Scaling limits of discrete optimal transport

Peter Gladbach[*]   Eva Kopfer[†]   Jan Maas[‡]

Abstract

We consider dynamical transport metrics for probability measures on discretisations of a bounded convex domain in $\mathbb{R}^d$. These metrics are natural discrete counterparts to the Kantorovich metric $\mathbb{W}_2$, defined using a Benamou–Brenier type formula. Under mild assumptions we prove an asymptotic upper bound for the discrete transport metric $\mathcal{W}_{\mathcal{T}}$ in terms of $\mathbb{W}_2$, as the size of the mesh $\mathcal{T}$ tends to 0. However, we show that the corresponding lower bound may fail in general, even on certain one-dimensional and symmetric two-dimensional meshes. In addition, we show that the asymptotic lower bound holds under an isotropy assumption on the mesh, which turns out to be essentially necessary. This assumption is satisfied, e.g., for tilings by convex regular polygons, and it implies Gromov–Hausdorff convergence of the transport metric.

Contents

[*]Mathematisches Institut, Universität Leipzig, Augustusplatz 10, 04109 Leipzig, Germany (`gladbach@math.uni-leipzig.de`)

[†]Institut für Angewandte Mathematik, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany (`eva.kopfer@iam.uni-bonn.de`)

[‡]Institute of Science and Technology Austria (IST Austria), Am Campus 1, 3400 Klosterneuburg, Austria (`jan.maas@ist.ac.at`)

1 Introduction

Over the last decades, optimal transport has become a vibrant research area at the interface of analysis, probability, and geometry. A central object is the 2-Kantorovich distance $\mathbb{W}_2$ (often called 2-Wasserstein metric), which plays a major role in non-smooth geometry and analysis, and in the theory of dissipative PDE. We refer to the monographs [1, 30, 32] for an overview of the theory and its applications.

More recently, discrete dynamical transport metrics have been introduced in the context of Markov chains [24], reaction-diffusion systems [26] and discrete Fokker–Planck equations [7]. These metrics are natural discrete counterparts to $\mathbb{W}_2$ in several ways: they have been used to obtain a gradient flow formulation for discrete evolution equations [12,25], and to develop a discrete theory of Ricci curvature that leads to various functional inequalities for discrete systems [11, 16, 27]. The geometry of geodesics for these metrics is currently actively studied, both from an analytic point of view [13, 18], and through numerical methods [14, 31]; see also [8, 23] for further recent developments involving discrete optimal transport.

It is natural to ask whether the discrete transport metrics converge to $\mathbb{W}_2$ under suitable assumptions. The first result of this type has been obtained in [20]. The authors approximated the continuous torus $\mathbb{T}^d$ by the discrete torus $\mathbb{T}_N^d = (\mathbb{Z}/N\mathbb{Z})^d$, and endowed the space of probability measures $\mathcal{P}(\mathbb{T}_N^d)$ with the discrete transport metric $\mathcal{W}_N$. The main result in [20] asserts that, under a natural rescaling, the metric spaces $(\mathcal{P}(\mathbb{T}_N^d), \mathcal{W}_N)$ converge to the L^2-Kantorovich space $(\mathcal{P}(\mathbb{T}^d), \mathbb{W}_2)$ in the sense of Gromov–Hausdorff as $N \to \infty$.

A different convergence result was subsequently obtained by Garcia Trillos [19]. Given a set $\mathcal{X}_N$ consisting of N distinct points in $\mathbb{T}^d$, Garcia Trillos considers the graph obtained by connecting all pairs of points that lie at distance less than ε, for a suitable ε depending on N. Under appropriate conditions on the uniformity of the point set, it is shown in [19] that the discrete transport metric converges to $\mathbb{W}_2$, provided that $\varepsilon = \varepsilon(N)$ decays sufficiently slow. While the result of [19] covers a wide range of settings, the latter assumption typically implies that the number of neighbours of a point in $\mathcal{X}_N$ tends to ∞ as $N \to \infty$; in particular, the result of [20] is not contained in [19].

The aim of this paper is to investigate Gromov–Hausdorff convergence for transport metrics on general finite volume discretisations of a bounded convex domain $\Omega \subseteq \mathbb{R}^d$. While our setting is different from [19], it corresponds in terms of scaling to the limiting regime in which the results of [19] fail to apply.

Setting of the paper

We informally present the main results of this paper. For precise definitions we refer to Section 2.

Let $\Omega \subseteq \mathbb{R}^d$ be a bounded convex open set. We endow the set of Borel probability measures $\mathcal{P}(\Omega)$ with the 2-Kantorovich metric, which can be expressed in terms of the *Benamou–Brenier formula*

$$\mathbb{W}_2(\mu_0, \mu_1) = \inf \left\{ \sqrt{\int_0^1 \mathbb{A}^*(\mu_t, \dot{\mu}_t) \, dt} \right\} ,$$

for $\mu_0, \mu_1 \in \mathcal{P}(\overline{\Omega})$. Here, the infimum is taken among all absolutely continuous curves $(\mu_t)_{t\in[0,1]}$ in $\mathcal{P}(\overline{\Omega})$ connecting μ_0 and μ_1, and

$$\mathbb{A}^*(\mu, w) = \sup_{\phi \in \mathcal{C}^1(\overline{\Omega})} \left\{ 2\langle \phi, w \rangle - \mathbb{A}(\mu, \phi) \right\}, \qquad \mathbb{A}(\mu, \phi) = \int_{\overline{\Omega}} |\nabla \phi|^2 \, \mathrm{d}\mu . \qquad (1.1)$$

We discretise the domain Ω using a finite volume discretisation, closely following the setup from [17]. An *admissible mesh* consists of a partition $\mathcal{T}$ of $\overline{\Omega}$ into sets K with non-empty and convex interior, together with a family of distinct points $\{x_K\}_{K\in\mathcal{T}}$ such that $x_K \in \overline{K}$ for all $K \in \mathcal{T}$. We write $(K|L) = \overline{K} \cap \overline{L}$ to denote the flat convex surface with $(d-1)$-dimensional Hausdorff measure $|(K|L)|$. We make the geometric assumption that the vector $x_K - x_L$ is orthogonal to $(K|L)$ if K and L are neighbouring cells and we write $d_{KL} := |x_K - x_L|$. In addition, we impose some mild regularity conditions on the mesh; see Definition 2.11 for the notion of ζ-regularity that is imposed in the sequel. We write $[\mathcal{T}] := \max_{K\in\mathcal{T}} \mathrm{diam}(K)$ to denote the mesh size of $\mathcal{T}$.

The discrete transport metric on $\mathcal{P}(\mathcal{T})$ is defined in terms of a discrete Benamou–Brenier formula: for $m_0, m_1 \in \mathcal{P}(\mathcal{T})$, we set

$$\mathcal{W}_{\mathcal{T}}(m_0, m_1) = \inf \left\{ \sqrt{\int_0^1 \mathcal{A}_{\mathcal{T}}^*(m_t, \dot{m}_t) \, \mathrm{d}t} \right\},$$

where the action functionals are defined using natural discrete counterparts to (1.1):

$$\mathcal{A}_{\mathcal{T}}^*(m, \sigma) = \sup_{\psi : \mathcal{T} \to \mathbb{R}} \left\{ 2\langle \psi, \sigma \rangle - \mathcal{A}_{\mathcal{T}}(m, \psi) \right\},$$

$$\mathcal{A}_{\mathcal{T}}(m, \psi) = \frac{1}{2} \sum_{K,L\in\mathcal{T}} S_{KL} \theta_{KL} \left(\frac{m(K)}{|K|}, \frac{m(L)}{|L|} \right) \big(\psi(K) - \psi(L)\big)^2 .$$

Here, the transmission coefficients S_{KL} are defined by $S_{KL} = \frac{|(K|L)|}{|x_K - x_L|}$. This choice ensures the formal consistency of the discrete and the continuous definitions; cf. Remark 2.14 below for a verification at the level of the associated Dirichlet forms. We refer to [17, Theorem 4.2] for a convergence result for the discrete heat equation to the continuous heat equation.

The functions $\theta_{KL} : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ are *admissible means*, i.e., θ_{KL} is a continuous function that is $\mathcal{C}^\infty$ on $(0, \infty) \times (0, \infty)$, positively 1-homogeneous, jointly concave, and normalised (i.e., $\theta(1, 1) = 1$); see Definition 2.8 below for further details. Furthermore, we impose the symmetry condition $\theta_{KL}(s, t) = \theta_{LK}(t, s)$. A common choice is the logarithmic mean $\theta_{KL}(s, t) := \int_0^1 s^{1-p} t^p \, \mathrm{d}p$ for all $K \sim L$, which naturally arises in the gradient flow structure of the discrete heat equation [7, 24, 26]. We write $\boldsymbol{\theta} = (\theta_{KL})$ to denote the collection of mean functions in the definition of $\mathcal{W}_{\mathcal{T}}$, and suppress the dependence of $\mathcal{W}_{\mathcal{T}}$ on $\boldsymbol{\theta}$ in the notation. The freedom to choose these mean functions is due to the discreteness of the problem. We will see that a careful choice is crucial in the sequel.

Statement of the main results

The goal of this paper is to analyse the limiting behaviour of the discrete transport metrics $\mathcal{W}_{\mathcal{T}}$ as $[\mathcal{T}] \to 0$. To formulate the main results we introduce the canonical projection operator $P_{\mathcal{T}} : \mathcal{P}(\overline{\Omega}) \to \mathcal{P}(\mathcal{T})$ given by

$$\big(P_{\mathcal{T}}\mu\big)(K) = \mu(K) \qquad \text{for } \mu \in \mathcal{P}(\overline{\Omega}) \text{ and } K \in \mathcal{T} . \qquad (1.2)$$

3

Our first main result establishes a one-sided asymptotic estimate for the discrete transport metric in great generality.

Theorem 1.1 (Asymptotic upper bound for $\mathcal{W}_{\mathcal{T}}$)**.** *Fix $\zeta \in (0,1]$, and let $\mu_0, \mu_1 \in \mathcal{P}(\overline{\Omega})$. For any family of ζ-regular meshes $\{\mathcal{T}\}$, and for any choice of mean functions $\boldsymbol{\theta}_{\mathcal{T}}$, we have*

$$\limsup_{[\mathcal{T}] \to 0} \mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_0, P_{\mathcal{T}}\mu_1) \leq \mathbb{W}_2(\mu_0, \mu_1) \ .$$

In view of the Gromov–Hausdorff convergence results from [20] and [19], one might expect that a corresponding asymptotic lower bound for $\mathcal{W}_{\mathcal{T}}$ in terms of $\mathbb{W}_2$ holds as well. However, convergence can fail, even in one dimension, as the following example shows.

Example (A; a one-dimensional period mesh). For $N \in \mathbb{N}$ and $r \in (0, \frac{1}{2})$, we consider a one-dimensional discretisation $\mathcal{T}_{r,N}$ of the unit interval $[0,1]$, obtained by alternatingly concatenating intervals of length $\frac{r}{N}$ and $\frac{1-r}{N}$.
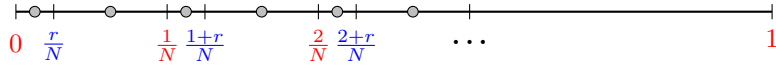

Figure 1: The mesh $\mathcal{T}_{r,N}$ on the interval $[0,1]$.

The next result shows that $\mathcal{W}_{\mathcal{T}}$ can *not* be bounded from below by $\mathbb{W}_2$ as $[\mathcal{T}] \to 0$.

Proposition 1.2 (Counterexample to the lower bound for $\mathcal{W}_{\mathcal{T}}$)**.** *Let $\Omega = (0,1)$ and let $\mathcal{T}_{r,N}$ be as in Example (A) above. Fix an admissible symmetric mean θ, and consider the transport metric $\mathcal{W}_{\mathcal{T}}$ defined by setting $\theta_{KL} = \theta$ for all $K, L \in \mathcal{T}$. Then there exist probability measures $\mu_0, \mu_1 \in \mathcal{P}(\overline{\Omega})$ such that, for each fixed $r \in (0, \frac{1}{2})$,*

$$\limsup_{N \to \infty} \mathcal{W}_{\mathcal{T}_{r,N}}(P_{\mathcal{T}_{r,N}}\mu_0, P_{\mathcal{T}_{r,N}}\mu_1) < \mathbb{W}_2(\mu_0, \mu_1) \ .$$

We stress that the discrete heat flow converges to the continuous heat flow in the setting of this counterexample.

The idea behind the one-dimensional counterexample is that an "unreasonably cheap" discrete transport can be constructed by introducing microscopic oscillations in the discrete density in such a way that most of the mass is assigned to small cells. We refer to Section 5 for more details.

In view of Proposition 1.2 it is natural to look for additional geometric assumptions on the mesh under which an asymptotic lower bound for $\mathcal{W}_{\mathcal{T}}$ in terms of $\mathbb{W}_2$ can be obtained. A *weight function* on $\mathcal{T}$ is a mapping $\boldsymbol{\lambda} : \mathcal{T} \times \mathcal{T} \to [0,1]$ satisfying $\lambda_{KL} + \lambda_{LK} = 1$ for all $K, L \in \mathcal{T}$. The following definition plays a central role in our investigations.

Definition 1.3 (Asymptotic isotropy)**.** *A family of admissible meshes $\{\mathcal{T}\}$ is said to be asymptotically isotropic with weight functions $\{\boldsymbol{\lambda}^{\mathcal{T}}\}$ if, for all $K \in \mathcal{T}$,*

$$\sum_{L} \lambda_{KL}^{\mathcal{T}} \frac{|(K|L)|}{d_{KL}} (x_K - x_L) \otimes (x_K - x_L) \leq |K|\big(1 + \eta_{\mathcal{T}}(K)\big) I_d \ , \tag{1.3}$$

where $\sup_{K \in \mathcal{T}} |\eta_{\mathcal{T}}(K)| \to 0$ as $[\mathcal{T}] \to 0$.

4

The asymptotic isotropy condition puts a strong geometric constraint on a family of meshes, although it will be shown in Section 5 that isotropy always holds on average on a macroscopic scale.

Remark (Centre-of-mass condition). In examples it is often convenient to verify the asymptotic isotropy condition by checking the following stronger condition: we say that $\mathcal{T}$ satifies the *centre-of-mass condition* with weight function $\boldsymbol{\lambda}$ if

$$\fint_{(K|L)} x \, \mathrm{d}S = \lambda_{LK} x_K + \lambda_{KL} x_L$$

for any pair of neighbouring cells $K, L \in \mathcal{T}$. As $\lambda_{KL} + \lambda_{LK} = 1$, this condition asserts that the centre of mass of the interface $(K|L)$ lies on the line segment connecting x_K and x_L. In the literature on finite volume methods this property is known as *superadmissibility* of the mesh; see [15, Lemma 2.1].

For all interior cells $K \in \mathcal{T}$, we claim that the centre-of-mass condition yields the asymptotic isotropy condition (1.3) with equality and $\eta_{\mathcal{T}}(K) = 0$. To see this, let $n = n(x)$ be the outward unit normal on ∂K, and note that

$$\int_{\partial K} (x - x_K) \otimes n \, \mathrm{d}S = |K| I_d \, , \tag{1.4}$$

as can be shown by applying Gauss's Theorem to the vector fields $\Phi^{ij} : \mathbb{R}^d \to \mathbb{R}^d$ given by $\Phi^{ij}(x) := \langle x - x_K, e_j \rangle e_i$ for $1 \leq i, j \leq d$. For all $L \sim K$, the centre-of-mass condition yields

$$\begin{aligned}
\int_{(K|L)} (x - x_K) \otimes n \, \mathrm{d}S &= \int_{(K|L)} \lambda_{KL}(x_L - x_K) \otimes n \, \mathrm{d}S \\
&= \lambda_{KL} \frac{|(K|L)|}{d_{KL}} (x_K - x_L) \otimes (x_K - x_L) \, .
\end{aligned} \tag{1.5}$$

The claim follows by summation over L.

Note that for boundary cells, the centre-of-mass condition does not imply asymptotic isotropy in general. If Ω is polygonal and $\mathcal{T}$ can be extended to a global mesh satisfying the centre-of-mass condition, then our claim holds also for boundary cells by positive semi-definiteness of (1.5). This is the case in several examples below.

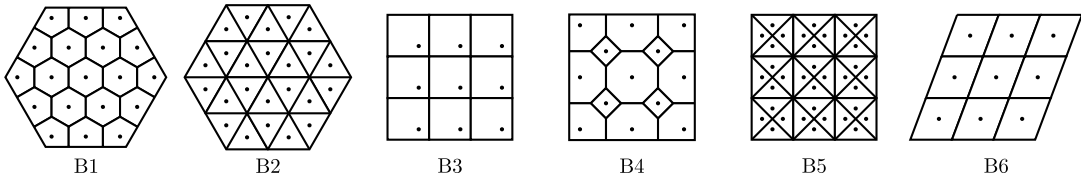Figure 2: Five admissible meshes (B1)–(B5) and a non-admissible mesh (B6).

Example (A; revisited). Clearly, the centre-of-mass condition holds in every one-dimensional mesh for an appropriate weight function $\boldsymbol{\lambda}$. For the one-dimensional periodic lattice $\mathcal{T}_{r,N}$ in Example (A), it is immediately checked that $\lambda_{KL} = r$ if K is small, and $\lambda_{KL} = 1 - r$ if K is large.

Example (B; Two dimensional meshes). The centre-of-mass condition holds for the regular hexagonal lattice (B1) and the equilateral triangular lattice (B2) in dimension 2, if the points x_K are placed at the centre of mass of the cells. In these examples we have $\lambda_{KL} = \frac{1}{2}$ for all K and L. The hexagonal lattice (B1) is truncated in such a way that all interior interface have equal size. More generally, it is immediate to see that the centre-of-mass condition holds for any tiling of the plane by convex regular polygons; cf. [22, Chapter 2] for many examples.

The centre-of-mass condition is clearly satisfied for rectangular grids in any dimension, if the points x_K are placed at the centre of the cells. The weights λ_{KL} will depend on the size of the rectangles. It is possible to put the points x_K at different positions, as is done in (B3). In that case, the centre-of-mass condition is violated, but the isotropy condition holds.

Another example for which the centre-of-mass condition holds is shown in (B4). The value of the weights λ_{KL} is determined by the length ratio of the edges.

The centre-of-mass condition fails for the lattice in (B5). Indeed, to satisfy this condition, the points x_K would have to be placed at the boundary of the cells, in a way that violates our assumption that the points x_K are all distinct. This would lead to infinite transmission coefficients S_{KL}. The isotropy condition fails to hold as well, as will be discussed in Section 5.

In each of the examples (B1)–(B4) it is readily checked that the isotropy condition also holds for boundary cells, by introducing suitable fake points outside of the domain.

The mesh in (B6) is not admissible, as the line segments connecting the points x_K are not orthogonal to the cell interfaces.

Our next main result provides an asymptotic lower bound for $\mathcal{W}_{\mathcal{T}}$ in terms of $\mathbb{W}_2$ under the assumption that the meshes $\{\mathcal{T}\}$ satisfy the asymptotic isotropy condition, and the means $\theta^{\mathcal{T}}_{KL}$ are carefully chosen to reflect this condition.

A mean function (θ_{KL}) is said to be *compatible* with a weight function (λ_{KL}) if, for any $K, L \in \mathcal{T}$ and all $a, b \geq 0$, we have

$$\theta_{KL}(a, b) \leq \lambda_{KL} a + \lambda_{LK} b \ ,$$

or equivalently, $\partial_1 \theta_{KL}(1, 1) = \lambda_{KL}$ for any K, L; cf. Section 2.2 for a more extensive discussion.

Remark. In the special case that $\lambda_{KL} = \frac{1}{2}$, the compatibility condition holds for any admissible mean that is symmetric (i.e., $\theta_{KL}(s, t) = \theta_{KL}(t, s)$ for all K, L and $s, t \geq 0$).

Theorem 1.4 (Asymptotic lower bound for $\mathcal{W}_{\mathcal{T}}$). *Fix $\zeta \in (0, 1]$, and let $\mu_0, \mu_1 \in \mathcal{P}(\overline{\Omega})$. Let $\{\mathcal{T}\}$ be a family of ζ-regular meshes satisfying the asymptotic isotropy condition with weights $(\lambda^{\mathcal{T}}_{KL})_{K,L \in \mathcal{T}}$, and let $(\theta^{\mathcal{T}}_{KL})_{K,L \in \mathcal{T}}$ be admissible means that are compatible with $(\lambda^{\mathcal{T}}_{KL})_{K,L \in \mathcal{T}}$. Then:*

$$\mathbb{W}_2(\mu_0, \mu_1) \leq \liminf_{[\mathcal{T}] \to 0} \mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}} \mu_0, P_{\mathcal{T}} \mu_1) \ .$$

Remark. As discussed above, the assumptions of the theorem are satisfied in the Examples (A) and (B1)–(B4).

In Section 6 we will prove slightly stronger versions of Theorems 1.1 and 1.4, that provide uniform error bounds in terms of μ_0 and μ_1. As a consequence, we obtain the following result.

Corollary 1.5 (Gromov–Hausdorff convergence of $\mathcal{W}_{\mathcal{T}}$). *Under the conditions of Theorem 1.4, we have convergence of metric space in the sense of Gromov–Hausdorff:*

$$(\mathcal{P}(\mathcal{T}), \mathcal{W}_{\mathcal{T}}) \to (\mathcal{P}(\overline{\Omega}), \mathbb{W}_2) \quad as \; [\mathcal{T}] \to 0 \; .$$

Another consequence is the following result on the behaviour of $\mathcal{W}_{\mathcal{T}}$-geodesics. Let $Q_{\mathcal{T}} : \mathcal{P}(\mathcal{T}) \to \mathcal{P}(\overline{\Omega})$ be the natural embedding defined in (3.1) below.

Corollary 1.6 (Convergence of geodesics). *Under the conditions of Theorem 1.4, let $\mu_i \in \mathcal{P}(\overline{\Omega})$ and $m_i^{\mathcal{T}} \in \mathcal{P}(\mathcal{T})$ be such that $Q_{\mathcal{T}} m_i^{\mathcal{T}} \rightharpoonup \mu_i$ as $[\mathcal{T}] \to 0$ for $i = 0, 1$. Then:*

$$\lim_{[\mathcal{T}] \to 0} \mathcal{W}_{\mathcal{T}}(m_0^{\mathcal{T}}, m_1^{\mathcal{T}}) = \mathbb{W}_2(\mu_0, \mu_1) \; .$$

Moreover, if $(m_t^{\mathcal{T}})_{t \in [0,1]}$ is a constant speed geodesic in $(\mathcal{P}(\mathcal{T}), \mathcal{W}_{\mathcal{T}})$ and $Q_{\mathcal{T}} m_t^{\mathcal{T}} \rightharpoonup \mu_t \in \mathcal{P}(\overline{\Omega})$ as $[\mathcal{T}] \to 0$ for every $t \in [0,1]$, then $(\mu_t)_{t \in [0,1]}$ is a constant speed geodesic in $(\mathcal{P}(\overline{\Omega}), \mathbb{W}_2)$.

Finally, we will show that the asymptotic isotropy condition is essentially necessary in Theorem 1.4 and Corollary 1.5. More precisely, the following result shows that the asymptotic lower bound for $\mathcal{W}_{\mathcal{T}}$ fails to hold if the asymptotic isotropy condition is locally violated at all scales. In this sense the asymptotic isotropy condition for $\{\mathcal{T}\}$ is essentially equivalent to Gromov–Hausdorff convergence of $\mathcal{W}_{\mathcal{T}}$ to $\mathbb{W}_2$. The result relies crucially on the smoothness of the mean functions θ_{KL}. In particular, the result does not apply to $\theta_{KL}(a, b) = \min\{a, b\}$.

