Analysis of the MBO scheme: from materials science to data clustering

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Jona Lelmi

aus

Borgo San Lorenzo, Italien

Bonn 2023

Angefertigt mit Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen
Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Tim Laux
2. Gutachter: Prof. Dr. Martin Rumpf

Erscheinungsjahr: 2023
Tag Der Promotion: 24.05.2023

...alle volate

ACKNOWLEDGEMENTS

I would like to thank Tim for his valuable guidance during my doctoral studies. He has been very patient and flexible when I felt the desire of changing research directions according to my interests: thanks to his remarkable intuition, he was able to be a guide in areas of mathematics which where new to both of us. He has also been a guide when I was not sure about continuing my studies and I partly owe to his supporting advice the fact that I am now writing this thesis. Starting a PhD in the middle of a pandemic is not an easy task, but with an advisor – which I now consider a friend – who cares so much about his students everything is much easier.

I would like to thank my family for the constant support during these years abroad, for being always ready to give me advice when I doubted my choices.

I would like to thank my colleagues Kerrek, Leon, Milan and Sebastian for providing such a nice work environment, for the many discussions, for the many lunches and coffees, and for the many beers together. I would like to thank again Sebastian for careful reading part of a previous version of this thesis.

Finally, a big thanks goes to Paolo, with whom I shared the whole life period in Bonn, from the beginning of the Master to the end of the PhD. I consider our friendship the greatest gift of these German years.

SUMMARY

The subject of this thesis is the rigorous analysis of the convergence properties of an algorithm – the Merriman, Bence, Osher (MBO) scheme, or thresholding scheme – which is used both for simulations in materials science and for clustering and classification in data science.

The original version of the algorithm developed by Merriman, Bence, and Osher is used to approximate the evolution by (multiphase) mean curvature flow, for which we use the abbreviation MCF. MCF evolves a closed surface by prescribing the normal velocity at every point to be given by $-H$, where H is the mean curvature of the surface. The multiphase version of this evolution models the dynamics of a partition of space: each interface separating two different phases evolves by MCF and a balance of forces condition at triple junctions – i.e. points where three distinct interfaces meet – is given as a boundary condition.

The MBO scheme produces successive approximations of the evolution by MCF by alternating between two steps: (i) convolution with a smooth kernel and (ii) pointwise thresholding. These correspond to an operator splitting for the Allen-Cahn Equation, which is known to approximate MCF [42]. Since its introduction, several works have tackled the question of the convergence of the algorithm. In the simple two-phase case, this dates back to the independent works of Evans [26], and Barles and Georgelin [6]. Twenty years later, the gradient flow structure revealed in the work of Esedoḡlu and Otto [25] allowed Laux and Otto to prove the conditional convergence of the scheme in the multiphase setting [53].

Several modifications of the MBO algorithm have been proposed in the literature. One of them, due to Esedoḡlu and Salvador [72], allows for more freedom in the choice of some physical parameters – surface tensions and mobilities – which are used when multiphase MCF is adopted as a model for the slow relaxation of grain boundaries in polycrystals. This modified version of the scheme retains the simplicity and structure of the original one, but prior to the work of Laux and the author of this thesis [51], no rigorous proof of its convergence was known. In Chapter 2 we present the paper [51], included in Appendix A, which contains the proof of the conditional convergence of the modified thresholding scheme to a De Giorgi's solution to multiphase MCF, a weak notion of solution based on the gradient flow structure.

The MBO scheme was adapted in data science by Bertozzi et al. [62, 61, 78]. In this context, it is used for data clustering – i.e. for finding a partition of a given data set into subsets of points similar to each other. The thresholding scheme is part of the family

of graph-based learning algorithms, which share the common idea of building a graph structure on top of the data set and exploiting it to detect the underlying geometry of the data. The MBO scheme for data clustering does this in the following way: given a proposed clustering, one updates the partition by (i) diffusing the partition through the graph heat operator, and (ii) updating the clustering by pointwise thresholding. The thresholding scheme has demonstrated a good benchmark in many clustering contexts, for instance in image segmentation, but prior to the content of this thesis little was known about its theoretical justifications. Here we present two papers [50, 52], written both by Laux and the author of this work, where we initiate and substantially advance the mathematical understanding of the algorithm in the large data limit – i.e. we characterize its behavior when the number of data points grows to infinity.

The thesis is organized as follows: after a detailed introduction that motivates this work and explains it on a conceptual level, two chapters separate the presentation of the results into two parts.

Chapter 2 is concerned with the analysis of the modified thresholding scheme of Esedoḡlu and Salvador [25], and contains a detailed overview of the results of the paper [51], included in Appendix A. We now briefly summarize the core content of the first part of the thesis in terms of new results and scientific achievements.

Building on the seminal work of Esedoḡlu and Otto [25] and on the work of Laux and Otto [55], we analyze the modified algorithm of Esedoḡlu and Salvador [25], in particular:

(a) **De Giorgi's solution**. We introduce a new definition of weak solution to multiphase MCF, which extends the one given in the two-phase setting in [55]. This solution concept – called De Giorgi's solution – is based on the general theory of gradient flows.

(b) **Convergence of the scheme**. We give the first proof of the (conditional) convergence of the modified thresholding scheme of Esedoḡlu and Salvador [72] to a De Giorgi's solution to multiphase MCF.

Chapter 3 is concerned with the analysis of the MBO scheme for data clustering in the large data limit and contains a detailed overview of the results in the papers [50, 52], included in Appendix B and Appendix C respectively. In machine learning, studying the large data limit of an algorithm aims at understanding the asymptotic behavior of its outcomes as the number of data points grows to infinity. In recent years, this question has attracted the attention of the mathematical community and is part of the broad goal of giving solid theoretical justifications to data science. Several recent works, e.g. [34, 36, 35, 33, 10, 11, 12, 13, 14, 15, 9, 23], have successfully addressed this problem for some graph-based learning algorithms, developing also some mathematical techniques that are by now standard in the field – mainly viscosity solution approaches and variational and optimal transport methods.

We now briefly summarize the core content of the second part of the thesis in terms of new results and scientific achievements. With Laux, we analyzed the large data limit for the MBO scheme in two works [50, 52] outlined in Chapter 3. In particular:

1. **Γ-convergence of the thresholding energies**. By the minimizing movements interpretation of Esedoḡlu and Otto [25] and Bertozzi et al. [78], outcomes of multiclass MBO can be seen as (local) minimizers of the graph thresholding

energy. One is thus interested in characterizing the asymptotic behavior of these local minimizers. Under the manifold assumption for the data, we have:

(a) **Discrete-to-nonlocal**. We prove that, almost surely, for each fixed step size in the scheme, the graph thresholding energies Γ-converge to the corresponding thresholding energy on the weighted data manifold, which in the simple two-class setting coincides with the heat content with respect to a suitable weighted heat kernel.

(b) **Nonlocal-to-local**. We prove that as the step size h goes to zero, the manifold thresholding energies Γ-converge to and are consistent with a weighted partition's perimeter on the manifold. These results were known only in the flat Euclidean setting [1, 67, 25]. The generalization to weighted manifolds is of independent interest and is non-trivial because in general the heat kernel does not enjoy any translation invariance property as the Euclidean one.

2. **Convergence of the dynamics**. We study the convergence of the dynamics of the MBO scheme in the two-class setting. In this case, one can use the comparison principle for MCF to work with viscosity solutions. We show that over any sequence of step sizes which go to zero sufficiently slow as the number of data points grows to infinity, the dynamics of the thresholding scheme converge to a viscosity solution to MCF. More precisely:

   (a) **Abstract convergence result**. We first concentrate on an abstract version of the thresholding scheme, in which on each graph, the heat kernel is substituted by an abstract operator S_n. We show that whenever S_n approximates the manifold heat semigroup in the large data limit well enough, the dynamics of the MBO scheme converge to a viscosity solution to MCF on the manifold.

   (b) **Random graphs convergence**. We show that, when the data points are randomly and independently sampled from the data manifold, the graph heat kernel or the composition of the graph heat kernel with the projection on the space of the first $K_n \in \mathbf{N}$ eigenvectors of the graph Laplacian are admissible choices for S_n, in the sense that with either of these choices, the assumptions in the abstract convergence result hold true with high probability. This implies that the dynamics of the MBO scheme on random geometric graphs converge almost surely to a viscosity solution to MCF.

Hence, the present thesis advances the understanding of multiphase MCF and the MBO scheme in materials science and data clustering.

The content of this work is rigorous and theoretical from a mathematical viewpoint, but the results contained in it are of interest also in an applied setting, either as mathematical foundations or as useful insights for tweaking the algorithms we analyzed. Furthermore, in the proofs of the main theorems, we combine several techniques originating from the calculus of variations, the theory of gradient flows, optimal transportation, geometric measure theory, and data science in ways that could be inspiring for tackling similar open questions in the literature.

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Outline

The Merriman, Bence, and Osher (MBO) scheme – also known as thresholding scheme – is a numerical method originally introduced to approximate the evolution by mean curvature flow (MCF). The algorithm produces a time approximation of MCF by iterating the following two steps: *(i)* convolution with a smooth kernel and *(ii)* thresholding. The scheme is computationally very efficient: indeed, the first step can be implemented using the Fast Fourier Transform, while the second one is just a pointwise operation.

Besides being mathematically interesting, MCF is used to model the slow relaxation of grain boundaries in polycrystals, and this motivates the interest in finding good numerical schemes to approximate it. Since its first introduction in [63, 64], the MBO scheme has been widely studied: it has been tweaked in several ways to better approximate the physical model of grain growth, and, correspondingly, its convergence properties have been studied from a rigorous mathematical point of view [26, 6, 53, 56, 55, 57]. More recently, Bertozzi et al. [62, 61, 78] adapted the scheme to data science. In this context, it is used for clustering, which is the task of splitting a given data set into *clusters* – subsets of points similar to each other, where the measure of similarity is the Euclidean distance. The object of this thesis is the rigorous mathematical understanding of *(i)* a new variant of the thresholding scheme used to better model grain growth in polycrystals; and *(ii)* the large data limit of its data clustering analog.

Before entering into the discussion on the core content of this work, let us take a step back and focus on MCF. MCF describes the evolution of a closed surface by its mean curvature. If Σ_0 denotes an initial configuration, the surface changes over time $t \mapsto \Sigma_t$ according to the law

$$V = -H \quad \text{on } \Sigma_t, \tag{1.1}$$

where V and H denote, respectively, the normal velocity and the mean curvature of the surface Σ_t. Equation (1.1) - although formally very simple - presents many challenges. First of all, it is a degenerate parabolic equation, so the existence of smooth solutions is already an interesting problem. Short-time existence of a smooth solution was addressed for the first time by Gage and Hamilton in [32]. Being a diffusion equation, MCF has the appealing instantaneous regularizing effect, but even smooth solutions may develop singularities in finite time (e.g. [37]), so that, in general,

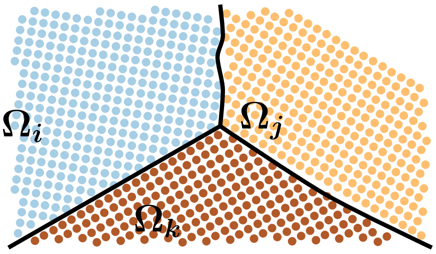

Figure 1.1: The ideal boundaries between different grains in a polycrystal.

long-time existence in the classical sense cannot be obtained. Considerable effort has thus been made in finding weak notions of solutions that can pass through topological changes. Successful attempts in this direction rely on two distinguished features of MCF: it satisfies a *comparison principle* and it has a *gradient flow* structure. Given that the MBO scheme is meant to approximate MCF, it is not too surprising that these two properties transfer to the algorithm, and are actually the main building blocks in any proof of its convergence.

The *comparison principle* for MCF asserts that for two domains with one included into the other, this inclusion will be preserved if their boundaries evolve by MCF. This observation allowed Chen, Giga, and Goto [18], and, independently, Evans and Spruck [27] to develop a viscosity solutions theory for MCF. Correspondingly, the *comparison principle* for the MBO scheme has been used to prove its convergence in the viscosity solutions setting in [26, 6, 68].

The *gradient flow* structure refers to the fact that MCF is the trajectory of steepest descent for the area functional: if $t \mapsto \Sigma_t$ denotes a smooth evolution of smooth surfaces, it can be shown that

$$\frac{d}{dt}\text{Area}(\Sigma_t) = \int_{\Sigma_t} VH dS, \tag{1.2}$$

from which it easily follows that among all possible choices for V with $\int_{\Sigma_t} V^2 dS = \int_{\Sigma_t} H^2 dS$, (1.1) makes (1.2) as small as possible. This structure was exploited for the introduction of Brakke's solutions [8] and of BV solutions in the sense of Luckhaus and Sturzenhecker [58]. It took more than twenty years before the gradient flow structure of the thresholding scheme was understood: in their seminal paper [25], Esedoḡlu and Otto gave a variational interpretation of the MBO scheme, characterizing each iteration as a step of minimizing movements for the *thresholding energy*, which in its simplest setting corresponds to the heat content. This observation made it possible to tackle the convergence of the algorithm in the setting of BV solutions [53, 56], Brakke's solutions [54] and De Giorgi's solutions [55, 51].

