

Lemma 2.6. *For any $\bar{\rho}_{N,0}, \bar{\rho}_{N,1} \in \mathscr{P}(\mathbf{T}^d_N)$ we have*

$$\mathcal{W}_N(\bar{\rho}_{N,0}, \bar{\rho}_{N,1})^2 = \inf \Big\{ \int_0^1 \mathcal{A}_N(\rho_{N,t}, V_{N,t}) \, \mathrm{d}t \Big\} \, , \tag{2.8}$$

where the infimum runs over all curves $[0,1] \ni t \mapsto (\rho_{N,t}, V_{N,t})$ such that:

(i) $\rho_{N,t} \in \mathscr{P}(\mathbf{T}_N^d)$ for any $t \in [0,1]$, and the function $t \mapsto \rho_{N,t}(\mathbf{a})$ is continuous for any $\mathbf{a} \in \mathbf{T}_N^d$ with $\rho_{N,0} = \bar{\rho}_{N,0}$, $\rho_{N,1} = \bar{\rho}_{N,1}$;

(ii) $V_{N,t} : \mathscr{R}^N \to \mathbf{R}$ for any $t \in [0,1]$, and the function $t \mapsto V_{N,t}(R)$ belongs to $L^1(0,1)$ for any $R \in \mathscr{R}^N$;

(iii) the discrete continuity equation

$$\frac{\mathrm{d}}{\mathrm{d}t} \rho_{N,t}(\mathbf{a}) + \frac{1}{2d} \sum_{i=1}^d \left(V_{N,t}(R_{\mathbf{a},i+}^N) - V_{N,t}(R_{\mathbf{a},i-}^N) \right) = 0 \tag{2.9}$$

holds for all $\mathbf{a} \in \mathbf{T}_N^d$ in the sense of distributions.

By analogy with Definition 2.2 we formulate the following discrete counterpart.

Definition 2.7 (Solutions to the continuity equation in the discrete torus)**.** *Let* $[0,1] \times \mathbf{T}_N^d \ni (t,\mathbf{a}) \mapsto \rho_{N,t}(\mathbf{a}) \in \mathbf{R}$ *and* $[0,1] \times \mathscr{R}^N \ni (t,R) \mapsto V_{N,t}(R) \in \mathbf{R}^d$. *We say that* $(\rho_{N,t}, V_{N,t})$ *is a solution to the discrete continuity equation* (2.9) *provided that* (i), (ii) *and* (iii) *in Lemma 2.6 are fulfilled.*

Finally, we recall a couple of properties of $\mathcal{W}_N$ that will be used in the sequel. We shall use the metric d_N on $\mathbf{T}_N^d$ defined by

$$\mathsf{d}_N(\mathbf{a}, \mathbf{b}) = \frac{1}{N} \sqrt{\sum_{i=1}^d |a_i - b_i|^2}$$

for $\mathbf{a}, \mathbf{b} \in \mathbf{T}_N^d$. We let

$$W_{2,N} \tag{2.10}$$

denote the standard 2-Wasserstein distance on $\mathscr{P}(\mathbf{T}_N^d)$ induced by the distance d_N on $\mathbf{T}_N^d$. In the following result we collect some basic properties of the metric $\mathcal{W}_N$.

Proposition 2.8. *The following assertions hold.*

(i) *The function* $(\rho, \sigma) \mapsto \mathcal{W}_N^2(\rho, \sigma)$ *is convex on* $\mathscr{P}(\mathbf{T}_N^d) \times \mathscr{P}(\mathbf{T}_N^d)$ *with respect to linear interpolation.*

(ii) *There exists a universal constant* $C > 0$ *such that*

$$\mathcal{W}_N \leq \frac{C}{\sqrt{d}} W_{2,N} \,.$$

In particular, the diameter of the spaces $(\mathscr{P}(\mathbf{T}_N^d), \mathcal{W}_N)$ *is bounded by a constant depending only on the dimension.*

Proof. The first assertion has been proved in [10, Proposition 2.8]. For the second assertion, we apply [10, Proposition 2.12] to obtain

$$\mathcal{W}_N \leq \frac{c}{dN} W_{2,N}' \,,$$

where $c \approx 0,78$ is a universal constant and $W_{2,N}'$ is the 2-Wasserstein distance on $\mathscr{P}(\mathbf{T}_N^d)$ induced by the distance d_N' on $\mathbf{T}_N^d$, defined by $d_N'(\mathbf{a}, \mathbf{b}) := \sum_i |\mathbf{a}_i - \mathbf{b}_i|$. Since $d_N'(\mathbf{a}, \mathbf{b}) \leq \sqrt{d} N \mathsf{d}_N(\mathbf{a}, \mathbf{b})$, we have $W_{2,N}' \leq \sqrt{d} N W_{2,N}$, which implies yields the desired estimate. Since the diameter of the spaces $(\mathbf{T}_N^d, \mathsf{d}_N)$ is uniformly bounded by a dimensional constant, the final assertion follows as well. $\square$

2.4. **Some properties of the heat semigroup on the discrete and continuous torus.**
We endow the continuous torus $\mathbf{T}^d$ with its natural Riemannian flat distance, and we denote
the Lebesgue measure by π.

Let $(\mathsf{H}_t)_{t\geq 0}$ be the heat semigroup on $\mathbf{T}^d$ with generator Δ, acting either on measures or
functions. The heat semigroup on $\mathbf{T}_N^d$ is the semigroup generated by the discrete Laplacian
Δ_N, and will be denoted by $(\mathsf{H}_t^N)_{t\geq 0}$.

Let h_t be the heat kernel on $\mathbf{T}^d$, i.e., the density of $\mathsf{H}_t(\delta_0)$ with respect to π. Similarly, h_t^N
will denote the heat kernel on $\mathbf{T}_N^d$, which is defined by $\mathsf{h}_t^N(x) = \mathsf{H}_t^N(N^d\mathbf{1}_{\{\mathbf{0}\}})(x)$. We thus
have the formulas

$$\mathsf{H}_t f(x) = \int_{\mathbf{T}^d} \mathsf{h}_t(x-y)f(y)\,\mathrm{d}\pi(y)\,, \qquad \mathsf{H}_t^N f_N(\mathbf{a}) = \frac{1}{N^d}\sum_{\mathbf{b}\in\mathbf{T}_N^d} \mathsf{h}_t^N(\mathbf{a}-\mathbf{b})f_N(\mathbf{b})\,,$$

valid for all L^1-functions $f : \mathbf{T}^d \to \mathbf{R}$ and $f_N : \mathbf{T}_N^d \to \mathbf{R}$.

The heat semigroup on $\mathbf{T}^d$ acts on vector fields as well coordinatewise. Similarly, the action
of H_t^N on a vector field $V_N : \mathscr{R}^N \to \mathbf{R}$ can be defined via

$$\mathsf{H}_t^N V_N(R_{\mathbf{a},i+}^N) := \frac{1}{N^d}\sum_{\mathbf{b}\in\mathbf{T}_N^d} \mathsf{h}_t^N(\mathbf{a}-\mathbf{b})V_N(R_{\mathbf{b},i+}^N)\,. \tag{2.11}$$

Given a function $f : \mathbf{T}^d \to \mathbf{R}$, its Lipschitz constant will be denoted by $\mathrm{Lip}(f)$. Similarly,
we define the Lipschitz constant of a function $f : \mathbf{T}_N^d \to \mathbf{R}$ by

$$\mathrm{Lip}_N(f) := \sup_{\mathbf{a}\neq\mathbf{b}} \frac{|f(\mathbf{a})-f(\mathbf{b})|}{\mathsf{d}_N(\mathbf{a},\mathbf{b})}\,.$$

The propositions below collect some basic properties of the heat flows that we will use in the
sequel.

Proposition 2.9 (Heat flow on the continuous torus)**.** *The following assertions hold for all*
$s > 0$.