Theorem 1.7 (Necessity of asymptotic isotropy). *Fix $\zeta \in (0, 1]$, and let $\{\mathcal{T}\}$ be a family of ζ-regular meshes on Ω. For each $\mathcal{T}$, let $\boldsymbol{\theta}^{\mathcal{T}}$ be a mean function on $\mathcal{T}$ satisfying the regularity condition*

$$\sup_{\mathcal{T}} \sup_{K, L \in \mathcal{T}} \|D^2 \theta_{KL}^{\mathcal{T}}\|_{L^\infty(B((1,1),s))} < \infty$$

for some $s > 0$. Consider the weight functions $\boldsymbol{\lambda}^{\mathcal{T}}$ defined by $\lambda_{KL} = \partial_1 \theta_{KL}^{\mathcal{T}}(1, 1)$, and assume that there exists a non-empty open subset $U \subseteq \Omega$, a unit vector $v \in S^{d-1}$ and $c > 0$, such that

$$\liminf_{[\mathcal{T}] \to 0} \sum_{K \in \mathcal{T}, K \subseteq V} \left(\left(\sum_L \lambda_{KL} |n_{KL} \cdot v|^2 |(K|L)| d_{KL} \right) - |K| \right)_+ \geq c|V| \; , \qquad (1.6)$$

for any non-empty open subset $V \subseteq U$. Then there exist $\mu_0, \mu_1 \in \mathcal{P}(\overline{\Omega})$ such that

$$\limsup_{[\mathcal{T}] \to 0} \mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}} \mu_0, P_{\mathcal{T}} \mu_1) < \mathbb{W}_2(\mu_0, \mu_1) \; .$$

The condition (1.6) can easily be verified in the setting of Examples (A) and (B5); cf. Section 5 for details.

Remark. On a technical level, our method of proof offers the advantage that the maps for which Gromov–Hausdorff convergence is proved are the canonical projections $\mathcal{P}_{\mathcal{T}}$, rather than regularised versions of these maps, as in [20]. Another advantage is that we do not require any regularisation argument at the discrete level, as was done both in [20] and [19]. All regularisation arguments are done at the continuous level. In particular, we do not require any lower Ricci curvature bounds for the Markov chain at the discrete level, which would be quite restrictive.

Let us finally discuss how the results in this paper relate to the discretisation of gradient flows. Indeed, it follows from results in [7, 24, 26] that the discrete heat equation on a mesh $\mathcal{T}$ is the entropy gradient flow with respect to $\mathcal{W}_\mathcal{T}$ *if and only if* the mean function θ_{KL} is the logarithmic mean for every pair of adjacent cells K, L. Moreover, for any vanishing sequence of regular meshes, solutions of the discrete heat equation converge to solutions of the continuous heat equation [17].

By contrast, our main results imply that if θ_{KL} is the logarithmic mean, the associated transport metrics do *not* converge to $\mathbb{W}_2$, unless the meshes satisfy (rather restrictive) isotropy conditions. Thus, in the non-isotropic setting, preservation of the gradient flow structure is incompatible with convergence of the associated transport distances.

For 1-dimensional isotropic meshes, evolutionary Γ-convergence of the entropic gradient flow structure for the discrete heat flow with respect to the transport distance $\mathcal{W}_\mathcal{T}$ has been proved in [10]. Loose speaking, this means that

$$\int_0^1 \mathcal{A}_\mathcal{T}^*(m_t, \dot{m}_t) + |\mathrm{grad}_{\mathcal{W}_\mathcal{T}}\mathrm{Ent}_\mathcal{T}(m_t)|^2 \, \mathrm{d}t \to \int_0^1 \mathbb{A}^*(\mu_t, \dot{\mu}_t) + |\mathrm{grad}_{\mathbb{W}_2}\mathrm{Ent}(\mu_t)|^2 \, \mathrm{d}t$$

in the sense of Γ-convergence. For gradient flow approximations to nonlinear parabolic problems, convergence results have been obtained as well; see, e.g., [5].

Structure of the paper

In Section 2 we recall some known facts about the Kantorovich distance $\mathbb{W}_2$ and the heat flow on convex bounded domains. Furthermore we introduce ζ-regular meshes and the associated discrete transport distance $\mathcal{W}_\mathcal{T}$. Section 3 contains several useful coarse *a priori* bounds for discrete optimal transport, and in Section 4 we obtain finite volume estimates for the action functionals $\mathcal{A}_\mathcal{T}$ and $\mathcal{A}_\mathcal{T}^*$. The failure of Gromov–Hausdorff convergence to $\mathbb{W}_2$ (Proposition 1.2 and Theorem 1.7) is established in Section 5. Finally, Section 6 contains the proofs of the lower bound (Theorem 1.1), the upper bound (Theorem 1.4), and the convergence results (Corollaries 1.5 and 1.6).

Acknowledgements

2 Preliminaries

2.1 The Kantorovich metric

Let $(\mathcal{X}, \mathsf{d})$ be a Polish space, and let $\mathcal{P}(\mathcal{X})$ be the set of Borel probability measures on $\mathcal{X}$. The class of Borel measures on $\mathcal{X}$ is denoted by $\mathcal{M}_+(\mathcal{X})$, and the class of signed Borel measures with mass 0 by $\mathcal{M}_0(\mathcal{X})$.

For $1 \leq p < \infty$, the *p-Kantorovich metric* (often called *p*-Wasserstein metric) is defined by

$$\mathbb{W}_p(\mu_0, \mu_1) = \inf_{\gamma \in \Gamma(\mu_0, \mu_1)} \left(\int_{\mathcal{X} \times \mathcal{X}} \mathsf{d}(x, y)^p \, \mathrm{d}\gamma(x, y) \right)^{1/p} \tag{2.1}$$

for $\mu_0, \mu_1 \in \mathcal{P}(\mathcal{X})$. Here, $\Gamma(\mu_0, \mu_1)$ denotes the set of all couplings (also called transport plans) between μ_0 and μ_1:

$$\Gamma(\mu_0, \mu_1) = \{\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{X}) : \pi^0_\# \gamma = \mu_0 \text{ and } \pi^1_\# \gamma = \mu_1\} \, ,$$

where $\pi^0(x, y) = x$, $\pi^1(x, y) = y$, and $\pi^i_\# \gamma$ denotes the push-forward of γ under π^i. If $\mathcal{W}(\mu_0, \mu_1) < \infty$, then the infimum in (2.1) is attained; see, e.g., [32] for basic properties of $\mathbb{W}_p$.

Let $\Omega \subseteq \mathbb{R}^d$ be a convex bounded open set. Let $\mathcal{D} = \mathcal{C}_c^\infty(\mathbb{R}^d)$ be the space of test functions and let $\mathcal{D}'$ be the space of distributions. In this paper we will make use of the dynamical formulation of the metric $\mathbb{W}_2$, which is given in terms of the action functional $\mathbb{A} : \mathcal{M}_+(\overline{\Omega}) \times \mathcal{C}^1(\overline{\Omega}) \to \mathbb{R}$ and its Legendre dual $\mathbb{A}^* : \mathcal{M}_+(\overline{\Omega}) \times \mathcal{D}' \to \mathbb{R} \cup \{+\infty\}$, where

$$\mathbb{A}(\mu, \phi) := \int_{\overline{\Omega}} |\nabla \phi|^2 \, \mathrm{d}\mu \, , \qquad \mathbb{A}^*(\mu, w) = \sup_{\phi \in \mathcal{D}} \left(2\langle \phi, w \rangle - \mathbb{A}(\mu, \phi) \right) . \tag{2.2}$$

The *Benamou–Brenier formula* [4] asserts that

$$\mathbb{W}_2(\mu_0, \mu_1) = \inf \left\{ \sqrt{\int_0^1 \mathbb{A}^*(\mu_t, \dot{\mu}_t) \, \mathrm{d}t} \; : \; (\mu_t)_{t \in [0,1]} \in \mathbb{AC}(\mu_0, \mu_1) \right\} . \tag{2.3}$$

Here, $\mathbb{AC}(\mu_0, \mu_1)$ denotes the class of all $\mathbb{W}_2$-absolutely continuous curves $(\mu_t)_{t \in [0,1]}$ in $\mathcal{P}(\overline{\Omega})$ connecting μ_0 and μ_1.

For fixed $\delta > 0$, it will be useful to introduce the set $\mathcal{P}_\delta(\overline{\Omega}) \subseteq \mathcal{P}(\overline{\Omega})$ consisting of all $\mu = u \, \mathrm{d}x$ such that $u : \overline{\Omega} \to \mathbb{R}$ is Lipschitz continuous with $\mathrm{Lip}(u) \leq \frac{1}{\delta}$ and $\min_{x \in \overline{\Omega}} u(x) \geq \delta$. For $\mu \in \mathcal{P}_\delta(\overline{\Omega})$ and $w \in L^2(\Omega)$ with $\int_\Omega w(x) \, \mathrm{d}x = 0$, the maximiser in the definition of $\mathbb{A}^*(\mu, w)$ is given by the unique distributional solution $\phi \in H^1(\Omega)$ to the elliptic problem

$$\begin{cases} -\operatorname{div}(u\nabla\phi) = w & \text{in } \Omega \\ \partial_n \phi = 0 & \text{on } \partial\Omega \end{cases}$$

satisfying $\int_\Omega \phi \, \mathrm{d}x = 0$. Moreover, $\mathbb{A}(\mu, \phi) = \mathbb{A}^*(\mu, w)$.

Let $(H_a)_{a \geq 0}$ be the heat semigroup associated to the Neumann Laplacian Δ on Ω. Since Ω is convex, $\overline{\Omega}$ is a $\mathrm{CD}(0, d)$ space in the sense of Bakry–Émery and Lott–Villani–Sturm. In particular, the heat semigroup satisfies the Bakry–Émery gradient estimate

$$|\nabla H_a \phi|^2 \leq H_a |\nabla \phi|^2 \tag{2.4}$$

for all sufficiently smooth functions $\phi : \overline{\Omega} \to \mathbb{R}$, and the local logarithmic Sobolev inequality

$$|\nabla \log H_a \phi|^2 \leq \frac{1}{a} \frac{H_a(\phi \log \phi) - H_a \phi \log H_a \phi}{H_a \phi} \, , \tag{2.5}$$

9

for all smooth and positive functions $\phi : \overline{\Omega} \to \mathbb{R}$; cf. [3, Theorem 5.5.2]. Moreover, it is well known that the Neumann heat kernel h_a satisfies the Gaussian bounds

$$(4\pi a)^{-d/2} e^{-\frac{|x-y|^2}{4a}} \leq h_a(x,y) \leq C\big(a^{-d/2} \vee 1\big) e^{-c\frac{|x-y|^2}{a}} \tag{2.6}$$

for all $a > 0$ and $x,y \in \overline{\Omega}$, with constants $c, C > 0$ depending on Ω; see [9, Theorems 3.2.9 and 5.6.1]. The following result asserts that the heat semigroups maps $\mathcal{P}(\overline{\Omega})$ into $\mathcal{P}_\delta(\overline{\Omega})$.

Lemma 2.1. *For all $a > 0$ there exists a constant $\delta > 0$, depending on Ω and a, such that for any $\mu \in \mathcal{P}(\overline{\Omega})$ the density u_a of $H_a\mu$ satisfies*

$$u_a(x) \geq \delta \text{ for all } x \in \overline{\Omega} \quad \text{and} \quad \mathrm{Lip}(u_a) \leq \frac{1}{\delta} \ .$$

Proof. The lower bound follows immediately from the Gaussian lower bound in (2.6).

To prove the Lipschitz bound, we use the pointwise gradient inequality

$$2a|\nabla H_a \phi|^2 \leq H_a(\phi^2) \ ,$$

which follows from the CD$(0,d)$ property of Ω; cf. [3, Theorem 4.7.2]. Together with (2.6) we obtain

$$|\nabla u_a(x)|^2 = |\nabla H_{a/2} u_{a/2}(x)|^2 \leq a^{-1} \|H_{a/2}(u_{a/2}^2)\|_\infty \leq a^{-1} \|u_{a/2}\|_\infty^2 \leq C \ ,$$

for some $C < \infty$ depending on a. This implies the result. $\qquad\square$

We collect some known properties of the Kantorovich metric that will be useful in the sequel. Let us remark that the convexity of the domain Ω is crucial for Part (ii) in the following lemma, as its proof relies on gradient estimates in the sense of Bakry and Émery (see, e.g., [2, Theorem 3.17] and [3]).

Lemma 2.2 (Bounds on the Kantorovich metric)**.** *The following assertions hold:*

(i) *(Monotonicity) For $i = 0, 1$, let $\mu_i \in \mathcal{M}_+(\overline{\Omega})$ with $\mu_0 \geq \mu_1$, and let $w \in \mathcal{D}'$. Then:*

$$\mathbb{A}^*(\mu_0, w) \leq \mathbb{A}^*(\mu_1, w) \ . \tag{2.7}$$

(ii) *(Contraction bounds) For any $\mu \in \mathcal{P}(\overline{\Omega})$, $\phi \in \mathcal{D}$, $w \in \mathcal{D}'$, and $a \geq 0$, we have*

$$\mathbb{A}(\mu, H_a\phi) \leq \mathbb{A}(H_a\mu, \phi) \quad \text{and} \quad \mathbb{A}^*(H_a\mu, H_a w) \leq \mathbb{A}^*(\mu, w) \ . \tag{2.8}$$

Consequently, the following contraction property holds for any $\mu_0, \mu_1 \in \mathcal{P}(\overline{\Omega})$:

$$\mathbb{W}_2(H_a\mu_0, H_a\mu_1) \leq \mathbb{W}_2(\mu_0, \mu_1) \ . \tag{2.9}$$

(iii) *(Hölder continuity of the heat flow) There exists a constant $C < \infty$ such that*

$$\mathbb{W}_2(\mu, H_a\mu) \leq C\sqrt{a}$$

for any $\mu \in \mathcal{P}(\overline{\Omega})$ and $a \geq 0$.

Proof. Assertion (i) follows immediately from the definitions.

The first part of Assertion (ii) is a consequence of the Bakry–Émery gradient estimate (2.4) and the selfadjointness of H_a. The second part follows immediately, and the inequality (2.9) holds by the Benamou–Brenier formula (2.3).

To show Assertion (iii), note first that $d\gamma(x, y) = h_a(x, y) \, d\mu(x) \, dy$ is a coupling between μ and $H_a\mu$. Thus using (2.6) we obtain

$$
\begin{aligned}
\mathbb{W}_2(\mu, H_a\mu)^2 &\leq \int_{\overline{\Omega}} \int_{\overline{\Omega}} |x - y|^2 h_a(x, y) \, d\mu(x) \, dy \\
&\leq C\big(a^{-d/2} \vee 1\big) \int_{\overline{\Omega}} \int_{\overline{\Omega}} |x - y|^2 e^{-c\frac{|x-y|^2}{a}} \, d\mu(x) \, dy \\
&\leq C\big(a^{-d/2} \vee 1\big) \int_{\overline{\Omega}} \int_{\mathbb{R}^d} |x - y|^2 e^{-c\frac{|x-y|^2}{a}} \, dy \, d\mu(x) \\
&= C\big(1 \vee a^{d/2}\big) a .
\end{aligned}
\tag{2.10}
$$

If $a \leq \operatorname{diam}(\Omega)^2$, we have $1 \vee a^{d/2} \leq C$ for some constant C depending on Ω, so we may absorb the factor $1 \vee a^{d/2}$ into the constant. If $a \geq \operatorname{diam}(\Omega)^2$, we trivially have $\mathbb{W}_2(\mu, H_a\mu) \leq \operatorname{diam}(\Omega) \leq \sqrt{a}$. $\qquad\square$

2.2 Discrete transport metrics

We briefly recall the definition of the discrete transport metrics introduced in [7, 24, 26].

Let $\mathcal{X}$ be a finite set, and let $R : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ be a non-negative function satisfying $R(x, x) = 0$ for all $x \in \mathcal{X}$. We interpret $R(x, y)$ as the transition rate from x to y for a continuous-time Markov chain on $\mathcal{X}$, which we assume to be irreducible. Under this assumption, there exists a unique invariant probability measure $\pi \in \mathcal{P}(\mathcal{X})$. We assume that π satisfies the *detailed balance condition*:

$$
\pi(x)R(x, y) = \pi(y)R(y, x) \quad \text{for all } x, y \in \mathcal{X} .
$$

To define the discrete transport metric we need to choose a family of admissible means in the sense of the following definition. Note that these assumptions slightly differ from [11, Assumption 2.1].

Definition 2.3 (Admissible mean). *An* admissible mean *is a continuous function* $\theta :$ $\mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ *that is* $\mathcal{C}^\infty$ *on* $(0, \infty) \times (0, \infty)$, *positively 1-homogeneous, jointly concave, and normalised (i.e.,* $\theta(1, 1) = 1$).

We collect some known properties of admissible means in the following result.

Lemma 2.4. *Let* θ *be an admissible mean.*

(i) *For* $a, b \geq 0$ *we have* $\theta(a, b) \geq \min\{a, b\}$.

(ii) *For any* $a, b \geq 0$ *and* $c, d > 0$ *we have*

$$
\theta(a, b) \leq a\partial_1\theta(c, d) + b\partial_2\theta(c, d) ,
\tag{2.11}
$$

with equality if $a = c$ *and* $b = d$.

Proof. To show (i), assume first that $a > b$, and write $(a, b) = \frac{b}{b+\varepsilon}(b+\varepsilon, b+\varepsilon) + \frac{\varepsilon}{b+\varepsilon}(M, 0)$, where $M = \frac{b+\varepsilon}{\varepsilon}(a - b)$. Using the concavity, 1-homogeneity and normalisation of θ, we obtain $\theta(a, b) \geq \frac{b}{b+\varepsilon}\theta(b + \varepsilon, b + \varepsilon) = b\theta(1, 1) = b$. The case $a < b$ can be treated analogously, and the case $a = b$ follows immediately by 1-homogeneity and normalisation.

We refer to [11, Lemma 2.2] for a proof of (ii). $\qquad\square$

It will be useful to associate a number $\lambda^{(\theta)}$ to an admissible mean θ that quantifies its asymmetry.

Definition 2.5 (Weight). *For an admissible mean θ, its weight $\lambda^{(\theta)} \in [0,1]$ is given by*

$$\lambda^{(\theta)} = \partial_1 \theta(1,1) \ .$$

If there is no danger of confusion, we will simply write $\lambda = \lambda^{(\theta)}$.

Remark 2.6. Note that $\lambda^{(\theta)}$ is indeed nonnegative, since θ is non-decreasing in each of its arguments. Applying (2.11) with $a = b = c = d = 1$, it follows that

$$\partial_1\theta(1,1) + \partial_2\theta(1,1) = 1 \ , \tag{2.12}$$

which implies that $\lambda^{(\theta)} \leq 1$.

Remark 2.7. It follows from (2.11) with $c = d = 1$, that

$$\theta(a,b) \leq \lambda^{(\theta)}a + (1 - \lambda^{(\theta)})b$$

for any $a, b \geq 0$. Moreover, this inequality characterises $\lambda^{(\theta)} \in [0,1]$ uniquely. If θ is symmetric, it follows from (2.12) that $\lambda^{(\theta)} = \frac{1}{2}$.

Examples of symmetric admissible means are the arithmetic mean $\theta_{\mathrm{arith}}(a,b) = \frac{a+b}{2}$, the geometric mean $\theta_{\mathrm{geom}}(a,b) = \sqrt{ab}$, the logarithmic mean $\theta_{\mathrm{log}}(a,b) = \int_0^1 a^{1-p}b^p \, \mathrm{d}p$, and the harmonic mean $\theta_{\mathrm{harm}}(a,b) = \frac{2ab}{a+b}$. For each of these means there exist natural generalisations with weights $\lambda \in [0,1]$, such as

$$\theta_{\mathrm{arith}}^{(\lambda)}(a,b) = \lambda a + (1-\lambda)b \ , \qquad \theta_{\mathrm{geom}}^{(\lambda)}(a,b) = a^\lambda b^{1-\lambda} \ ,$$
$$\theta_{\mathrm{log}}^{(\lambda)}(a,b) = \int_0^1 a^p b^{1-p} \, \tau(\mathrm{d}p) \ , \quad \theta_{\mathrm{harm}}^{(\lambda)}(a,b) = \frac{1}{\lambda/a + (1-\lambda)/b} \ . \tag{2.13}$$

Here, τ is an arbitrary Borel probability measure on $[0,1]$ with $\int_0^1 p\tau(\mathrm{d}p) = \lambda$.

Definition 2.8 (Mean and weight functions). *Let $\mathcal{X}$ be a finite set.*

(i) *A* mean function *is a family of admissible means $\boldsymbol{\theta} = (\theta_{xy})_{x,y \in \mathcal{X}}$ satisfying the symmetry condition $\theta_{xy}(a,b) = \theta_{yx}(b,a)$ for all $x, y \in \mathcal{X}$ and $a, b \geq 0$.*

(ii) *A* weight function *is a collection $\boldsymbol{\lambda} = (\lambda_{xy})_{x,y \in \mathcal{X}} \subseteq [0,1]$ satisfying $\lambda_{xy} + \lambda_{yx} = 1$ for all $x, y \in \mathcal{X}$,*

(iii) *For a mean function $\boldsymbol{\theta}$, its* induced weight function *$\boldsymbol{\lambda}^{(\theta)}$ is defined by $\lambda_{xy}^{(\theta)} = \partial_1\theta_{xy}(1,1)$ for all x,y. We say that a mean function $\boldsymbol{\theta}$ is* compatible *with a weight function $\boldsymbol{\lambda}$, if $\boldsymbol{\lambda}$ is induced by $\boldsymbol{\theta}$.*

It follows from (2.12) that the induced weight function is indeed a weight function.

We are now in a position to define the discrete transport metrics. Given a Markov chain $(\mathcal{X}, R, \pi)$ and a mean function $\boldsymbol{\theta}$, the discrete transport metric is defined using discrete analogues of (2.2). The action functional $\mathcal{A} : \mathcal{P}(\mathcal{X}) \times \mathbb{R}^{\mathcal{X}} \to \mathbb{R}$ and its dual $\mathcal{A}^* : \mathcal{P}(\mathcal{X}) \times \mathbb{R}^{\mathcal{X}} \to \mathbb{R} \cup \{+\infty\}$ are given by

$$\mathcal{A}(m,\psi) = \frac{1}{2} \sum_{x,y \in \mathcal{X}} \theta_{xy}\big(m(x)R(x,y), m(y)R(y,x)\big)\big(\psi(x) - \psi(y)\big)^2 \ ,$$
$$\frac{1}{2}\mathcal{A}^*(m,\sigma) = \sup_{\psi \in \mathbb{R}^{\mathcal{X}}} \left\{ \sum_{x \in \mathcal{X}} \psi(x)\sigma(x) - \frac{1}{2}\mathcal{A}(m,\psi) \right\} \ . \tag{2.14}$$

For $m_0, m_1 \in \mathcal{P}(\mathcal{X})$, the associated transport metric is given by

$$\mathcal{W}(m_0, m_1) = \inf \left\{ \sqrt{\int_0^1 \mathcal{A}^*(m_t, \dot{m}_t) \, dt} \; : \; (m_t)_{t \in [0,1]} \in \mathcal{AC}(m_0, m_1) \right\} .$$

Here, $\mathcal{AC}(m_0, m_1)$ denotes the class of all curves $(m_t)_{t \in [0,1]} \subseteq \mathcal{P}(\mathcal{X})$ connecting m_0 and m_1, with the property that $t \mapsto m_t(x)$ is absolutely continuous for all $x \in \mathcal{X}$.

Remark 2.9. In most of the previous papers dealing with discrete dynamical transport metrics, θ_{xy} has been chosen to be independent of x and y. In particular, to obtain a gradient flow structure for Markov chains [7,24,26], one chooses θ_{xy} to be the logarithmic mean for all x, y. We will see in this paper that the additional freedom can be important to obtain Gromov–Hausdorff convergence. In a noncommutative setting, a similar generalisation has been useful in the setting of Lindblad equations with a nontracial invariant state [6,28].