1.2 The MBO scheme for materials science

MCF is used in materials science to model the slow relaxation of (ideal) grain boundaries in *polycrystals* subject to heat treatment. A polycrystal is a material composed of several regions, called *crystals*, where the constituents are disposed in a highly ordered fashion. Grain boundaries are ideal borders separating two different grains, see Figure 1.1. When the material is heated, the constituents on the grain boundaries change orientation, causing the ideal boundaries to move. It was observed by Mullins [69] that the motion of a common boundary separating two crystals follows the dynamics

$$V = -\sigma\mu H, \tag{1.3}$$

where V and H are, respectively, the normal velocity and the mean curvature of the grain boundary. In equation (1.3), σ denotes the surface energy, or *surface tension*, which describes the free energy per unit area on the boundary, due to the different orientation of the atomic constituents in the two crystals. The quantity μ is referred to as *mobility*, which is a quantity that depends on the temperature of the grains. Equation (1.3) is the evolution by MCF (up to a constant).

When the material is composed of more than two phases, as in the case of Figure 1.1, the model prescribes equation (1.3) for each pair of neighboring phases, where the constants σ and μ depend on the pair that we are considering. Moreover, the model prescribes a balance of forces condition where three different grain boundaries meet, i.e. at *triple junctions*. If we denote by Σ_{ij} the boundary between phase Ω_i and phase Ω_j, we call evolution by *multiphase MCF* the following system of partial differential equations:

$$\begin{cases} V_{ij} = -\mu_{ij}\sigma_{ij}H_{ij} & \text{along the grain boundary } \Sigma_{ij}, \\ \sigma_{ij}\nu_{ij} + \sigma_{jk}\nu_{jk} + \sigma_{ki}\nu_{ki} = 0 & \text{along triple junctions } \Sigma_{ij} \cap \Sigma_{jk}. \end{cases} \tag{1.4}$$

Here σ_{ij} and μ_{ij} denote, respectively, the surface tension and the mobility coefficient between Ω_i and Ω_j, while ν_{ij} denotes the normal of Σ_{ij}, pointing from Ω_i in direction of Ω_j. For more information about the modeling, the interested reader may consult [48].

Given the relevance of system (1.4) for applications, many efforts have been made to develop efficient numerical methods for solving (1.4). The main difficulty in directly discretizing (1.4) are topological changes in the network of grain boundaries: for instance, some of the grains could disappear. The MBO scheme handles those topological changes in a natural way.

In view of the significance for applications, it is natural to study the convergence properties of the time discretizations produced by the algorithm. The first results in this direction date back to Evans [26], and Barles and Georgelin [6], where the authors prove, independently and with different techniques, that in the simple two-phase setting, the thresholding scheme converges to a viscosity solution of MCF.

The analysis of the MBO scheme in its multiphase version cannot be handled with these methods, because of the lack of a comparison principle in this vectorial setting. Laux and Otto gave the first proof of the convergence of the scheme in the multiphase setting in [53] by exploiting its gradient flow structure. The limiting equation (1.4) is interpreted in a weak sense using geometric measure theory techniques.

Laux and Otto's work heavily exploits the minimizing movements interpretation of the thresholding scheme, which was revealed in the fundamental paper [25]. There, Esedoḡlu and Otto prove that each step of the scheme is equivalent to a solution of the following variational problem

$$\min_{\Sigma} \left\{ E_h(\Sigma) + \frac{1}{2h} d_h^2(\Sigma, \Sigma^{n-1}) \right\}, \tag{1.5}$$

where $E_h(\Sigma)$ is the *thresholding energy*, a non-local energy which approximates the area functional, and $d_h(\Sigma, \Sigma^{n-1})$ is a suitable non-local distance between the configuration

Σ and the approximation of the grain boundary at the previous time step Σ^{n-1}. Since the work of Jordan, Kinderlehrer, and Otto [47], the importance of the formerly often neglected metric in such gradient-flow structures has been widely appreciated. It is well-known that, in the case of MCF, the dissipation mechanism in the gradient flow structure is completely degenerate [65]. This is why in the noteworthy minimizing movements scheme for MCF by Almgren, Taylor, and Wang [2], and Luckhaus and Sturzenhecker [58], the authors had to introduce a proxy for the metric term. For the same reason, the distance term appearing in (1.5) is a proxy for the dissipation in MCF. It is worth noticing that from a numerical analysis viewpoint, (1.5) implies that thresholding behaves like the *implicit* Euler scheme and is therefore unconditionally stable. Also in different frameworks, this variational viewpoint turned out to be useful, such as MCF in higher codimension [57] or the Muskat problem [45].

In the works of Laux and Otto, the authors study the convergence of the MBO scheme in the form of the generalization [25], which works as follows (we restrict to the torus, i.e. periodic boundary conditions).

Algorithm 1.2.1. *Let $h > 0$ be a given step-size. Let $\{\Omega_1^0, ..., \Omega_N^0\}$ be disjoint open subsets of $[0,1)^d$ such that $[0,1)^d = \cup_i \overline{\Omega_i^0}$. For $n \in \mathbf{N}$, to obtain the new collection $\{\Omega_1^{n+1}, ..., \Omega_N^{n+1}\}$ at time $t = h(n+1)$ from the collection $\{\Omega_1^n, ..., \Omega_N^n\}$ at time $t = hn$, perform the following two steps:*

1. *For any $i = 1, ..., N$ form the convolutions with the standard Gaussian kernel G_h of width h*

$$\psi_i^n = G_h * \left(\sum_{j=1}^N \sigma_{ij} \mathbf{1}_{\Omega_j^n} \right).$$

2. *Thresholding step: define*

$$\Omega_i^{n+1} := \left\{ x : \psi_i^n(x) < \min_{j \neq i} \psi_j^n(x) \right\}.$$

This version of the MBO scheme has the only downside of the somewhat unnatural implicit choice of mobilities $\mu_{ij} = \frac{1}{\sigma_{ij}}$. Only recently, Salvador and Esedoḡlu [72] have presented a strikingly simple way to incorporate a wide class of mobilities μ_{ij}. Their algorithm is based on the fact that although the same kernel appears in the energy and the metric, each term only uses certain properties of it, which can be tuned independently: starting from two Gaussian kernels G_γ and G_β of different widths, they find positive linear combinations $K_{ij} = a_{ij} G_\gamma + b_{ij} G_\beta$, whose effective mobility and surface tension match the given μ_{ij} and σ_{ij}, respectively. The convolution step in Algorithm 1.2.1 is then replaced by

$$\psi_i^n = \sum_{j \neq i} K_{ij} * \mathbf{1}_{\Omega_j^n}.$$

It is remarkable that this algorithm retains the same simplicity and structure as the previous ones [64, 25]. In Chapter 2 we present the paper [51] where, together with Laux, we prove the first convergence result for this new general scheme. We exploit the gradient-flow structure and show that under the natural assumption of energy convergence, any limit of thresholding satisfies De Giorgi's inequality, a weak notion of solution to multiphase MCF.

This result fits into the theory of general gradient flows and crucially depends on De Giorgi's abstract framework. This research direction at the level of MCF was initiated by Laux and Otto and appeared in the lecture notes [55]. There, De Giorgi's inequality is derived for the simple model case of two phases. In the first part of this thesis, we complete these ideas and use a careful localization argument to generalize this result to the multiphase case. A further particular novelty of this work is that for the first time, we prove the convergence of the new scheme for arbitrary mobilities [72]. De Giorgi's general strategy we are implementing in [51] is also related to the approaches by Sandier and Serfaty [73] and Mielke [66]. They provide sufficient conditions for gradient flows to converge in the same spirit as Γ-convergence of energy functionals implies the convergence of minimizers. In the dynamic situation, it is clear that one needs conditions on both energy and metric in order to verify such a convergence.

Our proof rests on the fact that thresholding, like any minimizing movements scheme, satisfies a sharp energy-dissipation inequality of the form

$$\begin{aligned} &E_h(\Sigma^h(T)) + \frac{1}{2}\int_0^T \left(\frac{1}{h^2} d_h^2(\Sigma^h(t), \Sigma^h(t-h)) + |\partial E_h|^2(\tilde{\Sigma}^h(t)) \right) dt \\ &\leq E_h(\Sigma(0)), \end{aligned} \tag{1.6}$$

where $\Sigma^h(t)$ denotes the piecewise constant interpolation in time of the MBO approximations, $\tilde{\Sigma}^h(t)$ denotes another, intrinsic interpolation in terms of the variational scheme, cf. Lemma A.1.4, and $|\partial E_h|$ is the metric slope of E_h, cf. (A.11).

Our main goal is to pass to the limit in (1.6) and obtain the sharp energy-dissipation relation for the limit, which in the simple two-phase case formally reads

$$\sigma \mathrm{Area}(\Sigma(T)) + \frac{1}{2}\int_0^T \int_{\Sigma(t)} \left(\frac{1}{\mu} V^2 + \mu\sigma^2 H^2 \right) dS\, dt \leq \sigma \mathrm{Area}(\Sigma(0)). \tag{1.7}$$

To this end, one needs sharp lower bounds for the terms on the left-hand side of (1.6). While the proof of the lower bound on the metric slope of the energy

$$\liminf_{h\downarrow 0} \int_0^T |\partial E_h|^2(\tilde{\Sigma}^h(t))\, dt \geq \mu\sigma^2 \int_0^T \int_{\Sigma(t)} H^2 dS\, dt$$

is a straightforward generalization of the argument in [55], the main novelty of our work lies in the sharp lower bound for the distance-term of the form

$$\liminf_{h\downarrow 0} \int_0^T \frac{1}{h^2} d_h^2(\Sigma^h(t), \Sigma^h(t-h))\, dt \geq \frac{1}{\mu} \int_0^T \int_{\Sigma(t)} V^2\, dS\, dt.$$

This requires us to work on a mesoscopic time scale $\tau \sim \sqrt{h}$, which is much larger than the microscopic time-step size h and which is natural in view of the parabolic nature of our problem.

It is remarkable that De Giorgi's inequality (1.7) in fact characterizes the solution of MCF under additional regularity assumptions. Indeed, if $\Sigma(t)$ evolves smoothly, this inequality can be rewritten as

$$\frac{1}{2}\int_0^T \int_{\Sigma(t)} \sigma \Big(\frac{1}{\sqrt{\mu\sigma}} V + \sqrt{\mu\sigma} H \Big)^2 dS\, dt \leq 0,$$

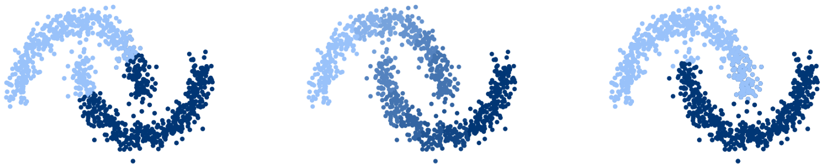

Figure 1.2: One iteration of the MBO scheme with relatively large time-step size on the two moons data set. The initial configuration (left) is diffused in the first step (middle picture). The thresholding step then turns this into a new clustering (right).

and therefore $V = -\mu\sigma H$. For the expository purpose, we focused here on the vanilla two-phase case. In the multiphase case, the resulting inequality implies both the PDEs and the balance of forces condition, cf. Remark A.2.3. An optimal energy-dissipation relation like the one here also plays a crucial role in the recent weak-strong uniqueness result for multiphase MCF by Fischer, Hensel, Laux, and Simon [29]. There, a new dynamic analog of calibrations is introduced and uniqueness is established in the following two steps: *(i)* any strong solution is a calibrated flow and *(ii)* every calibrated flow is unique in the class of weak solutions. In fact, Hensel and Laux recently showed in [40] that (a slightly weaker version of) De Giorgi's inequality is sufficient for weak-strong uniqueness.

1.3 The MBO scheme for data clustering

In the context of data science, the MBO scheme is a particular instance of a graph-based learning algorithm used for data clustering [62, 61, 78]. Graph-based learning algorithms build a weighted graph $G = (V, W)$ on top of a given data set $V = \{x_1, \ldots, x_n\} \subset \mathbf{R}^d$, with edges weighted by a non-increasing function of the distance of the data points, and try to exploit this structure to detect the underlying geometry of the data set to partition it into $P \in \mathbf{N}$ clusters. The MBO scheme produces successive proposed partitions by alternating the following two steps: *(i)* diffusion on the graph; *(ii)* thresholding. For the sake of simplifying the exposition, let us assume for the moment that $P = 2$, so that a clustering can be encoded by a function $\chi : V \to \{0, 1\}$ (in this case, χ is the indicator function of one of the two clusters). Each iteration of the algorithm works as follows.

Algorithm 1.3.1. *Let $\chi : V \to \{0, 1\}$ encode an initial guess for the clustering and let $h > 0$ be a given step size. To obtain a new partition of the data set, update the given one by performing the following two operations:*

1. ***Diffusion**. Solve the heat equation on the graph at time h with initial value χ, i.e. set*
$$u := e^{-h\Delta_G}\chi,$$
where Δ_G is a suitably defined graph Laplacian.

2. ***Thresholding**. Update the partition by defining $\chi : V \to \{0, 1\}$ according to*
$$\Big\{x \in V : \ \chi(x) = 1\Big\} = \Big\{x \in V : \ u(x) \geq \frac{1}{2}\Big\}.$$

The scheme can be implemented very efficiently in this case too, indeed, the first operation corresponds to solving the heat equation on a finite graph, which can be done by computing the matrix exponential of a graph Laplacian; while the second step is just a pointwise operation. See Figure 1.2 for one iteration of the MBO scheme on the two moons data set.

The MBO scheme has a very good benchmark in clustering problems, it can for instance be used for image segmentation [62, 61]. After its introduction, several authors found modifications of the algorithm with better accuracy in semisupervised learning tasks (i.e. classification problems with low labeling rates): for example, VolumeMBO [46] and PoissonMBO [12].