(i) *There exist constants $c(s) > 0$ and $C(s) < \infty$ such that for any $\mu \in \mathscr{P}(\mathbf{T}^d)$ the density*
ρ_s *of $\mathsf{H}_s\mu$ satisfies*

$$\rho_s \geq c(s) \qquad and \qquad \mathrm{Lip}(\rho_s) \leq C(s)\,.$$

Furthermore, there exists a dimensional constant $C < \infty$ such that

$$W_2(\mathsf{H}_s\mu,\mu) \leq C\sqrt{s}\,.$$

(ii) *There exists a constant $C(s) < \infty$ such that for any $f \in L^1(\mathbf{T}^d)$ we have*

$$\|\mathsf{H}_s f\|_{L^\infty} + \mathrm{Lip}(\mathsf{H}_s f) \leq C(s)\|f\|_{L^1}\,.$$

(iii) *Let $(\mu_t) \subset \mathscr{P}(\mathbf{T}^d)$ be a geodesic, let v_t be the corresponding velocity vector fields achieving*
the minimum in (2.1), and let $\rho_{s,t}$ and $V_{s,t}$ be the densities of $\mathsf{H}_s(\mu_t)$ and $\mathsf{H}_s(v_t\mu_t)$
respectively. Then, $t \mapsto (\rho_{s,t}, V_{s,t})$ is a solution to the continuity equation (2.3), and we
have

$$\int_0^1 \int_{\mathbf{T}^d} \frac{V_{s,t}^2(x)}{\rho_{s,t}(x)}\,\mathrm{d}x\,\mathrm{d}t \leq W_2^2(\mu_0,\mu_1)\,. \tag{2.12}$$

Proof. The first assertions in (i) are obvious. To prove the last claim in (i), notice that by the convexity of W_2^2 it is sufficient to prove the claim when μ is a Dirac mass. In this case the result follows from the fact that the heat kernel on the torus can be represented by periodization of the heat kernel on $\mathbf{R}^d$, and the parabolic scaling of the latter.

The result of (ii) is standard, and (iii) follows from the convexity of $\mathbf{R}^d \times \mathbf{R}^+ \ni (x, a) \mapsto \frac{x^2}{a}$ and the fact that H_s is a convolution operator, see, e.g., Lemma 8.1.10 in [1]. $\square$

Proposition 2.10 (Heat flow on the discrete torus)**.** *The following assertions hold for $s > 0$.*

 (i) There exists a dimensional constant $C > 0$ such that for any $\rho_N \in \mathscr{P}(\mathbf{T}_N^d)$ we have

$$\mathrm{Lip}_N(\mathsf{H}_s^N \rho_N) \le \min\left\{ Cs^{-(d+1)/2}, \ \mathrm{Lip}_N(\rho_N) \right\}.$$

 (ii) For any $\rho_N \in \mathscr{P}(\mathbf{T}_N^d)$ and any momentum vector field $V_N : \mathscr{R}^N \to \mathbf{R}^d$ we have

$$\mathcal{A}_N(\mathsf{H}_s^N \rho_N, \mathsf{H}_s^N V_N) \le \mathcal{A}_N(\rho_N, V_N).$$

Proof. The estimate $\mathrm{Lip}_N(\mathsf{H}_s^N \rho_N) \le \mathrm{Lip}_N(\rho_N)$ in (i) is a simple consequence of the fact that the heat semigroup consists of convolution operators. Taking the convexity of $(x, a, b) \mapsto \frac{x^2}{\theta(a,b)}$ into account, this also gives (ii).

To prove the remaining bound in (i), we note that for any probability density $\rho_N \in \mathscr{P}(\mathbf{T}_N^d)$,

$$|\mathsf{H}_s^N \rho_N(\mathbf{a}) - \mathsf{H}_s^N \rho_N(\mathbf{b})| = \frac{1}{N^d}\left| \sum_{\mathbf{c} \in \mathbf{T}_N^d} \left(\mathsf{h}_s^N(\mathbf{a} - \mathbf{c}) - \mathsf{h}_s^N(\mathbf{b} - \mathbf{c}) \right) \rho_N(\mathbf{c}) \right|$$

$$\le \frac{1}{N^d}\left(\sum_{\mathbf{c} \in \mathbf{T}_N^d} \rho_N(\mathbf{c}) \right) \sup_{\mathbf{c} \in \mathbf{T}_N^d} \left| \mathsf{h}_s^N(\mathbf{a} - \mathbf{c}) - \mathsf{h}_s^N(\mathbf{b} - \mathbf{c}) \right|$$

$$= \sup_{\mathbf{c} \in \mathbf{T}_N^d} \left| \mathsf{h}_s^N(\mathbf{a} - \mathbf{c}) - \mathsf{h}_s^N(\mathbf{b} - \mathbf{c}) \right|.$$

Since $\mathsf{h}_s^N(\mathbf{a}) = \mathsf{h}_s^{1,N}(a_1) \cdot \ldots \cdot \mathsf{h}_s^{1,N}(a_d)$, where $\mathsf{h}^{1,N}$ denotes the heat kernel in one dimension, we infer that

$$\left| \mathsf{h}_s^N(\mathbf{a}) - \mathsf{h}_s^N(\mathbf{b}) \right| \le \|\mathsf{h}_s^{1,N}\|_{L^\infty}^{d-1} \sum_{k=1}^d |\mathsf{h}_s^{1,N}(a_k) - \mathsf{h}_s^{1,N}(b_k)|$$

$$\le \sqrt{d}\mathsf{d}_N(\mathbf{a}, \mathbf{b}) \|\mathsf{h}_s^{1,N}\|_{L^\infty}^{d-1} \mathrm{Lip}_N(\mathsf{h}_s^{1,N}),$$

and therefore

$$\mathrm{Lip}_N(\mathsf{H}_s^N \rho_N) \le \sqrt{d} \, \|\mathsf{h}_s^{1,N}\|_{L^\infty}^{d-1} \mathrm{Lip}_N(\mathsf{h}_s^{1,N}), \tag{2.13}$$

so it remains to obtain bounds on the heat kernel in one dimension. These can be obtained using the well-known (and easy to check) fact that, if $d = 1$, the spectrum of the operator $-\Delta_N$ consists of the eigenvalues

$$\lambda_l = 2dN^2\big(1 - \cos(2\pi l/N)\big), \qquad l \in L_N := \left\{ z \in \mathbf{Z} \ : \ \left\lfloor -\frac{N}{2} \right\rfloor + 1 \le z \le \left\lfloor \frac{N}{2} \right\rfloor \right\}.$$

Note that $\lambda_l = \lambda_{-l}$. The corresponding eigenvectors v_l are given by

$$v_l(\mathbf{a}) = \exp\left(\frac{2\pi i l \mathbf{a}}{N} \right), \qquad l \in L_N.$$

As a consequence, the heat kernel $\mathsf{h}_s^{1,N}$ can be written explicitly as

$$\mathsf{h}_s^{1,N}(\mathbf{a}) = \sum_{l \in L_N} e^{-\lambda_l s} v_l(\mathbf{a}) \ .$$

We shall use the fact that there exists a constant $c > 0$ such that for all $N \geq 1$ and $l \in L_N$,

$$|\lambda_l| \geq cl^2 \ , \qquad \|v_l\|_{L^\infty} \leq 1 \ , \qquad \text{and} \qquad \text{Lip}_N(v_l) \leq cl \ .$$

It follows that for some constant $C > 0$ and all $\mathbf{a}, \mathbf{b} \in \mathbf{T}_N^d$,

$$\left| \mathsf{h}_s^{1,N}(\mathbf{a}) \right| \leq \sum_{l \in L_N} e^{-\lambda_l s} |v_l(\mathbf{a})| \leq \sum_{l \in L_N} e^{-cl^2 s} \leq \frac{C}{\sqrt{s}} \ ,$$

$$\left| \mathsf{h}_s^{1,N}(\mathbf{a}) - \mathsf{h}_s^{1,N}(\mathbf{b}) \right| \leq \sum_{l \in L_N} e^{-\lambda_l s} |v_l(\mathbf{a}) - v_l(\mathbf{b})| \leq C \sum_{l \in L_N} l e^{-cl^2 s} \mathsf{d}_N(\mathbf{a}, \mathbf{b}) \leq \frac{C}{s} \mathsf{d}_N(\mathbf{a}, \mathbf{b}) \ ,$$

so that $\|\mathsf{h}_s^{1,N}\|_{L^\infty} \leq C s^{-1/2}$ and $\text{Lip}_N(\mathsf{h}_s^{1,N}) \leq C s^{-1}$. Plugging these estimates into (2.13), we obtain the desired result. $\square$

3. Proof of the main result

3.1. **Ingredients and structure of the proof.** In order to prove the stated Gromov-Hausdorff convergence of the spaces $(\mathscr{P}(\mathbf{T}_N^d), \mathcal{W}_N)$, we will introduce the natural mappings from the continuous torus to the discrete one, and those going the other way around.