We will occasionally use an equivalent formulation for $\mathcal{W}$ given by

$$\mathcal{W}(m_0, m_1) = \inf \left\{ \sqrt{\int_0^1 \mathcal{K}(m_t, V_t) \, dt} : (m, V) \in \mathcal{CE}(m_0, m_1) \right\} .$$

Here, $\mathcal{CE}(m_0, m_1)$ denotes the class of pairs (m, V) satisfying

- $m : [0, T] \to \mathcal{P}(\mathcal{X})$ is continuous with $m|_{t=0} = m_0$, and $m|_{t=1} = m_1$;
- $V : [0, T] \times \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is locally integrable;
- the *continuity equation* holds in the sense of distributions:

$$\dot{m}_t(x) + \frac{1}{2} \sum_{y \in \mathcal{X}} \pi(x) R(x, y) \big(V_t(x, y) - V_t(y, x) \big) = 0 \qquad \text{for all } x \in \mathcal{X} . \quad (2.15)$$

Moreover,

$$\mathcal{K}(m, V) := \frac{1}{2} \sum_{x, y \in \mathcal{X}} \pi(x) R(x, y) K_{xy} \left(\frac{m(x)}{\pi(x)}, \frac{m(y)}{\pi(y)}, V(x, y) \right) , \quad (2.16)$$

where the convex and lower semicontinuous function $K : \mathbb{R}^3 \to [0, \infty]$ is given by

$$K_{xy}(a, b, w) := \begin{cases} \frac{w^2}{\theta_{xy}(a,b)} , & w \in \mathbb{R}, \; a, b > 0 , \\ 0 , & w = 0, \; a, b \geq 0 , \\ +\infty , & \text{otherwise} . \end{cases} \quad (2.17)$$

By Fenchel duality, $\mathcal{A}^*$ can be obtained from $\mathcal{K}$ by minimising over all solutions to the continuity equation

$$\mathcal{A}^*(m, \sigma) = \inf_V \left\{ \mathcal{K}(m, V) : \right.$$
$$\left. \sigma(x) + \frac{1}{2} \sum_{y \in \mathcal{X}} \pi(x) R(x, y) \big(V(x, y) - V(y, x) \big) = 0 \quad \forall x \in \mathcal{X} \right\} . \quad (2.18)$$

Here the infimum runs over all vector fields $V : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Without loss of generality we may impose the anti-symmetry condition $V(x, y) = -V(y, x)$ for all $x, y \in \mathcal{X}$; see [11, Section 2] for more details.

13

2.3 Admissible and ζ-regular meshes

Following [17, Section 3.1.2], we introduce the notion of an admissible mesh.

Definition 2.10 (Mesh). *A mesh of Ω is a pair $(\mathcal{T}, \{x_K\}_{K\in\mathcal{T}})$ where*

- *$\mathcal{T} = \{K\}_{K\in\mathcal{T}}$ is a finite partition (i.e., a pairwise disjoint covering) of $\overline{\Omega}$ into sets (called cells) with nonempty and convex interior.*
- *$\{x_K\}_{K\in\mathcal{T}} \subseteq \overline{\Omega}$ is a family of distinct points with $x_K \in \overline{K}$ for every $K \in \mathcal{T}$.*

Note that all interior cells K are polytopes. Throughout the paper we will use the following notation:

- $|K|$ denotes the Lebesgue measure of a cell $K \in \mathcal{T}$.
- $(K|L) = \overline{K} \cap \overline{L}$ is the flat convex surface with $(d-1)$-dimensional Hausdorff measure $|(K|L)|$. Two cells $K, L \in \mathcal{T}$ with $K \neq L$ are called *nearest neighbours* if $|(K|L)| > 0$. In this case we write $K \sim L$. We write $K \simeq L$ if $K = L$ or $K \sim L$.
- $d_{KL} = |x_K - x_L|$.
- $[\mathcal{T}] = \max_{K\in\mathcal{T}} \operatorname{diam}(K)$ denotes the *mesh size* of $\mathcal{T}$.

Definition 2.11 (Admissible and ζ-regular mesh). *A mesh $(\mathcal{T}, \{x_K\}_{K\in\mathcal{T}})$ is called admissible if $x_K - x_L \perp (K|L)$ whenever $K, L \in \mathcal{T}$ are nearest neigbours. An admissible mesh $\mathcal{T}$ is called ζ-regular for $\zeta \in (0,1]$ if the following conditions hold:*

- *(inner ball condition) $B(x_K, \zeta[\mathcal{T}]) \subseteq K$ for every $K \in \mathcal{T}$;*
- *(area bound) $|(K|L)| \geq \zeta[\mathcal{T}]^{d-1}$ for every $K, L \in \mathcal{T}$ with $K \sim L$.*

In view of Definition 2.11, we stress that the ζ-regularity of a mesh implies its admissibility.

Every Voronoi tesselation yields an admissible mesh. Another example is obtained by slicing $\overline{\Omega}$ multiple times in the cardinal directions. In this case there are several degrees of freedom for placing the points $\{x_K\}_{K\in\mathcal{T}}$, so these points are not necessarily uniquely determined by $\mathcal{T}$. We refer to [17] for more information.

In the following result we collect some basic geometric properties of ζ-regular meshes that will be useful in the sequel.

Lemma 2.12. *Let $\mathcal{T}$ be a ζ-regular mesh of Ω for some $\zeta \in (0,1]$. Then there exists a constant $C < \infty$ depending only on Ω and ζ such that the following assertions hold:*

(i) The number of nearest neighbours of any cell is bounded by C.

(ii) Any pair of cells $K, L \in \mathcal{T}$ can be connected (for some $N \geq 0$) by a path $(K_i)_{i=1}^N \subseteq \mathcal{T}$ with $K_0 = K$, $K_N = L$, $K_{i-1} \sim K_i$ for $1 \leq i \leq N$, and

$$\sum_{i=1}^{N} d_{K_{i-1},K_i} \leq C d_{KL} \ . \tag{2.19}$$

(iii) For all $K \sim L$ we have the following estimates:

$$\operatorname{diam}(K) \leq C d_{KL} \ , \tag{2.20}$$

$$C^{-1}|K| \leq d_{KL}|(K|L)| \leq C|K| \ , \tag{2.21}$$

$$|K| \leq C|L| \ . \tag{2.22}$$

14

Proof. (i): Write $h = [\mathcal{T}]$ for brevity. By the inner ball condition we have

$$\bigcup_{L:L\sim K} B(x_L, \zeta h) \subseteq \bigcup_{L:L\sim K} L \subseteq B(x_K, 2h) \ .$$

Hence, Assertion (i) follows by comparing volumes.

(ii): Let ℓ be the line segment from x_K to x_L. By path-connectedness, there exists a continuous curve γ in the $[\mathcal{T}]$-neighbourhood of ℓ that connects x_K to x_L and avoids the boundaries of the cell interfaces. Let $\{K_i\}_{i=0}^N$ be the N cells that successively intersect γ. By ζ-regularity, each of the balls $B(x_{K_i}, \zeta[\mathcal{T}])$ is contained in K_i. In turn, each of the cells K_i is contained in the cylinder of radius $2[\mathcal{T}]$, whose central axis is obtained by extending ℓ by a distance $[\mathcal{T}]$ in both directions.

The number of disjoint balls of radius r that can be packed into a cylinder of length s and radius R is bounded by $C_d s \frac{R^{d-1}}{r^d}$, where C_d is a dimensional constant. Therefore, if follows that

$$N \le C \frac{|x_K - x_L| + 2[\mathcal{T}]}{[\mathcal{T}]} \ ,$$

where C depends on ζ and d. As $\zeta[\mathcal{T}] \le \min_{K' \ne L'} |x_{K'} - x_{L'}|$, if follows that $N \le C|x_K - x_L|/[\mathcal{T}]$. As $d_{K_{i-1}, K_i} \le 2[\mathcal{T}]$, this yields the desired bound.

(iii): To prove (2.20), we use the inner ball condition to obtain $d_{KL} \ge \zeta h \ge \zeta \operatorname{diam}(K)$.

Since $x_K - x_L \perp (K|L)$, the volume of the bipyramid spanned by $(K|L)$, x_K and x_L is given by $\frac{1}{d}|(K|L)|d_{KL}$. Using the inner ball condition we obtain

$$\frac{1}{d}d_{KL}|(K|L)| \le |K \cup L| \le |B(x_K, 2h)| \le C|B(x_K, \zeta h)| \le C|K| \ ,$$

which proves the upper bound in (2.21). To prove the corresponding lower bound, note that $\zeta h \le d_{KL}$, hence by ζ-regularity,

$$|K| \le |B(x_K, h)| \le Ch^d \le Ch|(K|L)| \le Cd_{KL}|(K|L)| \ .$$

The inequality (2.22) follows from (2.21). $\qquad\square$

The ζ-regularity condition allows us to control the constants in several useful inequalities. Most notably, we will use the Poincaré inequality [29]

$$\int_K \phi^2 \, \mathrm{d}x \le \frac{\operatorname{diam}(K)^2}{\pi^2} \int_K |\nabla\phi|^2 \, \mathrm{d}x \tag{2.23}$$

and the trace inequality

$$\int_{\partial K} \phi^2 \, \mathrm{d}S \le C \operatorname{diam}(K) \int_K |\nabla\phi|^2 \, \mathrm{d}x \ , \tag{2.24}$$

both of which are valid for all $K \in \mathcal{T}$ and all $\phi \in H^1(K)$ with $\int_K \phi \, \mathrm{d}x = 0$; cf. [21, Theorem 1.5.1.10]. The constant $C < \infty$ depends only on ζ and the dimension d. For the convenience of the reader we record a simple consequence that we will use below.

Lemma 2.13. *Let $\zeta \in (0, 1]$. There exists a constant $C < \infty$ depending on ζ and Ω, such that, for any $K \in \mathcal{T}$ and any $\phi \in H^1(K)$,*

$$\fint_{\partial K} |\phi| \, \mathrm{d}S \le \fint_K |\phi| \, \mathrm{d}x + C[\mathcal{T}] \sqrt{\fint_K |\nabla\phi|^2 \, \mathrm{d}x} \ . \tag{2.25}$$

Moreover, for any convex subset $B \subseteq K$ with $|B| \geq \zeta |K|$, we have

$$\left| \fint_K \phi \, \mathrm{d}x - \fint_B \phi \, \mathrm{d}x \right| \leq C[\mathcal{T}] \sqrt{\fint_K |\nabla \phi|^2 \, \mathrm{d}x} \ . \tag{2.26}$$

Proof. Write $\bar{\phi} = \fint_K \phi \, \mathrm{d}x$. Using (2.24) and (2.21) we obtain

$$\|\phi\|_{L^1(\partial K)} \leq |\partial K| \, |\bar{\phi}| + \|\phi - \bar{\phi}\|_{L^1(\partial K)}$$
$$\leq |\partial K| \, |\bar{\phi}| + C\sqrt{|\partial K| \operatorname{diam}(K)} \|\nabla \phi\|_{L^2(K)} \ .$$

As Lemma 2.12 implies $|K| \leq C|\partial K|[\mathcal{T}]$, we have $\operatorname{diam}(K) \leq C\frac{|\partial K|}{|K|}[\mathcal{T}]^2$, which yields (2.25).

To prove (2.26), we write $\overline{\phi}_K = \fint_K \phi \, \mathrm{d}x$ and $\overline{\phi}_B = \fint_B \phi \, \mathrm{d}x$ for brevity. Using the Poincaré inequality (2.23) we obtain

$$|K| \, |\overline{\phi}_B - \overline{\phi}_K|^2 \leq C|B| \, |\overline{\phi}_B - \overline{\phi}_K|^2$$
$$\leq C \int_B |\overline{\phi}_B - \phi|^2 \, \mathrm{d}x + C \int_K |\phi - \overline{\phi}_K|^2 \, \mathrm{d}x$$
$$\leq C[\mathcal{T}]^2 \int_K |\nabla \phi|^2 \, \mathrm{d}x \ ,$$

which implies (2.26). $\qquad\square$

2.4 Discrete optimal transport on admissible meshes

Given an admissible mesh $\mathcal{T}$ of Ω we consider an irreducible Markov chain on $\mathcal{T}$ with transition rates

$$R(K,L) = \frac{|(K|L)|}{|K| d_{KL}} \quad \text{if } K \sim L \tag{2.27}$$

and $R(K,L) = 0$ otherwise.

Remark 2.14 (Formal consistency). This choice of the transition rates $R(K,L)$ is motivated by the following formal consistency computation for the Dirichlet energy associated to our problem. Let $\mathcal{U}$ be the uniform probability measure on Ω. For a smooth function $\phi : \Omega \to \mathbb{R}$, the continuous action functional satisfies

$$\mathbb{A}(\mathcal{U}, \phi) = \frac{1}{|\Omega|} \int_\Omega |\nabla \phi|^2 \, \mathrm{d}x = -\frac{1}{|\Omega|} \sum_K \int_K \phi \Delta \phi \, \mathrm{d}x \approx -\frac{1}{|\Omega|} \sum_K \phi(x_K) \int_{\partial K} \partial_n \phi \, \mathrm{d}S$$
$$\approx -\frac{1}{|\Omega|} \sum_{K,L} \phi(x_K) |(K|L)| \frac{\phi(x_L) - \phi(x_K)}{|x_K - x_L|} = \frac{1}{2|\Omega|} \sum_{K,L} \frac{|(K|L)|}{|x_K - x_L|} \big(\phi(x_K) - \phi(x_L) \big)^2 \ .$$

Let $U_\mathcal{T} \in \mathcal{P}(\mathcal{T})$ be the canonical discretisation of $\mathcal{U}$ given by $U_\mathcal{T} = P_\mathcal{T} \mathcal{U}$, and let $\psi_\mathcal{T} : \mathcal{T} \to \mathbb{R}$ be given by $\psi_\mathcal{T}(K) = \phi(x_K)$. Then the latter expression is indeed of the form $\mathcal{A}(U_\mathcal{T}, \psi_\mathcal{T})$ defined in (2.14), provided that the coefficients $R(K,L)$ are defined by (2.27).

It is immediate to check that the unique invariant probability measure π on $\mathcal{T}$ is given by $\pi(K) = \frac{|K|}{|\Omega|}$. Moreover, the detailed balance condition $\pi(K)R(K,L) = \pi(L)R(L,K)$ holds.

For this Markov chain, the discrete action functional $\mathcal{A}_{\mathcal{T}} : \mathcal{P}(\mathcal{T}) \times \mathbb{R}^{\mathcal{T}} \to \mathbb{R}$ and its dual $\mathcal{A}_{\mathcal{T}}^* : \mathcal{P}(\mathcal{T}) \times \mathbb{R}^{\mathcal{T}} \to \mathbb{R} \cup \{+\infty\}$ defined in (2.14) are given by

$$\mathcal{A}_{\mathcal{T}}(m, \psi) = \frac{1}{2} \sum_{K,L \in \mathcal{T}} \frac{|(K|L)|}{d_{KL}} \theta_{KL} \left(\frac{m(K)}{|K|}, \frac{m(L)}{|L|} \right) \big(\psi(K) - \psi(L) \big)^2 \,,$$

$$\frac{1}{2} \mathcal{A}_{\mathcal{T}}^*(m, \sigma) = \sup_{\psi \in \mathbb{R}^{\mathcal{T}}} \left(\sum_{K \in \mathcal{T}} \psi(K) \sigma(K) - \frac{1}{2} \mathcal{A}_{\mathcal{T}}(m, \psi) \right) \,.$$

The main object of study in this paper is the associated transport metric, defined for $m_0, m_1 \in \mathcal{P}(\mathcal{T})$ by

$$\mathcal{W}_{\mathcal{T}}(m_0, m_1) = \inf \left\{ \sqrt{\int_0^1 \mathcal{A}_{\mathcal{T}}^*(m_t, \dot{m}_t) \, \mathrm{d}t} \ : \ (m_t)_{t \in [0,1]} \in \mathcal{AC}(m_0, m_1) \right\} \,.$$

3   A priori estimates

In this section we prove the necessary *a priori* estimates. Throughout this section, we fix a convex bounded open set $\Omega \subseteq \mathbb{R}^d$ and a ζ-regular mesh $\mathcal{T}$ for some $\zeta \in (0, 1]$. Moreover, we use the convention that the constants C appearing in this section (which are allowed to change from line to line) may depend on Ω and ζ, but not on other properties of $\mathcal{T}$.

Lemma 3.1. *There exists a constant $C < \infty$ such that for any $K, L \in \mathcal{T}$ with $K \sim L$,*

$$\mathcal{W}_{\mathcal{T}}(\delta_K, \delta_L) \leq C d_{KL} \,.$$

Proof. It follows from [11, Lemma 2.13] that

$$\mathcal{W}_{\mathcal{T}}(\delta_K, \delta_L)^2 \leq c \frac{d_{KL}}{|(K|L)|} \max\{|K|, |L|\}$$

for some universal constant $c < \infty$. The claim follows by ζ-regularity in view of (2.21). $\square$

To compare discrete and continuous measures, we use the canonical projection operator $P_{\mathcal{T}}$ defined in (1.2). The associated embedding operator $Q_{\mathcal{T}} : \mathcal{P}(\mathcal{T}) \to \mathcal{P}(\overline{\Omega})$ is given by

$$Q_{\mathcal{T}} m = \sum_{K \in \mathcal{T}} m(K) \mathcal{U}_K \qquad \text{for } m \in \mathcal{P}(\mathcal{T}) \,, \tag{3.1}$$

where $\mathcal{U}_K$ denotes the uniform probability measure on K. Note that $P_{\mathcal{T}} \circ Q_{\mathcal{T}}$ is the identity on $\mathcal{P}(\mathcal{T})$. The following lemma quantifies how close $Q_{\mathcal{T}} \circ P_{\mathcal{T}}$ is to the identity on $\mathcal{P}(\overline{\Omega})$.

Lemma 3.2 (Consistency)**.** *For all $\mu \in \mathcal{P}(\overline{\Omega})$ we have*

$$\mathbb{W}_2(\mu, Q_{\mathcal{T}} P_{\mathcal{T}} \mu) \leq [\mathcal{T}] \,.$$

Proof. For $U \in \mathcal{T}$, let $\gamma_K \in \mathcal{P}(\overline{\Omega} \times \overline{\Omega})$ be any coupling between $\tilde{\mu}_K$, the normalised restriction of μ to K, and $\mathcal{U}_K$, the uniform probability measure on K. It then follows that $\gamma := \sum_{K \in \mathcal{T}} \mu(K) \gamma_K$ belongs to $\Gamma(\mu, Q_{\mathcal{T}} P_{\mathcal{T}} \mu)$. Consequently,

$$\mathbb{W}_2(\mu, Q_{\mathcal{T}} P_{\mathcal{T}} \mu)^2 \leq \int_{\overline{\Omega} \times \overline{\Omega}} |x - y|^2 \, \mathrm{d}\gamma(x, y)$$

$$= \sum_K \mu(K) \int_{K \times K} |x - y|^2 \, \mathrm{d}\gamma_K(x, y) \leq \sum_K \mu(K) \, \mathrm{diam}(K)^2 \leq [\mathcal{T}]^2 \,.$$

$\square$

The following result provides a coarse bound for the discrete distance $\mathcal{W}_\mathcal{T}$ in terms of the Kantorovich distance $\mathbb{W}_2$.

Lemma 3.3 (Upper bound for $\mathcal{W}_\mathcal{T}$)**.** *For all* $m_0, m_1 \in \mathcal{P}(\mathcal{T})$ *we have*

$$\mathcal{W}_\mathcal{T}(m_0, m_1) \leq C\Big(\mathbb{W}_2(Q_\mathcal{T} m_0, Q_\mathcal{T} m_1) + [\mathcal{T}]\Big) . \tag{3.2}$$

Proof. Let $\gamma \in \Gamma_o(Qm_0, Qm_1)$ be an optimal plan for $\mathbb{W}_2$, and set $\gamma_{KL} := \gamma(K \times L)$ for brevity. Observe that $m_0 = \sum_{K,L} \gamma_{KL} \delta_K$ and $m_1 = \sum_{K,L} \gamma_{KL} \delta_L$. Therefore, by convexity of the squared distance (cf. [11, Proposition 2.11]) we have

$$\mathcal{W}_\mathcal{T}(m_0, m_1)^2 \leq \sum_{K,L} \gamma_{KL} \mathcal{W}_\mathcal{T}(\delta_K, \delta_L)^2 .$$

For $K, L \in \mathcal{T}$, take a connecting path $\{K_i\}_{i=0}^N \subseteq \mathcal{T}$ with $K_0 = K$ and $K_N = L$ satisfying the ζ-regularity estimate (2.19). Using Lemma 3.1 we obtain

$$\mathcal{W}_\mathcal{T}(\delta_K, \delta_L) \leq \sum_{i=1}^N \mathcal{W}_\mathcal{T}(\delta_{K_{i-1}}, \delta_{K_i}) \leq C \sum_{i=1}^N d_{K_{i-1}, K_i} \leq C d_{KL} .$$

For all $x \in K$ and $y \in L$ we have $d_{KL}^2 = |x_K - x_L|^2 \leq |x - y|^2 + C[\mathcal{T}]$. Combining these estimates, the result follows. $\qquad\square$

The following lemma provides a coarse lower bound for the discrete dual action functional in terms of its continuous counterpart.

Lemma 3.4 (Bound for the dual action functional)**.** *There exists a constant* $C < \infty$ *such that for any* $m \in \mathcal{P}(\mathcal{T})$, *and any* $\sigma \in \mathcal{M}_0(\mathcal{T})$, *we have*

$$\mathbb{A}^*(H_{[\mathcal{T}]} Q_\mathcal{T} m, H_{[\mathcal{T}]} Q_\mathcal{T} \sigma) \leq C \mathcal{A}_\mathcal{T}^*(m, \sigma) .$$

Proof. Let $m \in \mathcal{P}(\mathcal{T})$ and $\sigma \in \mathcal{M}_0(\mathcal{T})$ be such that $\mathcal{A}_\mathcal{T}^*(m, \sigma) < \infty$. In view of (2.18) there exists an anti-symmetric momentum vector field $V : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ solving the discrete continuity equation

$$\sigma(K) + \sum_L \frac{|(K|L)|}{d_{KL}} V(K, L) = 0 . \tag{3.3}$$

For $K \in \mathcal{T}$, define $g : \partial K \to \mathbb{R}$ by $g(x) = \frac{V(K,L)}{d_{KL}}$ for $x \in (K|L)$ with $K \sim L$, and $g(x) = 0$ for $x \in \partial\Omega$. It then follows that $\sigma(K) = -\int_{\partial K} g \, dS$. Therefore, denoting the outward unit normal on ∂K by n, the Neumann problem

$$\begin{cases} -\Delta \phi_K = \sigma(K)/|K| & \text{in } K \\ \partial_n \phi_K = g & \text{on } \partial K \end{cases} \tag{3.4}$$

has a unique solution $\phi_K \in H^1(K)$ with $\int_K \phi_K \, dx = 0$. Since $\sigma(K) \leq \|g\|_{L^1(\partial K)}$ by (3.3), we obtain

$$\begin{aligned}
\int_K |\nabla \phi_K|^2 \, dx &= \frac{1}{|K|} \int_K \sigma(K) \phi_K \, dx + \int_{\partial K} g \phi_K \, dS \\
&\leq \frac{|\sigma(K)|}{|K|} \|\phi_K\|_{L^1(K)} + \|g\|_{L^2(\partial K)} \|\phi_K\|_{L^2(\partial K)} \\
&\leq \|g\|_{L^2(\partial K)} \left(\sqrt{\frac{|\partial K|}{|K|}} \|\phi_K\|_{L^2(K)} + \|\phi_K\|_{L^2(\partial K)} \right) .
\end{aligned} \tag{3.5}$$

18

Since $\int_K \phi_K \, \mathrm{d}x = 0$, the right-hand side can be bounded in terms of $\|\nabla \phi_K\|_{L^2(K)}$ using the Poincaré inequality (2.23) and the trace inequality (2.24). Moreover, using the ζ-regularity inequalities (2.20) (2.21), and the bound on the number of neighbours from Lemma 2.12, we obtain $|\partial K| \operatorname{diam}(K) \leq C|K|$. Consequently,

$$\int_K |\nabla \phi_K|^2 \, \mathrm{d}x \leq C \operatorname{diam}(K) \int_{\partial K} g^2 \, \mathrm{d}S \ .$$

Using the ζ-regularity inequality (2.20) we obtain,

$$\int_{\partial K} g^2 \, \mathrm{d}S = \sum_{L \in \mathcal{T}} \frac{|(K|L)|}{d_{KL}^2} V(K,L)^2 \leq \frac{C}{\operatorname{diam}(K)} \sum_{L \in \mathcal{T}} \frac{|(K|L)|}{d_{KL}} V(K,L)^2 \ ,$$

and therefore,

$$\int_K |\nabla \phi_K|^2 \, \mathrm{d}x \leq C \sum_{L \in \mathcal{T}} \frac{|(K|L)|}{d_{KL}} V(K,L)^2 \ .$$

Let us now define the vector field $\Phi \in L^2(\overline{\Omega}; \mathbb{R}^d)$ by $\Phi(x) = \nabla \phi_K(x)$ for $x \in K$. By (3.4) and the anti-symmetry of the Neumann boundary values, one has

$$Q_{\mathcal{T}} \sigma + \operatorname{div} \Phi = 0 \tag{3.6}$$

in the sense of distributions, and $\Phi \cdot n = 0$ on $\partial \Omega$.