In recent years, considerable effort has been made to give theoretical justifications and mathematical foundations to data science algorithms used in practice. An emerging area of research focuses on characterizing the large data limit of the outcomes of graph-based learning methods: given a sequence of weighted graphs $G_n = (V_n, W_n)$ such that $V_n \subset V_{n+1}$ and $|V_n| = n$ for each $n \in \mathbf{N}$, one considers the outcome of a given algorithm on each of the graphs and tries to identify its continuum limit as the number n of data-points grows to infinity. In order to have a well-posed mathematical problem, one requires the data points to satisfy the *manifold assumption*: even if they lay in a high dimensional feature space, i.e. $\{x_i\}_{i=1}^{+\infty} \subset \mathbf{R}^d$, their intrinsic dimension k is lower, because – as usual in applications – many of the features will be correlated. One then usually assumes that $\{x_i\}_{i=1}^{+\infty} \subset M$, for a k-dimensional closed Riemannian submanifold $M \subset \mathbf{R}^d$. Under this assumptions – and given a sequence of positive localization parameters $\{\epsilon_n\}_{n\in\mathbf{N}}$ – one constructs the family of weighted graphs $G_n = (V_n, W_n)$ by setting:

1. $V_n = \{x_1, ..., x_n\}$;
2. $(W_n)_{ii} = 0$ and for $1 \leq i \neq j \leq n$

$$(W_n)_{ij} = \frac{1}{\epsilon_n^k} \eta\left(\frac{|x_i - x_j|}{\epsilon_n}\right),$$

   where η is a non-increasing, sufficiently regular function, and $|\cdot|$ denotes the d-dimensional Euclidean distance.

When we additionally assume that the data points $\{x_i\}_{i=1}^{+\infty}$ are independently sampled on M, with a distribution given by $\nu := \rho\,\mathrm{Vol}_M$, for a smooth, positive function $\rho \in C^\infty(M)$, the graphs just constructed are referred to as *random geometric graphs*. In this context, two predominant methods have emerged as extremely suited for identifying large data limits: variational methods and viscosity solutions techniques. The former have been used to study various graph cut algorithms [36, 34], p-Laplacian regularization [76], and spectral clustering [35], just to name a few. The latter have for example been used to understand the consistency of Lipschitz learning [10].

Even if the study of large data limits may seem very abstract, the theoretical results often lead to insights into the choice of parameters for a given algorithm and its regime of validity. For instance, Laplace learning (also known as Label Propagation) is a widely used algorithm for classification tasks and has been successfully used in a semi-supervised setting: one is given a data set made of a few labeled points and many unlabeled ones, and the goal is to *propagate* the labels to the whole data set. Laplace learning does this by extending the classification function – initially defined only on the labeled points – to a globally defined harmonic function on the graph. It was observed in [70] that the algorithm may actually behave badly when the number of labeled points is too small. In [15], the authors rigorously identify the continuum limit for Laplace learning, and – by doing this – they precisely relate the regime of validity of the algorithm in the semi-supervised setting to the labeling rate used.

When working with the MBO scheme for data clustering in practice, one has to tune two parameters: the step size and the number of eigenvalues of the graph Laplacian used to approximate the heat operator. It can be observed numerically that poor choices of these quantities lead to poor accuracy. For instance, on the one hand, if the step size is chosen too small compared to the characteristic length scale of the graph, the algorithm will be *pinned*, because in the first step, there would not be enough diffusion for any data point to change its label. On the other hand, if the step size is chosen too large, the algorithm will immediately jump to a trivial state. It should also be remarked that the actual implementation of the continuum MBO scheme really corresponds to performing a graph MBO scheme for a very particular graph: a regular grid. Indeed, computing the convolution of two functions by the Fast Fourier Transform corresponds to discretizing space by using a regular grid. For all these reasons, a theoretical analysis of the MBO scheme in its data science formulation is not only natural but also desirable.

The first rigorous analysis of the graph MBO scheme on regular grids is [68], where the authors prove the convergence of the space-time discretization of the two-class algorithm in the viscosity solutions setting. In the general data science framework, the distribution of the data points is unknown and far from being a regular grid. In this thesis, we will present two works [50, 52] where, together with Laux, we use both variational methods and viscosity solutions techniques to study the large data limit of the MBO scheme for data clustering.

The starting point is once again the minimizing movements interpretation of Esedoḡlu and Otto [25], which was rephrased in the graph context by Bertozzi et al. in [78]. It says that in each iteration, the scheme produces a new clustering χ_{new} starting from a clustering χ_{old} by solving the following minimization problem

$$\chi_{new} \in \operatorname*{argmin}_{u:V\to[0,1]} \left\{ E_h^G(u) + \frac{1}{2h} d_{h,G}^2(u, \chi_{old}) \right\}, \tag{1.8}$$

where $h > 0$ is the chosen step size, E_h^G is the graph-thresholding energy – which can be thought of as a nonlocal energy approximating the graph-cut – and $d_{h,G}$ is a suitable distance. By successively applying the algorithm, one produces a sequence $\{\chi^l\}_{l=0}^{+\infty}$ of clusterings of the data set. The interpretation (1.8) implies that for $N \in \mathbf{N}$ large enough, the clustering χ^N can be thought of as being close to a local minimizer of the graph thresholding energy. Two questions are then natural:

(i) What is the asymptotic behavior of these local minimizers?

(ii) What is the asymptotic behavior of the dynamics produced by the algorithm?

We give a qualitative answer to Question (i) in [50], and a quantitative answer to Question (ii) in [52]. Both papers are presented in detail in Chapter 3. Here, we outline their content.

1.3.1 Γ-convergence of the thresholding energies

To address Question (i), the right tool is Γ-convergence – a notion of convergence for sequences of functionals that ensures the convergence of local minimizers, which was originally introduced by De Giorgi and Franzoni in [21]. In [50] we study the Γ-convergence of the thresholding energies when the number of data points goes to infinity and the step size h goes to zero. In the setting of random geometric graphs, we show that it holds almost surely that the first limit – i.e. when we let the number of data points go to infinity – is the continuum thresholding energy on the manifold M, which for measurable functions $u : M \to [0,1]$ is defined as

$$E_h^M(u) = \frac{1}{\sqrt{h}} \int_M (1-u) e^{-h\Delta_{\rho^2}} u \rho^2 d\,\mathrm{Vol}_M\,.$$

Here $e^{-h\Delta_{\rho^2}}$ is the heat operator on the manifold associated to the weighted Laplacian with weight ρ^2 (which on smooth functions $f \in C^\infty(M)$ acts as $\Delta_{\rho^2} f := -\frac{1}{\rho^2}\,\mathrm{div}(\rho^2 \nabla f)$). This Γ-convergence takes place in the weak-TL^2 sense, a notion of convergence based on optimal transport allowing one to compare functions that are in L^2-spaces with respect to two different probability measures. The second limit – i.e. when we let $h \downarrow 0$ – is understood in the sense of Γ-convergence with respect to the strong $L^1(M)$-topology. We prove that when $h \downarrow 0$, the energies E_h^M Γ-converge to a constant multiple of the weighted perimeter functional, i.e.

$$\Gamma(L^1(M)) - \lim_{h\downarrow 0} E_h^M = F,$$

where for a measurable function $u : M \to [0,1]$ we define

$$F(u) := \begin{cases} \frac{1}{\sqrt{\pi}} \int_{\partial^*\{u=1\}} \rho^2 \mathcal{H}^{k-1}(dx) & \text{if } u \in BV(M,\{0,1\}), \\ +\infty & \text{otherwise.} \end{cases}$$

These two results say that, qualitatively, the outcomes of the MBO scheme can be thought of as local minimizers of the weighted perimeter functional. This formalizes the intuitive idea that the MBO scheme is driven by minimizing a nonlocal version of the graph-cut. The results that we just stated are valid also when we consider an arbitrary number of classes $P \geq 2$, in which case the thresholding energy will take into account the pairwise interaction of the different classes, and the continuum interpretation of the outcomes of the scheme will correspond to an optimal partition problem. These facts will be described in detail in Chapter 3, but let us briefly comment on the key ingredients for proving them. The main point for proving that the thresholding energies on the graph converge to the nonlocal thresholding energy on the manifold as the number of data points goes to infinity is the convergence and regularization properties of the heat operators: we prove that whenever u_n are functions defined on the first n data points which converge to $u \in L^2(M)$ weakly with respect to the $TL^2(M)$-convergence, then for each fixed $h > 0$ the corresponding heat operators $e^{-h\Delta} u_n$ converge (up to constants in the weighted Laplacian) to $e^{-h\Delta_{\rho^2}} u$ *strongly* in $TL^2(M)$. The convergence of the thresholding energies is then just a corollary that uses the fact that products of weakly convergent functions and strongly convergent ones converge to the product of their limits.

For the second limit, i.e., letting the step size h go to zero, the main difficulty is that the heat kernel on a manifold does not enjoy any translation invariance property as the Euclidean one. To overcome this problem, we use a careful localization argument in space-time which allows to approximate the manifold heat kernel with its Euclidean version on the tangent bundle - this is achieved by means of the asymptotic expansion for the heat kernel. The Γ-lim sup is obtained by localizing on the reduced boundary of the set of finite perimeter, while the Γ-lim inf inequality is obtained by means of the blow-up method of Fonseca and Müller [30], see also [1, 4].

1.3.2 Convergence of the dynamics

The selection of local minimizers of the thresholding energy strongly depends on the dynamics of the algorithm, which is the main motivation for studying Question (ii) and is the object of [52]. There, together with Laux, we study the convergence of the dynamics of the two-class MBO scheme. We restrict to this case because in this setting one can use the comparison principle for MCF, and thus it allows for the use of viscosity solutions. When dealing with the multiclass MBO scheme, the lack of a comparison principle makes the problem much harder and requires different techniques. After the first works on viscosity solutions [20], the machinery has proven to be a solid way to develop a theory of weak solutions for many problems satisfying a maximum principle – and its use is the basis for many fundamental contributions in geometric PDEs [18, 27] numerical analysis [6, 43] and, more recently, for new results in theoretical data science [10, 11, 9].

We work with a sequence of weighted geometric graphs $G_n = (V_n, W_n)$ constructed as before. In this setting, we study the convergence of the sequence of dynamics of the MBO scheme as the data size n goes to infinity.

The paper [52] contains two main results. For the first one, in the MBO scheme, we replace the heat operators on the graphs with abstract operators $S_n : (0, +\infty) \times \mathcal{V}_n \to \mathcal{V}_n$ which are linear in the second variable (here $\mathcal{V}_n$ is the space of real-valued functions defined on the vertex set V_n) and we show that if the sequence $\{S_n\}_{n\in\mathbf{N}}$ approximates well enough the heat kernel corresponding to a weighted Laplace–Beltrami operator on the manifold, then we have the convergence of the dynamics of the MBO scheme on the graphs to the viscosity solution of MCF on the manifold. The conditions that the operators $\{S_n\}$ have to satisfy are three: (i) they should satisfy an approximate maximum principle, (ii) they should approximate the action of the heat kernel on smooth functions in a uniform sense, and (iii) their action on the constant function $\mathbf{1}$ should be close enough to the constant $\mathbf{1}$. All these properties are made quantitatively precise in Theorem C.2.2 in Appendix C.

In the second main result of [52] we check that (i), (ii) and (iii) are satisfied with high probability on random geometric graphs when S_n are chosen to be the heat operators on the graphs or the operators obtained by cutting off frequencies higher than a threshold K_n, defined precisely in Item (iv) in Theorem C.2.4 in Appendix C. Let us stress that the latter result is crucial for applications. Indeed, when one implements the MBO scheme on a large dataset, computing the full heat kernel is intractable, and thus one usually works with an approximate version of it, obtained by cutting off high frequencies in precisely the way described above. Our result gives a solid mathematical justification for this procedure, proving that the scheme converges in the large data limit to the viscosity solution to MCF provided the frequency cut-off is chosen according to $K_n \geq (\log(n))^q$ where q is a suitable positive real number and n is the number of data points. We also notice that our result gives sufficient conditions on how to choose the length scale ϵ_n and the time-step size h_n in order to ensure convergence of the scheme. In particular, the choice of h_n is not anymore based solely on rules of thumb but has theoretical foundations. Previously, only a negative result ensuring pinning of the scheme was known [78, Theorem 4.2]. However, we point out that the conditions on ϵ_n and h_n are only sufficient, but not sharp. Indeed, we expect that the convergence of the scheme should hold true whenever $\epsilon_n = o(h_n)$, while our conditions imply that $\epsilon_n = o(h_n^{3/2})$. The sharp rate $\epsilon_n = o(h_n)$ was verified in the simple setting of the deterministic two-dimensional regular grid $\mathbf{Z}^2$ in [68], and is based on the explicit expression for the heat kernel on regular grids. But an extension to the general setting in which we are working requires different techniques. Let us spend a few words on the strategy of the proofs used in [52].