First we construct discrete measures by integration over cubes, and discrete vector fields by integration over facets:

Definition 3.1 (From $\mathbf{T}^d$ to $\mathbf{T}_N^d$). *Given a probability measure $\mu \in \mathscr{P}(\mathbf{T}^d)$ and $N \in \mathbf{N}$ the probability density $\mathcal{P}_N(\mu) \in \mathscr{P}(\mathbf{T}_N^d)$ is defined as*

$$\mathcal{P}_N(\mu)(\mathbf{a}) := N^d \mu(Q_\mathbf{a}^N) \ .$$

Similarly, given a continuous momentum vector field $V = (V_1, \ldots, V_d) : \mathbf{T}^d \to \mathbf{R}^d$ we define $\mathcal{P}_N(V) : \mathscr{R}^N \to \mathbf{R}$ by

$$\mathcal{P}_N(V)(R) := 2dN^d \int_R V_i(x) \, \mathrm{d}x \ , \qquad R = R_{\mathbf{a},i\pm}^N \in \mathscr{R}^N \ .$$

Probability densities on $\mathbf{T}^d$ are defined by piecewise constant extensions of densities on $\mathbf{T}_N^d$, and vector fields on $\mathbf{T}^d$ are defined by linear interpolation.

Definition 3.2 (From $\mathbf{T}_N^d$ to $\mathbf{T}^d$). *Given a probability density $\rho^N \in \mathscr{P}(\mathbf{T}_N^d)$ and a momentum vector field $V^N : \mathscr{R}^N \to \mathbf{R}^d$, the probability measure $\mathcal{Q}_N(\rho^N)\pi \in \mathscr{P}(\mathbf{T}^d)$ and the momentum vector field $\mathcal{Q}_N(V^N) : \mathbf{T}^d \to \mathbf{R}^d$ are defined as*

$$\mathcal{Q}_N(\rho^N)(x) := N^{-d} \rho^N(\mathbf{a}) \ ,$$

$$\mathcal{Q}_N(V^N)_i(x) := \frac{1}{2dN} \Big((1 - Nx_i + a_i) V^N(R_{\mathbf{a},i-}^N) + (Nx_i - a_i) V^N(R_{\mathbf{a},i+}^N) \Big) \ ,$$

where $\mathbf{a} = (a_1, \ldots, a_d) \in \mathbf{T}_N^d$ is uniquely determined by the condition $x = (x_1, \ldots, x_d) \in Q_\mathbf{a}^N$.

The maps $\mathcal{P}_N$, $\mathcal{Q}_N$ will be the ones that we use to prove Gromov-Hausdorff convergence. They are constructed in such a way that ensures that solutions of the continuity equation are mapped to solutions of the continuity equation.

Proposition 3.3. *The following assertions hold:*

(1) *Let (ρ_t, V_t) be a solution to the continuity equation (2.3) such that the mapping $x \mapsto V_t(x)$ is continuous for almost every t. Then $(\mathcal{P}_N(\rho_t), \mathcal{P}_N(V_t))$ solves the discrete continuity equation (2.9).*

(2) *Vice versa, let $(\rho_{N,t}, V_{N,t})$ be a solution to the discrete continuity equation (2.9). Then $(\mathcal{Q}_N(\rho_{N,t}), \mathcal{Q}_N(V_{N,t}))$ solves the continuity equation (2.3).*

Proof. These statements are direct consequences of the definitions and the Gauss–Green Theorem. $\square$

It follows from the definitions that $\mathcal{P}_N \circ \mathcal{Q}_N$ is the identity operator on $\mathscr{P}(\mathbf{T}_N^d)$. On the other hand, $\mathcal{Q}_N \circ \mathcal{P}_N$ is a good approximation of the identity in the following sense.

Lemma 3.4. *For all $\mu \in \mathscr{P}(\mathbf{T}^d)$ and all $N \geq 2$ we have*

$$W_2(\mathcal{Q}_N(\mathcal{P}_N(\mu)), \mu) \leq \frac{\sqrt{d}}{N} \ . \tag{3.1}$$

Proof. Since both measures agree on each cube $Q_{\mathbf{a}}^N$, it follows that

$$W_2(\mathcal{Q}_N(\mathcal{P}_N(\mu)), \mu)^2 \leq \sum_{\mathbf{a} \in \mathbf{T}_N^d} \mu(Q_{\mathbf{a}}^N) \operatorname{diam}(Q_{\mathbf{a}}^N)^2 \ .$$

Taking into account that the diameter of each $Q_{\mathbf{a}}^N$ equals $\sqrt{d}/N$, the result follows. $\square$

The following simple result allows us to compare the 2-Wasserstein distances on $\mathscr{P}(\mathbf{T}^d)$ and $\mathscr{P}(\mathbf{T}_N^d)$. Recall that $W_{2,N}$ has been defined in (2.10).

Lemma 3.5. *For all $\mu_0, \mu_1 \in \mathscr{P}(\mathbf{T}^d)$ we have*

$$W_{2,N}(\mathcal{P}_N(\mu_0), \mathcal{P}_N(\mu_1)) \leq \sqrt{2} W_2(\mu_0, \mu_1) + \frac{\sqrt{2d}}{N} \ .$$

Proof. Define $T_N : \mathbf{T}^d \to \mathbf{T}_N^d$ by $T_N(x) := \mathbf{a}$ whenever $x \in Q_{\mathbf{a}}^N$. Since $|(T_N x)_i - (T_N y)_i| \leq 1 + N|x_i - y_i|$ for $x, y \in \mathbf{T}^d$, we have

$$\mathsf{d}_N(T_N x, T_N y) \leq |x - y| + \frac{\sqrt{d}}{N} \ .$$

Using the fact that $\mathcal{P}_N(\mu_i) = (T_N)_{\#} \mu_i$, the result follows. $\square$

In order to carry out our estimates, we will sometimes need some regularity on the probability densities involved. For this reason, we introduce the following set.

Definition 3.6 (Regular densities). *Let $\delta > 0$. Then the set $\mathscr{P}_\delta(\mathbf{T}_N^d) \subset \mathscr{P}(\mathbf{T}_N^d)$ is the set of probability densities $\rho_N \in \mathscr{P}(\mathbf{T}_N^d)$ such that*

$$\min_{\mathbf{a} \in \mathbf{T}_N^d} \rho_N(\mathbf{a}) \geq \delta \ , \qquad \operatorname{Lip}_N(\rho_N) \leq \delta^{-1} \ .$$

Notice that the projections $\mathcal{P}_N$ preserve this sort of regularity, i.e.,

$$\operatorname{Lip}_N(\mathcal{P}_N(\rho)) \leq \operatorname{Lip}(\rho) \ , \qquad \min_{\mathbf{a} \in \mathbf{T}_N^d} \mathcal{P}_N(\rho)(\mathbf{a}) \geq \inf_{x \in \mathbf{T}^d} \rho(x) \ , \tag{3.2}$$

as is readily checked from the definitions.

The set $\mathscr{P}_\delta(\mathbf{T}_N^d)$ is endowed with the following distance, which is obtained by minimizing the action functional over all paths in the space of regular densities.

Definition 3.7 (The distance $\mathcal{W}_{N,\delta}$). *Let $\delta > 0$ and $\rho_{N,0}, \rho_{N,1} \in \mathscr{P}_\delta(\mathbf{T}_N^d)$. The distance $\mathcal{W}_{N,\delta}(\rho_{N,0}, \rho_{N,1})$ is defined as*

$$\left(\mathcal{W}_{N,\delta}(\rho_{N,0}, \rho_{N,1})\right)^2 := \inf \left\{ \int_0^1 \mathcal{A}_N(\rho_{N,t}, V_{N,t}) \, \mathrm{d}t \right\},$$

the infimum being taken among all solutions $(\rho_{N,t}, V_{N,t})$ of the continuity equation (2.9) such that $\rho_{N,t} \in \mathscr{P}_\delta(\mathbf{T}_N^d)$ for any $t \in [0,1]$.