Take now $m \in \mathcal{P}(\mathcal{T})$. Writing $\rho(K) = \frac{m(K)}{|K|}$, we obtain

$$
\begin{aligned}
\mathbb{A}^*(Q_{\mathcal{T}} m, Q_{\mathcal{T}} \sigma) &\leq \sum_{K \in \mathcal{T}} \int_K \frac{|\nabla \phi_K|^2}{\rho(K)} \, \mathrm{d}x \\
&\leq C \sum_{K,L \in \mathcal{T}} \frac{|(K|L)|}{d_{KL}} \frac{V(K,L)^2}{\theta_{\mathrm{harm}}(\rho(K), \rho(L))} \ ,
\end{aligned}
\tag{3.7}
$$

where $\theta_{\mathrm{harm}}(a,b) = \frac{2ab}{a+b}$ is the harmonic mean.

We would like to obtain a similar estimate involving the means θ_{KL}, but as ρ is in general not bounded away from 0, we cannot bound the harmonic mean by a multiple of θ_{KL}. To remedy this issue, we perform an additional regularisation step. Consider the function $\tilde{\rho} : \mathcal{T} \to \mathbb{R}_+$ given by $\tilde{\rho}(K) = \rho(K) + \sum_{L \sim K} \rho(L)$, and set $\tilde{m}(K) = \tilde{\rho}(K)|K|$. In view of (2.7), Lemma 3.5 below, and (2.8), we obtain, for $a = [\mathcal{T}]$,

$$
\begin{aligned}
\mathbb{A}^*(H_a Q_{\mathcal{T}} m, H_a Q_{\mathcal{T}} \sigma) &\leq C \mathbb{A}^*(H_a Q_{\mathcal{T}} \tilde{m}, H_a Q_{\mathcal{T}} \sigma) \\
&\leq C \mathbb{A}^*(Q_{\mathcal{T}} \tilde{m}, Q_{\mathcal{T}} \sigma)
\end{aligned}
$$

We stress that to obtain the first inequality, the choice $a = [\mathcal{T}]$ is crucial; cf. Remark 3.6. Moreover, applying (3.7) to $\tilde{m}$, we obtain

$$\mathbb{A}^*(Q_{\mathcal{T}} \tilde{m}, Q_{\mathcal{T}} \sigma) \leq C \sum_{K,L \in \mathcal{T}} \frac{|(K|L)|}{d_{KL}} \frac{V(K,L)^2}{\theta_{\mathrm{harm}}(\tilde{\rho}(K), \tilde{\rho}(L))} \ .$$

Since $\tilde{\rho}(K), \tilde{\rho}(L) \geq \rho(K) + \rho(L)$ whenever $K \sim L$, we have $\theta_{\mathrm{harm}}(\tilde{\rho}(K), \tilde{\rho}(L)) \geq \rho(K) + \rho(L) \geq \theta_{KL}(\rho(K), \rho(L))$. Therefore, combining the estimates above, we obtain

$$\mathbb{A}^*(H_a Q_{\mathcal{T}} m, H_a Q_{\mathcal{T}} \sigma) \leq C \sum_{K,L \in \mathcal{T}} \frac{|(K|L)|}{d_{KL}} \frac{V(K,L)^2}{\theta_{KL}(\rho(K), \rho(L))} \ .$$

Taking the infimum over all V satisfying (3.3), the result follows using (2.18). $\qquad \square$

19

The following result was used in the proof of Lemma 3.4. Recall that we write $K \simeq L$ if $K \sim L$ or $K = L$.

Lemma 3.5. *For $m \in \mathcal{P}(\mathcal{T})$ we define $\tilde{m} : K \to \mathbb{R}_+$ by $\tilde{m}(K) = |K|\tilde{\rho}(K)$, where $\tilde{\rho}(K) = \sum_{L \simeq K} \rho(L)$ and $\rho(K) = \frac{m(K)}{|K|}$. Then, for every $a > 0$ and $x \in \overline{\Omega}$, the inequality*

$$H_a Q_{\mathcal{T}} m(x) \leq H_a Q_{\mathcal{T}} \tilde{m}(x) \leq C e^{C\sqrt{\eta(a)}[\mathcal{T}]} H_a Q_{\mathcal{T}} m(x) \ , \tag{3.8}$$

holds, with $\eta(a) = a^{-2} + (a^{-1} \log a) \vee 0$.

Remark 3.6. It is crucial in our application that by choosing $a = [\mathcal{T}]$ (as is done in Lemma 3.4), the exponent $\sqrt{\eta(a)}[\mathcal{T}]$ remains bounded as $[\mathcal{T}] \to 0$.

Proof. Since $\tilde{m} \geq m$, the first inequality follows from the positivity of H_a, so it remains to prove the second inequality.

To prove the second inequality, note that

$$H_a Q_{\mathcal{T}} m(x) = \sum_K \rho(K) \int_K h_a(x, y)\, \mathrm{d}y \quad \text{and}$$

$$H_a Q_{\mathcal{T}} \tilde{m}(x) = \sum_K \rho(K) \sum_{L : L \simeq K} \int_L h_a(x, y)\, \mathrm{d}y \ .$$

We claim that there exists a constant $C < \infty$ such that

$$h_a(x, y) \leq e^{C\sqrt{\eta(a)}|y-z|} h_a(x, z) \tag{3.9}$$

for $x, y, z \in \overline{\Omega}$.

Let T_{KL} be a (not necessarily optimal) transport map between the uniform probability measures on neighbouring cells K and L. As $|T_{KL}(z) - z| \leq 2[\mathcal{T}]$ for $z \in K$, the claim yields

$$\frac{1}{|L|} \int_L h_a(x, y)\, \mathrm{d}y = \frac{1}{|K|} \int_K h_a(x, T_{KL}(z))\, \mathrm{d}z \leq e^{C\sqrt{\eta(a)}[\mathcal{T}]} \frac{1}{|K|} \int_K h_a(x, z)\, \mathrm{d}z \ .$$

Since $|L| \leq C|K|$ by the ζ-regularity estimate (2.22), and since $\#\{L : L \simeq K\} \leq C$ by Lemma 2.12, we obtain

$$\sum_{L : L \simeq K} \int_L h_a(x, y)\, \mathrm{d}y \leq C e^{C\sqrt{\eta(a)}[\mathcal{T}]} \int_K h_a(x, z)\, \mathrm{d}z \ ,$$

which yields the result.

It remains to prove the claim (3.9). For this purpose, note that by the heat kernel bounds (2.6) there exist $\overline{\Omega}$-dependent constants $c, C > 0$ such that, for all $a > 0$,

$$\sup_{x,y\in\Omega} h_{a/2}(x, y) \leq C(1 \vee a^{-d/2}) \quad \text{and} \quad \inf_{x,y\in\Omega} h_a(x, y) \geq C a^{-d/2} e^{-c/a} \ .$$

For any smooth function $\phi : \Omega \to (0, M]$ with $M \in (0, \infty)$, the local logarithmic Sobolev inequality (2.5) implies that

$$|\nabla \log H_{a/2}\phi|^2 \leq \frac{2}{a} \log\left(\frac{M}{H_{a/2}\phi}\right) \ .$$

Applying this inequality with $\phi = h_{a/2}(x, \cdot)$, we obtain using the semigroup property,

$$\sup_{x,y\in\Omega} |\nabla_y \log h_a(x, y)|^2 \leq \frac{2}{a} \log\left(\frac{\sup_{x,y\in\Omega} h_{a/2}(x, y)}{\inf_{x,y\in\Omega} h_a(x, y)}\right) \leq C\eta(a) \ , \tag{3.10}$$

which implies (3.9). $\qquad \square$

4 Finite volume estimates for discrete optimal transport

The goal of this section is to show that the dual action functional $\mathcal{A}_{\mathcal{T}}^*$ for the discrete transport problems is a good approximation to its continuous counterpart $\mathbb{A}^*$. This will be shown in Proposition 4.6.

To obtain this result, we first show an error estimate for a discrete elliptic problem, in the spirit of [17, Theorem 3.5]. In our application, we think of w as being $\dot\mu_t$ at some fixed time t, so that the elliptic equation below is the continuity equation at time t.

As in Section 3, we fix a convex bounded open set $\Omega \subseteq \mathbb{R}^d$ and a ζ-regular mesh $\mathcal{T}$ for some $\zeta \in (0, 1]$.

Proposition 4.1 (Weighted H^1-error bound). *Let $w \in L^2(\Omega)$ with $\int_\Omega w(x)\,\mathrm{d}x = 0$ and $\mu = u\,\mathrm{d}x \in \mathcal{P}_\delta(\overline\Omega)$ be given, and let $\phi \in H^2(\Omega)$ be the unique variational solution to*

$$\begin{cases} -\operatorname{div}(u\nabla\phi) = w & in\ \Omega \\ \partial_n\phi = 0 & on\ \partial\Omega \end{cases} \tag{4.1}$$

satisfying $\int_\Omega \phi\,\mathrm{d}x = 0$.

Define $m \in \mathcal{P}(\mathcal{T})$ by $m = P_{\mathcal{T}}\mu$, and $\sigma \in \mathcal{M}_0(\mathcal{T})$ by $\sigma := P_{\mathcal{T}}w$. We write $\rho(K) := m(K)/|K|$ and $\widehat\rho(K,L) := \theta_{KL}(\rho(K), \rho(L))$. Let $\psi : \mathcal{T} \to \mathbb{R}$ be the unique solution to the corresponding discrete elliptic problem

$$-\sum_{L \in \mathcal{T}} \frac{|(K|L)|}{d_{KL}}\widehat\rho(K,L)\big(\psi(L) - \psi(K)\big) = \sigma(K) \tag{4.2}$$

satisfying $\sum_{K \in \mathcal{T}} |K|\psi(K) = 0$.

Set $\overline\phi(K) = \fint_{B_K} \phi\,\mathrm{d}x$, where $B_K = B(x_K, \zeta[\mathcal{T}])$ denotes the closed ball of radius $\zeta[\mathcal{T}]$ around x_K, and set

$$e(K) := \overline\phi(K) - \psi(K)\ .$$

Then there exists a constant $C < \infty$ depending only on δ, Ω, and ζ, such that

$$\mathcal{A}_{\mathcal{T}}(m, e) \leq C[\mathcal{T}]^2\,\|w\|_{L^2(\Omega)}^2\ . \tag{4.3}$$

Remark 4.2. The existence of a unique variational solution ϕ to the Neumann problem (4.1) in $H^1(\Omega)$ follows from the Lax–Milgram theorem. The existence of a unique solution to (4.2) follows from elementary linear algebra, cf. [24]. In both cases, uniqueness holds up to an additive constant.

Remark 4.3. Crucial for the proof is the *a priori* estimate $\|\phi\|_{H^2(\Omega)} \leq C\|w\|_{L^2(\Omega)}$ with $C < \infty$ depending only on Ω and δ; see [21, Theorem 3.1.2.3].

Remark 4.4. The error estimate (4.3) is similar to the H^2-error estimate in [17], except that we use an averaged sample $\fint_{B_K} \phi$ instead of the pointwise sample $\phi(x_K)$ to define the error term. This change is required to be able to deal with dimensions $d \geq 4$, where $H^2(\Omega)$ does not embed into the space of continuous functions.

Proof of Proposition 4.1. Integration of (4.1) over $K \in \mathcal{T}$ yields

$$-\sum_{L:L\sim K} \int_{(K|L)} u\partial_n\phi\,\mathrm{d}S = \int_K w\,\mathrm{d}x = \sigma(K)\ , \tag{4.4}$$

where n denotes the outward unit normal on ∂K. We define

$$R(K,L) := \frac{1}{\widehat{\rho}(K,L)} \fint_{(K|L)} u \partial_n \phi \, \mathrm{d}S - \frac{\overline{\phi}(L) - \overline{\phi}(K)}{d_{KL}} \, ,$$

and note that, by (4.2) and (4.4),

$$\sum_{L:L \sim K} \frac{|(K|L)|}{d_{KL}} \widehat{\rho}(K,L)\big(e(K) - e(L)\big)$$

$$= \sum_{L:L \sim K} \left[|(K|L)| \widehat{\rho}(K,L)\Big(R(K,L) - \frac{1}{\widehat{\rho}(K,L)} \fint_{(K|L)} u \partial_n \phi \, \mathrm{d}S \Big) \right] - \sigma(K)$$

$$= \sum_{L:L \sim K} |(K|L)| \widehat{\rho}(K,L) R(K,L) \, .$$

Multiplying this expression by $e(K)$, and using the symmetry of $\widehat{\rho}(K,L)$ and the antisymmetry of $R(K,L)$, we obtain

$$\mathcal{A}_{\mathcal{T}}(m,e) = \frac{1}{2} \sum_{K,L} \frac{|(K|L)|}{d_{KL}} \widehat{\rho}(K,L)\big(e(K) - e(L)\big)^2$$

$$= \sum_K e(K) \sum_{L:L \sim K} \frac{|(K|L)|}{d_{KL}} \widehat{\rho}(K,L)\big(e(K) - e(L)\big)$$

$$= \sum_K e(K) \sum_{L:L \sim K} |(K|L)| \widehat{\rho}(K,L) R(K,L)$$

$$= \frac{1}{2} \sum_{K,L} |(K|L)| \widehat{\rho}(K,L) R(K,L)\big(e(K) - e(L)\big)$$

$$\leq \sqrt{\frac{1}{2} \mathcal{A}_{\mathcal{T}}(m,e) \sum_{K,L} d_{KL} |(K|L)| \widehat{\rho}(K,L) R(K,L)^2} \, .$$

Consequently,

$$\mathcal{A}_{\mathcal{T}}(m,e) \leq \frac{1}{2} \sum_{K,L} d_{KL} |(K|L)| \widehat{\rho}(K,L) R(K,L)^2 \, . \tag{4.5}$$

Observe that

$$|R(K,L)| \leq \fint_{(K|L)} \left| \frac{u}{\widehat{\rho}(K,L)} - 1 \right| |\partial_n \phi| \, \mathrm{d}S + \left| \fint_{(K|L)} \partial_n \phi \, \mathrm{d}S - \frac{\overline{\phi}(L) - \overline{\phi}(K)}{d_{KL}} \right| \, . \tag{4.6}$$

To estimate the first term on the right-hand side of (4.6), we note that the function u is Lipschitz since $\mu \in \mathcal{P}_\delta(\overline{\Omega})$, and the mean θ_{KL} is Lipschitz on $[\delta, \infty)^2$. Using this observation followed by Lemma 2.13, we obtain

$$\fint_{(K|L)} \left| \frac{u}{\widehat{\rho}(K,L)} - 1 \right| |\partial_n \phi| \, \mathrm{d}S \leq C[\mathcal{T}] \fint_{(K|L)} |\partial_n \phi| \, \mathrm{d}S$$

$$\leq C[\mathcal{T}] \left(\fint_K |\nabla \phi| \, \mathrm{d}x + [\mathcal{T}] \sqrt{\fint_K |D^2 \phi|^2 \, \mathrm{d}x} \right) \, , \tag{4.7}$$

for some constant $C < \infty$ depending on Ω, δ and ζ.

To estimate the second integral in (4.6) for neigbouring cells $K \sim L$, write $x_{KL} = x_L - x_K$, so that $n = \frac{x_{KL}}{d_{KL}}$. We set

$$M_K := \fint_K \nabla \phi \cdot n \ , \qquad M_{KL} := \fint_{K \cup L} \nabla \phi \cdot n \ .$$

Using the trace inequality (2.24) we have

$$\left| \fint_{(K|L)} \partial_n \phi \, \mathrm{d}S - M_K \right|^2 \le \fint_{(K|L)} (\partial_n \phi - M_K)^2 \, \mathrm{d}S \le C[\mathcal{T}] \frac{|K|}{|\partial K|} \fint_K |D^2 \phi|^2 \, \mathrm{d}x \ . \qquad (4.8)$$

Arguing as in (2.26), we obtain

$$|M_K - M_{KL}| \le C[\mathcal{T}] \sqrt{\fint_{K \cup L} |D^2 \phi|^2 \, \mathrm{d}x} \ . \qquad (4.9)$$

Furthermore, writing $B_K = B(x_K, \zeta[\mathcal{T}])$ as before, the fundamental theorem of calculus yields

$$\frac{\overline{\phi}(L) - \overline{\phi}(K)}{d_{KL}} = \fint_{B_K} \frac{\phi(y + x_{KL}) - \phi(y)}{d_{KL}} \, \mathrm{d}y$$

$$= \fint_{B_K} \int_0^1 \partial_n \phi(y + t x_{KL}) \, \mathrm{d}t \, \mathrm{d}y = \fint_{K \cup L} \partial_n \phi(x) f(x) \, \mathrm{d}x \ ,$$

where f is a nonnegative function satisfying $\fint_{B_K} f(x) \, \mathrm{d}x = 1$ and $\|f\|_{L^\infty} \le C < \infty$. Therefore, using the Poincaré inequality (2.23),

$$\left| \frac{\overline{\phi}(L) - \overline{\phi}(K)}{d_{KL}} - M_{KL} \right| \le \fint_{K \cup L} |\partial_n \phi(x) - M_{KL}| f(x) \, \mathrm{d}x$$

$$\le \|f\|_{L^\infty} \sqrt{\fint_{K \cup L} |\partial_n \phi(x) - M_{KL}|^2 \, \mathrm{d}x} \qquad (4.10)$$

$$\le C[\mathcal{T}] \sqrt{\fint_{K \cup L} |D^2 \phi|^2 \, \mathrm{d}x} \ .$$

Combining the inequalities (4.8), (4.9) and (4.10), we obtain, using the ζ-regularity once more,

$$\left| \fint_{(K|L)} \partial_n \phi \, \mathrm{d}S - \frac{\overline{\phi}(L) - \overline{\phi}(K)}{d_{KL}} \right| \le C[\mathcal{T}] \sqrt{\fint_{K \cup L} |D^2 \phi|^2 \, \mathrm{d}x} \ .$$

Together with (4.7), the latter estimate yields

$$|R(K, L)| \le C[\mathcal{T}] \sqrt{\fint_{K \cup L} |\nabla \phi|^2 + |D^2 \phi|^2 \, \mathrm{d}x} \ .$$

Thus, using (4.5) we find

$$\mathcal{A}_{\mathcal{T}}(m, e) \le \frac{1}{2} \sum_{K, L} d_{KL} |(K|L)| \widehat{\rho}(K, L) R(K, L)^2$$

$$\le C[\mathcal{T}]^2 \sum_{K, L} \frac{d_{KL} |(K|L)|}{|K \cup L|} \int_{K \cup L} |\nabla \phi|^2 + |D^2 \phi|^2 \, \mathrm{d}x \qquad (4.11)$$

$$\le C[\mathcal{T}]^2 \|\phi\|_{H^2(\Omega)}^2 \ ,$$

23

since $d_{KL}(K|L)/|K \cup L| \leq C$, and the maximum number of neigbours of any cell is bounded in view of Lemma 2.12. The result thus follows using the *a priori* bound from Remark 4.3. $\qquad \square$

In the following result it suffices to assume that $\mathcal{T}$ is admissible. We do not need to require ζ-regularity.

Proposition 4.5 (Discrete weighted Poincaré inequality)**.** *There exists a constant $C < \infty$ depending only on Ω such that for all $\psi : \mathcal{T} \to \mathbb{R}$ satisfying $\sum_K |K|\psi(K) = 0$, and all $m \in \mathcal{P}(\mathcal{T})$ with $\rho(K) := \frac{m(K)}{|K|} \geq \delta > 0$ for all $K \in \mathcal{T}$, we have*

$$\sum_{K \in \mathcal{T}} |K|\psi(K)^2 \leq \frac{C}{\delta} \mathcal{A}_{\mathcal{T}}(m, \psi) .$$

Proof. This is a straightforward modification of the proof in [17, Lemma 3.7]. Define $\phi : \overline{\Omega} \to \mathbb{R}$ by $\phi = \sum_K \chi_K \psi(K)$ and set $\widehat{\rho}(K, L) = \theta_{KL}(\rho(K), \rho(L))$. We need to show that

$$\frac{1}{2|\Omega|} \int_{\Omega} \int_{\Omega} |\phi(x) - \phi(y)|^2 \, \mathrm{d}x \, \mathrm{d}y \leq \frac{C}{2\delta} \sum_{K,L} \frac{|(K|L)|}{d_{KL}} \widehat{\rho}(K, L)(\psi(K) - \psi(L))^2 .$$

For $K \sim L$ and $x, y \in \mathbb{R}^d$, put $\chi_{(K|L)}(x, y) = 1$ if x, y belong to Ω, $(K|L)$ intersects the straight line segment connecting x and y, and $(y - x) \cdot (x_L - x_K) > 0$. Otherwise, we set $\chi_{(K|L)}(x, y) = 0$. For $K \sim L$ and $z \in \mathbb{R}^d$, we set $c_{K,L;z} := \frac{z}{|z|} \cdot \frac{x_L - x_K}{d_{KL}}$. As Ω is convex, we obtain for a.e. $x, y \in \Omega$,

$$|\phi(x) - \phi(y)| \leq \sum_{K,L} |\psi(L) - \psi(K)| \chi_{(K|L)}(x, y) .$$

Note that $c_{K,L;y-x} > 0$ whenever $\chi_{(K|L)}(x, y) > 0$. Using this fact, the Cauchy–Schwarz inequality yields

$$|\phi(x) - \phi(y)|^2 \leq \left(\sum_{K,L} \frac{|\psi(L) - \psi(K)|^2}{c_{K,L;y-x}} \frac{\widehat{\rho}(K, L)}{d_{KL}} \chi_{(K|L)}(x, y) \right)$$
$$\times \left(\sum_{K,L} c_{K,L;y-x} \frac{d_{KL}}{\widehat{\rho}(K, L)} \chi_{(K|L)}(x, y) \right) .$$

For fixed x and y, let $K_0, \ldots, K_N$ be the subsequent cells intersecting the line segment $\{(1 - t)x + ty\}_{t \in [0,1]}$ as t ranges from 0 to 1. By definition, $\chi_{(K|L)}(x, y)$ vanishes, unless $(K|L) = (K_{i-1}|K_i)$ for some $i = 1, \ldots, N$. We thus have

$$\sum_{K,L} c_{K,L;y-x} \frac{d_{KL}}{\widehat{\rho}(K, L)} \chi_{(K|L)}(x, y) = \sum_{i=1}^{N} \frac{c_{K_{i-1},K_i;y-x} |x_{K_i} - x_{K_{i-1}}|}{\widehat{\rho}(K_{i-1}, K_i)}$$
$$\leq \delta^{-1} \sum_{i=1}^{N} c_{K_{i-1},K_i;y-x} |x_{K_i} - x_{K_{i-1}}| = \delta^{-1} \sum_{i=1}^{N} \frac{y - x}{|y - x|} \cdot (x_{K_i} - x_{K_{i-1}})$$
$$= \delta^{-1} \frac{y - x}{|y - x|} \cdot (x_{K_N} - x_{K_0}) \leq \delta^{-1} R .$$

where $R = \operatorname{diam}(\Omega)$. Let B_R denote the ball of radius R around the origin. Using a change of variables, we observe that

$$\int_\Omega \int_\Omega \frac{\chi_{(K|L)}(x,y)}{c_{K,L;y-x}} \,\mathrm{d}x \,\mathrm{d}y \leq \int_{B_R} \frac{1}{c_{K,L;z}} \int_\Omega \chi_{(K|L)}(x, x+z) \,\mathrm{d}x \,\mathrm{d}z$$

$$\leq \int_{B_R} |(K|L)| \, |z| \,\mathrm{d}z = C |(K|L)|$$

for a dimensional constant $C < \infty$. Therefore,

$$\int_\Omega \int_\Omega |\phi(x) - \phi(y)|^2 \,\mathrm{d}x \,\mathrm{d}y$$

$$\leq \delta^{-1} R \int_\Omega \int_\Omega \sum_{K,L} \frac{|\psi(L) - \psi(K)|^2}{c_{K,L;y-x}} \frac{\widehat{\rho}(K,L)}{d_{KL}} \chi_{(K|L)}(x,y) \,\mathrm{d}x \,\mathrm{d}y$$

$$\leq C\delta^{-1} \sum_{K,L} |\psi(L) - \psi(K)|^2 \widehat{\rho}(K,L) \frac{|(K|L)|}{d_{KL}}$$

$$= C\delta^{-1} \mathcal{A}_{\mathcal{T}}(m, \psi) \,,$$

for some Ω-dependent constant $C < \infty$, which completes the proof. $\qquad\square$

Now we are ready to prove the main result of this section.