For the abstract convergence result, we follow the general scheme of proof of Barles and Georgelin [6], also used in [68], where the authors prove the convergence of the classical MBO scheme to a viscosity solution to MCF in the Euclidean space. Given a smooth open set $\Omega \subset M$, the idea is to prove that the upper semicontinuous envelope u^* and the lower semicontinuous envelope u_* of the piecewise constant in time interpolations of outcomes of the MBO scheme (with initial values $\Omega \cap G_n$) are, respectively, a viscosity subsolution and a viscosity supersolution to MCF on the manifold. After doing that, one can use the comparison principle for viscosity solutions to MCF on weighted manifolds, due to Illmanen [41], to compare u^* and u_* with the unique viscosity solution u to MCF with initial value Ω, to show that $\text{sign}_*(u) \leq u_*$ and $\text{sign}^*(u) \geq u^*$. In order to check that u^* and u_* are, respectively, a viscosity subsolution and a viscosity supersolution to MCF we have to adapt the strategy in [6] to our setting: we need to carefully identify admissible error terms for the argument of [6]. Finally, to apply the comparison principle, it is crucial to show an ordering of the initial values in the sense that $\text{sign}_*(u(0,\cdot)) \leq u_*(0,\cdot)$ and $\text{sign}^*(u(0,\cdot)) \geq u^*(0,\cdot)$. We verify this in the general case of a weighted manifold by carefully checking that one iteration of the MBO scheme with step size h produces a set whose normal distance from the previous one is of order h. This issue seems to have been overlooked in the literature and we believe that our proof fills an important gap in the previous works, even in the Euclidean setting.

For the result on random geometric graphs, we draw inspiration from [24]. There, the authors work on a fixed graph with points sampled independently from a weighted manifold and consider the error in a uniform sense between the restriction of the manifold heat kernel to the graph and the operator obtained by considering the first K frequencies of the graph heat kernel. Their estimate, however, cannot be applied in our setting because, since we want to take the number of data points to infinity, we have to be able to take the frequency cut-off K to infinity jointly with it. For this reason, a careful interplay between the chosen rates of convergence for K, the step size h, and the localization parameter ϵ is needed. In [52] we thus obtain a new estimate giving precise conditions on the relation between the frequency cut-off and the number of data points. To get this, we make use of recent results on the convergence of spectra of graph Laplacians [33, 13, 14].

The rest of the thesis is organized as follows: in Chapter 2 we summarize the results of the paper [51], which we include, without introduction, in Appendix A. Chapter 3 contains a summary of the results of the papers [50, 52], which we include, without introduction, in Appendix B and Appendix C respectively.

CHAPTER 2

THE MBO SCHEME FOR MATERIALS SCIENCE

In this chapter, we present paper [51] containing the proof of the convergence of a modified thresholding scheme to a De Giorgi's solution to multiphase MCF. The paper was written jointly by Laux and the author of the current thesis, and it was published in

Calc. Var. Partial Differ. Equ. 61(1):Paper No. 35, 42, 2022.

In Appendix A the paper is reproduced, without introduction, in the form it appeared on the ArXiv at https://arxiv.org/abs/2101.11663.

2.1 The modified thresholding scheme

In this section, we introduce the relevant background on the modified thresholding scheme of Salvador and Esedoḡlu introduced in [25], which is the object of our analysis. As explained in the introduction, the algorithm is used to approximate the evolution by multiphase MCF, and its main novelty is that it allows for great freedom in the choice of mobilities.

Assume that we are given surface tensions $\sigma := (\sigma_{ij})_{ij} \in \mathbf{R}^{N\times N}$ and mobilities $\mu = (\mu_{ij})_{ij} \in \mathbf{R}^{N\times N}$, where we assume that $\sigma_{ii} = \mu_{ii} = 0$, $\sigma_{ij} = \sigma_{ji}$, and $\mu_{ij} = \mu_{ji}$. Assume that $\gamma, \beta \in \mathbf{R}$ are two positive constants such that $\gamma > \beta > 0$. For any pair $1 \le i \ne j \le N$, define $a_{ij}, b_{ij} \in \mathbf{R}$ as the solution to the linear system

$$\begin{cases} \sigma_{ij} &= \frac{a_{ij}\sqrt{\gamma}}{\sqrt{\pi}} + \frac{b_{ij}\sqrt{\beta}}{\sqrt{\pi}}, \\ \mu_{ij}^{-1} &= \frac{a_{ij}}{\sqrt{\pi}\sqrt{\gamma}} + \frac{b_{ij}}{\sqrt{\pi}\sqrt{\beta}}. \end{cases}$$

We then define the kernel K_{ij} as

$$K_{ij}(z) := a_{ij}G_\gamma(z) + b_{ij}G_\beta(z),$$

where for any $t > 0$ the Gaussian G_t is defined as

$$G_t(z) := \frac{1}{\sqrt{4\pi t}^d} \exp\left(-\frac{|z|^2}{4t}\right), \quad z \in \mathbf{R}^d.$$

For any $h > 0$, the rescaled kernel K_{ij}^h is defined as

$$K_{ij}^h(z) := \frac{1}{\sqrt{h}^d} K_{ij}\left(\frac{z}{\sqrt{h}}\right).$$

The algorithm of Salvador and Esedoḡlu used to approximate the evolution of a partition $\{\Omega_1^0, ..., \Omega_N^0\}$ of the torus $[0,1)^d$ under multiphase MCF with mobilities μ and surface tensions σ is then as follows.

Algorithm 2.1.1 (Algorithm A.1.1)**.** *Let* $\{\Omega_1^0, ..., \Omega_N^0\}$ *be disjoint open subsets of* $[0,1)^d$ *such that* $[0,1)^d = \cup_i \overline{\Omega_i^0}$, *to obtain the new collection* $\{\Omega_1^{n+1}, ..., \Omega_N^{n+1}\}$ *at time* $t = h(n+1)$ *from the collection* $\{\Omega_1^n, ..., \Omega_N^n\}$ *at time* $t = hn$ *perform the following operations:*

1. *For any* $i = 1, \dots, N$ *form the convolutions*

$$\psi_i^n = \sum_{j \neq i} K_{ij}^h * \mathbf{1}_{\Omega_j^n}.$$

2. *Thresholding step, define*

$$\Omega_i^{n+1} := \left\{ x : \psi_i^n(x) < \min_{j \neq i} \psi_j^n(x) \right\}.$$

Before presenting the results of [51], let us pause for a moment to heuristically explain why Algorithm 2.1.1 actually works.

Heuristics. Assume that the initial configuration $\{\Omega_1^0, \dots, \Omega_N^0\}$ is made of smooth open sets. We want to provide evidence that one step of Algorithm 2.1.1 actually approximates the evolution by multiphase MCF, at least away from triple junctions. To this end, pick $1 \leq i \neq j \leq N$ and define $\Sigma_{ij}^0 := \partial\Omega_i^0 \cap \partial\Omega_j^0$. Assume that $x \in \Sigma_{ij}^0$ is away from any triple junction. One can then show that with an exponentially small in h error, near x we have

$$\begin{aligned}
\psi_i^0 &\approx K_{ij}^h * \mathbf{1}_{\Omega_j^0}(x), \\
\psi_j^0 &\approx K_{ji}^h * \mathbf{1}_{\Omega_i^0}(x), \\
\psi_k^0 &\approx K_{ki}^h * \mathbf{1}_{\Omega_i^0}(x) + K_{kj}^h * \mathbf{1}_{\Omega_j^0}(x), \quad \forall k \notin \{i,j\}.
\end{aligned}$$

In particular, if $a_{ij} < a_{ik} + a_{ki}$ and $b_{ij} < b_{ik} + b_{kj}$ for all $k \notin \{i,j\}$ we have that the updated interface Σ_{ij}^1 is the set of points y close to x where

$$K_{ij}^h * \mathbf{1}_{\Omega_j^0}(y) \approx K_{ji}^h * \mathbf{1}_{\Omega_i^0}(y). \tag{2.1}$$

Now, by a careful Taylor expansion around x, which can be found in a heuristic form in [60] and in a rigorous form in [31], we have that if $\nu_{ij}(x)$ denotes the unit normal of Σ_{ij}^0 pointing from Ω_i^0 to Ω_j^0, for $r \in \mathbf{R}$ small enough

$$\begin{aligned}
K_{ij}^h * \mathbf{1}_{\Omega_i^0}(x + r\nu_{ij}(x)) \approx \frac{a_{ij} + b_{ij}}{2} &- \left(\frac{a_{ij}}{\sqrt{\gamma}} + \frac{b_{ij}}{\sqrt{\beta}}\right) \frac{r}{\sqrt{4\pi h}} \\
&- \left(a_{ij}\sqrt{\gamma} + b_{ij}\sqrt{\beta}\right) \frac{H_{ij}(x)}{\sqrt{4\pi}} + O(h).
\end{aligned}$$

Inserting into (2.1) and solving for r we obtain

$$r \approx -hH_{ij}(x)\sqrt{\gamma\beta}\frac{a_{ij}\sqrt{\gamma}+b_{ij}\sqrt{\beta}}{a_{ij}\sqrt{\beta}+b_{ij}\sqrt{\gamma}}.$$

It can be checked that

$$\sqrt{\gamma\beta}\frac{a_{ij}\sqrt{\gamma}+b_{ij}\sqrt{\beta}}{a_{ij}\sqrt{\beta}+b_{ij}\sqrt{\gamma}} = \sigma_{ij}\mu_{ij}.$$

In particular, we have

$$r \approx -h\sigma_{ij}\mu_{ij}H_{ij}(x).$$

This says that the normal movement of the point x is approximately equal to the evolution by MCF. In [51] we rigorously prove the convergence of the scheme to a De Giorgi's solution to multiphase MCF.

2.1.1 Main results

In [51], we prove the first convergence result for Algorithm 2.1.1. This is the first proof of De Giorgi's inequality in the multiphase case. We exploit the gradient-flow structure and show that under the natural assumption of energy convergence, any limit of thresholding satisfies De Giorgi's inequality, a weak notion of solution to multiphase MCF. This assumption is inspired by the fundamental work of Luckhaus-Sturzenhecker [58] and has appeared in the context of thresholding in [53, 54].

Before stating the main result of the paper, we present the notion of De Giorgi's solution for multiphase MCF that we will use. Hereafter, we denote by $\mathcal{A}$ the class of measurable partitions of the torus and by $\mathcal{M}$ its convex relaxation, i.e.

$$\mathcal{A} := \left\{ \chi : [0,1)^d \to \{0,1\}^P \middle| \sum_{i=1}^N \chi_i = 1 \right\},$$
$$\mathcal{M} := \left\{ u : [0,1)^d \to [0,1]^P \ \middle| \ \sum_{i=1}^N u_i = 1 \right\}.$$

If $\chi \in \mathcal{A}$, we set $\Omega_i := \{\chi_i = 1\}$ for every $1 \le i \le N$. If χ is such that $\nabla\chi$ is a bounded measure, we denote by $\Sigma_{ij} := \partial^*\Omega_i \cap \partial^*\Omega_j$, the intersection between the reduced boundaries $\partial^*\Omega_i$ and $\partial^*\Omega_j$. For each $h > 0$ we denote by E_h the thresholding energy, defined for any $u \in \mathcal{M}$ as

$$E_h(u) := \frac{1}{\sqrt{h}} \sum_{i,j=1}^N \sigma_{ij} \int_{[0,1)^d} u_i K_{ij}^h * u_j dx. \tag{2.2}$$

The following is a slightly simplified version of Definition A.2.2.

Definition 2.1.2. Given $\chi^0 \in \mathcal{A}$ and such that $\nabla\chi^0$ is a bounded measure, a map $\chi : [0,1)^d \times (0,T) \to \{0,1\}^N$ such that $\sum_i \chi_i = 1$ and $\chi \in L^1((0,T), BV([0,1)^d))^N$ is called a De Giorgi solution to the multiphase MCF with surface tensions σ_{ij} and mobilities μ_{ij} provided the following three facts hold:

1. There exist $H_{ij} \in L^2(\mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx)dt)$ which are mean curvatures in the weak sense, i.e., such that for any test vector field $\xi \in C_c^\infty([0,1)^d \times (0,T))^d$

$$\begin{aligned}\sum_{i,j} \sigma_{ij} \int_{[0,1)^d\times(0,T)} (\nabla \cdot \xi - \nu_{ij} \cdot \nabla\xi\nu_{ij}) \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt \\ = -\sum_{i,j} \sigma_{ij} \int_{[0,1)^d\times(0,T)} H_{ij}\nu_{ij} \cdot \xi \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt.\end{aligned} \tag{2.3}$$

2. There exist normal velocities $V_{ij} \in L^2(\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt)$ with

$$\begin{aligned}\int_{[0,1)^d} \eta(t=0)\chi_i^0 dx + \int_{[0,1)^d\times(0,T)} \partial_t\eta\ \chi_i\ dxdt \\ + \sum_{k\neq i} \int_{[0,1)^d\times(0,T)} \eta V_{ik}\ \mathcal{H}^{d-1}_{|\Sigma_{ik}(t)}(dx)dt = 0\end{aligned}$$

for all $\eta \in C_c^\infty([0,1)^d \times [0,T))$.

3. De Giorgi's inequality is satisfied, i.e., for almost every $t \in (0,T)$

$$\begin{aligned}\sum_{i,j} \sigma_{ij}\mathcal{H}^{d-1}(\Sigma_{ij}(t)) + \frac{1}{2}\sum_{i,j} \int_{[0,1)^d\times(0,t)} \left(\frac{V_{ij}^2}{\mu_{ij}} + \mu_{ij}\sigma_{ij}^2 H_{ij}^2\right) \mathcal{H}^{d-1}_{|\Sigma_{ij}(s)}(dx)ds \\ \leq \sum_{i,j} \sigma_{ij}\mathcal{H}^{d-1}(\Sigma_{ij}^0).\end{aligned} \tag{2.4}$$

Definition 2.1.2 is motivated by the fact that for smooth solutions, De Giorgi's inequality (2.4) characterizes the evolution by MCF away from triple junctions and (2.3) prescribes the boundary condition at triple junctions. For more on this we refer to Remark A.2.3. The definition finds inspiration in the general gradient flows framework [5, 73], and it generalizes the previous two-phase version [55].