The last tool that we need is a variant of the distance $\mathcal{W}_N$ on $\mathscr{P}(\mathbf{T}_N^d)$, where instead of the logarithmic mean θ one considers the harmonic mean $\tilde{\theta}$ given by

$$\tilde{\theta}(a,b) := \frac{2ab}{a+b}$$

for any $a, b > 0$. If $a = 0$ or $b = 0$, we set $\tilde{\theta}(a,b) = 0$. For $\rho_N \in \mathscr{P}(\mathbf{T}_N^d)$ and $R = R_{\mathbf{a},i+}^N \in \mathscr{R}^N$ we put

$$\tilde{\rho}_N(R) := \tilde{\theta}(\rho_N(\mathbf{a}), \rho_N(\mathbf{a} + \mathbf{e}_i)).$$

Definition 3.8 (The distance $\widetilde{\mathcal{W}}_N$). *For $\rho_{N,0}, \rho_{N,1} \in \mathscr{P}(\mathbf{T}_N^d)$, the metric $\widetilde{\mathcal{W}}_N(\rho_{N,0}, \rho_{N,1})$ is defined as*

$$\left(\widetilde{\mathcal{W}}_N(\rho_{N,0}, \rho_{N,1})\right)^2 := \inf \left\{ \int_0^1 \frac{1}{4d^2 N^{d+2}} \sum_{R \in \mathscr{R}^N} \frac{V_{N,t}(R)^2}{\tilde{\rho}_{N,t}(R)} \, \mathrm{d}t \right\},$$

the infimum being taken among all solutions $(\rho_{N,t}, V_{N,t})$ of the continuity equation (2.9).

Distances of this form have already been introduced in [13]. Notice that since $\tilde{\theta}(a,b) \leq \theta(a,b)$ for any $a, b \geq 0$, it follows immediately that $\widetilde{\mathcal{W}}_N \geq \mathcal{W}_N$.

Let us now describe our strategy to prove Theorem 1.1. We start with two measures $\mu_0, \mu_1 \in \mathscr{P}(\mathbf{T}^d)$, regularize them a bit using the heat flow for a short time $s > 0$, and then show (Proposition 3.10) that for some constant $C(s) < \infty$ (independent on μ_0, μ_1) we have

$$\mathcal{W}_N(\mathcal{P}_N(\mathsf{H}_s(\mu_0)), \mathcal{P}_N(\mathsf{H}_s(\mu_1))) \leq W_2(\mu_0, \mu_1) + \frac{C(s)}{\sqrt{N}}.$$

This will follow quite easily. The converse inequality will be harder to achieve, as the natural inequality that one obtains for $\rho_{N,0}, \rho_{N,1} \in \mathscr{P}(\mathbf{T}_N^d)$ (in Proposition 3.11) involves the harmonic mean rather than the logarithmic mean, i.e., we prove that

$$W_2(\mathcal{Q}_N(\rho_0^N), \mathcal{Q}_N(\rho_1^N)) \leq \widetilde{\mathcal{W}}_N(\rho_0^N, \rho_1^N).$$

Thus the problem becomes to bound $\widetilde{\mathcal{W}}_N$ from above in terms of $\mathcal{W}_N$ plus a small error. Unfortunately, the harmonic-logarithmic mean inequality $\tilde{\theta}(a,b) \leq \theta(a,b)$ goes in the 'wrong' direction, but the elementary inequality

$$\frac{1}{\tilde{\theta}(a,b)} - \frac{1}{\theta(a,b)} \leq \frac{(b-a)^2}{ab} \frac{1}{\tilde{\theta}(a,b)}$$

that we establish in Proposition 3.12, allows us to obtain an estimate for all regular densities, i.e.,

$$\widetilde{\mathcal{W}}_N(\rho_0^N, \rho_1^N) \leq \left(1 - \frac{1}{\delta^4 N^2}\right)^{-\frac{1}{2}} \mathcal{W}_{N,\delta}(\rho_0^N, \rho_1^N).$$

for $\rho_0^N, \rho_1^N \in \mathscr{P}_\delta(\mathbf{T}_N^d)$,

Thus at the end everything reduces to prove that $\mathcal{W}_{N,\delta}$ can be bounded above, up to a small error, by $\mathcal{W}_N$. Clearly, this is false without some additional assumptions on the measures we want to interpolate. The idea is then to notice that the measures on the discrete torus that we produced in our first step, using $\mathcal{P}_N$ after an application of the heat flow, belong to $\mathscr{P}_\delta(\mathbf{T}_N^d)$ for some $\delta > 0$. We then show in Proposition 3.13, which is technically the most involved, that given $\varepsilon, \delta > 0$, there exists $\bar{\delta} > 0$ such that the bound

$$\mathcal{W}_{N,\bar{\delta}}(\rho_{N,0}, \rho_{N,1}) \leq \mathcal{W}_N(\rho_{N,0}, \rho_{N,1}) + \varepsilon$$

holds for any $\rho_{N,0}, \rho_{N,1} \in \mathscr{P}_\delta(\mathbf{T}_N^d)$. This will be enough to complete the argument.

3.2. **Estimates.** Here we collect all the estimates that we need to implement the strategy outlined above. We start by observing the effect of $\mathcal{P}_N$ on the action of vector fields.

Lemma 3.9. *Let $\mu = \rho\pi \in \mathscr{P}(\mathbf{T}^d)$ be a probability measure and $V : \mathbf{T}^d \to \mathbf{R}^d$ a momentum vector field. Assume that both ρ and V are Lipschitz and that $\min \rho > 0$. Put $\rho^N := \mathcal{P}_N(\mu)$ and $V^N := \mathcal{P}_N(V)$. Then there exists a universal constant $C > 0$, such that for any $N \geq 1$ we have the bound*

$$\mathcal{A}_N(\rho^N, V^N) \leq \int_{\mathbf{T}^d} \frac{|V(x)|^2}{\rho(x)} \, \mathrm{d}x + \frac{Cd}{N} \left(\frac{\|V\|_{L^\infty} \mathrm{Lip}(V)}{\min \rho} + \frac{\|V\|_{L^\infty}^2 \mathrm{Lip}(\rho)}{(\min \rho)^2} \right). \tag{3.3}$$

Proof. We apply Jensen's inequality to the convex function $(x, y, z) \mapsto \frac{x^2}{\theta(y,z)}$ to obtain for $R = R_{\mathbf{a},i\pm}^N \in \mathscr{R}^N$,

$$\frac{1}{2d^2 N^{d+2}} \frac{V^N(R)^2}{\hat{\rho}^N(R)} = \frac{2}{N^2} \frac{\left(\int_R V_i(r) \, \mathrm{d}r \right)^2}{\theta\left(\int_R \int_0^{1/N} \rho(r - h\mathbf{e}_i) \, \mathrm{d}h \, \mathrm{d}r \,, \, \int_R \int_0^{1/N} \rho(r + h\mathbf{e}_i) \, \mathrm{d}h \, \mathrm{d}r \right)}$$

$$\leq 2 \int_R \int_0^{\frac{1}{N}} \frac{|V_i(r)|^2}{\theta\left(\rho(r - h\mathbf{e}_i) \,, \, \rho(r + h\mathbf{e}_i) \right)} \, \mathrm{d}h \, \mathrm{d}r \tag{3.4}$$

$$= \int_R \int_{-\frac{1}{N}}^{\frac{1}{N}} \frac{|V_i(r)|^2}{\theta\left(\rho(r - h\mathbf{e}_i) \,, \, \rho(r + h\mathbf{e}_i) \right)} \, \mathrm{d}h \, \mathrm{d}r \,.$$

Using the elementary fact that for $x, \tilde{x} \in \mathbf{R}$ and $y \geq \tilde{y} > 0$,

$$\left| \frac{x^2}{y} - \frac{\tilde{x}^2}{\tilde{y}} \right| \leq \frac{|x + \tilde{x}|}{\tilde{y}} |x - \tilde{x}| + \frac{x^2}{\tilde{y}^2} |y - \tilde{y}| \,,$$

we obtain for $r \in R$ and $|h| \leq \frac{1}{N}$,

$$\left| \frac{|V_i(r)|^2}{\theta\left(\rho(r - h\mathbf{e}_i) \,, \, \rho(r + h\mathbf{e}_i) \right)} - \frac{|V_i(r + h\mathbf{e}_i)|^2}{\rho(r + h\mathbf{e}_i)} \right| \leq \frac{C}{N} \left(\frac{\|V\|_{L^\infty} \mathrm{Lip}(V)}{\min \rho} + \frac{\|V\|_{L^\infty}^2 \mathrm{Lip}(\rho)}{(\min \rho)^2} \right),$$

for some universal constant $C > 0$. Combining this bound with (3.4), and summing over all $R \in \mathscr{R}^N$ the result follows. $\square$

The previous result can be used to obtain the following lower bound for the Wasserstein metric W_2.