Proposition 4.6 (Comparison of the dual action functionals)**.** *Let $\delta > 0$. For all $\mu \in \mathcal{P}_\delta(\overline{\Omega})$ and $w \in L^2(\Omega)$ with $\int_\Omega w(x) \,\mathrm{d}x = 0$ we have*

$$\left| \mathcal{A}_{\mathcal{T}}^*(P_{\mathcal{T}}\mu, P_{\mathcal{T}}w) - \mathbb{A}^*(\mu, w) \right| \leq C[\mathcal{T}] \, \|w\|_{L^2(\Omega)}^2 \,,$$

where $C < \infty$ depends only on Ω, ζ, and δ.

Proof. We use the notation from Proposition 4.1. By Remark 4.3 there exists a function $\phi \in H^2(\Omega)$ with $\|\phi\|_{H^2(\Omega)} \leq C\|w\|_{L^2(\Omega)}$ such that

$$-\operatorname{div}(u\nabla\phi) = w \quad \text{and} \quad \partial_n \phi = 0 \,. \tag{4.12}$$

Let $m = P_{\mathcal{T}}\mu$ and $\sigma = P_{\mathcal{T}}w$. As noted in Remark 4.2 there exists $\psi : \mathcal{T} \to \mathbb{R}$ solving

$$-\sum_{L \in \mathcal{T}} \frac{|(K|L)|}{d_{KL}} \widehat{\rho}(K,L)\big(\psi(L) - \psi(K)\big) = \sigma(K) \,. \tag{4.13}$$

Recall that $\mathbb{A}^*(\mu, w) = \mathbb{A}(\mu, \phi)$ and $\mathcal{A}_{\mathcal{T}}^*(m, \sigma) = \mathcal{A}_{\mathcal{T}}(m, \psi)$. Using (4.13) and exploiting symmetry, we obtain

$$\mathcal{A}_{\mathcal{T}}(m, \psi) = \frac{1}{2} \sum_{K,L} \frac{|(K|L)|}{d_{KL}} \widehat{\rho}(K,L)\big(\psi(K) - \psi(L)\big)^2$$

$$= \sum_K \psi(K)\sigma(K)$$

$$= \int_\Omega \phi w \,\mathrm{d}x + \left(\sum_K \overline{\phi}(K)\sigma(K) - \int_\Omega \phi w \,\mathrm{d}x \right) + \sum_K \big(\psi(K) - \overline{\phi}(K)\big)\sigma(K)$$

$$= \mathbb{A}(\mu, \phi) + \sum_K \int_K (\overline{\phi}(K) - \phi) w \,\mathrm{d}x - \sum_K e(K)\sigma(K) \,.$$

It remains to estimate the latter two terms.

To bound the first term, let $\widetilde{\phi}(K) = \fint_K \phi \, dx$, and observe that, using Lemma 2.13 and the Poincaré inequality,

$$\|\overline{\phi}(K) - \phi\|_{L^2(K)} \leq \sqrt{|K|} \, |\overline{\phi}(K) - \widetilde{\phi}(K)| + \|\widetilde{\phi}(K) - \phi\|_{L^2(K)}$$
$$\leq C[\mathcal{T}] \|\nabla \phi\|_{L^2(K)} \ .$$

Therefore, the first term can be bounded by

$$\left| \sum_K \int_K \left(\overline{\phi}(K) - \phi\right) w \, dx \right| \leq \sum_K \|\overline{\phi}(K) - \phi\|_{L^2(K)} \|w\|_{L^2(K)}$$
$$\leq C[\mathcal{T}] \, \|\phi\|_{H^1(\Omega)} \|w\|_{L^2(\Omega)} \ .$$

To estimate the second term, we use Proposition 4.5 and Proposition 4.1 to obtain

$$\sum_K e(K) \sigma(K) \leq \sqrt{\sum_K \frac{\sigma(K)^2}{|K|}} \sqrt{\sum_K |K| e(K)^2}$$
$$\leq C \|w\|_{L^2(\Omega)} \sqrt{\mathcal{A}_{\mathcal{T}}(\rho, e)}$$
$$\leq C[\mathcal{T}] \, \|w\|_{L^2(\Omega)}^2 \ .$$

Combining these estimates yields the result. $\qquad\square$

5 Counterexamples to Gromov–Hausdorff convergence

In this section we show that if the asymptotic isotropy condition fails sufficiently often, then the discrete transport metric $\mathcal{W}_{\mathcal{T}}$ does *not* converge to the 2-Kantorovich metric $\mathbb{W}_2$, in spite of the fact that the discrete heat flow converges to the continuous heat flow; see, e.g., [17, Theorem 4.2]. In fact, in the one-dimensional example below, even evolutionary Γ-convergence has been proved for the entropic gradient flow structure of the discrete heat flow with respect to the transport distance $\mathcal{W}_{\mathcal{T}}$; cf. [10].

5.1 A one-dimensional counterexample

We present a one-dimensional example to illustrate the non-convergence to $\mathbb{W}_2$ in the simplest possible setting.

We start with a well-known result on the existence of smooth $\mathbb{W}_2$-geodesics in the one-dimensional case. For the convenience of the reader we include a direct proof. We write $\mathcal{I} = [0, 1]$ for brevity.

Lemma 5.1. *Let $\delta > 0$, and let $\mu_0, \mu_1 \in \mathcal{P}_\delta(\mathcal{I})$ with densities $u_0, u_1 \in \mathcal{C}^0(\mathcal{I})$ respectively. Then there exist constants $\tilde{\delta} > 0$ and $C < \infty$ depending only on $\delta > 0$, such that the unique $\mathbb{W}_2$-geodesic $(\mu_t)_{t \in [0,1]}$ connecting μ_0 and μ_1 satisfies $\mu_t \in \mathcal{P}_{\tilde{\delta}}(\mathcal{I})$ and $d\mu_t(x) = u_t(x) \, dx$ for all $t \in [0, 1]$, with $\sup_{t \in [0,1]} \|\dot{u}_t\|_{\mathcal{C}^0(\mathcal{I})} \leq C$.*

Proof. Let F_i denote the distribution function of μ_i given by $F_i(x) = \int_a^x u_i(y) \, dy$, which is readily seen to be invertible. The unique optimal transport map T between μ_0 and μ_1 is then given by $T = F_1^{-1} \circ F_0$. By the inverse function theorem, $T \in \mathcal{C}^1(\mathcal{I})$ and $T'(x) \in [M^{-1}, M]$ for all $x \in \mathcal{I}$, where $M > 1$ depends on δ. The unique $\mathbb{W}_2$-geodesic

between μ_0 and μ_1 is given by $\mu_t = (T_t)_{\#}\mu_0$, where $T_t(x) = (1-t)x + tT(x)$, hence the density u_t of μ_t satisfies

$$u_t(x) = \frac{u_0(T_t^{-1}(x))}{T_t'(T_t^{-1}(x))} \ .$$

The result follows directly from this explicit expression. $\qquad\square$

For $N \in \mathbb{N}$ and $r \in (0, \frac{1}{2})$, we consider the $\frac{1}{N}$-periodic mesh $\mathcal{T}_{r,N}$ of $\mathcal{I} = [0, 1]$ from Figure 1, given by

$$\mathcal{T}_{r,N} = \left\{ \left[\frac{k}{N}, \frac{k+r}{N} \right), \left[\frac{k+r}{N}, \frac{k+1}{N} \right) \ : \ 0 \le k < N-1 \right\} \ .$$

The cells in $\mathcal{T}_{r,N}$ will be denoted K_k for $k = 0, \ldots, 2N-1$ according to their natural ordering. To make sure that $\mathcal{T}_{r,N}$ is a partition of $[0, 1]$, one should add the point 1 to the set K_{2N-1}, but this will be irrelevant in what follows. Let $x_k = \frac{r+k}{2N}$ be the midpoints of K_k, so that $d_{k,k+1} = \frac{1}{2N}$ and $[\mathcal{T}_{r,N}] = \frac{1-r}{N}$. (For notational simplicity we write $d_{k,k+1}$ instead of $d_{K_k,K_{k+1}}$. Similarly, we write $P_{r,N}$ instead of $P_{\mathcal{T}_{r,N}}$ etc.) According to (2.27), the transition rates $R_{k,k\pm1}$ from cell k to cell $k \pm 1$ are given by

$$R_{k,k\pm1} = \begin{cases} \frac{2N^2}{1-r}, & k \text{ is odd} , \\ \frac{2N^2}{r}, & k \text{ is even} . \end{cases}$$

with the understanding that $R_{0,-1} = R_{N,N+1} = 0$.

We fix an admissible mean θ (in the sense of Definition 2.3) that is assumed to be symmetric, i.e., $\theta(a, b) = \theta(b, a)$, and consider the transport metric $\mathcal{W}_{r,N}$ defined by setting $\theta_{KL} = \theta$ for all $K \sim L$. For each fixed $r \in (0, \frac{1}{2})$, the next result implies that the distances $\mathcal{W}_{r,N}$ do not Gromov–Hausdorff converge to $\mathbb{W}_2$. The idea of the proof is to add a suitable energy-reducing oscillation to the density of a smooth competitor; see Figure 3 below.

In Section 6 we will show that Gromov–Hausdorff convergence holds if one takes a different (non-symmetric) mean adapted to the inhomogeneity of the mesh.

Proposition 5.2. *Fix $r \in (0, \frac{1}{2})$ and $\delta > 0$. Then there exists a constant $\varepsilon \in (0, 1)$ depending only on r and δ, such that for any $\mu_0, \mu_1 \in \mathcal{P}_\delta(\mathcal{I})$,*

$$\limsup_{N\to\infty} \mathcal{W}_{r,N}(P_{r,N}\mu_0, P_{r,N}\mu_1) \le (1-\varepsilon)\mathbb{W}_2(\mu_0, \mu_1) \ . \tag{5.1}$$

Proof. We divide the proof into several steps.

Step 1. Fix $r \in (0, \frac{1}{2})$, $\delta \in (0, 1)$, and $N \ge 1$. For $\mu \in \mathcal{P}_\delta(\mathcal{I})$, set $m = P_{r,N}\mu$, and let ρ be its density given by $\rho(K) = \frac{m(K)}{|K|}$. For $\eta \in (0, \delta)$ we define $m^\eta \in \mathcal{P}(\mathcal{T}_{r,N})$ by

$$m^\eta(K_k) := \begin{cases} m(K_k) + \frac{r(1-r)}{N}\eta & k \text{ is even} , \\ m(K_k) - \frac{r(1-r)}{N}\eta & k \text{ is odd} , \end{cases}$$

so that its density is given by

$$\rho^\eta(K_k) := \begin{cases} \rho(K_k) + (1-r)\eta & k \text{ is even} , \\ \rho(K_k) - r\eta & k \text{ is odd} . \end{cases}$$

If r is small, the density ρ^η increases substantially with η in the small (even) cells, whereas it decreases only moderately in the large (odd) cells.

We claim that, for any $\delta > 0$ and $r \in (0, \frac{1}{2})$, there exists $\eta' > 0$ and $N' < \infty$, such that for any pair of neighbouring cells K and L, and any $\mu \in \mathcal{P}_\delta(\mathcal{I})$,

$$\widehat{\rho^\eta}(K, L) \geq \widehat{\rho}(K, L) + \tfrac{1}{2}\eta(\tfrac{1}{2} - r) , \tag{5.2}$$

whenever $\eta \leq \eta'$ and $N \geq N'$. Here, we write $\widehat{\rho^\eta}(K, L) = \theta(\rho^\eta(K), \rho^\eta(L))$ as usual.

To show this, we assume without loss of generality that K is small and L is large; thus $|K| = \frac{r}{N}$ and $|L| = \frac{1-r}{N}$. Define $f(\eta) := \widehat{\rho^\eta}(K, L)$. The concavity of θ implies that

$$f(\eta) \geq f(0) + \eta f'(\eta) ,$$

thus it suffices to show that $f'(\eta) \geq \frac{1}{2}(\frac{1}{2} - r)$. Since θ is 1-homogeneous, we have $\partial_1\theta(a, a) = \frac{1}{2} = \partial_2\theta(a, a)$ and $\partial_1\theta(a, b) = \partial_1\theta(a/b, 1)$ for all $a, b > 0$. Therefore,

$$f'(\eta) = (1 - r)\partial_1\theta(\rho^\eta(K), \rho^\eta(L)) - r\partial_2\theta(\rho^\eta(K), \rho^\eta(L))$$
$$= (1 - r)\partial_1\theta\left(\frac{\rho^\eta(K)}{\rho^\eta(L)}, 1\right) - r\partial_2\theta\left(1, \frac{\rho^\eta(L)}{\rho^\eta(K)}\right) .$$

Set $\varepsilon := \frac{1}{2}(\frac{1}{2} - r)$ and choose $h > 0$ so small that $|\partial_1\theta(a, 1) - \frac{1}{2}| \leq \varepsilon$ whenever $|a - 1| \leq h$. If η and N^{-1} are chosen sufficiently small (depending on δ and r), we obtain, since $\mu \in \mathcal{P}_\delta(\mathcal{I})$,

$$\left|\frac{\rho^\eta(K)}{\rho^\eta(L)} - 1\right| \leq \frac{|\rho(K) - \rho(L)| + \eta}{\rho^\eta(L)} \leq \frac{N^{-1}\delta^{-1} + \eta}{\delta} \leq h ,$$

and similarly, $\left|\frac{\rho^\eta(L)}{\rho^\eta(K)} - 1\right| \leq h$. Therefore,

$$f'(\eta) \geq (1 - r)\left(\tfrac{1}{2} - \varepsilon\right) - r\left(\tfrac{1}{2} + \varepsilon\right) = \tfrac{1}{2} - r - \varepsilon = \tfrac{1}{2}\left(\tfrac{1}{2} - r\right) ,$$

which proves the claim.

Since there is a constant $C = C(\delta) < \infty$ such that $\widehat{\rho^\eta}(K, L) \leq C$, it follows from the claim that there exists a constant $c = c(\delta) \in (0, 1)$ such that

$$\frac{1}{\widehat{\rho^\eta}(K, L)} \leq \frac{1 - c\eta(\frac{1}{2} - r)}{\widehat{\rho}(K, L)} .$$

Thus, for any $V : \mathcal{T}_{r,N} \times \mathcal{T}_{r,N} \to \mathbb{R}$ we have

$$2N \sum_{k=0}^{2N-1} \frac{V^2(K_k, K_{k+1})}{\widehat{\rho^\eta}(K_k, K_{k+1})} \leq 2N\left(1 - c\eta\left(\tfrac{1}{2} - r\right)\right) \sum_{k=0}^{2N-1} \frac{V^2(K_k, K_{k+1})}{\widehat{\rho}(K_k, K_{k+1})} .$$

Using the notation from (2.16), this means that

$$\mathcal{K}_{r,N}(m^\eta, V) \leq \left(1 - c\eta\left(\tfrac{1}{2} - r\right)\right)\mathcal{K}_{r,N}(m, V) . \tag{5.3}$$

Step 2. Take $\mu_0, \mu_1 \in \mathcal{P}_\delta(\mathcal{I})$ for some $\delta > 0$, and let $(\mu_t)_{t \in [0,1]}$ be the constant speed geodesic connecting μ_0 and μ_1. By Lemma 5.1, there exists $\tilde{\delta} > 0$ such that

$\mu_t \in \mathcal{P}_{\tilde{\delta}}(\mathcal{I})$, and the density u_t of μ_t satisfies $\sup_{t \in [0,1]} \|\dot{u}_t\|_{L^\infty(\mathcal{I})} < \infty$. It then follows from Proposition 4.6 that $m_t := P_{r,N}\mu_t$ satisfies

$$\mathcal{A}_{r,N}^*(m_t, \dot{m}_t) \le \mathbb{A}^*(\mu_t, \dot{\mu}_t) + \frac{C}{N} \ , \tag{5.4}$$

with $C < \infty$ depending on r and δ.

Let $V : \mathcal{T} \times \mathcal{T} \times (0,1) \to \mathbb{R}$ be an anti-symmetric function satisfying the continuity equation (2.15), given by

$$\dot{m}_t(K_k) + 2N\Big(V_t(K_k, K_{k+1}) - V_t(K_{k-1}, K_k)\Big) = 0$$

for all $k = 0, \ldots, 2N-1$, with $V_{0,-1} = V_{2N-1,2N} = 0$. Since $\dot{m}_t^\eta = \dot{m}_t$ for all η, it follows that (m^η, V) solves the continuity equation as well. Therefore, (5.3) yields

$$\mathcal{A}_{r,N}^*(m_t^\eta, \dot{m}_t^\eta) \le \big(1 - c\eta\big(\tfrac{1}{2} - r\big)\big)\mathcal{A}_{r,N}^*(m_t, \dot{m}_t)$$

Combining this bound with (5.4), we infer that there exists a constant $c \in (0,1)$ depending only on $\tilde{\delta}$, such that

$$\limsup_{N \to \infty} \mathcal{W}_{r,N}(m_0^\eta, m_1^\eta) \le \big(1 - c\eta\big(\tfrac{1}{2} - r\big)\big)\mathbb{W}_2(\mu_0, \mu_1) \ ,$$

provided η is chosen sufficiently small depending on $\tilde{\delta}$ and r.

To finish the argument, we note that Lemma 3.3 yields, for $i = 0,1$,

$$\mathcal{W}_{r,N}(m_i, m_i^\eta) \le C\Big(\mathbb{W}_2(Q_{r,N}m_i, Q_{r,N}m_i^\eta) + \frac{1}{N}\Big) \le C\Big(\frac{1}{N}\sqrt{r\eta} + \frac{1}{N}\Big) \le \frac{C}{N} \ ,$$

where $C < \infty$ depends on δ and r. Consequently, by the triangle inequality,

$$\limsup_{N \to \infty} \mathcal{W}_{r,N}(m_0, m_1) \le \big(1 - c\eta\big(\tfrac{1}{2} - r\big)\big)\mathbb{W}_2(\mu_0, \mu_1) \ ,$$

where $c \in (0,1)$ depends only on $\tilde{\delta}$. This implies (5.1). $\qquad\square$

The construction in the proof of Proposition 5.2 breaks down if we choose mean functions $\theta_{k,k+1}$ adapted to the inhomogeneity of the grid, instead of a fixed symmetric mean θ. Indeed, suppose that $\theta_{2k,2k\pm1}$ is a smooth mean function with weight r in the sense of Definition 2.5, so that $\partial_1\theta_{2k,2k\pm1}(1,1) = r$ and $\partial_2\theta_{2k,2k\pm1}(1,1) = 1 - r$. Typical examples are given in (2.13) with $r = \lambda$. By homogeneity of $\theta_{2k,2k\pm1}$ we have, for any $a > 0$:

$$\partial_\eta\big|_{\eta=0}\theta_{2k,2k\pm1}(a + (1-r)\eta, a - r\eta) = (1-r)\partial_1\theta_{2k,2k\pm1}(a,a) - r\partial_2\theta_{2k,2k\pm1}(a,a) = 0 \ ,$$

hence the concave function $\eta \mapsto \theta_{2k,2k\pm1}(a+(1-r)\eta, a-r\eta)$ attains its maximum at $\eta = 0$. This argument shows that one cannot increase the mean density (and thus decrease the energy) by introducing microscopic density oscillations. This is in sharp contrast to (5.2).
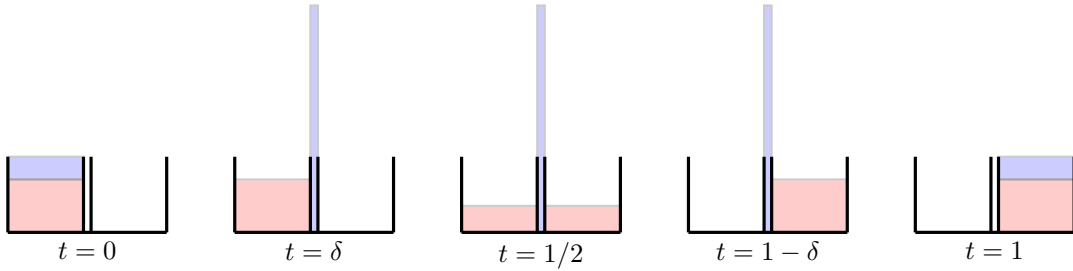
$t = 0$ $t = \delta$ $t = 1/2$ $t = 1 - \delta$ $t = 1$

Figure 3: The picture illustrates the principle behind the proof of Theorem 5.2. It shows how an "unreasonably cheap" transport can be constructed between the measures at time 0 and 1. First, in the interval $[0, \delta]$, a bit of mass is moved into the short interval, so that the mean of the densities in the left and the middle cells increases. Then the bulk of the mass is moved during the interval $[\delta, 1 - \delta]$, after which the short intervals are emptied in the interval $[1 - \delta, 1]$. The first and the final phase are cheap because very little mass is moved, and the middle phase is cheap since the mean of the densities is kept high by the mass in the middle cell.

5.2   Necessity of the asymptotic isotropy condition

Our next aim is to show that for any family of meshes $\{\mathcal{T}\}$ for which the asymptotic isotropy condition fails at every scale, the distance $\mathcal{W}_\mathcal{T}$ is asymptotically strictly smaller than $\mathbb{W}_2$.

We start with a lemma that guarantees the existence of certain smooth $\mathbb{W}_2$-geodesics that transport mass in a parallel fashion.