To state our main result, we need to introduce some notation: let $\chi^0 \in \mathcal{A}$ be an initial partition of the torus. For each $1 \leq i \leq N$ define $\Omega_i^0 := \{\chi_i^0 = 1\}$. Fix $h > 0$ and iteratively apply the modified thresholding scheme in Algorithm 2.1.1 with initial value $\{\Omega_1^0, \ldots, \Omega_N^0\}$. For each $n \in \mathbf{N}$, denote by $\chi^n \in \mathcal{A}$ the outcome of each step, i.e. for each $1 \leq i \leq N$

$$\chi_i^n(x) = 1 \Leftrightarrow x \in \Omega_i^n.$$

We then define the piecewise constant in time interpolation χ^h as

$$\chi^h(x,t) = \chi^n(x), \quad t \in [nh, (n+1)h) \text{ for } n \in \mathbf{N}.$$

We are now ready to state our conditional convergence result. The energy convergence assumption (2.6) appearing in the theorem below is motivated by a similar one on the implicit time discretization of Luckhaus and Sturzenhecker [58], and has also appeared in previous work on the thresholding scheme [53, 54, 55]. As of now, this assumption can be verified only in particular cases, such as before the first singularity [77] or for certain types of singularities, namely mean convex ones [22, 31].

Theorem 2.1.3 (Theorem A.2.1)**.** *Given $\chi^0 \in \mathcal{A}$ and such that $\nabla\chi^0$ is a bounded measure and a sequence $h \downarrow 0$, assume that there exists $\chi : [0,1)^d \times (0,T) \to [0,1]^N$ such that*

$$\chi^h \rightharpoonup \chi \text{ in } L^1([0,1)^d \times (0,T)). \tag{2.5}$$

Then $\chi \in \{0,1\}^N$ almost everywhere, $\sum_i \chi_i = 1$ and $\chi \in L^1((0,T), BV([0,1)^d))^N$.

If we assume that

$$\limsup_{h\downarrow 0} \int_0^T E_h(\chi^h(t))dt \le \sum_{i,j} \sigma_{ij} \int_0^T \mathcal{H}^{d-1}(\Sigma_{ij}(t))dt, \tag{2.6}$$

then χ is a De Giorgi solution in the sense of Definition 2.1.2.

2.1.2 Outline of the proof

To prove Theorem 2.1.3 we start from the minimizing movements interpretation of Esedoḡlu and Otto, which says that for each $h > 0$ and for any $t > 0$

$$\chi^h(t) \in \operatorname*{argmin}_{u\in\mathcal{M}} \left\{ E_h(u) + \frac{1}{2h} d_h^2(u, \chi^h(t-h)) \right\}, \tag{2.7}$$

where E_h is defined in (2.2), while d_h is a distance on $\mathcal{M}$, defined as

$$d_h^2(u,v) := -2\sqrt{h} \sum_{i,j} \int (u_i - v_i) K_{ij}^h * (u_j - v_j)dx, \quad u,v \in \mathcal{M}.$$

It turns out that by the general theory of gradient flows [5], any minimizing movements dynamics like (2.7) satisfies the following abstract energy-dissipation inequality: for any $T \in \mathbf{N}h$

$$\begin{aligned} &E_h(\chi^h(T)) + \frac{1}{2}\int_0^T \left(\frac{1}{h^2} d_h^2(\chi^h(t), \chi^h(t-h)) + |\partial E_h(u^h(t))|^2 \right) dt \\ &\le E_h(\chi^h(0)). \end{aligned} \tag{2.8}$$

Here $u^h(t)$ is the so-called variational interpolation, which for $n \in \mathbf{N}$ and $t \in ((nh,(n+1)h]$ is defined by

$$u^h(t) \in \operatorname*{argmin}_{u\in\mathcal{M}} \left\{ E_h(u) + \frac{d_h^2(u,\chi^n)}{2(t-nh)} \right\},$$

and $|\partial E_h|(u^h(t))$ is the metric slope defined by

$$|\partial E_h|(u^h(t)) := \lim_{d_h(u^h(t),v)\to 0} \frac{(E_h(u^h(t)) - E_h(v))_+}{d_h(u^h(t),v)} \in [0,\infty]. \tag{2.9}$$

To prove Theorem 2.1.3 we pass to the limit $h \downarrow 0$ into (2.8):

1. **Compactness, Lemma A.2.4**. By using that the right hand side is bounded in h, one can prove that the family of partitions $\{\chi^h\}_{h>0}$ is precompact in the strong $L([0,1)^d \times (0,T))^N$ topology. Moreover, any weak limit of the family belongs to $\mathcal{A} \cap BV([0,1)^d)^N$. In what follows, we can thus pick a sequence χ^h converging to $\chi \in \mathcal{A} \cap BV([0,1)^d)^N$, and denote by Σ_{ij} the interfaces of the limit.

2. **Energy terms**. For the first left hand side term and for the right-hand side term we use the Γ-convergence and consistency of the energies E_h. It is shown in [25] that E_h Γ-converge in the $L^1(\mathbf{R}^d)$ topology to the energy E, defined for $u \in \mathcal{M}$ as

$$E(u) := \begin{cases} \sum_{i,j} \sigma_{ij} \mathcal{H}^{d-1}(\Sigma_{ij}) & \text{if } u \in \mathcal{A} \cap BV([0,1)^d)^N, \\ +\infty & \text{otherwise.} \end{cases}$$

Moreover, for any $\chi \in \mathcal{A} \cap BV([0,1)^d)^N$, $E_h(\chi)$ converges to $E(\chi)$ as $h \downarrow 0$.

3. **Metric slope term, Proposition A.2.6**. It can be shown that for each $i \neq j$ there exists a mean curvature $H_{ij} \in L^2(\mathcal{H}^{d-1}_{\Sigma_{ij}(t)}(dx)dt)$ in the sense of (A.17). Moreover the following inequality is true:

$$\liminf_{h\downarrow 0} \int_0^T |\partial E_h|^2(u^h(t))dt \geq \sum_{i,j} \mu_{ij}\sigma_{ij}^2 \int_0^T \int_{\Sigma_{ij}(t)} |H_{ij}(x,t)|^2 \mathcal{H}^{d-1}(dx)dt. \quad (2.10)$$

4. **Velocity term, Proposition A.2.5**. One shows that for any $1 \leq k \leq N$ there exists $V_k \in L^2(|\nabla\chi_k|dt)$ such that

$$\partial_t \chi_k = V_k |\nabla \chi_k| dt \quad (2.11)$$

in the sense of distributions, and if we define $V_{ij}(x,t) := V_i(x,t)|_{\Sigma_{ij}(t)}$ then we have

$$\liminf_{h\downarrow 0} \int_0^T \frac{1}{h^2} d_h^2(\chi^h(t), \chi^h(t-h))dt \geq \sum_{i,j} \frac{1}{\mu_{ij}} \int_0^T \int_{\Sigma_{ij}(t)} |V_{ij}(x,t)|^2 \mathcal{H}^{d-1}(dx)dt.$$

These steps clearly allow to pass to the limit in the abstract inequality (2.8) and to derive Theorem 2.1.3. To prove (2.10) and (2.11) we build on the ideas of [51], where the analogous inequalities are derived in the two-phase case. The main idea in extending Laux and Otto's strategy is a careful localization argument, which allows us to look at the dynamics of χ locally on each interface Σ_{ij}. In this way, the evolution of Σ_{ij} away from triple junctions can be treated as in the two-phase case up to error terms controlled by our localization method.

CHAPTER 3

LARGE DATA LIMIT OF THE MBO SCHEME FOR DATA CLUSTERING

In this chapter, we present papers [50, 52] about the large data limit of the MBO scheme for data clustering. Both papers have been written jointly by Laux and the author of the current thesis.

Section 3.2 describes the content of [50], which is reproduced in Appendix B without introduction. The paper is currently under review, and has appeared online on the ArXiv at https://arxiv.org/abs/2112.06737.

Section 3.3 describes the content of [52], which is reproduced in Appendix C without introduction. The paper is currently under review, and has appeared online on the ArXiv at https://arxiv.org/abs/2209.05837.

A summary of the content of the two papers will also appear as a PAMM report in [49].

3.1 The large data limit framework

In this section, we introduce the common mathematical framework for [50, 52], which is the by now standard setting for large data limit results in graph-based learning.

Let $M \subset \mathbf{R}^d$ be a closed, k-dimensional Riemannian submanifold of $\mathbf{R}^d$. We assume that we are given a sequence of data points $\{x_i\}_{i=1}^{+\infty} \subset M$, and a sequence of positive localization parameters $\{\epsilon_n\}_{n=1}^{+\infty}$. We then construct a sequence of weighted graphs $G_n = (V_n, W_n)$ where for each $n \in \mathbf{N}$:

(i) The vertex set V_n is given by $V_n := \{x_1, \ldots, x_n\}$;

(ii) The weight matrix W_n has zeros on the diagonal, and, for every pair $1 \le i \neq j \le n$

$$(W_n)_{ij} = \frac{1}{\epsilon_n^k} \eta\left(\frac{|x_i - x_j|}{\epsilon_n}\right),$$

where $|\cdot|$ denotes the d-Euclidean distance, and $\eta : [0, +\infty) \to [0, +\infty)$ is a sufficiently regular, non-increasing function, such that $\eta(0) = 1$.

We now introduce the random walk graph Laplacian. First of all, for each $n \in \mathbf{N}$ we define the degree of the node $i \in \{1, ..., n\}$ as

$$d_n(x_i) := \frac{1}{n} \sum_{j=1, j \neq i}^{n} (W_n)_{ij},$$

and we denote by D_n the diagonal matrix $D_n := \operatorname{diag}(d_n(x_1), ..., d_n(x_n))$. The random walk Laplacian on G_n is the $(n \times n)$-matrix Δ_{G_n} defined as

$$\Delta_{G_n} := \frac{1}{\epsilon_n^2} \left(\mathbb{I}_n - \frac{1}{n} D_n^{-1} W_n \right).$$

The random walk Laplacian is a positive definite, self-adjoint operator with respect to the scalar product $\langle \cdot, \cdot \rangle_{\mathcal{V}_n}$, which is defined as $\langle u, v \rangle_{\mathcal{V}_n} := \frac{1}{n} \sum_{i=1}^{n} d_n(x_i) u(x_i) v(x_i)$ for $u, v \in \mathcal{V}_n := \{u | u : V_n \to \mathbf{R}\}$.

When working with random geometric graphs, i.e. when the points $\{x_i\}_{i \in \mathbf{N}}$ are drawn independently according to some probability measure $\nu := \rho \operatorname{Vol}_M$ for some smooth density ρ, we assume

$$\epsilon_n \gg \left(\frac{\log(n)}{n} \right)^{1/k},$$

so that the graphs become almost surely connected for n large enough. In particular, the degrees are positive and the random walk Laplacian is well-defined.

3.2 Γ-convergence of the thresholding energies

The object of study in [50] is the multiclass version of the MBO scheme, which allows to perform data clustering with any number $P \in \mathbf{N}$ of classes. In this section, we work with random geometric graphs $G_n = (V_n, W_n)$ as previously constructed.

3.2.1 The multiclass MBO scheme

Given $P \in \mathbf{N}$, a clustering of a data set V_n can be encoded by a function $\chi : V_n \to \{0, 1\}^P$ such that $\sum_{m=1}^{P} \chi_m = 1$, where χ_m is the m-th component of χ, for $1 \leq m \leq P$. The MBO scheme works as follows.

Algorithm 3.2.1. *Let $h > 0$ be a given step-size, and let $\chi : V_n \to \{0, 1\}^P$ be a proposed partition of the data set into P clusters. To obtain a new clustering, perform the following steps:*

1. ***Diffusion**. For every $m = 1, ..., P$ define*

$$u_m := \sum_{l \neq m} e^{-h \Delta_{G_n}} \chi_l.$$

2. ***Thresholding**. For every $m = 1, ..., P$ update the clusters by setting*

$$\{\chi_m = 1\} := \left\{ x \in V_n : \ u_m(x) < \min_{l \neq m} u_l(x) \right\}.$$

The minimizing movements interpretation of Esedoḡlu and Otto states that the two steps in Algorithm 3.2.1 are equivalent to solving the following variational problem

$$\chi \in \underset{u:V_n\to[0,1]^P,\sum_{i=1}^P u_i=1}{\operatorname{argmin}} \left\{ E^h_{n,\epsilon_n}(u) + \frac{1}{2h} d_h^2(u,\chi) \right\}, \tag{3.1}$$

where E^h_{n,ϵ_n} is the *thresholding* energy, defined for $u : V_n \to [0,1]^P$ as

$$E^h_{n,\epsilon_n}(u) = \frac{1}{\sqrt{h}} \sum_{i\neq j} \langle u_i, e^{-h\Delta_{G_n}} u_j \rangle_{\mathcal{V}_n},$$

and d_h is a suitable distance on the space of $[0,1]^P$-valued functions on V_n. Because of (3.1), by iterating sufficiently many times Algorithm 3.2.1, we can think of the final outcome as being close to a (local) minimizer of the thresholding energy. The object of [50] is to study the asymptotic behavior of these (local) minimizers.

3.2.2 Main results

The paper [50] draws a rigorous connection between outcomes of the MBO scheme and local minimizers of an optimal partition problem on the data manifold. Firstly, we relate the outcomes of the algorithm with local minimizers of a nonlocal energy on the manifold by studying the large data limit of the MBO scheme with a fixed step size $h > 0$. Secondly, we study the convergence property of the latter by letting the step size converge to zero.