Proposition 3.10. *Let $s > 0$. There exists a dimensional constant $C(s) < \infty$ such that for all probability measures $\mu_0, \mu_1 \in \mathscr{P}(\mathbf{T}^d)$ and for all $N \geq 1$ we have*

$$\mathcal{W}_N(\mathcal{P}_N(\mathsf{H}_s(\mu_0)), \mathcal{P}_N(\mathsf{H}_s(\mu_1))) \leq W_2(\mu_0, \mu_1) + \frac{C(s)}{\sqrt{N}} \ .$$

Proof. Let (μ_t) be a constant speed geodesic connecting μ_0 to μ_1 in $(\mathscr{P}(\mathbf{T}^d), W_2)$, and let (v_t) denote the corresponding velocity vector field achieving the minimum in (2.1). For $s > 0$, let $\rho_{s,t}$ and $V_{s,t}$ be the densities with respect to π of $\mathsf{H}_s(\mu_t)$ and $\mathsf{H}_s(v_t \mu_t)$ respectively. According to (iv) of Proposition 2.9, for given $s > 0$, the curve $t \mapsto (\rho_{s,t}, V_{s,t})$ is a solution to the continuity equation (2.3) and we have

$$\int_0^1 \int_{\mathbf{T}^d} \frac{|V_{s,t}(x)|^2}{\rho_{s,t}(x)} \, \mathrm{d}t \, \mathrm{d}x \leq W_2^2(\rho_0, \rho_1) \ . \tag{3.5}$$

By (i) and (ii) of Proposition 2.9 we also know that there exists constants $c(s) > 0$ and $C(s) < \infty$ such that for all $t \in [0,1]$,

$$\inf_{x \in \mathbf{T}^d} \rho_{s,t}(x) \geq c(s) \ , \quad \mathrm{Lip}(\rho_{s,t}) \leq C(s) \ , \quad \|V_{s,t}\|_{L^\infty} + \mathrm{Lip}(V_{s,t}) \leq C(s)\|V_{s/2,t}\|_{L^1} \ . \tag{3.6}$$

Set $t \mapsto \eta_{N,t} := \mathcal{P}_N(\mathsf{H}_s(\mu_t))$ and $t \mapsto W_{N,t} := \mathcal{P}_N(V_{s,t})$. By Proposition 3.3 the curve $(\eta_{N,t}, W_{N,t})$ solves the continuity equation (2.9). Applying Lemma 3.9, (3.6) and (3.5), we obtain for some (different) dimensional constant $C(s) < \infty$,

$$\mathcal{W}_N(\mathcal{P}_N(\mathsf{H}_s(\mu_0)), \mathcal{P}_N(\mathsf{H}_s(\mu_1)))^2$$

$$\leq \int_0^1 \mathcal{A}_N(\eta_{N,t}, W_{N,t}) \, \mathrm{d}t$$

$$\leq \int_0^1 \left[\int_{\mathbf{T}^d} \frac{|V_{s,t}(x)|^2}{\rho_{s,t}(x)} \, \mathrm{d}x + \frac{Cd}{N} \left(\frac{\|V_{s,t}\|_{L^\infty} \mathrm{Lip}(V_{s,t})}{\min \rho_{s,t}} + \frac{\|V_{s,t}\|_{L^\infty}^2 \mathrm{Lip}(\rho_{s,t})}{(\min \rho_{s,t})^2} \right) \right] \mathrm{d}t$$

$$\leq W_2^2(\rho_0, \rho_1) + \frac{C(s)}{N} \int_0^1 \|V_{s/2,t}\|_{L^1}^2 \, \mathrm{d}t \ .$$

Applying the Cauchy-Schwarz inequality in the form

$$\|V_{s/2,t}\|_{L^1}^2 \leq \int_{\mathbf{T}^d} \frac{|V_{s/2,t}(x)|^2}{\rho_{s/2,t}(x)} \, \mathrm{d}x \ ,$$

together with (3.5), we obtain

$$\mathcal{W}_N(\mathcal{P}_N(\mathsf{H}_s(\mu_0)), \mathcal{P}_N(\mathsf{H}_s(\mu_1)))^2 \leq W_2(\rho_0, \rho_1)^2 + \frac{C(s)}{N} \int_0^1 \int_{\mathbf{T}^d} \frac{|V_{s/2,t}(x)|^2}{\rho_{s/2,t}(x)} \, \mathrm{d}x \, \mathrm{d}t$$

$$\leq W_2(\rho_0, \rho_1)^2 + \frac{C(s)}{N} W_2(\rho_0, \rho_1)^2 \ .$$

Taking into account that $(\mathscr{P}(\mathbf{T}^d), W_2)$ has finite diameter, we obtain the the result by taking square roots and using that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$. $\qquad\square$

The next result provides a lower bound for W_2. Recall that $\widetilde{\mathcal{W}}_N$ is defined using the harmonic mean instead of the logarithmic mean.

Proposition 3.11. *Let $N \geq 1$ and $\rho_0^N, \rho_1^N \in \mathscr{P}(\mathbf{T}_N^d)$. Then*

$$W_2(\mathcal{Q}_N(\rho_0^N), \mathcal{Q}_N(\rho_1^N)) \leq \widetilde{\mathcal{W}}_N(\rho_0^N, \rho_1^N) \ . \tag{3.7}$$

Proof. Let $t \mapsto (\rho_t^N, V_t^N)$ be a solution to the continuity equation (2.9). Define $\rho_t := \mathcal{Q}_N(\rho_t^N)$ and $V_t := \mathcal{Q}_N(V_t^N)$. Then, for every $t \in [0,1]$ we have

$$\int_{\mathbf{T}^d} \frac{|V_t(x)|^2}{\rho_t(x)} \, \mathrm{d}x = \sum_{\mathbf{a} \in \mathbf{T}_N^d} \int_{Q_{\mathbf{a}}^N} \frac{|V_t(x)|^2}{\rho_t(x)} \, \mathrm{d}x$$

$$= \frac{1}{N^{d-1}} \sum_{\mathbf{a},i} \frac{1}{\rho_t^N(\mathbf{a})} \int_{\frac{a_i}{N}}^{\frac{a_i+1}{N}} \left| \frac{1 - Nx_i + a_i}{2dN} V_t^N(R_{\mathbf{a},i-}^N) + \frac{Nx_i - a_i}{2dN} V_t^N(R_{\mathbf{a},i+}^N) \right|^2 \mathrm{d}x_i$$

$$= \frac{1}{4d^2 N^{d+2}} \sum_{\mathbf{a},i} \frac{1}{\rho_t^N(\mathbf{a})} \int_0^1 \left| (1-y) V_t^N(R_{\mathbf{a},i-}^N) + y V_t^N(R_{\mathbf{a},i+}^N) \right|^2 \mathrm{d}y$$

$$\leq \frac{1}{4d^2 N^{d+2}} \sum_{\mathbf{a},i} \frac{V^N(R_{\mathbf{a},i-}^N)^2 + V^N(R_{\mathbf{a},i+}^N)^2}{2\rho^N(\mathbf{a})}$$

$$= \frac{1}{4d^2 N^{d+2}} \sum_{\mathbf{a},i} \frac{V^N(R_{\mathbf{a},i+}^N)^2}{2} \left(\frac{1}{\rho^N(\mathbf{a})} + \frac{1}{\rho^N(\mathbf{a} + \mathbf{e}_i)} \right)$$

$$= \frac{1}{4d^2 N^{d+2}} \sum_{\mathbf{a},i} \frac{V^N(R_{\mathbf{a},i+}^N)^2}{\tilde{\rho}^N(R_{\mathbf{a},i+}^N)} \ .$$

Since from Proposition 3.3 we know that $t \mapsto (\rho_t, V_t)$ solves the continuity equation, we obtain

$$W_2^2(\rho_0, \rho_1) \leq \int_0^1 \int_{\mathbf{T}^d} \frac{|V_t(x)|^2}{\rho_t(x)} \, \mathrm{d}x \, \mathrm{d}t \leq \frac{1}{4d^2 N^{d+2}} \sum_{R \in \mathcal{R}^N} \int_0^1 \frac{V_t^N(R)^2}{\tilde{\rho}_t^N(R)} \, \mathrm{d}t \ .$$

Taking the infimum over all the solutions (ρ_t^N, V_t^N) of (2.9) and recalling the Definition 3.8 of $\widetilde{\mathcal{W}}_N$ we get the result. $\qquad \square$

For regular densities, the following result compares the distances defined using the harmonic and the logarithmic means. Note that the reverse inequality $\mathcal{W}_N \leq \widetilde{\mathcal{W}}_N$ follows directly from the harmonic-logarithmic mean inequality.