Lemma 5.3. *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded open set with Lipschitz boundary, let $x_0 \in \Omega$, and $v \in S^{d-1}$. Then there exist $r > 0$, $\delta > 0$, $\kappa > 0$, and a $\mathbb{W}_2$-geodesic $(\mu_t)_{t \in [0,1]} \subseteq \mathcal{P}_\delta(\overline{\Omega})$ with the following properties:*

(i) the continuity equation

$$\begin{cases} \partial_t \mu_t + \operatorname{div}(\mu_t \nabla \phi_t) = 0 & in \ \Omega \ , \\ \nabla \phi_t \cdot n = 0 & on \ \partial\Omega \ , \end{cases} \tag{5.5}$$

holds for some vector field $\phi \in \mathcal{C}^1([0,1] \times \overline{\Omega})$ satisfying $\nabla \phi_t(x) = \kappa v$ for all $t \in [0, 1]$ and $x \in B(x_0, r)$;

(ii) $\partial_t \mu_t \in \mathcal{C}^0(\overline{\Omega})$ for all $t \in [0, 1]$, and $\sup_{t \in [0,1]} \|\partial_t \mu_t\|_{\mathcal{C}^0(\Omega)} < \infty$.

Proof. Fix an open ball $B = B(x_0, r) \subseteq \Omega$ and let $\eta \in \mathcal{C}_c^\infty(\mathbb{R}^d)$ be a nonnegative function, supported in the unit ball $B(0, 1)$, satisfying $\eta(x) = 1$ for $x \in B(0, \frac{1}{2})$. We define

$$\phi_0(x) = v \cdot (x - x_0) \, \eta\left(\frac{x - x_0}{r}\right) \,,$$

so that $\phi_0 \in \mathcal{C}^\infty(\overline{\Omega})$ with support contained in $B(x_0, r)$, and $\nabla \phi_0(x) = v$ for all $x \in B(x_0, \frac{r}{2})$.

Since ϕ_0 is smooth, there exists $T > 0$ such that the Hamilton–Jacobi equation $\partial_t \phi_t + \frac{1}{2} |\nabla \phi_t|^2 = 0$ has a unique solution in $\mathcal{C}^1([0, T] \times \mathbb{R}^d)$ with initial value ϕ_0. It follows from the Hopf–Lax formula $\phi_t(x) = \inf_y \{\phi_0(y) + \frac{1}{2t} |x - y|^2\}$ that the following properties hold for all $t \in [0, T]$, provided $T > 0$ is sufficiently small:

30

- supp $\phi_t \subseteq B(x_0, r)$;
- $\nabla \phi_t(x) = v$ for all $x \in B(x_0, \frac{r}{4})$.

Let $\mu_0 \in \mathcal{P}(\overline{\Omega})$ be the normalised Lebesgue measure, and set $\mu_t = (I_d + t\nabla\phi_0)_{\#}\mu_0$ for $t \in [0, T]$. Then $(\mu_t, \phi_t)_t$ solves the continuity equation (5.5). Moreover, the density ρ_t of μ_t solves the Monge-Ampère equation

$$\rho_0(x) = \rho_t(x + t\nabla\phi_0(x)) \det(I + tD^2\phi_0(x)) .$$

It follows from this expression that there exist $T > 0$ and $\delta > 0$ such that $\mu_t \in \mathcal{P}_\delta(\Omega)$ and $\partial_t \mu_t \in \mathcal{C}^0(\Omega)$ for all $t \in [0, T]$, with $\sup_{t \in [0,T]} \|\partial_t \mu_t\|_{\mathcal{C}^0} < \infty$.

To obtain the result, it remains to rescale the geodesic in time. In doing so, we replace ϕ_t by $T\phi_t$, so that $\nabla\phi_t(x) = \kappa v$ for $x \in B(x_0, \frac{r}{4})$ and $t \in [0, 1]$, with $\kappa = T$. Replacing $\frac{r}{4}$ by r, the result follows. $\square$

The following lemma asserts that, at the macroscopic scale, the isotropy condition holds without any assumption on the mesh. For $A \subseteq \mathbb{R}^d$ and $r > 0$, we let $B(A, r) := \bigcup_{x \in A} B(x, r)$ denote the r-neigbourhood of A.

Lemma 5.4. *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded convex domain. Let $\mathcal{T}$ be an admissible mesh on Ω, and let $\boldsymbol{\lambda}$ a weight function on $\mathcal{T}$. For any open subset $U \subseteq \Omega$ and any unit vector $v \in S^{d-1}$, we have*

$$\left| \left(\sum_{K,L \in \mathcal{T}; K, L \subseteq U} \lambda_{KL}(v \cdot n_{KL})^2 |(K|L)| d_{KL} \right) - |U| \right| \leq \left| B(\partial U, 4[\mathcal{T}]) \right| . \tag{5.6}$$

Note that by the symmetry of the summand, the left-hand side does not depend on the choice of the weight function $\boldsymbol{\lambda}$.

Proof. We consider the cells $C_{KL} = C_{LK} \subseteq \mathbb{R}^d$ defined by

$$C_{KL} = \{x \in (K|L) + \mathbb{R}v \subseteq \mathbb{R}^d \ : \ x \cdot v \in \text{conv}(x_K \cdot v, x_L \cdot v)\} .$$

Observe that these sets have pairwise disjoint interiors (up to the symmetry condition $C_{KL} = C_{LK}$). Set $U^- = U \setminus B(\partial U, 4[\mathcal{T}])$ and $U^+ = B(U, 4[\mathcal{T}])$. It then follows that

$$U^- \subseteq \bigcup_{K,L \subseteq U} C_{KL} \subseteq U^+ ,$$

hence $|U^-| \leq \sum_{K,L \subseteq U} \lambda_{KL} |C_{KL}| \leq |U^+|$ since $\lambda_{KL} + \lambda_{LK} = 1$. The result follows, as the area formula yields $|C_{KL}| = (v \cdot n_{KL})^2 |(K|L)| d_{KL}$. $\square$

As the right-hand side in the previous result is small, the contribution of the term $\sum_{L \in \mathcal{T}; L \subseteq U} \lambda_{KL}(v \cdot n_{KL})^2 |(K|L)| d_{KL}$ is equal to $|K|$ on average, up to a microscopically small error. However, it may happen that the isotropy condition fails at the microscopic scale, in the sense that the microscopically small error in (5.6) results from a cancellation of positive and negative contributions of macroscopic size. The following definition makes this intuition precise.

Definition 5.5 (Local anisotropy). *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded convex domain, and let $U \subseteq \Omega$ be a non-empty open subset. Let $\{\mathcal{T}\}$ be a family of ζ-regular meshes on Ω for some $\zeta > 0$, and for each $\mathcal{T}$, let $\boldsymbol{\lambda}^{\mathcal{T}}$ be a weight function on $\mathcal{T}$. We say that $\{\mathcal{T}\}$ is*

locally asymptotically $\{\boldsymbol{\lambda}^{\mathcal{T}}\}$-anisotropic *on U, if there exists a unit vector $v \in S^{d-1}$ and a constant $c > 0$, such that*

$$\liminf_{[\mathcal{T}] \to 0} \sum_{K \in \mathcal{T}; K \subseteq V} \left(\left(\sum_{L \in \mathcal{T}; L \subseteq V} \lambda_{KL}(v \cdot n_{KL})^2 |(K|L)| d_{KL} \right) - |K| \right)_+ \geq c|V| , \qquad (5.7)$$

for any open cube $V \subseteq U$.

Example 5.6 (Anisotropy in one dimension). For $r \in (0, \frac{1}{2})$ and $N \geq 1$, consider the one-dimensional periodic mesh $\mathcal{T}_{r,N}$ from Section 5.1. We fix $s \in [0, 1]$ and define a weight function $\boldsymbol{\lambda}$ on $\mathcal{T}_{r,N}$ by setting $\lambda_{KL} = s$ is K is small, and $\lambda_{KL} = 1 - s$ if K is large. As large and small cells alternate, this indeed defines a weight function.

Fix an interval $V = (a, b)$ for some $0 < a < b < 1$, and set $v = 1$. For any $K \in \mathcal{T}_{r,N}$ and $[\mathcal{T}]$ sufficiently small, it follows that

$$S_K := \sum_{L \in \mathcal{T}_{r,N}; L \subseteq V} \lambda_{KL}(v \cdot n_{KL})^2 |(K|L)| d_{KL} = \begin{cases} \frac{s}{N} & \text{if } K \text{ is small} , \\ \frac{1-s}{N} & \text{if } K \text{ is large} . \end{cases}$$

Note that for any neighbouring pair K, L, we have $S_K + S_L = \frac{1}{N} = |K| + |L|$. This means that the isotropy condition holds on average, in accordance with Proposition 5.4. However, it follows that

$$\liminf_{N \to \infty} \sum_{K \in \mathcal{T}_{r,N}; K \subseteq V} \left(S_K - |K| \right)_+ = |s - r|(b - a) .$$

Therefore, the local anisotropy condition (5.7) holds whenever $r \neq s$. If $r = s$, we have already seen in the introduction that the asymptotic isotropy condition (and in fact the centre-of-mass condition) holds.

Example 5.7 (Anisotropy in a 2-dimensional example). Consider the crossed square grid from Figure 4 with $[\mathcal{T}] = \frac{1}{N}$. It follows that $|(K|L)| = \frac{1}{N}$ in the coordinate directions, and $|(K|L)| = \frac{1}{N\sqrt{2}}$ in diagonal directions. We fix $r \in (0, \frac{1}{2})$ and choose the points x_K in such a way that $d_{KL} = \frac{2r}{N}$ if n_{KL} points in one of the coordinate directions. If n_{KL} is in one of the diagonal directions, we then have $d_{KL} = (\frac{1}{2} - r)\frac{\sqrt{2}}{N}$. By symmetry, it is natural to choose $\lambda_{KL} = \frac{1}{2}$ for all $K \sim L$.

For each interior cell K we compute $M_K := \sum_{L \in \mathcal{T}; L \sim K} \lambda_{KL} |(K|L)| d_{KL} n_{KL} \otimes n_{KL}$. Denoting the cells by N, E, S and W, we have

$$M_N = M_S = \frac{1}{2} \frac{2r}{N^2} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{2} \frac{\frac{1}{2} - r}{N^2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \frac{1}{4N^2} \begin{bmatrix} 1 - 2r & 0 \\ 0 & 1 + 2r \end{bmatrix} .$$

An analogous computation shows that

$$M_E = M_W = \frac{1}{4N^2} \begin{bmatrix} 1 + 2r & 0 \\ 0 & 1 - 2r \end{bmatrix} .$$

We thus find that

$$M_N + M_E + M_S + M_W = \frac{1}{N^2} I = \left(|N| + |E| + |S| + |W| \right) I ,$$

in accordance to the fact that isotropy holds on average, for any $r \in (0, \frac{1}{2})$.

To show that the family $\{\mathcal{T}_{r,N}\}_N$ is locally anisotropic, we fix $v = (1, 0)$. Then:

$$(v \cdot M_K v - |K|)_+ = \begin{cases} 0 & \text{if } K = N \text{ or } K = S \ , \\ \frac{r}{2N^2} & \text{if } K = E \text{ or } K = W \ . \end{cases}$$

It follows that, for any cube V,

$$\liminf_{N \to \infty} \sum_{K \in \mathcal{T}_{r,N}; K \subseteq V} \left(\left(\sum_{L \in \mathcal{T}_{r,N}; L \subseteq V} \lambda_{KL} (v \cdot n_{KL})^2 |(K|L)| d_{KL} \right) - |K| \right)_+ = r|V| \ ,$$

which shows that the mesh is everywhere locally anisotropic, for any $r \in (0, \frac{1}{2})$.

The following proposition shows that if the mesh is locally $\{\boldsymbol{\lambda}^{\mathcal{T}}\}$-anisotropic, and if the mean functions $\theta^{\mathcal{T}}$ are chosen accordingly, then the discrete transport distances are asymptotically strictly smaller than $\mathbb{W}_2$.

Theorem 5.8 (Necessity of asymptotic isotropy)**.** *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded convex domain. Let $\{\mathcal{T}\}$ be a family of ζ-regular meshes on Ω for some $\zeta > 0$, and assume that $\{\mathcal{T}\}$ is locally anisotropic on U for some weight functions $\{\boldsymbol{\lambda}^{\mathcal{T}}\}$. Let $\{\boldsymbol{\theta}^{\mathcal{T}}\}$ be a family of mean functions satisfying $\partial_1 \theta_{KL}^{\mathcal{T}}(1, 1) = \lambda_{KL}$ for any $\mathcal{T}$ and any $K, L \in \mathcal{T}$, and suppose that the regularity condition*

$$\sup_{\mathcal{T}} \sup_{K, L \in \mathcal{T}} \|D^2 \theta_{KL}^{\mathcal{T}}\|_{L^\infty(B((1,1),s))} < \infty \tag{5.8}$$

holds for some $s > 0$. Then there exist $\mu_0, \mu_1 \in \mathcal{P}(\overline{\Omega})$ such that

$$\limsup_{[\mathcal{T}] \to 0} \mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}} \mu_0, P_{\mathcal{T}} \mu_1) < \mathbb{W}_2(\mu_0, \mu_1) \ . \tag{5.9}$$

Remark 5.9. Note that all examples from Section 2.2 satisfy (5.8). However, this condition excludes certain smooth mean functions approximating $\theta(a, b) = \min(a, b)$.

Proof. We fix $x_0 \in \Omega$ and $v \in S^{d-1}$. Using Lemma 5.3 we obtain $\delta > 0$, $r > 0$, and a geodesic $(\mu_t)_{t \in [0,1]} \subseteq \mathcal{P}_\delta(\overline{\Omega})$, solving the continuity equation $\dot{\mu} + \operatorname{div}(\mu \nabla \phi) = 0$ where the velocity vector field $\nabla \phi_t \in L^2(\Omega)$ satisfies $\nabla \phi_t(x) = \kappa v$ for some $\kappa > 0$, for all $t \in [0, 1]$ and all x in the ball $B = B(x_0, r)$. For brevity we write $\tilde{B} = B(x_0, \frac{r}{2})$.

Fix $\ell > 0$, and consider the collection of open cubes given by

$$\mathscr{Q}_\ell := \{\ell(p + (0, 1)^d) \subseteq \tilde{B} \ : \ p \in \mathbb{Z}^d\} \ .$$

For $Q \in \mathscr{Q}_\ell$ we define $\mathcal{T}_Q := \{K \in \mathcal{T} \ : \ K \subseteq Q\}$, and for $K \in \mathcal{T}_Q$ we set

$$S_K := \sum_{L \in \mathcal{T}_Q} \lambda_{KL} (v \cdot n_{KL})^2 |(K|L)| d_{KL} \ .$$

We define the subsets $\mathcal{T}_Q^+, \mathcal{T}_Q^- \subseteq \mathcal{T}$ by

$$\mathcal{T}_Q^\pm := \{K \in \mathcal{T}_Q \ : (S_K - |K|)_\pm > 0\} \ .$$

It follows directly from (5.7) that, for all $Q \in \mathscr{Q}_\ell$,

$$\liminf_{[\mathcal{T}] \to 0} \sum_{K \in \mathcal{T}_Q^+} (S_K - |K|) \geq c|Q| \ . \tag{5.10}$$

Combining this bound with Lemma 5.4, we also find

$$\liminf_{[\mathcal{T}]\to 0} \sum_{K\in\mathcal{T}_Q^-} \big(|K| - S_K\big) \geq c|Q| \, . \tag{5.11}$$

In particular, if $[\mathcal{T}]$ is sufficiently small, both $\mathcal{T}_Q^+$ and $\mathcal{T}_Q^-$ are non-empty.

We define a variation $\nu_\ell^{\mathcal{T}} : \mathcal{T} \to \mathbb{R}$ by

$$\nu_\ell^{\mathcal{T}}(K) = \sum_{Q\in\mathcal{Q}_\ell} \left(\alpha_Q^{\mathcal{T}}\mathbf{1}_{\{K\in\mathcal{T}_Q^+\}} - \beta_Q^{\mathcal{T}}\mathbf{1}_{\{K\in\mathcal{T}_Q^-\}}\right) \, ,$$

where $\alpha_Q^{\mathcal{T}}, \beta_Q^{\mathcal{T}} \in (0,1]$ are the unique numbers such that $\sum_{K\in\mathcal{T}_Q} \nu_\ell^{\mathcal{T}}(K)|K| = 0$ and $\max\{\alpha_Q^{\mathcal{T}},\beta_Q^{\mathcal{T}}\} = 1$ for all $Q \in \mathcal{Q}_\ell$.

Set $m_t^{\mathcal{T}} = P_{\mathcal{T}}\mu_t \in \mathcal{P}(\mathcal{T})$, and let $\rho_t^{\mathcal{T}}(K) = m_t^{\mathcal{T}}(K)/|K|$ be its density as usual. We consider the perturbed measure $m_{\varepsilon,t}^{\mathcal{T}}$ with density $\rho_{\varepsilon,t}^{\mathcal{T}}$ given by

$$\rho_{\varepsilon,t}^{\mathcal{T}}(K) = \rho_t^{\mathcal{T}}(K) + \varepsilon\nu_\ell^{\mathcal{T}}(K) \, ,$$

where we suppress the dependence of $\rho_{\varepsilon,t}^{\mathcal{T}}$ on ℓ in the notation. Note that $m_{\varepsilon,t}^{\mathcal{T}}$ belongs to $\mathcal{P}(\mathcal{T})$ if $0 < \varepsilon < \delta$, since $\mu_t \in \mathcal{P}_\delta(\overline{\Omega})$. Write $\widehat{\rho_{\varepsilon,t}^{\mathcal{T}}}(K,L) = \theta_{KL}\big(\rho_{\varepsilon,t}^{\mathcal{T}}(K), \rho_{\varepsilon,t}^{\mathcal{T}}(L)\big)$. In view of the regularity assumption (5.8) on θ_{KL} and the fact that $\mu_t \in \mathcal{P}_\delta(\overline{\Omega})$, a Taylor expansion yields

$$\frac{1}{\widehat{\rho_{\varepsilon,t}^{\mathcal{T}}}(K,L)} \leq \frac{1}{\widehat{\rho_t^{\mathcal{T}}}(K,L)} - \frac{\gamma_{KL}\varepsilon}{\widehat{\rho_t^{\mathcal{T}}}^2(K,L)} + C\varepsilon\big(\varepsilon + [\mathcal{T}]\big) \tag{5.12}$$

for all $\varepsilon < \varepsilon_0(\delta,s)$, where $\gamma_{KL} = \lambda_{KL}\nu_\ell^{\mathcal{T}}(K) + \lambda_{LK}\nu_\ell^{\mathcal{T}}(L)$, and C depends on δ and s.

Let $\psi_t^{\mathcal{T}} : \mathcal{T} \to \mathbb{R}$ be the solution to the discrete elliptic problem (4.2), and consider the associated momentum vector field

$$V_t^{\mathcal{T}}(K,L) = \widehat{\rho_t^{\mathcal{T}}}(K,L)\big(\psi_t^{\mathcal{T}}(L) - \psi_t^{\mathcal{T}}(K)\big) \, ,$$

so that $\mathcal{A}_{\mathcal{T}}(m_t^{\mathcal{T}}, \psi_t^{\mathcal{T}}) = \mathcal{K}_{\mathcal{T}}(m_t^{\mathcal{T}}, V_t^{\mathcal{T}})$. Using (5.12) we obtain

$$\begin{aligned}
\mathcal{K}_{\mathcal{T}}(m_{\varepsilon,t}^{\mathcal{T}}, V_t^{\mathcal{T}}) \leq{}& \mathcal{K}_{\mathcal{T}}(m_t^{\mathcal{T}}, V_t^{\mathcal{T}}) - \frac{\varepsilon}{2}\sum_{K,L}\gamma_{KL}\frac{|(K|L)|}{d_{KL}}\big(\psi_t^{\mathcal{T}}(L) - \psi_t^{\mathcal{T}}(K)\big)^2 \\
&+ C\varepsilon\big(\varepsilon + [\mathcal{T}]\big)\sum_{K,L}\frac{|(K|L)|}{d_{KL}}V_t^{\mathcal{T}}(K,L)^2 \, .
\end{aligned} \tag{5.13}$$

We will estimate the three terms on the right-hand side separately.

To bound the first term, we apply Proposition 4.6 to obtain

$$\mathcal{K}_{\mathcal{T}}(m_t^{\mathcal{T}}, V_t^{\mathcal{T}}) = \mathcal{A}_{\mathcal{T}}^*(m_t^{\mathcal{T}}, \dot{m}_t^{\mathcal{T}}) \leq \mathbb{A}^*(\mu_t, \dot{\mu}_t) + C[\mathcal{T}]\|\dot{\mu}_t\|_{L^2(\Omega)}^2 \, .$$

Together with the uniform L^2-bound on $\dot{\mu}_t$ from Lemma 5.3, this implies

$$\limsup_{[\mathcal{T}]\to 0} \int_0^1 \mathcal{K}_{\mathcal{T}}(m_t^{\mathcal{T}}, V_t^{\mathcal{T}})\,\mathrm{d}t \leq \limsup_{[\mathcal{T}]\to 0} \int_0^1 \mathbb{A}^*(\mu_t, \dot{\mu}_t)\,\mathrm{d}t = \mathbb{W}_2(\mu_0, \mu_1)^2 \, . \tag{5.14}$$

34

Moreover, as $\mu_t \in \mathcal{P}_\delta(\overline{\Omega})$, we have

$$\frac{1}{2}\sum_{K,L}\frac{|(K|L)|}{d_{KL}}V_t^{\mathcal{T}}(K,L)^2 \leq C\mathcal{K}_{\mathcal{T}}(m_t^{\mathcal{T}}, V_t^{\mathcal{T}}) \,,$$

for some $C < \infty$ depending on δ. Using this estimate and (5.14) we obtain

$$\limsup_{[\mathcal{T}]\to 0}\int_0^1\sum_{K,L}\frac{|(K|L)|}{d_{KL}}V_t^{\mathcal{T}}(K,L)^2\,\mathrm{d}t \leq \mathbb{W}_2(\mu_0,\mu_1)^2 \,, \tag{5.15}$$

which bounds the third term in (5.13).