To precisely state our results, let us introduce, for any $h > 0$, the thresholding energy E_h, defined for measurable functions $u : M \to [0,1]^P$, such that $\sum_{i=1}^P u_i = 1$, as

$$E_h(u) := \frac{1}{\sqrt{h}} \sum_{i\neq j} \int_M u_i e^{-\kappa h \Delta_{\rho^2}} u_j \rho^2 d \operatorname{Vol}_M,$$

where $\kappa := \kappa(\eta)$ is a constant depending on the choice of η, and which hereafter we assume to be equal to one, and $e^{-\kappa h \Delta_{\rho^2}}$ denotes the heat semigroup associated to the weighted Laplacian Δ_{ρ^2}, defined by its action on smooth functions $f \in C^\infty(M)$ by

$$\Delta_{\rho^2} f := -\frac{1}{\rho^2} \operatorname{div}\left(\rho^2 \nabla f\right).$$

The first main result of [50] is the following discrete-to-nonlocal convergence.

Theorem 3.2.2 (Theorem B.2.1)**.** *Under the scaling regime* $\left(\frac{\log(n)}{n}\right)^{\frac{1}{k+2}} \ll \epsilon_n \ll 1$ *it holds almost surely that for every* $h > 0$

$$\Gamma(weak - TL^2(M)) - \lim_{n\to+\infty} E^h_{n,\epsilon_n} = E_h.$$

The second main result of [50] relates the nonlocal energy E_h to an optimal partition problem. Before stating the result, let us introduce some notation. If $u : M \to \mathbf{R}$ is a $BV(M)$ function, we denote by $|Du|_{\rho^2}$ its ρ^2-weighted total variation. If $u \in BV(M, \{0,1\}^P)$, for every $i = 1,\dots,P$ we set $\Omega_i := \{u_i = 1\}$, and for $i \neq j$ we set $\Sigma_{ij} = \partial^*\Omega_i \cap \partial^*\Omega_j$, where $\partial^*\Omega_i$ denotes the reduced boundary of Ω_i. We then have the following result.

Theorem 3.2.3 (Theorem B.2.5). *It holds that*

$$\Gamma(L^1(M)) - \lim_{h \downarrow 0} E_h = E,$$

where the energy E is defined on measurable functions $u : M \to [0,1]^P$, such that $\sum_{i=1}^P u_i = 1$, as

$$E(u) := \begin{cases} \frac{1}{\sqrt{\pi}} \sum_{i \neq j} |Du_i|_{\rho^2}(\Sigma_{ij}) & \text{if } u \in BV(M, \{0,1\}^P), \\ +\infty & \text{otherwise.} \end{cases}$$

3.2.3 Outline of the proofs

We now give the main ideas for the proof of Theorem 3.2.2 and of Theorem 3.2.3.

Ideas for Theorem 3.2.2. We show that under the stated scaling regime, it holds almost surely that if $u^n \in \mathcal{V}_n$ is a sequence of $[0,1]^P$-valued functions converging weakly in $TL^2(M)$ to a function $u : M \to [0,1]^P$, then

$$\lim_{n \to +\infty} E^h_{n,\epsilon_n}(u^n) = E_h(u). \tag{3.2}$$

This corresponds to proving continuous convergence of the functionals E^h_{n,ϵ_n} with respect to weak-$TL^2(M)$ convergence, which implies Γ-convergence. To see that (3.2) holds, we first observe that if we have a sequence of functions $v^n \in \mathcal{V}_n$ converging to v weakly in $TL^2(M)$ and a sequence of functions $w^n \in \mathcal{V}_n$ converging to w strongly in $TL^2(M)$, then

$$\lim_{n \to +\infty} \langle v^n, w^n \rangle_{\mathcal{V}_n} = \int_M vw\rho^2 d\,\mathrm{Vol}_M\,.$$

We then spell out the definition of the graph thresholding energy

$$E^h_{n,\epsilon_n}(u^n) = \frac{1}{\sqrt{h}} \sum_{i \neq j} \langle u^n_i, e^{-h\Delta_{G_n}} u^n_j \rangle_{\mathcal{V}_n},$$

and we observe that to conclude, it thus suffices to show that the heat operators on the graphs upgrade weak-$TL^2(M)$ convergence to strong-$TL^2(M)$ convergence. In fact, we prove the following result.

Theorem 3.2.4 (Theorem B.2.2). *Under the assumptions of Theorem 3.2.2, it holds almost surely that if $u^n \in \mathcal{V}_n$ is a sequence of functions converging weakly in $TL^2(M)$ to u, then for every $h > 0$ the sequence $e^{-h\Delta_{G_n}} u^n$ converges strongly to $e^{-h\Delta_{\rho^2}} u$ in $TL^2(M)$.*

The proof of Theorem 3.2.4 relies on the energy-dissipation inequality satisfied by any gradient flow: by a standard argument, one can reduce to the case in which $u \in C^\infty(M)$ and the sequence u^n is obtained by restricting u to the n-th graph. Denoting by $v^n := e^{-t\Delta_{G_n}} u^n$, it holds that for all $t > 0$

$$E_n[v^n(t)] + \frac{1}{2} \int_0^t |\Delta_{G_n} v^n(s)|^2_{\mathcal{V}_n} ds + \frac{1}{2} \int_0^t \left| \frac{d}{ds} v^n(s) \right|^2_{\mathcal{V}_n} ds \leq E_n[u^n], \tag{3.3}$$

where E_n is the Dirichlet energy on the graph. Then one observes:

1. We have $\sup_n E_n[u^n] < +\infty$.

2. The energies E_n Γ-converge almost surely with respect to the strong TL^2-convergence to the Dirichlet energy E_{Dir} on the weighted manifold. Moreover, they satisfy a Γ-compactness principle ([34, Theorem 1.4] and Theorem B.5.16). In particular, up to a subsequence, $v^n(t)$ converges strongly in $TL^2(M)$ to a function $v(t) \in L^2(M)$.

3. The sequences $\Delta_{G_n} v^n$ and $\frac{d}{dt} v^n$ converge in weak-$TL^2(M)$ to $\Delta_{\rho^2} v$ and, respectively, $\frac{d}{dt} v$.

In particular, one can pass to the limit in (3.3) to obtain

$$\begin{aligned} &E_{Dir}[v(t)] + \frac{1}{2}\int_0^t \int_M |\Delta_{\rho^2} v|^2 \rho^2 d\,\mathrm{Vol}_M\, ds + \frac{1}{2}\int_0^t \int_M |\frac{dv}{ds}|^2 \rho^2 d\,\mathrm{Vol}_M\, ds \\ &\leq E_{Dir}[u], \end{aligned} \tag{3.4}$$

where E_{Dir} denotes the ρ^2-weighted Dirichlet energy on the manifold. Inequality (3.4) is the energy-dissipation inequality for the heat-flow on M. This completely characterizes the limit v as the solution to the heat equation on the manifold with initial value $u \in C^\infty(M)$ and concludes the argument for Theorem 3.2.4.

Ideas for Theorem 3.2.3. The proof relies on a careful space-time localization argument and on the asymptotic expansion of the heat kernel.

Γ-lim sup. For the Γ-lim sup inequality, we show that whenever $u \in BV(M, \{0,1\}^P)$, we have

$$\lim_{h \downarrow 0} E_h(u) = E(u).$$

By writing out the definition of $E_h(u)$, one checks that this is a consequence of the following: Assume that $E, F \subset M$ are sets of finite perimeter, then

$$\lim_{h\downarrow 0} \frac{1}{\sqrt{h}} \int_M \chi_F \left(\chi_E - e^{-h\Delta_{\rho^2}} \chi_E\right) d\mu = \frac{1}{\sqrt{\pi}} \int_{\partial^* E \cap \partial^* F} \langle \sigma_E(x), \sigma_F(x)\rangle_x |D\chi_F|(x), \tag{3.5}$$

where $\langle \sigma_E(x), \sigma_F(x)\rangle_x$ denotes the inner product on the tangent space $T_x M$ between the inner unit normals $\sigma_E(x)$ and $\sigma_F(x)$ of E and, respectively, F at the point $x \in \partial^* E \cap \partial^* F$. One then proves (3.5) in four steps:

1. Rewriting $\chi_E - e^{-h\Delta_{\rho^2}}\chi_E$ as $-\int_0^h \frac{d}{dt} e^{-t\Delta_{\rho^2}} \chi_E dt$, one checks that (3.5) is equivalent to proving that for every $x \in \partial^* E \cap \partial^* F$

$$\frac{1}{\sqrt{\pi}} \langle \sigma_F(x), \sigma_E(x)\rangle = \lim_{h\downarrow 0} \left\langle \sigma_F(x), \frac{1}{\sqrt{h}} \int_0^h \nabla e^{-t\Delta_{\rho^2}} \chi_E(x) dt \right\rangle. \tag{3.6}$$

2. Using Gaussian upper bounds for the manifold heat kernel, we may localize in space around x: after this step, (3.7) reduces to proving

$$\frac{1}{\sqrt{\pi}} \langle \sigma_F(x), \sigma_E(x)\rangle = \lim_{h\downarrow 0} \left\langle \sigma_F(x), \frac{1}{\sqrt{h}} \int_0^h \nabla e^{-t\Delta_{\rho^2}} \chi_{E\cap B_{h^s}(x)}(x) dt \right\rangle, \tag{3.7}$$

for some $s < \frac{1}{2}$.

3. If we denote by p the heat kernel for Δ_{ρ^2} we have that

$$\frac{1}{\sqrt{h}}\int_0^h \nabla e^{-t\Delta_{\rho^2}}\chi_{E\cap B_{h^s(x)}}(x)dt = \frac{1}{\sqrt{h}}\int_0^h\int_{B_{h^s(x)}} \nabla_x p(t,x,y)\chi_E(y)\rho^2 d\,\mathrm{Vol}_M\,dt.$$

   Since we are integrating on a small ball around x, we can use the asymptotic expansion for the heat kernel. We then show that as $h \downarrow 0$, the only relevant term in the expansion is the zero-th order one. In particular, we are left to proving that

$$\frac{1}{\sqrt{\pi}}\langle \sigma_F(x), \sigma_E(x)\rangle = \tag{3.8}$$
$$\lim_{h\downarrow 0}\frac{1}{\sqrt{h}}\left\langle \sigma_F(x), \int_0^h\int_M \nabla_x\left(\frac{e^{-\frac{d(x,y)^2}{4t}}}{(4\pi t)^{k/2}}v^0(x,y)\right)\chi_{E\cap B_{h^s(x)}}(y)\rho^2(y)d\,\mathrm{Vol}_M\,dt\right\rangle,$$

   where v^0 is the zero-th order coefficient in the asymptotic expansion for the heat kernel.

4. Using normal coordinates around x, one can reduce (3.8) to the analogous statement on the Euclidean space. This is then easily checked by using De Giorgi's structure theorem for sets of finite perimeter.

Γ-lim inf. The proof of the Γ-lim inf inequality uses the blow-up technique of Fonseca and Müller [30]: this allows for a localization in space which, in turn, allows for the use of the asymptotic expansion of the heat kernel that reduces the problem to an analogous one in the Euclidean space.

3.2.4 Extensions

The results in [50] can be extended in various directions:

1. *Surface tensions.* In the algorithm, it is possible to use different "surface tensions" between different classes, which are additional parameters one can use to penalize certain interfaces between labels more than others.

2. *Forcing.* It is possible to adapt the scheme to a semi-supervised setting, i.e. when same of the data-points are already labeled. In this case, the labeling information is used to produce a forcing/drift $f^n\colon V_n \to \mathbf{R}^P$ which is then used to slightly change the threshold value as $\left\{\chi_i^{q+1} = 1\right\} := \left\{u_i - \sqrt{h}f_i < u_j - \sqrt{h}f_j, \forall j \neq i\right\}$. The resulting scheme is just as efficient as the original one. The analysis is almost unchanged as the forcing only leads to a continuous perturbation in the energy.

3. *Volume constraints.* These Γ-convergence techniques immediately apply to variants with volume constraints on the clusters and give a convergence result for VolumeMBO.

4. *Other graph Laplacians.* The results apply to a larger variety of graph Laplacians and data dependent weights.

3.3 Convergence of the dynamics

The object of study in [52] is the two-class MBO scheme. In [52] we are interested in the asymptotic behavior of the sequences of partitions obtained by iteratively applying the MBO scheme: in other words, instead of just focusing on the final outcomes of the algorithm, we study the MBO dynamics. The analysis is restricted to the two-class case (i.e. $P = 2$), because our proof relies on the viscosity solutions framework, which is not available in the multiclass setting.

3.3.1 Two-class MBO

The two-class MBO scheme is a particular instance of Algorithm 3.2.1. In this case, a clustering can be encoded by a function $\chi^{G_n} : V_n \to \{0, 1\}$, which is the characteristic function of one of the two clusters. The two-class thresholding scheme is then as follows.