Proposition 3.12. *Let $\delta > 0$, $N > \delta^{-2}$ and $\rho_0^N \rho_1^N \in \mathscr{P}_\delta(\mathbf{T}_N^d)$. Then the following estimate holds:*

$$\widetilde{\mathcal{W}}_N(\rho_0^N, \rho_1^N) \leq \left(1 - \frac{1}{\delta^4 N^2} \right)^{-\frac{1}{2}} \mathcal{W}_{N,\delta}(\rho_0^N, \rho_1^N) \ . \tag{3.8}$$

Proof. Let $b \geq a > 0$ and, as before, let $\tilde{\theta}(a,b) := \frac{2}{\frac{1}{a} + \frac{1}{b}}$ be the harmonic mean. Set $f(t) = ((1-t)a + tb)^{-1}$ and notice that

$$\frac{1}{\theta(a,b)} = \int_0^1 f(t) \, \mathrm{d}t \ , \qquad \frac{1}{\tilde{\theta}(a,b)} = \frac{1}{2}(f(0) + f(1)) \ .$$

Since f is convex and non-increasing, we obtain

$$\frac{1}{\tilde{\theta}(a,b)} - \frac{1}{\theta(a,b)} = \frac{1}{2} \int_0^1 f(0) - f(t) \, \mathrm{d}t + \frac{1}{2} \int_0^1 f(1) - f(t) \, \mathrm{d}t$$

$$\leq \frac{1}{2} \left(f'(1) - f'(0) \right) = \frac{b-a}{2} \left(\frac{1}{a^2} - \frac{1}{b^2} \right) = \frac{(b-a)^2}{ab} \frac{1}{\tilde{\theta}(a,b)} \ .$$

Therefore, for $\rho^N \in \mathscr{P}_\delta(\mathbf{T}_N^d)$ and $R \in \mathscr{R}^N$ we have

$$\frac{1}{\tilde{\rho}^N(R)} \leq \left(1 - \frac{1}{\delta^4 N^2}\right)^{-1} \frac{1}{\hat{\rho}^N(R)} \ ,$$

and the result follows applying this inequality along a geodesic in $(\mathscr{P}_\delta(\mathbf{T}_N^d), \mathcal{W}_{N,\delta})$ connecting ρ_0^N to ρ_1^N. $\qquad\square$

The final proposition in this subsection shows that regular densities can be connected by a curve consisting of (a bit less) regular densities, for which the action functional is almost optimal.

Proposition 3.13. *Let $\varepsilon, \delta \in (0,1)$. Then there exists $\bar{\delta} > 0$ such that for any $N \geq 4$ and $\rho_{N,0}, \rho_{N,1} \in \mathscr{P}_\delta(\mathbf{T}_N^d)$, we have the bound*

$$\mathcal{W}_{N,\bar{\delta}}(\rho_{N,0}, \rho_{N,1}) \leq \mathcal{W}_N(\rho_{N,0}, \rho_{N,1}) + \varepsilon \ . \tag{3.9}$$

Proof. Let $a, b \in (0, \delta)$ to be fixed later and $t \mapsto (\rho_{N,t}, V_{N,t})$ be a $\mathcal{W}_N$-geodesic connecting $\rho_{N,0}$ to $\rho_{N,1}$. Define the curves $t \mapsto (\rho_{N,t}^1, V_{N,t}^1)$ and $t \mapsto (\rho_{N,t}^2, V_{N,t}^2)$ by

$$\rho_{N,t}^1 := (1-a)\rho_{N,t} + a \ , \qquad\qquad V_{N,t}^1 := (1-a)V_{N,t} \ , \tag{3.10}$$
$$\rho_{N,t}^2 := \mathsf{H}_b^N(\rho_{N,t}^1) \ , \qquad\qquad V_{N,t}^2 := \mathsf{H}_b^N(V_{N,t}^1) \ . \tag{3.11}$$

The latter expression should be interpreted in the sense of (2.11).

Step 1: From $\rho_{N,j}$ to $\rho_{N,j}^1$ for $j = 0, 1$.

For $j = 0, 1$, we define $s \mapsto \eta_{N,s,j}$ as the linear interpolation between $\rho_{N,j}$ and $\rho_{N,j}^1$, i.e.,

$$\eta_{N,s,j}(\mathbf{a}) := (1-s)\rho_{N,j}(\mathbf{a}) + s\rho_{N,j}^1(\mathbf{a}) = \rho_{N,j}(\mathbf{a}) + sa\big(1 - \rho_{N,j}(\mathbf{a})\big) \ .$$

Notice that since $\sum_{\mathbf{a} \in \mathbf{T}_N^d} 1 - \rho_{N,j}(\mathbf{a}) = 0$, it makes sense to define

$$W_{N,s,j}(R_{\mathbf{a},i\pm}^N) := \mp 2adN^2\Big(\Delta_N^{-1}(\mathbf{1} - \rho_{N,j})(\mathbf{a} \pm \mathbf{e}_i) - \Delta_N^{-1}(\mathbf{1} - \rho_{N,j})(\mathbf{a})\Big) \ ,$$

with $\mathbf{1}$ being the density constantly equal to one. A direct computation shows that $s \mapsto (\eta_{N,s,j}, W_{N,s,j})$ is a solution to the continuity equation (2.9). Notice that actually $W_{N,s,j}$ does not depend on s. Taking into account that

$$\eta_{N,s,j}(\mathbf{a}) \geq a \ , \qquad \mathbf{a} \in \mathbf{T}_N^d \ , \ s \in [0,1] \ , \ j = 0, 1 \ , \tag{3.12}$$

recalling the Poincaré inequality (Proposition 2.4), and using the trivial bound

$$\mathcal{E}_N(1 - \rho_{N,j}) \leq d\,(\mathrm{Lip}_N(\rho_{N,j}))^2 \leq d\delta^{-2} \ ,$$

we obtain

$$
\begin{aligned}
\mathcal{A}_N(\eta_{N,s,j}, W_{N,s,j}) &= \frac{1}{4d^2 N^{d+2}} \sum_{\mathbf{a} \in \mathbf{T}_N^d} \sum_{i=1}^d \frac{\left(W_{N,s,j}(R_{\mathbf{a},i+}^N)\right)^2}{\hat{\eta}_{N,s,j}(R_{\mathbf{a},i+}^N)} \\
&\leq \frac{a}{N^{d-2}} \sum_{\mathbf{a} \in \mathbf{T}_N^d} \sum_{i=1}^d \left(\Delta_N^{-1}(\mathbf{1} - \rho_{N,j})(\mathbf{a} + \mathbf{e}_i) - \Delta_N^{-1}(\mathbf{1} - \rho_{N,j})(\mathbf{a})\right)^2 \\
&= a\, \mathcal{E}_N(\Delta_N^{-1}(\mathbf{1} - \rho_{N,j})) \\
&\leq \frac{a}{\kappa} \|\mathbf{1} - \rho_{N,j}\|_{L_N^2}^2 \\
&\leq \frac{a}{\kappa^2} \mathcal{E}_N(\mathbf{1} - \rho_{N,j}) \\
&\leq \frac{ad}{\kappa^2 \delta^2} \, ,
\end{aligned}
\tag{3.13}
$$

where $\kappa := \inf_{N \geq 4} 2N^2(1 - \cos(2\pi/N)) > 0$. Notice also that

$$
\mathrm{Lip}_N(\eta_{N,s,j}) \leq \mathrm{Lip}_N(\rho_{N,j}) \leq \delta^{-1} \, , \qquad s \in [0,1] \, , \; j = 0, 1 \, .
\tag{3.14}
$$

Step 2: From $\rho_{N,j}^1$ to $\rho_{N,j}^2$ for $j = 0, 1$.