To treat the second term, we write $\overline{\phi}_K = \fint_{B_K}\phi\,\mathrm{d}x$, where $B_K = B(x_K, \zeta[\mathcal{T}])$. Since $\sup_{K,L}\gamma_{KL} \leq 1$ and $\mu_t \in \mathcal{P}_\delta(\overline{\Omega})$, Proposition 4.1 yields

$$\sum_{K,L}\gamma_{KL}\frac{|(K|L)|}{d_{KL}}\big((\psi_t^{\mathcal{T}}(L) - \psi_t^{\mathcal{T}}(K)) - (\overline{\phi}_L - \overline{\phi}_K)\big)^2$$

$$\leq C\sum_{K,L}\frac{|(K|L)|}{d_{KL}}\widehat{\rho_t^{\mathcal{T}}}(K,L)\big((\psi_t^{\mathcal{T}}(L) - \psi_t^{\mathcal{T}}(K)) - (\overline{\phi}_L - \overline{\phi}_K)\big)^2 \tag{5.16}$$

$$\leq C[\mathcal{T}]^2\|\dot{\mu}_t\|_{L^2(\Omega)}^2 \,,$$

where C depends on Ω, ζ, and δ. Furthermore, for $K \in \mathcal{T}_Q^\pm$ and $L \sim K$, we have $\nabla\phi = \kappa v$ on $K \cup L$, which implies that $\overline{\phi}_L - \overline{\phi}_K = \kappa v \cdot (x_L - x_K)$. Therefore, using the fact that $\sum_{K\in\mathcal{T}}\nu_\ell^{\mathcal{T}}(K)|K| = 0$,

$$\sum_{K,L}\gamma_{KL}\frac{|(K|L)|}{d_{KL}}(\overline{\phi}_L - \overline{\phi}_K)^2 = \frac{\kappa^2}{2}\sum_K\nu_\ell^{\mathcal{T}}(K)\sum_L\lambda_{KL}\frac{|(K|L)|}{d_{KL}}(v \cdot (x_L - x_K))^2$$

$$= \frac{\kappa^2}{2}\sum_K\nu_\ell^{\mathcal{T}}(K)(S_K - |K|)$$

$$= \frac{\kappa^2}{2}\sum_{Q\in\mathscr{Q}_\ell}\bigg(\alpha_Q^{\mathcal{T}}\sum_{K\in\mathcal{T}_Q^+}(S_K - |K|) + \beta_Q^{\mathcal{T}}\sum_{K\in\mathcal{T}_Q^-}(|K| - S_K)\bigg) \,.$$

Using (5.10) and (5.11), this identity yields

$$\liminf_{[\mathcal{T}]\to 0}\sum_{K,L}\gamma_{KL}\frac{|(K|L)|}{d_{KL}}(\overline{\phi}_L - \overline{\phi}_K)^2 \geq \frac{c\kappa^2}{2}\sum_{Q\in\mathscr{Q}_\ell}|Q| \geq \frac{c\kappa^2}{4}|\tilde{B}| \,,$$

provided ℓ is sufficiently small. Together with (5.16), it follows that

$$\liminf_{[\mathcal{T}]\to 0}\int_0^1\sum_{K,L}\gamma_{KL}\frac{|(K|L)|}{d_{KL}}(\psi_t^{\mathcal{T}}(L) - \psi_t^{\mathcal{T}}(K))^2\,\mathrm{d}t \geq \frac{c\kappa^2}{4}|\tilde{B}| \,. \tag{5.17}$$

Inserting the three estimates (5.14), (5.15) and (5.17) into (5.13), we obtain

$$\limsup_{[\mathcal{T}]\to 0}\mathcal{W}(m_{\varepsilon,0}^{\mathcal{T}}, m_{\varepsilon,1}^{\mathcal{T}})^2 \leq \limsup_{[\mathcal{T}]\to 0}\int_0^1\mathcal{K}_{\mathcal{T}}(m_{\varepsilon,t}^{\mathcal{T}}, V_t^{\mathcal{T}})\,\mathrm{d}t \leq \mathbb{W}_2(\mu_0,\mu_1)^2 - c|\tilde{B}|\varepsilon + C\varepsilon^2 \,,$$
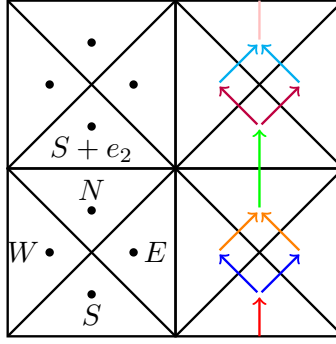
for suitable constants $c > 0$ and $C < \infty$.

Figure 4: We can decrease the action by adding ε to ρ in all N and S triangles and subtracting ε in all E and W triangles. If ρ is the same in E and W, we can choose a momentum vector field V that transports no mass over vertical lines with $V_{SE} = V_{SW}$ and $V_{EN} = V_{WN}$.

On the other hand, since $\sum_{K \in \mathcal{T}_Q} \nu_\ell^{\mathcal{T}}(K)|K| = 0$ for all $Q \in \mathscr{Q}_\ell$, it follows from Lemma 3.3 that

$$\mathcal{W}_{\mathcal{T}}(m_0^{\mathcal{T}}, m_{\varepsilon,0}^{\mathcal{T}}) \leq C\big(\mathbb{W}_2(\mu_0, Q_{\mathcal{T}} m_{\varepsilon,0}^{\mathcal{T}}) + [\mathcal{T}]\big)$$
$$\leq C(\ell + [\mathcal{T}]) ,$$

and the same holds at $t = 1$. In summary, we obtain

$$\limsup_{[\mathcal{T}] \to 0} \mathcal{W}(m_0^{\mathcal{T}}, m_1^{\mathcal{T}})^2 \leq \mathbb{W}_2(\mu_0, \mu_1)^2 - c|\tilde{B}|\varepsilon + C\varepsilon^2 + C\ell^2 , \qquad (5.18)$$

for all $\varepsilon < \varepsilon_0$. As $\ell > 0$ is arbitrary, this yields the result. $\qquad\square$

Remark 5.10. For the mesh in Figure 4, the construction in the proof of Theorem 5.8 can be somewhat simplified: as discussed in Example 5.7, isotropy fails to hold in the coordinate directions $v = \pm e_1, \pm e_2$. Picking a $\mathbb{W}_2$-geodesic transporting mass in direction e_2 in some open ball B, we notice that the discretisation of that geodesic transports no mass over vertical edges in B. The variation $\pi^{\mathcal{T}}$ can then be set to $\pi^{\mathcal{T}} = 1$ for all N and S cells and $\pi^{\mathcal{T}} = -1$ for all E and W cells. Along diagonal edges, the change in θ_{KL} is $o(\varepsilon)$, whereas for horizontal edges, the change in $\theta_{N,S+e_2}$ is $-\varepsilon + o(\varepsilon)$. For vertical edges, the change in $\theta_{E,W+e_1}$ is $\varepsilon + o(\varepsilon)$, which would be costly, but here the momentum vector field $V(E, W + e_1)$ vanishes.

6 Gromov–Hausdorff convergence

In this section we prove Theorems 1.1 and 1.4, as well as Corollaries 1.5 and 1.6. Let us start by stating the definition of Gromov–Hausdorff convergence.

Definition 6.1 (Gromov–Hausdorff convergence)**.** *We say that a sequence of compact metric spaces $(\mathcal{X}_n, \mathsf{d}_n)_{n \geq 1}$ converges in the sense of Gromov–Hausdorff to a compact metric space $(\mathcal{X}, \mathsf{d})$, if there exist maps $f_n \colon \mathcal{X} \to \mathcal{X}_n$ which are*

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

$$|\mathsf{d}_n(f_n(x), f_n(y)) - \mathsf{d}(x, y)| \leq \varepsilon_n \qquad (6.1)$$

  and

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

$$\mathsf{d}_n(f_n(x), x_n) \leq \varepsilon_n \qquad (6.2)$$

for some sequence $(\varepsilon_n)_n$ with $\varepsilon_n \to 0$ as $n \to \infty$.

Our main task is to show that the mappings $\mathcal{P}_\mathcal{T}$ are ε-isometric. We divide the argument into two parts: an upper bound for the discrete transport metric (Theorem 6.4) will be proved in Section 6.1. This result is valid for any sequence of ζ-regular meshes. Under strong additional symmetry assumptions, we will prove a corresponding lower bound (Theorem 6.8) in Section 6.2. The argument will be completed in Section 6.3.

We start with a useful time-regularisation result along the lines of [11, Lemma 2.9].

Lemma 6.2 (Time-regularisation). *Let* $(m_t)_{t \in [0,1]}$ *be a curve in* $\mathcal{P}(\mathcal{T})$ *with*

$$\int_0^1 \mathcal{A}_\mathcal{T}^*(m_t, \dot{m}_t) \, \mathrm{d}t < \infty \ .$$

For $\delta \in (0, \frac{1}{2})$ *we consider the "compressed" curve* $(m_t^\delta)_{t \in [-\delta, 1+\delta]}$ *in* $\mathcal{P}(\mathcal{T})$ *given by*

$$\tilde{m}_t^\delta := \begin{cases} m_0 & \text{for } t \in [-\delta, \delta] \ , \\ m_{(t-\delta)/(1-2\delta)} & \text{for } t \in (\delta, 1-\delta) \ , \\ m_1 & \text{for } t \in [1-\delta, 1+\delta] \ . \end{cases} \qquad (6.3)$$

Let $\eta : \mathbb{R} \to \mathbb{R}_+$ *be infinitely differentiable, symmetric, and supported in* $[-1, 1]$ *with* $\int_\mathbb{R} \eta(t) \, \mathrm{d}t = 1$, *and define* $m_t^\delta := \int_\mathbb{R} \eta(s/\delta) \tilde{m}_{t-s}^\delta \, \mathrm{d}s$. *Then the following assertions hold:*

(i) *The curve* $(m_t^\delta)_{t \in [0,1]}$ *is infinitely differentiable, it satisfies* $m_0^\delta = m_0$ *and* $m_1^\delta = m_1$, *and*

$$\int_0^1 \mathcal{A}_\mathcal{T}^*(m_t^\delta, \dot{m}_t^\delta) \, \mathrm{d}t \leq \frac{1}{1-2\delta} \int_0^1 \mathcal{A}_\mathcal{T}^*(m_t, \dot{m}_t) \, \mathrm{d}t \ .$$

(ii) *Let* $\{\mathcal{T}\}$ *be a sequence of a meshes. For each* $\mathcal{T}$, *let* $(m_t^\mathcal{T})_{t \in \mathcal{T}}$ *be a curve in* $\mathcal{P}(\mathcal{T})$, *and suppose that, for all* $t \in [0, 1]$, *there exists a probability measure* $\mu_t \in \mathcal{P}(\overline{\Omega})$ *such that* $Q_\mathcal{T} m_t^\mathcal{T} \rightharpoonup \mu_t$ *as* $[\mathcal{T}] \to 0$. *Then, for all* $t \in [0, 1]$,

$$Q_\mathcal{T} m_t^{\mathcal{T},\delta} \rightharpoonup \mu_t^\delta \quad \text{and} \quad Q_\mathcal{T} \dot{m}_t^{\mathcal{T},\delta} \rightharpoonup \dot{\mu}_t^\delta \quad \text{as } [\mathcal{T}] \to 0 \ .$$

Proof. Using the joint convexity of the mapping $(m, \sigma) \mapsto \mathcal{A}_\mathcal{T}^*(m, \sigma)$ we obtain

$$\int_0^1 \mathcal{A}_\mathcal{T}^*(m_t^\delta, \dot{m}_t^\delta) \, \mathrm{d}t \leq \int_0^1 \int_{-\delta}^{\delta} \eta(s/\delta) \mathcal{A}_\mathcal{T}^*(\tilde{m}_{t-s}^\delta, \dot{\tilde{m}}_{t-s}^\delta) \, \mathrm{d}s \, \mathrm{d}t$$

$$\leq \int_0^1 \mathcal{A}_\mathcal{T}^*(\tilde{m}_t^\delta, \dot{\tilde{m}}_t^\delta) \, \mathrm{d}t = \frac{1}{1-2\delta} \int_0^1 \mathcal{A}_\mathcal{T}^*(m_t, \dot{m}_t) \, \mathrm{d}t \ .$$

Since $\tilde{m}_t^{\mathcal{T},\delta} \rightharpoonup \tilde{\mu}_t^\delta$ for all $t \in [-\delta, 1+\delta]$ (where $t \mapsto \tilde{\mu}_t^\delta$ denotes the time-compressed version of $t \mapsto \tilde{\mu}_t$), the second part of the result follows using dominated convergence. $\qquad \square$

Clearly, a completely analogous result holds in the continuous setting.

6.1 Upper bound for the discrete transport metric

In this subsection we prove an upper bound for $\mathcal{W}_\mathcal{T}$, that relies on the finite volume bounds obtained in Section 4. Since these bounds require an ellipticity condition on the densities, we use a regularisation argument involving the heat flow. Lemma 6.3 contains the desired bound for the regularised measures. The regularisation is removed in Theorem 6.4.

We emphasise that these results hold under the mere assumption of ζ-regularity, and do not require the additional symmetry assumptions that we will impose in Section 6.2.

Lemma 6.3. *Fix $\zeta \in (0,1]$ and $a > 0$. There exists a constant $C < \infty$, depending on Ω, ζ, and a, such that for any ζ-regular mesh $\mathcal{T}$ of Ω, and $\mu_0, \mu_1 \in \mathcal{P}(\overline{\Omega})$ the following estimate holds:*

$$\mathcal{W}_\mathcal{T}(P_\mathcal{T} H_a \mu_0, P_\mathcal{T} H_a \mu_1)^2 \leq \mathbb{W}_2(\mu_0, \mu_1)^2 + C[\mathcal{T}] \ .$$

Proof. Let $(\mu_t)_{t \in [0,1]}$ be a geodesic connecting μ_0 and μ_1. Take $\eta : \mathbb{R} \to \mathbb{R}$ as in Lemma 6.2 and define, for $\delta \in (0, \frac{1}{2})$,

$$\mu_t^{a,\delta} := \int_\mathbb{R} \eta\Big(\frac{s}{\delta}\Big) H_a \tilde{\mu}_{t-s}^\delta \, \mathrm{d}s \ ,$$

where $(\tilde{\mu}_t^\delta)_{t \in [-\delta, 1+\delta]}$ is the compression of $(\mu_t)_{t \in [0,1]}$ as in (6.3). By Lemma 2.2, the density $u_t^{a,\delta}$ of $\mu_t^{a,\delta}$ satisfies $\mathrm{Lip}(u_t^{a,\delta}) \leq C$ and $u_t^{a,\delta}(x) \geq C^{-1} > 0$ for all $x \in \overline{\Omega}$, where $C < \infty$ depends only on a (and not on t or δ). Proposition 4.6 yields

$$\mathcal{A}_\mathcal{T}^*(P_\mathcal{T} \mu_t^{a,\delta}, P_\mathcal{T} \dot{\mu}_t^{a,\delta}) \leq \mathbb{A}^*(\mu_t^{a,\delta}, \dot{\mu}_t^{a,\delta}) + C[\mathcal{T}] \|\dot{u}_t^{a,\delta}\|_{L^2(\Omega)}^2 \ . \tag{6.4}$$

where $C < \infty$ depends on Ω, a, and ζ.

Denoting the density of $\tilde{\mu}_t^\delta$ by $\tilde{u}_t^\delta$, we observe that

$$\dot{u}_t^{a,\delta} = \frac{1}{\delta} \int_\mathbb{R} \eta'\Big(\frac{s}{\delta}\Big) H_a \tilde{u}_{t-s}^\delta \, \mathrm{d}s \ .$$

The heat kernel upper bound (2.6) yields

$$\|H_a \tilde{u}_t^\delta\|_{L^2(\Omega)} \leq C \sup_{x,y} h_a(x,y) \leq C(a^{-d/2} \vee 1) \ ,$$

where $C < \infty$ depends only on Ω. Consequently,

$$\|\dot{u}_t^{a,\delta}\|_{L^2(\Omega)} \leq C \|\eta'\|_{L^1(\mathbb{R})} (a^{-d/2} \vee 1) \ .$$

Integrating (6.4) over $[0,1]$ we obtain

$$\mathcal{W}_\mathcal{T}(P_\mathcal{T} H_a \mu_0, P_\mathcal{T} H_a \mu_1)^2 \leq \int_0^1 \mathcal{A}_\mathcal{T}^*(P_\mathcal{T} \mu_t^{a,\delta}, P_\mathcal{T} \dot{\mu}_t^{a,\delta}) \, \mathrm{d}t$$

$$\leq \int_0^1 \mathbb{A}^*(\mu_t^{a,\delta}, \dot{\mu}_t^{a,\delta}) \, \mathrm{d}t + C[\mathcal{T}](a^{-d} \vee 1)$$

where the η-dependence is absorbed in the constant C. Furthermore, using the convexity of $(\mu, w) \mapsto \mathbb{A}^*(\mu, w)$ as in Lemma 6.2, and the contraction bound from Lemma 2.2(ii), we obtain

$$\int_0^1 \mathbb{A}^*(\mu_t^{a,\delta}, \dot{\mu}_t^{a,\delta}) \, \mathrm{d}t \leq \frac{1}{1 - 2\delta} \int_0^1 \mathbb{A}^*(H_a \mu_t, H_a \dot{\mu}_t) \, \mathrm{d}t$$

$$\leq \frac{1}{1 - 2\delta} \int_0^1 \mathbb{A}^*(\mu_t, \dot{\mu}_t) \, \mathrm{d}t = \frac{1}{1 - 2\delta} \mathbb{W}_2(\mu_0, \mu_1)^2 \ .$$

Since the constant C does not depend on δ, we obtain the desired result by passing to the limit $\delta \to 0$ (and absorbing the factor $a^{-d} \vee 1$ into the constant C). $\qquad\square$

Theorem 6.4 (Upper bound for $\mathcal{W}_{\mathcal{T}}$)**.** *Fix $\zeta \in (0,1]$. For any $\varepsilon > 0$ there exists $h > 0$ such that for any ζ-regular mesh $\mathcal{T}$ with $[\mathcal{T}] \leq h$, we have*

$$\mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_0, P_{\mathcal{T}}\mu_1) \leq \mathbb{W}_2(\mu_0, \mu_1) + \varepsilon \tag{6.5}$$

for all $\mu_0, \mu_1 \in \mathcal{P}(\overline{\Omega})$.

Proof. Using the triangle inequality we estimate

$$\mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_0, P_{\mathcal{T}}\mu_1) \leq \mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_0, P_{\mathcal{T}}H_a\mu_0)$$
$$+ \mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}H_a\mu_0, P_{\mathcal{T}}H_a\mu_1) + \mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}H_a\mu_1, P_{\mathcal{T}}\mu_1)$$

for any $\mu_0, \mu_1 \in \mathcal{P}(\overline{\Omega})$ and $a > 0$. Lemma 6.3 yields

$$\mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}H_a\mu_0, P_{\mathcal{T}}H_a\mu_1)^2 \leq \mathbb{W}_2(\mu_0, \mu_1)^2 + C_1(a)[\mathcal{T}] \ ,$$

where $C_1(a) < \infty$ depends on Ω, ζ and a. Using the *a priori* estimate from Lemma 3.3 followed by Lemma 3.2 and Lemma 2.2, we obtain

$$\mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_i, P_{\mathcal{T}}H_a\mu_i)$$
$$\leq C_2\Big(\mathbb{W}_2(Q_{\mathcal{T}}P_{\mathcal{T}}\mu_i, Q_{\mathcal{T}}P_{\mathcal{T}}H_a\mu_i) + [\mathcal{T}]\Big)$$
$$\leq C_2\Big(\mathbb{W}_2(Q_{\mathcal{T}}P_{\mathcal{T}}\mu_i, \mu_i) + \mathbb{W}_2(\mu_i, H_a\mu_i) + \mathbb{W}_2(H_a\mu_i, Q_{\mathcal{T}}P_{\mathcal{T}}H_a\mu_i) + [\mathcal{T}]\Big)$$
$$\leq C_2\Big([\mathcal{T}] + \sqrt{a}\Big)$$

for $i = 0, 1$, where the constant $C_2 < \infty$ depends on Ω and ζ, but not on a. Combining these estimates we find

$$\mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_0, P_{\mathcal{T}}\mu_1) \leq \sqrt{\mathbb{W}_2(\mu_0, \mu_1)^2 + C_1(a)[\mathcal{T}]} + C_2\Big([\mathcal{T}] + \sqrt{a}\Big) \ .$$

Let now $\varepsilon > 0$, and choose a sufficiently small, so that $C_2\sqrt{a} \leq \varepsilon/2$. Then there exists $h > 0$ such that, whenever $[\mathcal{T}] \leq h$, we have

$$\mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_0, P_{\mathcal{T}}\mu_1) \leq \mathbb{W}_2(\mu_0, \mu_1) + \varepsilon \tag{6.6}$$

for all $\mu_0, \mu_1 \in \mathcal{P}(\overline{\Omega})$, which implies the result. $\qquad\square$

6.2   Lower bound for the discrete transport metric under isotropy conditions

Since the counterexamples in Section 5 show that $\mathcal{W}_{\mathcal{T}}$ does not Gromov–Hausdorff converge to $\mathbb{W}_2$ in general, we will impose an additional condition on the mesh. Let I_d denote the $d \times d$ identity matrix.

Definition 6.5. *Let $\{\mathcal{T}\}$ be a family of admissible meshes such that $[\mathcal{T}] \to 0$. We say that $\{\mathcal{T}\}$ satisfies the* asymptotic isotropy condition *with weight functions $\{\boldsymbol{\lambda}^{\mathcal{T}}\}$ if, for all $K \in \mathcal{T}$,*

$$\sum_L \lambda_{KL}^{\mathcal{T}} \frac{|(K|L)|}{d_{KL}}(x_K - x_L) \otimes (x_K - x_L) \leq |K|\big(1 + \eta_{\mathcal{T}}(K)\big)I_d \ , \tag{6.7}$$

where $\sup_{K \in \mathcal{T}} |\eta_{\mathcal{T}}(K)| \to 0$ as $[\mathcal{T}] \to 0$.

The following proposition contains the crucial bounds on the action functionals $\mathcal{A}_\mathcal{T}$ and their duals $\mathcal{A}_\mathcal{T}^*$. To obtain the result, one needs to carefully choose the means θ_{KL} according to the geometry of the mesh.

We say that a sequence of signed measures $\{w_\mathcal{T}\}_\mathcal{T} \subseteq \mathcal{M}_0(\overline{\Omega})$ converges weakly to a signed measure w if $\langle \phi, w_\mathcal{T} \rangle \to \langle \phi, w \rangle$ for all $\phi \in \mathcal{C}^0(\overline{\Omega})$. In this case, we write $w_\mathcal{T} \rightharpoonup w$.

Proposition 6.6 (Action bounds)**.** *Let $\{\mathcal{T}\}$ be a family of ζ-regular meshes satisfying the asymptotic isotropy condition with weight functions $\{\boldsymbol{\lambda}^\mathcal{T}\}$, and let $\{\boldsymbol{\theta}^\mathcal{T}\}$ be a family of weight functions that are compatible with $\{\boldsymbol{\lambda}^\mathcal{T}\}$.*

Suppose that $m_\mathcal{T} \in \mathcal{P}(\mathcal{T})$ satisfies $Q_\mathcal{T} m_\mathcal{T} \rightharpoonup \mu$ as $[\mathcal{T}] \to 0$ for some $\mu \in \mathcal{P}(\overline{\Omega})$.

(i) Let $\phi \in \mathcal{C}^1(\overline{\Omega})$ and define $\psi_\mathcal{T} : \mathcal{T} \to \mathbb{R}$ by $\psi_\mathcal{T}(K) := \phi(x_K)$. Then:

$$\limsup_{[\mathcal{T}] \to 0} \mathcal{A}_\mathcal{T}(m_\mathcal{T}, \psi_\mathcal{T}) \le \mathbb{A}(\mu, \phi) \ . \tag{6.8}$$

(ii) Let $\sigma_\mathcal{T} \in \mathcal{M}_0(\mathcal{T})$ and assume that there exists $w \in \mathcal{M}_0(\overline{\Omega})$ such that $Q_\mathcal{T} \sigma_\mathcal{T} \rightharpoonup w$ as $[\mathcal{T}] \to 0$. Then:

$$\mathbb{A}^*(\mu, w) \le \liminf_{[\mathcal{T}] \to 0} \mathcal{A}_\mathcal{T}^*(m_\mathcal{T}, \sigma_\mathcal{T}) \ . \tag{6.9}$$

Remark 6.7. We emphasise that it is important to assume in (6.9) that w in (6.9) is a signed measure, and not an arbitrary distribution.