Algorithm 3.3.1. *Let $h_n > 0$ be a chosen step-size, and let $\chi^{G_n} : V_n \to \{0, 1\}$ be a proposed partition of the data-set into two clusters. To obtain a new clustering, perform the following steps:*

1. ***Diffusion**. For every $m = 1, \dots, P$ define*
$$u_n = e^{-h\Delta_{G_n}} \chi^{G_n}.$$
2. ***Thresholding**. Update the clusters by setting*
$$\{\chi^{G_n} = 1\} := \left\{ x \in V_n : u_n(x) \geq \frac{1}{2} \right\}.$$

When implementing the scheme in practice, it is often intractable to compute the whole matrix exponential $e^{-h\Delta_{G_n}}$: usually, one computes an approximate version of the graph heat kernel by first applying to its argument the operator P_{K_n}, the projection on the subspace generated by the first $K_n \in \mathbf{N}$ eigenvectors of the graph Laplacian Δ_{G_n}. If we denote by $S_n : [0, +\infty) \times \mathcal{V}_n \to \mathcal{V}_n$ an abstract operator, linear in the second variable, the two-class MBO scheme can be written in an abstract version:

Algorithm 3.3.2. *Let $h_n > 0$ be a chosen step-size, and let $\chi^{G_n} : V_n \to \{0, 1\}$ be a proposed partition of the data set into two clusters. To obtain a new clustering, perform the following steps:*

1. ***Diffusion**. For every $m = 1, \dots, P$ define*
$$u_n = S_n(h_n, \chi^{G_n}).$$
2. ***Thresholding**. Update the clusters by setting*
$$\{\chi^{G_n} = 1\} := \left\{ x \in V_n : u_n(x) \geq \frac{1}{2} \right\}.$$

Iterating the algorithm produces a sequence of partitions $\{\chi^{l,G_n}\}_{l=1}^{+\infty}$ of the data-points V_n. We can then define piecewise constant in time interpolations by

$$u^{h_n,G_n}(t, x) := 2\chi^{l,G_n} - 1, \quad x \in V_n, t \in [lh_n, (l + 1)h_n).$$

We are interested in the asymptotic behavior of u^{h_n,G_n} as the number of data points goes to infinity.

3.3.2 Main results

Before presenting the main results of [52], let us informally introduce MCF on a weighted manifold (M, g, ρ^2). MCF is defined as the trajectory of steepest descent for the weighted area functional: one can prove that a smooth evolution $[0, +\infty) \ni t \to \Omega_t$ of smooth open sets evolve by MCF if

$$g(V, \nu) = -\frac{1}{\rho^2}\operatorname{div}(\rho^2 \nu),$$

where V is the velocity vector field of the evolution and ν is the outer unit normal field. If $u : [0, +\infty) \times M \to \mathbf{R}$ is a smooth function such that for any $s \in \mathbf{R}$ the sets $\Omega_t^s := \{x \in M : u(t, x) > s\}$ evolve by MCF, then u solves

$$\partial_t u = \left\langle g - \frac{Du \otimes Du}{|Du|^2}, D^2 u \right\rangle + g\left(\frac{\nabla \rho^2}{\rho^2}, \nabla u\right), \tag{3.9}$$

which is called the level set formulation of MCF. MCF satisfies a comparison principle that allows to interpret equation (3.9) in the viscosity sense.

Let $\Omega \subset M$ be a smooth open set and let Γ_0 be its boundary. We denote by $u : [0, +\infty) \times M \to \mathbf{R}$ the unique viscosity solution of the level set formulation of MCF with density ρ^2 (see Section C.3 for the details) with initial value $sd(\cdot, \Gamma_0) = d_M(x, \Omega^c) - d_M(x, \Omega)$, the signed distance function from Γ_0. For any $t > 0$ we also define

$$\Omega_t := \{x \in M \mid u(t, x) > 0\}, \ \Gamma_t = \{x \in M \mid u(t, x) = 0\}. \tag{3.10}$$

For each $n \in \mathbf{N}$, we consider the function u^{h_n, G_n} obtained as in the previous section by starting the MBO scheme on G_n with initial value given by $\chi^{1,G_n} = \mathbf{1}_{V_n \cap \Omega}$. We introduce the upper semicontinuous limit and the lower semicontinuos limit of u^{h_n, G_n} as

$$\begin{aligned} u^*(t, x) := \sup\Bigg\{ \limsup_{n\to+\infty} u^{h_n,G_n}(t_n, x_n) \,\Bigg|\, & t_n > 0, \lim_{n\to+\infty} t_n = t, \\ & x_n \in G_n, \lim_{n\to+\infty} x_n = x \Bigg\}, \\ u_*(t, x) := \inf\Bigg\{ \liminf_{n\to+\infty} u^{h_n,G_n}(t_n, x_n) \,\Bigg|\, & t_n > 0, \lim_{n\to+\infty} t_n = t, \\ & x_n \in G_n, \lim_{n\to+\infty} x_n = x \Bigg\}. \end{aligned}$$

We will prove that on random geometric graphs, it holds almost surely that the dynamics of the two-class MBO scheme converge to a viscosity solution to MCF in the following sense: for each $t > 0$ it holds

$$\begin{aligned} u_*(x, t) &= 1 \quad \text{if } x \in \Omega_t, \\ u^*(x, t) &= -1 \,\text{if } x \in (\Omega_t \cup \Gamma_t)^c. \end{aligned}$$

We get to this result in two steps:

1. First, we show the convergence of the MBO scheme in the abstract setting. Let u^{h_n, G_n} constructed as before but starting from the outcomes of the abstract

thresholding scheme (where in the diffusion step the heat operator is replaced by an abstract operator S_n). We show that *if* the operators S_n approximate the heat semigroup on the manifold well-enough, then the dynamics of the MBO scheme converge to the unique viscosity solution to MCF. For this result, the graphs are not assumed to be random.

2. Secondly, we show that when we work with random geometric graphs and S_n is either the heat kernel or a suitable approximation of it, then the assumptions for the previous result hold true with high probability. This allows to conclude almost sure convergence of the dynamics of the MBO scheme.

More precisely, we have the following two results.

Theorem 3.3.3 (Theorem C.2.2)**.** *Assume that:*

(i) The operators S_n satisfy the maximum principle up to errors $h_n^{3/2}$, i.e., for n large enough and for each $u, v \in \mathcal{V}_n$ it holds

$$u \leq v \Rightarrow S_n(h_n, u) \leq S_n(h_n, v) + \left(\max_{V_n} |u| + \max_{V_n} |v|\right) O(h_n^{3/2});$$

(ii) The operators S_n approximate the heat operator on the manifold, i.e. there exists a constant $\kappa > 0$ such that for every function $f \in C^\infty(M)$ we have

$$\max_{x \in V_n} \left|S(h_n, f)(x) - e^{-h\kappa\Delta_{\rho^2}} f(x)\right| = (\sup |f|)\, o(\sqrt{h_n}) + \mathrm{Lip}(f)\, O(h_n^{3/2});$$

where the functions $o(\sqrt{h_n}), O(h_n^{3/2})$ are independent of f.

(iii) The operators S_n almost preserve the total mass in the sense that

$$\max_{x \in V_n} |S_n(h_n, \mathbf{1}_{G_n})(x) - 1| = O(h_n^{3/2});$$

then the dynamics of the MBO scheme converge to the unique viscosity solution to MCF on (M, g, ρ^2).

Theorem 3.3.4 (Theorem C.2.4 and Corollary C.2.6)**.** *Let G_n be random geometric graphs. Assume that $q, \alpha, \beta > 0$ are suitably chosen and*

(i) $h_n \gg (\log(n))^{-\alpha}$;

(ii) $\left(\frac{\log(n)}{n}\right)^{\frac{1}{k+4}} \ll \epsilon_n \ll (\log(n))^{-\beta}$;

(iii) $K_n \geq (\log(n))^q$;

(iv) The eigenvalues of Δ_{ρ^2} satisfy $\inf_{i \in \mathbf{N}}(\lambda_i - \lambda_{i-1}) > 0$.

Then, for n sufficiently large, the operators $e^{-t\Delta_n}$ and $e^{-t\Delta_n} P_{K_n}$ satisfy conditions (i)*,* (ii)*, and* (iii) *in Theorem 3.3.3 with high probability. In particular, the dynamics of the MBO scheme converge almost surely to the unique viscosity solution to MCF on (M, g, ρ^2) whenever the number of eigenvectors K_n used for approximating the heat operator on G_n satisfies* (iii)*.*

3.3.3 Outline of the proofs

We now give the main ideas for the proof of Theorem 3.3.3 and of Theorem 3.3.4.

Ideas for Theorem 3.3.3. One shows the following two items:

1. The functions u^* and u_* are a viscosity subsolution and, respectively, a viscosity supersolution of the level set formulation of MCF on the weighted manifold (M, g, ρ^2). This is proved by following the strategy adopted by Barles and Georgelin in [6], where the result is proved in the Euclidean space for the continuum MBO scheme. The difficulty in extending their result to the graph setting lays on the fact that we want to substitute the continuum heat semigroup with the discrete abstract operators S_n. Properties (i), (ii), and (iii) in Theorem 3.3.3 ensure that the errors we make in performing this operation are negligible in the limit.

2. For the initial values, it holds that
$$u^*(0, x) \leq \text{sign}^*(u(0, x)),$$
$$u_*(0, x) \geq \text{sign}_*(u(0, x)),$$
where sign^* and sign_* are, respectively, the upper semi-continuous envelope and the lower semi-continuous envelope of the sign function. This step seems to have been overlooked in the literature.

The result then follows by an application of the comparison principle for MCF on the weighted manifold (M, g, ρ^2), which informally says that whenever v and w are viscosity subsolution and, respectively, supersolution of MCF, then

$$v(0, x) \leq w(0, x),\ x \in M \Rightarrow v(t, x) \leq w(t, x),\ t > 0,\ x \in M.$$

.

Ideas for Theorem 3.3.4. One needs to show that with the choice $S_n(h_n, \cdot) = e^{-h_n \Delta_{G_n}}$ or $S_n(h_n, \cdot) = e^{-h_n \Delta_{G_n}} P_{K_n}$ the assumptions of Theorem 3.3.3 hold true with high probability. We concentrate on the second choice, the other is analogous. We denote by $\{v_n^l\}_{1 \leq l \leq n}$ an orthonormal basis (with respect to the inner product $\langle \cdot, \cdot \rangle_{\mathcal{V}_n}$) made of eigenvectors for the Laplacian Δ_{G_n} corresponding to the eigenvalues $\{\lambda_n^l\}_{1 \leq l \leq n}$, which are ordered in the following way

$$0 = \lambda_n^1 < \lambda_n^2 < ... < \lambda_n^n.$$

The operator $S_n(h_n, \cdot) = e^{-h_n \Delta_{G_n}} P_{K_n}$ is then identified with the matrix

$$H_{\epsilon_n}^{K_n}(h_n, x, y) = \sum_{l=1}^{K_n} e^{-t\lambda_n^l} v_n^l(x) v_n^l(y) \frac{d_n(y)}{n}, \quad x, y \in V_n.$$

The full graph heat kernel is instead identified with the matrix

$$H_{\epsilon_n}^{n}(h_n, x, y) = \sum_{l=1}^{n} e^{-t\lambda_n^l} v_n^l(x) v_n^l(y) \frac{d_n(y)}{n}, \quad x, y \in V_n.$$

Properties (i) and (iii) hold exactly for the full graph heat kernel, thus one just has to prove that the difference $H_n^n - H_n^{K_n}$ is of the correct order.

Property (ii) is the main technical part in Theorem 3.3.3 and is proved in three steps:

1. Denoting by H the manifold heat kernel, one shows that with high probability

$$\max_{x,y\in V_n}\left|H^{K_n}_{\epsilon_n}(h_n,x,y)-\frac{\rho(y)}{n}H(h_n,x,y)\right| = o\left(\frac{\sqrt{h_n}}{n}\right).$$

   This corresponds to Lemma C.2.7 and improves a similar estimate obtained in [24]. It is proved by exploiting recent results on the convergence of the spectrum of the graph Laplacian to the spectrum of the manifold Laplacian. In particular, we use the fact that with high probability, the eigenvectors of the graph Laplacian converge in the L^∞ sense to the corresponding eigenfunction of the continuum Laplacian.

2. After choosing an optimal transport map T_n for the ∞-Wasserstein distance θ_n between ν and the empirical probability measure for the data points, one shows that with high probability, we have for every $f \in C^\infty(M)$,

$$\begin{aligned}\max_{x\in V_n}\left|S_n(h_n,f)(x)-e^{-h_n\Delta_{\rho^2}}f(x)\right| \leq & L_1 \sup_M |f|\frac{\theta_n}{\sqrt{h_n}}e^{\frac{2\theta_n \operatorname{diam}(M)}{h_n}} \\ & +\sup_M |f| o(\sqrt{h_n}) + L_2\left(\sup_M |f| + \operatorname{Lip}(f)\right)\theta_n,\end{aligned} \tag{3.11}$$

   where the constants L_1, L_2 and the function in $o(\sqrt{h_n})$ depend only on M.

   To show this, one first spells out the definition of $S_n(h_n,f)(x)$ at a point $x\in V_n$,

$$S_n(h_n,f)(x) = \sum_{y\in V_n} H^{K_n}_{\epsilon_n}(h_n,x,y)f(y).$$

   Then, the result in the previous step can be used to substitute $H^{K_n}_{\epsilon_n}(h_n,x,y)$ with the continuum quantity $\frac{\rho(y)}{n}H(h_n,x,y)$. This is the source of the error $\sup_M |f|o(\sqrt{h_n})$ in (3.11). One then has to compare

$$\sum_{y\in V_n}\frac{\rho(y)}{n}H(h_n,x,y)f(y), \tag{3.12}$$

   with

$$\int_M H(h_n,x,y)f(y)\rho^2(y)d\operatorname{Vol}_M. \tag{3.13}$$

   This can be done rewriting (3.12) as an integral w.r.t. the empirical probability measure associated to the data points. One then uses the transport map T_n to convert this into an integral with respect to the absolutely continuous measure $\rho^2\operatorname{Vol}_M$, and estimates the difference with (3.13) by using the Lipschitz property of f and the definition of θ_n. This is the source of the other errors in (3.11).