For $j = 0, 1$ we interpolate from $\rho_{N,j}^1$ and $\rho_{N,j}^2$ using the heat flow, i.e., we define $s \mapsto (\sigma_{N,s,j}, Z_{N,s,j})$ by

$$
\sigma_{N,s,j}(\mathbf{a}) := \mathsf{H}_{sb}^N(\rho_{N,j}^1) \, ,
$$
$$
Z_{N,s,j}(R_{\mathbf{a},i\pm}^N) := \mp 2bdN^2\left(\sigma_{N,s,j}(\mathbf{a} \pm \mathbf{e}_i) - \sigma_{N,s,j}(\mathbf{a})\right) \, .
$$

We then obtain

$$
\begin{aligned}
&\mathcal{A}_N(\sigma_{N,s,j}, Z_{N,s,j}) \\
&= \frac{1}{4d^2 N^{d+2}} \sum_{\mathbf{a} \in \mathbf{T}_N^d} \sum_{i=1}^d \frac{Z_{N,s,j}(R_{\mathbf{a},i+}^N)^2}{\hat{\sigma}_{N,s,j}(R_{\mathbf{a},i+}^N)} \\
&= \frac{b^2}{N^{d-2}} \sum_{\mathbf{a},i} \frac{\left(\sigma_{N,s,j}(\mathbf{a} + \mathbf{e}_i) - \sigma_{N,s,j}(\mathbf{a})\right)^2}{\hat{\sigma}_{N,s,j}(R_{\mathbf{a},i+}^N)} \\
&= \frac{b^2}{N^{d-2}} \sum_{\mathbf{a},i} \left(\sigma_{N,s,j}(\mathbf{a} + \mathbf{e}_i) - \sigma_{N,s,j}(\mathbf{a})\right)\left(\log(\sigma_{N,s,j}(\mathbf{a} + \mathbf{e}_i)) - \log(\sigma_{N,s,j}(\mathbf{a}))\right) \\
&= b^2 \mathcal{E}_N\left(\sigma_{N,s,j}, \log(\sigma_{N,s,j})\right) \, .
\end{aligned}
$$

In view of Proposition 2.10(i) we obtain by construction,

$$
\sigma_{N,s,j}(\mathbf{a}) \geq a \, , \qquad\qquad \mathbf{a} \in \mathbf{T}_N^d, \; s \in [0,1], \; j = 0, 1 \, ,
\tag{3.15}
$$
$$
\mathrm{Lip}_N(\sigma_{N,s,j}) \leq \mathrm{Lip}_N(\rho_{N,j}^1) \leq \delta^{-1} \, , \qquad s \in [0,1], \; j = 0, 1 \, .
\tag{3.16}
$$

Hence $\mathrm{Lip}_N(\log(\sigma_{N,s,j})) \leq \frac{\mathrm{Lip}_N(\sigma_{N,s,j})}{\min \sigma_{N,s,j}} \leq \delta^{-2}$. Since $|\mathcal{E}_N(f,g)| \leq d \, \mathrm{Lip}_N(f) \, \mathrm{Lip}_N(g)$ we obtain

$$
\mathcal{A}_N(\sigma_{N,s,j}, Z_{N,s,j}) \leq \frac{db^2}{\delta^3} \, .
\tag{3.17}
$$

Step 3: From $\rho_{N,0}^2$ to $\rho_{N,1}^2$.

From the convexity of the function $(x, a, b) \mapsto \frac{x^2}{\theta(a,b)}$ we get

$$\mathcal{A}_N(\rho_{N,t}^1, V_{N,t}^1) \leq (1-a)\mathcal{A}_N(\rho_{N,t}, V_{N,t}) = (1-a)\mathcal{W}_N(\rho_{N,0}, \rho_{N,1})^2 \, ,$$

for any $t \in [0,1]$. Using again the convexity of $(x, a, b) \mapsto \frac{x^2}{\theta(a,b)}$ and the fact that H acts as a convolution semigroup, we also get

$$\mathcal{A}_N(\rho_{N,t}^2, V_{N,t}^2) \leq \mathcal{A}_N(\rho_{N,t}^1, V_{N,t}^1)$$

for any $t \in [0,1]$. Combining these two inequalities and integrating we get

$$\int_0^1 \mathcal{A}_N(\rho_{N,t}^2, V_{N,t}^2) \, \mathrm{d}t \leq \int_0^1 \mathcal{A}_N(\rho_{N,t}^1, V_{N,t}^1) \leq (1-a)\mathcal{W}_N(\rho_{N,0}, \rho_{N,1})^2 \, . \tag{3.18}$$

Since the heat semigroup preserves positivity, we obtain

$$\rho_{N,t}^2(\mathbf{a}) \geq a \, , \qquad \mathbf{a} \in \mathbf{T}_N^d, \ t \in [0,1] \, , \tag{3.19}$$

and by (i) of Proposition 2.10 we have

$$\mathrm{Lip}_N(\rho_{N,t}^2) \leq Cb^{-(d+1)/2} \, , \qquad t \in [0,1] \, , \tag{3.20}$$

for some universal constant $C > 0$.

Step 4: Gluing the pieces.

Let $\ell \in (0, 1/4)$ to be fixed later. We define the curve $t \mapsto (\rho_{N,t}^3, V_{N,t}^3)$ on $[0,1]$ by gluing the pieces together, that is,

$$(\rho_{N,t}^3, V_{N,t}^3) := \begin{cases} (\eta_{N,\frac{t}{\ell},0} & , \ell^{-1}W_{N,\frac{t}{\ell},0}) & t \in [0,\ell] \, , \\ (\sigma_{N,\frac{t-\ell}{\ell},0} & , \ell^{-1}Z_{N,\frac{t-\ell}{\ell},0}) & t \in (\ell, 2\ell) \, , \\ (\rho_{N,\frac{t-2\ell}{1-4\ell}}^2 & , (1-4\ell)^{-1}V_{N,\frac{t-2\ell}{1-4\ell}}^2) & t \in [2\ell, 1-2\ell] \, , \\ (\sigma_{N,\frac{1-\ell-t}{\ell},1} & , \ell^{-1}Z_{N,\frac{1-\ell-t}{\ell},1}) & t \in (1-2\ell, 1-\ell) \, , \\ (\eta_{N,\frac{1-t}{\ell},1} & , \ell^{-1}W_{N,\frac{1-t}{\ell},1}) & t \in [1-\ell, 1] \, . \end{cases}$$

Clearly, $t \mapsto (\rho_{N,t}^3, V_{N,t}^3)$ is a solution to the continuity equation (2.9). From (3.13), (3.17) and (3.18) we get, taking the scaling factors into account,

$$\int_0^1 \mathcal{A}_N(\rho_{N,t}^3, V_{N,t}^3) \leq \frac{2ad}{\ell\kappa^2\delta^2} + \frac{2db^2}{\ell\delta^3} + \frac{1-a}{1-4\ell}\mathcal{W}_N(\rho_{N,0}, \rho_{N,1})^2 \, .$$

It remains to fix the constants $a, b \in (0, \delta)$ and $\ell \in (0, 1/4)$ as functions of δ and ε. From (ii) of Proposition 2.8 we know that the diameter of $(\mathscr{P}(\mathbf{T}_N^d), \mathcal{W}_N)$ is bounded by a constant $D > 0$ depending only on d. Choose now $\ell > 0$ so small that $\frac{1}{1-4\ell} \leq 1 + \frac{\varepsilon^2}{3D^2}$, and then $a, b > 0$ so small that

$$\frac{2ad}{\ell\kappa^2\delta^2} \leq \frac{\varepsilon^2}{3} \, , \qquad \frac{2db^2}{\ell\delta^3} \leq \frac{\varepsilon^2}{3} \, .$$

With these choices we get

$$\int_0^1 \mathcal{A}_N(\rho_{N,t}^3, V_{N,t}^3) \leq \varepsilon^2 + \mathcal{W}_N(\rho_{N,0}, \rho_{N,1})^2 \, . \tag{3.21}$$

Furthermore, the inequalities (3.12), (3.15), and (3.19) and the inequalities (3.14), (3.16) and (3.20) imply that

$$\min \rho_{N,t}^3 \geq a \, , \qquad \mathrm{Lip}_N(\rho_{N,t}^3) \leq \max\{\delta^{-1}, Cb^{-(d+1)/2}\} \, ,$$

hence $\rho_{N,t}^3$ belongs to $\mathscr{P}_{\bar{\delta}}(\mathbf{T}_N^d)$ for some $\bar{\delta}$ depending on a, b and δ. The result follows in view of Definition 3.7 of $\mathcal{W}_{N,\bar{\delta}}$.   □

3.3. Wrap up and conclusion of the argument.

Finally we shall prove Theorem 1.1. Let us first recall one of the equivalent characterisations of Gromov-Hausdorff convergence, which we formulate here as a definition. We refer to, e.g., [18, Definition 27.6 and (27.4)]) for more details.

Definition 3.14 (Gromov-Hausdorff Convergence)**.** *We say that a sequence of compact metric spaces $(\mathcal{X}_n, d_n)$ converges in the sense of Gromov-Hausdorff to a compact metric space $(\mathcal{X}, d)$, if there exists a sequence of maps $f_n : \mathcal{X} \to \mathcal{X}_n$ which are*

(i) ε_n*-isometric, i.e., for all $x, y \in \mathcal{X}$,*

$$|d_n(f_n(x), f_n(y)) - d(x, y)| \le \varepsilon_n ; \qquad\qquad and$$

(ii) ε_n*-surjective, i.e., for all $x_n \in \mathcal{X}_n$ there exists $x \in \mathcal{X}$ with*

$$d(f_n(x), x_n) \le \varepsilon_n ,$$

for some sequence $\varepsilon_n \to 0$.