Proof. We will first prove (6.8). For $K \in \mathcal{T}$ set $v_K := \nabla \phi(x_K)$ and write $\rho_\mathcal{T}(K) = m_\mathcal{T}(K)/|K|$. Let $\omega \colon [0, \infty) \to [0, \infty)$ be the modulus of continuity of $\nabla \phi$. Then

$$\big(\phi(x_K) - \phi(x_L)\big)^2 \le \big(v_K \cdot (x_L - x_K)\big)^2 + 4\|\nabla \phi\|_{L^\infty} \omega(2[\mathcal{T}]) d_{KL}^2$$

whenever $L \sim K$. By Remark 2.7, we have

$$\theta_{KL}^\mathcal{T}(\rho_\mathcal{T}(K), \rho_\mathcal{T}(L)) \le \lambda_{KL}^\mathcal{T} \rho_\mathcal{T}(K) + \lambda_{LK}^\mathcal{T} \rho_\mathcal{T}(L) \ .$$

Using these estimates we obtain

$$\begin{aligned}
\mathcal{A}_\mathcal{T}(m_\mathcal{T}, \psi_\mathcal{T}) &\le \frac{1}{2} \sum_{K,L} \frac{|(K|L)|}{d_{KL}} \Big(\lambda_{KL}^\mathcal{T} \rho_\mathcal{T}(K) + \lambda_{LK}^\mathcal{T} \rho_\mathcal{T}(L)\Big) \big(\phi(x_K) - \phi(x_L)\big)^2 \\
&= \sum_K \rho_\mathcal{T}(K) \sum_L \lambda_{KL}^\mathcal{T} \frac{|(K|L)|}{d_{KL}} \big(\phi(x_K) - \phi(x_L)\big)^2 \\
&\le \sum_K \rho_\mathcal{T}(K) \sum_L \lambda_{KL}^\mathcal{T} \frac{|(K|L)|}{d_{KL}} \big(v_K \cdot (x_L - x_K)\big)^2 \\
&\quad + 4\|\nabla \phi\|_{L^\infty} \omega(2[\mathcal{T}]) \sum_K \rho_\mathcal{T}(K) \sum_L d_{KL} |(K|L)| \ .
\end{aligned} \tag{6.10}$$

In view of Lemma 2.12(i) and (2.21), we observe that

$$\sum_L d_{KL} |(K|L)| \le C|K| \ ,$$

where $C < \infty$ depends on Ω and ζ. In the former term we use the asymptotic isotropy condition (6.7) to write

$$\sum_L \lambda_{KL}^{\mathcal{T}} \frac{|(K|L)|}{d_{KL}} \big(v_K \cdot (x_L - x_K)\big)^2 \leq \big(1 + \eta_{\mathcal{T}}(K)\big)|K|\,|v_K|^2,$$

where $\eta_{\mathcal{T}}(K)$ is the error term in (6.7), which converges to 0, uniformly in K, as $[\mathcal{T}] \to 0$.

Summing up all contributions, we obtain

$$\mathcal{A}_{\mathcal{T}}(m_{\mathcal{T}}, \psi_{\mathcal{T}}) \leq \sum_K \rho_{\mathcal{T}}(K)|K|\Big(\big(1 + \eta_{\mathcal{T}}(K)\big)|v_K|^2 + 4C\|\nabla\phi\|_{L^\infty}\omega(2[\mathcal{T}])\Big).$$

Writing $\mu_{\mathcal{T}} := Q_{\mathcal{T}} m_{\mathcal{T}}$ and $\xi_{\mathcal{T}} = \sum_K \chi_K \big(1 + \eta_{\mathcal{T}}(K)\big)|v_K|^2$ we have

$$\sum_K \rho_{\mathcal{T}}(K)|K|\big(1 + \eta_{\mathcal{T}}(K)\big)|v_K|^2$$

$$= \langle \xi_{\mathcal{T}}, \mu_{\mathcal{T}} \rangle = \int |\nabla\phi|^2 \, d\mu + \int |\nabla\phi|^2 \, d(\mu_{\mathcal{T}} - \mu) + \int \xi_{\mathcal{T}} - |\nabla\phi|^2 \, d\mu_{\mathcal{T}}.$$

Since $\mu_{\mathcal{T}}$ converges weakly to μ and $\|\xi_{\mathcal{T}} - |\nabla\phi|^2\|_{L^\infty} \to 0$ as $[\mathcal{T}] \to 0$, we obtain (6.8).

Let us now prove (6.9). Take $\phi \in \mathcal{C}^1(\overline{\Omega})$ and define $\psi_{\mathcal{T}} : \mathcal{T} \to \mathbb{R}$ by $\psi_{\mathcal{T}}(K) = \phi(x_K)$. We claim that $\langle \phi, w \rangle = \lim_{[\mathcal{T}] \to 0} \langle \psi_{\mathcal{T}}, \sigma_{\mathcal{T}} \rangle$. To show this, set $w_{\mathcal{T}} := Q_{\mathcal{T}} \sigma_{\mathcal{T}}$ and $\phi_{\mathcal{T}} := \sum_K \psi_{\mathcal{T}}(K)\chi_K$, and note that $\langle \psi_{\mathcal{T}}, \sigma_{\mathcal{T}} \rangle = \langle \phi_{\mathcal{T}}, w_{\mathcal{T}} \rangle$. Therefore,

$$\langle \psi_{\mathcal{T}}, \sigma_{\mathcal{T}} \rangle - \langle \phi, w \rangle = \langle \phi_{\mathcal{T}} - \phi, w_{\mathcal{T}} \rangle + \langle \phi, w_{\mathcal{T}} - w \rangle.$$

Since $w_{\mathcal{T}} \rightharpoonup w$, the Banach–Steinhaus Theorem implies that $\sup_{\mathcal{T}} \|w_{\mathcal{T}}\|_{\mathrm{TV}} < \infty$. Together with the bound $\|\phi_{\mathcal{T}} - \phi\|_{L^\infty} \leq C[\mathcal{T}]$, this yields the claim.

Suppose first that $\mathbb{A}^*(\mu, w)$ is finite. Fix $\varepsilon > 0$ and choose $\phi \in \mathcal{C}^1(\overline{\Omega})$ such that

$$\frac{1}{2}\mathbb{A}^*(\mu, w) \leq \langle \phi, w \rangle - \frac{1}{2}\mathbb{A}(\mu, \phi) + \varepsilon.$$

Using the claim and (6.8), it follows that

$$\frac{1}{2}\mathbb{A}^*(\mu, w) \leq \liminf_{[\mathcal{T}] \to 0} \Big(\langle \psi_{\mathcal{T}}, \sigma_{\mathcal{T}} \rangle - \frac{1}{2}\mathcal{A}_{\mathcal{T}}(m_{\mathcal{T}}, \psi_{\mathcal{T}})\Big) + \varepsilon$$

$$\leq \liminf_{[\mathcal{T}] \to 0} \frac{1}{2}\mathcal{A}_{\mathcal{T}}^*(m_{\mathcal{T}}, \sigma_{\mathcal{T}}) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the result follows.

Suppose next that $\mathbb{A}^*(\mu, w) = \infty$. Then, for each $N > 0$, there exists a function $\phi_N \in \mathcal{C}^1(\overline{\Omega})$ such that

$$\langle \phi_N, w \rangle - \frac{1}{2}\mathbb{A}(\mu, \phi_N) \geq N.$$

Define $\psi_{\mathcal{T}}^N : \mathcal{T} \to \mathbb{R}$ by $\psi_{\mathcal{T}}^N(K) := \phi_N(x_K)$. Using the claim and (6.8) once more, we obtain

$$N \leq \liminf_{[\mathcal{T}] \to 0} \Big(\langle \psi_{\mathcal{T}}^N, \sigma_{\mathcal{T}} \rangle - \frac{1}{2}\mathcal{A}_{\mathcal{T}}(m_{\mathcal{T}}, \psi_{\mathcal{T}}^N)\Big)$$

$$\leq \liminf_{[\mathcal{T}] \to 0} \frac{1}{2}\mathcal{A}_{\mathcal{T}}^*(m_{\mathcal{T}}, \sigma_{\mathcal{T}}),$$

which implies that $\liminf_{[\mathcal{T}] \to 0} \mathcal{A}_{\mathcal{T}}^*(m_{\mathcal{T}}, \sigma_{\mathcal{T}}) = \infty$. $\qquad\square$

Theorem 6.8 (Lower bound for $\mathcal{W}_{\mathcal{T}}$). *Fix* $\zeta \in (0, 1]$. *For any* $\varepsilon > 0$, *there exists* $h > 0$ *such that the following holds: for any family of* ζ-*regular meshes* $\{\mathcal{T}\}$ *satisfying the asymptotic isotropy condition with weight functions* $\{\boldsymbol{\lambda}^{\mathcal{T}}\}$, *and any family of mean functions* $\{\boldsymbol{\theta}^{\mathcal{T}}\}$ *that are compatible with* $\{\boldsymbol{\lambda}^{\mathcal{T}}\}$, *we have*

$$\mathbb{W}_2(\mu_0, \mu_1) \leq \mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_0, P_{\mathcal{T}}\mu_1) + \varepsilon \tag{6.11}$$

for all $\mu_0, \mu_1 \in \mathcal{P}(\overline{\Omega})$, *whenever* $[\mathcal{T}] \leq h$.

Proof. To obtain a contradiction, we suppose that the opposite holds, i.e., there exists $\varepsilon > 0$, a sequence of meshes $\{\mathcal{T}\}$ with $[\mathcal{T}] \to 0$, and probability measures $\mu_0^{\mathcal{T}}, \mu_1^{\mathcal{T}} \in \mathcal{P}(\mathcal{T})$, such that

$$\mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_0^{\mathcal{T}}, P_{\mathcal{T}}\mu_1^{\mathcal{T}}) < \mathbb{W}_2(\mu_0^{\mathcal{T}}, \mu_1^{\mathcal{T}}) - \varepsilon \ . \tag{6.12}$$

Let $(m_t^{\mathcal{T}})_{t \in [0,1]} \subseteq \mathcal{P}(\mathcal{T})$ be a constant speed geodesic connecting $P_{\mathcal{T}}\mu_0^{\mathcal{T}}$ and $P_{\mathcal{T}}\mu_1^{\mathcal{T}}$, so that

$$\int_0^1 \mathcal{A}_{\mathcal{T}}^*(m_t^{\mathcal{T}}, \dot{m}_t^{\mathcal{T}}) \, \mathrm{d}t = \mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_0^{\mathcal{T}}, P_{\mathcal{T}}\mu_1^{\mathcal{T}})^2 \ . \tag{6.13}$$

Set $\tilde{\mu}_t^{\mathcal{T}} := H_{[\mathcal{T}]}Q_{\mathcal{T}}m_t^{\mathcal{T}}$. Then, for almost every $t \in [0, 1]$, Lemma 3.4 yields

$$\mathbb{A}^*(\tilde{\mu}_t^{\mathcal{T}}, \dot{\tilde{\mu}}_t^{\mathcal{T}}) \leq C \mathcal{A}_{\mathcal{T}}^*(m_t^{\mathcal{T}}, \dot{m}_t^{\mathcal{T}}) \ ,$$

where the constant $C < \infty$ does not depend on $\mathcal{T}$ or t. Consequently, for $0 \leq t_1 \leq t_2 \leq 1$,

$$\begin{aligned}
\mathbb{W}_2(\tilde{\mu}_{t_1}^{\mathcal{T}}, \tilde{\mu}_{t_2}^{\mathcal{T}}) &\leq \int_{t_1}^{t_2} \sqrt{\mathbb{A}^*(\tilde{\mu}_t^{\mathcal{T}}, \dot{\tilde{\mu}}_t^{\mathcal{T}})} \, \mathrm{d}t \\
&\leq C \int_{t_1}^{t_2} \sqrt{\mathcal{A}_{\mathcal{T}}^*(m_t^{\mathcal{T}}, \dot{m}_t^{\mathcal{T}})} \, \mathrm{d}t \leq C \sqrt{t_2 - t_1} \mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_0^{\mathcal{T}}, P_{\mathcal{T}}\mu_1^{\mathcal{T}}) \ ,
\end{aligned} \tag{6.14}$$

and hence by (6.12) the family of curves $\{(\tilde{\mu}_t^{\mathcal{T}})_t\}_{\mathcal{T}}$ is equicontinuous. Since $(\mathcal{P}(\overline{\Omega}), \mathbb{W}_2)$ is compact, the Arzelà–Ascoli Theorem yields a subsequence of meshes and a $\mathbb{W}_2$-continuous curve of probability measures $(\lambda_t)_{t \in [0,1]}$ in $\mathcal{P}(\overline{\Omega})$ such that $\mathbb{W}_2(\tilde{\mu}_t^{\mathcal{T}}, \lambda_t) \to 0$ for all $t \in [0, 1]$ as $[\mathcal{T}] \to 0$. Moreover, since $\mathbb{W}_2(\tilde{\mu}_t^{\mathcal{T}}, Q_{\mathcal{T}}m_t^{\mathcal{T}}) \leq C\sqrt{[\mathcal{T}]}$ by Lemma 2.2, it follows that $Q_{\mathcal{T}}m_t^{\mathcal{T}} \to \lambda_t$ in $(\mathcal{P}(\overline{\Omega}), \mathbb{W}_2)$ as $[\mathcal{T}] \to 0$.

For $\delta \in (0, \frac{1}{2})$, let $t \mapsto m_t^{\mathcal{T}, \delta}$ be a time-regularised version of $t \mapsto m_t^{\mathcal{T}}$, as defined in Lemma 6.2. It follows from this lemma that $Q_{\mathcal{T}}m_t^{\mathcal{T}, \delta} \rightharpoonup \lambda_t^{\delta}$ and $Q_{\mathcal{T}}\dot{m}_t^{\mathcal{T}, \delta} \rightharpoonup \dot{\lambda}_t^{\delta}$ for all $t \in [0, 1]$ as $[\mathcal{T}] \to 0$. Moreover, by Lemma 3.2,

$$\lambda_0^{\delta} = \lambda_0 = \lim_{[\mathcal{T}] \to 0} Q_{\mathcal{T}}m_0^{\mathcal{T}} = \lim_{[\mathcal{T}] \to 0} Q_{\mathcal{T}}(P_{\mathcal{T}}\mu_0^{\mathcal{T}}) = \lim_{[\mathcal{T}] \to 0} \mu_0^{\mathcal{T}} \ ,$$

and similarly $\lim_{[\mathcal{T}] \to 0} \mu_1^{\mathcal{T}} = \lambda_1^{\delta}$, where the convergence is with respect to $\mathbb{W}_2$. Consequently,

$$\lim_{[\mathcal{T}] \to 0} \mathbb{W}_2(\mu_0^{\mathcal{T}}, \mu_1^{\mathcal{T}}) = \mathbb{W}_2(\lambda_0, \lambda_1) \ .$$

Using Proposition 6.6 and Fatou's Lemma, Lemma 6.2, and (6.13), it follows that

$$\begin{aligned}
\mathbb{W}_2(\lambda_0, \lambda_1)^2 &\leq \int_0^1 \mathbb{A}^*(\lambda_t^{\delta}, \dot{\lambda}_t^{\delta}) \, \mathrm{d}t \leq \liminf_{[\mathcal{T}] \to 0} \int_0^1 \mathcal{A}_{\mathcal{T}}^*(m_t^{\mathcal{T}, \delta}, \dot{m}_t^{\mathcal{T}, \delta}) \, \mathrm{d}t \\
&\leq \frac{1}{1 - 2\delta} \liminf_{[\mathcal{T}] \to 0} \int_0^1 \mathcal{A}_{\mathcal{T}}^*(m_t^{\mathcal{T}}, \dot{m}_t^{\mathcal{T}}) \, \mathrm{d}t \\
&= \frac{1}{1 - 2\delta} \liminf_{[\mathcal{T}] \to 0} \mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_0, P_{\mathcal{T}}\mu_1)^2 \ .
\end{aligned}$$

42

Since $\delta \in (0, \frac{1}{2})$ is arbitrary, we obtain

$$\lim_{[\mathcal{T}] \to 0} \mathbb{W}_2(\mu_0^{\mathcal{T}}, \mu_1^{\mathcal{T}}) \leq \liminf_{[\mathcal{T}] \to 0} \mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_0, P_{\mathcal{T}}\mu_1) \ ,$$

which is the desired contradiction to (6.12). $\qquad\square$

6.3 Proof of the Gromov–Hausdorff convergence

It remains to prove the corollaries stated in the introduction.

Proof of Corollary 1.5. Fix $\varepsilon > 0$. We will check that there exists $h > 0$ such that the map $P_{\mathcal{T}} : \mathcal{P}(\overline{\Omega}) \to \mathcal{P}(\mathcal{T})$ is ε-isometric and ε-surjective whenever $[\mathcal{T}] \leq h$.

The ε-surjectivity holds trivially, as $\mathcal{P}_{\mathcal{T}}$ is even surjective. To show that $P_{\mathcal{T}}$ is ε-isometric, we combine Theorem 6.4 and Theorem 6.8 to infer that there exists $h > 0$ such that

$$|\mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_0, P_{\mathcal{T}}\mu_1) - \mathbb{W}_2(\mu_0, \mu_1)| \leq \varepsilon$$

for all $\mu_0, \mu_1 \in \mathcal{P}(\overline{\Omega})$, whenever $[\mathcal{T}] \leq h$. This yields the result. $\qquad\square$

Proof of Corollary 1.6. For $i = 0, 1$, let $\mu_i \in \mathcal{P}(\overline{\Omega})$ and $m_i^{\mathcal{T}} \in \mathcal{P}(\mathcal{T})$ be such that $Q_{\mathcal{T}}m_i^{\mathcal{T}} \rightharpoonup \mu_i$ as $[\mathcal{T}] \to 0$. Lemmas 3.3 and 3.2 imply that, for some constant $C < \infty$ depending only on Ω and ζ (that changes from line to line),

$$\mathcal{W}_{\mathcal{T}}(m_i^{\mathcal{T}}, P_{\mathcal{T}}\mu_i) \leq C\Big(\mathbb{W}_2(Q_{\mathcal{T}}m_i^{\mathcal{T}}, Q_{\mathcal{T}}P_{\mathcal{T}}\mu_i) + [\mathcal{T}]\Big)$$
$$\leq C\Big(\mathbb{W}_2(Q_{\mathcal{T}}m_i^{\mathcal{T}}, \mu_i) + [\mathcal{T}]\Big) \ .$$

As $Q_{\mathcal{T}}m_i^{\mathcal{T}} \rightharpoonup \mu_i$, we have $\mathbb{W}_2(Q_{\mathcal{T}}m_i^{\mathcal{T}}, \mu_i) \to 0$, and therefore $\mathcal{W}_{\mathcal{T}}(m_i^{\mathcal{T}}, P_{\mathcal{T}}\mu_i) \to 0$. The triangle inequality then yields

$$\big|\mathcal{W}_{\mathcal{T}}(m_0^{\mathcal{T}}, m_1^{\mathcal{T}}) - \mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_0, P_{\mathcal{T}}\mu_1)\big| \leq \mathcal{W}_{\mathcal{T}}(m_0^{\mathcal{T}}, P_{\mathcal{T}}\mu_0) + \mathcal{W}_{\mathcal{T}}(m_1^{\mathcal{T}}, P_{\mathcal{T}}\mu_1) \to 0 \ .$$

Since $\mathcal{W}_{\mathcal{T}}(P_{\mathcal{T}}\mu_0, P_{\mathcal{T}}\mu_1) \to \mathbb{W}_2(\mu_0, \mu_1)$ by Theorems 1.1 and 1.4, we obtain the desired convergence $\mathcal{W}_{\mathcal{T}}(m_0^{\mathcal{T}}, m_1^{\mathcal{T}}) \to \mathbb{W}_2(\mu_0, \mu_1)$.

The final claim is now straightforward: for $0 \leq s \leq t \leq 1$ we have

$$\mathbb{W}_2(\mu_s, \mu_t) = \lim_{[\mathcal{T}] \to 0} \mathcal{W}_{\mathcal{T}}(m_s^{\mathcal{T}}, m_t^{\mathcal{T}}) = (t - s) \lim_{[\mathcal{T}] \to 0} \mathcal{W}_{\mathcal{T}}(m_0^{\mathcal{T}}, m_1^{\mathcal{T}}) = (t - s)\mathbb{W}_2(\mu_0, \mu_1) \ ,$$

which yields the result. $\qquad\square$

Remark 6.9. Clearly, it follows from the proof of Corollary 1.6 that the one-sided estimate $\limsup_{[\mathcal{T}] \to 0} \mathcal{W}_{\mathcal{T}}(m_0^{\mathcal{T}}, m_1^{\mathcal{T}}) \leq \mathbb{W}_2(\mu_0, \mu_1)$ holds under the conditions of Theorem 1.1, even when the asymptotic isotropy condition fails.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the Space of Probabiliy Measures.* Birkhäuser, Basel, 2005.

[2] L. Ambrosio, N. Gigli, and G. Savaré. Bakry-Émery curvature-dimension condition and Riemannian Ricci curvature bounds. *Ann. Probab.*, 43(1):339–404, 2015.

[3] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften*. Springer, Cham, 2014.

[4] 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.

[5] C. Cancès and C. Guichard. Numerical analysis of a robust free energy diminishing finite volume scheme for parabolic equations with gradient structure. *Found. Comput. Math.*, 17(6):1525–1584, 2017.

[6] E. A. Carlen and J. Maas. Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance. *J. Funct. Anal.*, 273(5):1810–1869, 2017.

[7] 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. Ration. Mech. Anal.*, 203(3):969–1008, 2012.

[8] S.-N. Chow, W. Li, and H. Zhou. A discrete Schrödinger equation via optimal transport on graphs. *J. Funct. Anal.*, 276(8):2440–2469, 2019.

[9] E. B. Davies. *Heat kernels and spectral theory*, volume 92 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989.

[10] K. Disser and M. Liero. On gradient structures for Markov chains and the passage to Wasserstein gradient flows. *Netw. Heterog. Media*, 10(2):233–253, 2015.

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

[12] M. Erbar and J. Maas. Gradient flow structures for discrete porous medium equations. *Discrete Contin. Dyn. Syst.*, 34(4):1355–1374, 2014.

[13] M. Erbar, J. Maas, and M. Wirth. On the geometry of geodesics in discrete optimal transport. *Calc. Var. Partial Differential Equations*, 58(1):58:19, 2019.

[14] M. Erbar, M. Rumpf, B. Schmitzer, and S. Simon. Computation of optimal transport on discrete metric measure spaces. *Numer. Math.*, 144(1):157–200, 2020.

[15] R. Eymard, T. Gallouët, and R. Herbin. Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes SUSHI: a scheme using stabilization and hybrid interfaces. *IMA J. Numer. Anal.*, 30(4):1009–1043, 2010.

[16] M. Fathi and J. Maas. Entropic Ricci curvature bounds for discrete interacting systems. *Ann. Appl. Probab.*, 26(3):1774–1806, 2016.

[17] R. Eymard, T. Gallouët and R. Herbin. The finite volume method. In *Handbook of Numerical Analysis*, volume VII, pages 715–1022. North Holland, Amsterdam, 2000.

[18] W. Gangbo, W. Li, and C. Mou. Geodesics of minimal length in the set of probability measures on graphs. *ESAIM Control Optim. Calc. Var.*, 25:Art. 78, 36, 2019.

[19] N. Garcia Trillos. Gromov-Hausdorff limit of Wasserstein spaces on point clouds. *arXiv preprint arXiv:1702.03464*, 2017.

[20] N. Gigli and J. Maas. Gromov-Hausdorff convergence of discrete transportation metrics. *SIAM J. Math. Anal.*, 45:879–899, 2013.

[21] P. Grisvard. *Elliptic problems in nonsmooth domains*, volume 69. SIAM, 2011.

[22] B. Grünbaum and G. C. Shephard. *Tilings and patterns*. W. H. Freeman and Company, New York, 1987.

[23] W. Li, P. Yin, and S. Osher. Computations of optimal transport distance with Fisher information regularization. *J. Sci. Comput.*, 75(3):1581–1595, 2018.

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

[25] J. Maas and D. Matthes. Long-time behavior of a finite volume discretization for a fourth order diffusion equation. *Nonlinearity*, 29(7):1992–2023, 2016.

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

[27] A. Mielke. Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Partial Differential Equations*, 48:1–31, 2013.

[28] M. Mittnenzweig and A. Mielke. An entropic gradient structure for Lindblad equations and couplings of quantum systems to macroscopic models. *J. Stat. Phys.*, 167(2):205–233, 2017.

[29] L. E. Payne and H. F. Weinberger. An optimal Poincaré inequality for convex domains. *Arch. Rational Mech. Anal.*, 5:286–292 (1960), 1960.

[30] F. Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser, 2015.

[31] J. Solomon, R. Rustamov, L. Guibas, and A. Butscher. Continuous-flow graph transportation distances. *arXiv preprint arXiv:1603.06927*, 2016.

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