Now we are ready to prove our main result Theorem 1.1, which we restate for the convenience of the reader.

Theorem. *Let $d \ge 1$. Then the metric spaces $(\mathscr{P}(\mathbf{T}_N^d), \mathcal{W}_N)$ converge to $(\mathscr{P}(\mathbf{T}^d), W_2)$ in the sense of Gromov-Hausdorff as $N \to \infty$.*

Proof. For $s > 0$ and $N \ge 1$ we consider the map from $\mathscr{P}(\mathbf{T}^d)$ to $\mathscr{P}(\mathbf{T}_N^d)$ given by

$$\mu \mapsto \mathcal{P}_N(\mathsf{H}_s \mu) .$$

We claim that for each $s > 0$ there exists $\bar{N}(s) \ge 1$ such that for all $N \ge \bar{N}(s)$ this map is both $\varepsilon(s)$-isometric and $\varepsilon(s)$-surjective, for some sequence $\varepsilon(s) \downarrow 0$ as $s \downarrow 0$. This suffices to prove the theorem.

$\varepsilon(s)$**-isometry.** Let $\mu_0, \mu_1 \in \mathscr{P}(\mathbf{T}^d)$. Part (i) of Proposition 2.9 in conjunction with (3.2) yields that $\mathcal{P}_N(\mathsf{H}_s \mu_0)$ and $\mathcal{P}_N(\mathsf{H}_s \mu_1)$ belong to $\mathscr{P}_{\delta(s)}(\mathbf{T}_N^d)$ for some $\delta(s) > 0$ and for any $N \ge 1$. Let $\eta > 0$. From Proposition 3.13 we then get the existence of $\bar{\delta}(\eta, s) > 0$ such that

$$\mathcal{W}_{N,\bar{\delta}(\eta,s)}\big(\mathcal{P}_N(\mathsf{H}_s \mu_0), \mathcal{P}_N(\mathsf{H}_s \mu_1)\big) \le \mathcal{W}_N\big(\mathcal{P}_N(\mathsf{H}_s \mu_0), \mathcal{P}_N(\mathsf{H}_s \mu_1)\big) + \eta .$$

From Proposition 3.12 we infer that

$$\widetilde{\mathcal{W}}_N\big(\mathcal{P}_N(\mathsf{H}_s \mu_0), \mathcal{P}_N(\mathsf{H}_s \mu_1)\big) \le \left(1 - \frac{1}{\bar{\delta}(\eta,s)^4 N^2}\right)^{-\frac{1}{2}} \mathcal{W}_{N,\bar{\delta}(\eta,s)}\big(\mathcal{P}_N(\mathsf{H}_s \mu_0), \mathcal{P}_N(\mathsf{H}_s \mu_1)\big) ,$$

and then from Proposition 3.11 that

$$W_2\big(\mathcal{Q}_N(\mathcal{P}_N(\mathsf{H}_s \mu_0)), \mathcal{Q}_N(\mathcal{P}_N(\mathsf{H}_s \mu_1))\big) \le \widetilde{\mathcal{W}}_N\big(\mathcal{P}_N(\mathsf{H}_s \mu_0), \mathcal{P}_N(\mathsf{H}_s \mu_1)\big) .$$

Lemma 3.4 and Proposition 2.9(i) yield

$$W_2(\mu_0, \mu_1) \le W_2\big(\mathcal{Q}_N(\mathcal{P}_N(\mathsf{H}_s \mu_0)), \mathcal{Q}_N(\mathcal{P}_N(\mathsf{H}_s \mu_1))\big) + 2C\sqrt{s} + 2\frac{\sqrt{d}}{N} .$$

Combining these four inequalities, we obtain

$$W_2(\mu_0, \mu_1) \le \left(1 - \frac{1}{\bar{\delta}(\eta,s)^4 N^2}\right)^{-\frac{1}{2}} \left(\mathcal{W}_N\big(\mathcal{P}_N(\mathsf{H}_s \mu_0), \mathcal{P}_N(\mathsf{H}_s \mu_1)\big) + \eta\right) + 2C\sqrt{s} + 2\frac{\sqrt{d}}{N} .$$

On the other hand, Proposition 3.10 grants that

$$\mathcal{W}_N\big(\mathcal{P}_N(\mathsf{H}_s\mu_0), \mathcal{P}_N(\mathsf{H}_s\mu_1)\big) \le W_2(\mu_0, \mu_1) + \frac{C(s)}{\sqrt{N}}\ .$$

Taking Proposition 2.8(ii) into account, the latter two inequalities yield that for $\bar{N} = \bar{N}(s)$ sufficiently large and $\eta = \eta(s)$ sufficiently small, we have for all $N \ge \bar{N}(s)$,

$$\left| W_2(\mu_0, \mu_1) - \mathcal{W}_N\big(\mathcal{P}_N(\mathsf{H}_s\mu_0), \mathcal{P}_N(\mathsf{H}_s\mu_1)\big)\right| \le \varepsilon(s)$$

for some $\varepsilon(s) \downarrow 0$ as $s \downarrow 0$.

$\varepsilon(s)$-**surjectivity**. Let $\rho^N \in \mathscr{P}(\mathbf{T}_N^d)$ and set $\rho_s^N := \mathsf{H}_s\mathcal{Q}_N(\rho^N)$. Then, for some dimensional constant $C < \infty$ which may change from line to line, we obtain using Proposition 2.8(ii), Lemma 3.5, and Proposition 2.9(i),

$$\begin{aligned}
\mathcal{W}_N\big(\rho^N, \mathcal{P}_N(\rho_s^N)\big) &= \mathcal{W}_N\big(\mathcal{P}_N(\mathcal{Q}_N(\rho^N)), \mathcal{P}_N(\rho_s^N)\big) \\
&\le C W_{2,N}\big(\mathcal{P}_N(\mathcal{Q}_N(\rho^N)), \mathcal{P}_N(\rho_s^N)\big) \\
&\le C W_2\big(\mathcal{Q}_N(\rho^N), \rho_s^N\big) + \frac{C}{N} \\
&\le C\Big(\sqrt{s} + \frac{1}{N}\Big)\ .
\end{aligned}$$

Taking, say, $N = 1/\sqrt{s}$, we infer that $\mathcal{P}_N \circ \mathsf{H}_s$ is $2C\sqrt{s}$-surjective, which completes the proof. $\qquad\square$

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[2] L. Ambrosio, N. Gigli, and G. Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. *Preprint at arXiv:1109.0222*, 2011.

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

[4] E. Carlen and J. Maas. An analog of the 2-Wasserstein metric in non-commutative probability under which the fermionic Fokker-Planck equation is gradient flow for the entropy. *arXiv:1203.5377*, 2012.

[5] J. A. Carrillo, R. J. McCann, and C. Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamericana*, 19(3):971–1018, 2003.

[6] S.-N. Chow, W. Huang, Y. Li, and H. Zhou. Fokker-Planck equations for a free energy functional or Markov process on a graph. *Arch. Rational Mech. Anal.*, 203(3):969–1008, 2012.

[7] M. Colombo and M. Gobbino. Passing to the limit in maximal slope curves: from a regularized Perona-Malik equation to the total variation flow. *Preprint at arXiv:1103.3365*, 2011.

[8] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 1996.

[9] M. Erbar. The heat equation on manifolds as a gradient flow in the Wasserstein space. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(1):1–23, 2010.

[10] M. Erbar and J. Maas. Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal., to appear. arXiv:1111.2687*, 2012+.

[11] N. Gigli. On the heat flow on metric measure spaces: existence, uniqueness and stability. *Calc. Var. Partial Differential Equations*, 39(1-2):101–120, 2010.

[12] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[13] J. Maas. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.*, 261(8):2250–2292, 2011.

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

[15] A. Mielke. Dissipative quantum mechanics using GENERIC. *Preprint*, 2012.

[16] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[17] S. Serfaty. Gamma-convergence of gradient flows on Hilbert and metric spaces and applications. *Discrete Contin. Dyn. Syst.*, 31(4):1427–1451, 2011.

[18] C. Villani. *Optimal transport, Old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2009.

Université de Nice-Sophia Antipolis, Labo. J.-A. Dieudonné, UMR 6621, Parc Valrose, 06108 Nice Cedex 02, France
E-mail address: gigli@unice.fr

Institute for Applied Mathematics, University of Bonn, Endenicher Allee 60, 53115 Bonn, Germany
E-mail address: maas@uni-bonn.de