3. The final step consists in showing that the right hand side of (3.11) is indeed of the correct order. This is just a technical calculation.

APPENDIX A

DE GIORGI'S INEQUALITY FOR THE THRESHOLDING SCHEME WITH ARBITRARY MOBILITIES AND SURFACE TENSIONS

A.1 Setup and the modified thresholding scheme

Here and in the rest of the paper, $[0,1)^d$ denotes the d-dimensional torus. Thus when we deal with functions $u : [0,1)^d \to \mathbf{R}$ we always assume that they have periodic boundary conditions. In particular they can be extended periodically on $\mathbf{R}^d$. In general if u is a function as before and $f : \mathbf{R}^d \to \mathbf{R}$ then by $f * u$ we mean the convolution on $\mathbf{R}^d$ between f and the periodic extension of u, i.e.

$$f * u(x) := \int_{\mathbf{R}^d} f(z)u(x-z)dz, \ x \in \mathbf{R}^d$$

when this expression makes sense.

A.1.1 The modified algorithm

We start by describing the algorithm proposed by Salvador and Esedoḡlu in [72]. Let the symmetric matrix $\sigma = (\sigma_{ij})_{ij} \in \mathbf{R}^{N\times N}$ of surface tensions and the symmetric matrix $\mu = (\mu_{ij})_{ij}$ of mobilities be given. In this work we define for notational convenience $\sigma_{ii} = \mu_{ii} = 0$. Let $\gamma > \beta > 0$ be given. Define the matrices $\mathbb{A} = (-a_{ij})_{ij} \in \mathbf{R}^{N\times N}$ and $\mathbb{B} = (-b_{ij})_{ij} \in \mathbf{R}^{N\times N}$ by

$$a_{ij} = \frac{\sqrt{\pi}\sqrt{\gamma}}{\gamma-\beta}(\sigma_{ij} - \beta\mu_{ij}^{-1}),$$
$$b_{ij} = \frac{\sqrt{\pi}\sqrt{\beta}}{\gamma-\beta}(-\sigma_{ij} + \gamma\mu_{ij}^{-1}),$$

for $i \neq j$ and $a_{ii} = b_{ii} = 0$. Then a_{ij}, b_{ij} are uniquely determined as solutions of the following linear system

$$\begin{cases} \sigma_{ij} = \frac{a_{ij}\sqrt{\gamma}}{\sqrt{\pi}} + \frac{b_{ij}\sqrt{\beta}}{\sqrt{\pi}}, \\ \mu_{ij}^{-1} = \frac{a_{ij}}{\sqrt{\pi}\sqrt{\gamma}} + \frac{b_{ij}}{\sqrt{\pi}\sqrt{\beta}}. \end{cases}$$

The algorithm introduced by Salvador and Esedoḡlu is as follows. Let the time step size $h > 0$ be fixed. Hereafter $G^h_\gamma := G^{(d)}_{\gamma h}$ denotes the d-dimensional heat kernel (A.3) at time γh.

Algorithm A.1.1 (Modified thresholding scheme). *Let $\{\Omega^0_1, ..., \Omega^0_N\}$ be disjoint open subsets of $[0,1)^d$ such that $[0,1)^d = \cup_i \overline{\Omega^0_i}$, to obtain the new collection $\{\Omega^{n+1}_1, ..., \Omega^{n+1}_N\}$ at time $t = h(n+1)$ from the collection $\{\Omega^n_1, ..., \Omega^n_N\}$ at time $t = hn$*

1. *For any $i = 1, ..., N$ form the convolutions*

$$\phi^n_{1,i} = G^h_\gamma * \mathbf{1}_{\Omega^n_i}, \ \phi^n_{2,i} = G^h_\beta * \mathbf{1}_{\Omega^n_i}$$

2. *For any $i = 1, ..., N$ form the comparison functions*

$$\psi^n_i = \sum_{j\neq i} a_{ij}\phi^n_{1,j} + b_{ij}\phi^n_{2,j}.$$

3. *Thresholding step, define*

$$\Omega^{n+1}_i := \left\{ x : \psi^n_i(x) < \min_{j\neq i} \psi^n_j(x) \right\}.$$

We will assume the following:

$$\text{The coefficients } a_{ij}, b_{ij} \text{ satisfy the strict triangle inequality.} \tag{A.1}$$

$$\text{The matrices } \mathbb{A} \text{ and } \mathbb{B} \text{ are positive definite on } (1, ..., 1)^\perp. \tag{A.2}$$

In particular, for $v \in (1, ..., 1)^\perp$ we can define norms

$$|v|^2_{\mathbb{A}} = v \cdot \mathbb{A}v, \ |v|^2_{\mathbb{B}} = v \cdot \mathbb{B}v.$$

We remark that we need the matrices $\mathbb{A}, \mathbb{B}$ to be positive definite on $(1, ..., 1)^\perp$ to guarantee that the functional defined in (A.8) is a distance, see the comment following (A.8) below.

Observe that condition (A.1) is always satisfied if we choose γ large and β small provided the surface tensions and the inverse of the mobilities satisfy the strict triangle inequality. Indeed, define

$$m_\sigma = \min_{i,j,k}\{\sigma_{ik} + \sigma_{kj} - \sigma_{ij}\} \text{ and } M_\sigma = \max_{i,j,k}\{\sigma_{ik} + \sigma_{kj} - \sigma_{ij}\},$$

where i, j, k range over all triples of distinct indices $1 \le i, j, k \le N$. Define $m_{\frac{1}{\mu}}$ and $M_{\frac{1}{\mu}}$ in a similar way. Then a computation shows that a_{ij} and b_{ij} satisfy the (strict) triangle inequality if

$$\beta < \frac{m_\sigma}{M_{\frac{1}{\mu}}} \text{ and } \gamma > \frac{M_\sigma}{m_{\frac{1}{\mu}}},$$

which can always be achieved since $\gamma > \beta > 0$ are arbitrary. For the second condition (A.2), we have the following result of Salvador and Esedoḡlu [72].

Lemma A.1.2. *Let the matrix σ of the surface tensions and the matrix $\frac{1}{\mu}$ of the inverse mobilities (for the diagonal we set inverses to be zeros) be negative definite on $(1,...,1)^\perp$. Let $\gamma > \beta$ be such that*

$$\gamma > \frac{\min_{i=1,\ldots,N-1} s_i}{\max_{i=1,\ldots,N-1} m_i}, \ \beta < \frac{\max_{i=1,\ldots,N-1} s_i}{\min_{i=1,\ldots,N-1} m_i}$$

where s_i and m_i are the nonzero eigenvalues of $J\sigma J$ and $J\frac{1}{\mu}J$ respectively, where the matrix J has components $J_{ij} = \delta_{ij} - \frac{1}{N}$. Then $\mathbb{A}$ and $\mathbb{B}$ are positive definite on $(1,...,1)^\perp$.

In particular, if we choose γ large enough and β small enough, condition (A.2) on the matrices $\mathbb{A}, \mathbb{B}$ is satisfied provided the matrices σ and $\frac{1}{\mu}$ are negative definite on $(1,...,1)^\perp$. By a classical result of Schoenberg [74] this is the case if and only if $\sqrt{\sigma_{ij}}$ and $1/\sqrt{\mu_{ij}}$ are ℓ^2 embeddable. In particular, this holds for the choice of Read-Shockley surface tensions and equal mobilities.

For $1 \le i \ne j \le N$ define the kernels

$$K_{ij}(z) = a_{ij} G_\gamma(z) + b_{ij} G_\beta(z)$$

where, for a given $t > 0$, we define $G_t^{(d)}$ as the heat kernel in $\mathbf{R}^d$, i.e.,

$$G_t^{(d)}(z) = \frac{e^{-\frac{|z|^2}{4t}}}{\sqrt{4\pi t}^d}. \tag{A.3}$$

If the dimension d is clear from the context, we suppress the superscript (d) in (A.3). We recall here some basic properties of the heat kernel.

$$\begin{aligned}
G_t(z) &> 0 && \text{(non-negativity)},\\
G_t(z) &= G_t(Rz)\ \forall R \in O(d) && \text{(symmetry)}, \quad \text{(A.4)}\\
G_t(z) &= \frac{1}{\sqrt{t}^d} G_1\left(\frac{z}{\sqrt{t}}\right) && \text{(scaling)}, \quad \text{(A.5)}\\
G_t * G_s &= G_{t+s} && \text{(semigroup property)}, \quad \text{(A.6)}\\
G_t^{(d)}(z) &= \prod_{i=1}^d G_t^{(1)}(z_i) && \text{(factorization property)}. \quad \text{(A.7)}
\end{aligned}$$

We observe that the kernels K_{ij} are positive, with positive Fourier transform $\hat{K}_{ij}$ provided $\gamma > \max_{i,j} \sigma_{i,j}\mu_{i,j}$ and $\beta < \min_{i,j} \sigma_{i,j}\mu_{i,j}$. In particular assuming

1. σ_{ij} and $\frac{1}{\mu_{ij}}$ satisfy the strict triangle inequality,

2. σ and $\frac{1}{\mu}$ are negative definite on $(1,...,1)^\perp$,

we can always achieve the conditions posed on $\mathbb{A}, \mathbb{B}$ and the positivity of the kernels K_{ij} by choosing γ large and β small.

Given any $h > 0$ we define the scaled kernels

$$K_{ij}^h(z) = \frac{1}{\sqrt{h}^d} K_{ij}\left(\frac{z}{\sqrt{h}}\right),$$

then the first and the second step in Algorithm A.1.1 may be compactly rewritten as follows

$$\psi_i^n = \sum_{j\neq i} K_{ij}^h * \mathbf{1}_{\Omega_j^n}.$$

For later use, we also introduce the kernel

$$K(z) = \frac{1}{2}G_\gamma(z) + \frac{1}{2}G_\beta(z).$$

A.1.2 Connection to De Giorgi's minimizing movements

The first observation is that Algorithm A.1.1 has a minimizing movements interpretation. To explain this, let us introduce the class

$$\mathcal{A} := \left\{ \chi : [0,1)^d \to \{0,1\}^N \,\middle|\, \sum_{k=1}^N \chi_k = 1 \right\}$$

and its relaxation

$$\mathcal{M} := \left\{ u : [0,1)^d \to [0,1]^N \,\middle|\, \sum_{k=1}^N u_k = 1 \right\}.$$

If $\chi \in \mathcal{A}\cap BV([0,1)^d)^N$, then each of the sets $\Omega_i := \{\chi_i = 1\}$ is a set of finite perimeter. We denote by $\partial^*\Omega_i$ the reduced boundary of the set Ω_i, and for any pair $1 \le i \neq j \le N$ we denote by $\Sigma_{ij} := \partial^*\Omega_i \cap \partial^*\Omega_j$ the interface between the sets. For $u \in \mathcal{M}$ we define

$$E(u) := \begin{cases} \sum_{i,j} \sigma_{ij}\mathcal{H}^{d-1}(\Sigma_{ij}) & \text{if } u \in \mathcal{A}\cap BV([0,1)^d)^N, \\ +\infty & \text{otherwise.} \end{cases}$$

For $h > 0$ fixed we define the approximate energy E_h for $u \in \mathcal{M}$

$$E_h(u) = \sum_{i,j} \frac{1}{\sqrt{h}} \int_{[0,1)^d} u_i K_{ij}^h * u_j dx.$$

For $u, v \in \mathcal{M}$ and $h > 0$ we also define the distance

$$\begin{aligned} d_h^2(u,v) &:= -2hE_h(u-v) = -2\sqrt{h}\sum_{i,j}\int (u_i - v_i)K_{ij}^h * (u_j - v_j)dx \qquad \text{(A.8)} \\ &= 2\sqrt{h}\int |G_\gamma^{h/2} * (u-v)|_{\mathbb{A}}^2 + |G_\beta^{h/2} * (u-v)|_{\mathbb{B}}^2 \, dx, \end{aligned}$$

where we used the semigroup property (A.6) and the symmetry (A.4) to derive the last equality. We also point out that since $\sum_i u_i = \sum_i v_i = 1$ a.e., we have $G_\gamma^{h/2} * (u - v), G_\beta^{h/2} * (u-v) \in (1, ..., 1)^\perp$. Hence the assumptions on $\mathbb{A}$ and $\mathbb{B}$ guarantee that d_h defines a distance on $\mathcal{M}$ (and on $\mathcal{A}$).

Lemma A.1.3. *The pair $(\mathcal{M}, d_h)$ is a compact metric space. The function E_h is continuous with respect to d_h. For every $1 \le i \le N$ and $n \in \mathbf{N}$ define $\chi_i^n = \mathbf{1}_{\Omega_i^n}$,* *where* $\Omega_1^n, ..., \Omega_N^n$ *are obtained from* $\Omega_1^{n-1}, ..., \Omega_N^{n-1}$ *by the thresholding scheme. Then* χ^n *minimizes*

$$\frac{1}{2h} d_h^2(u, \chi^{n-1}) + E_h(u) \text{ among all } u \in \mathcal{M}. \tag{A.9}$$

Proof. For $u, v \in \mathcal{M}$ definition (A.8) and the fact that $\mathbb{A}$ and $\mathbb{B}$ are positive definite imply that d_h is a distance on $\mathcal{M}$. The fact that $(\mathcal{M}, d_h)$ is compact and E_h is continuous is just a consequence of the fact that d_h metrizes the weak convergence in L^2 on $\mathcal{M}$, the interested reader may find the details of the reasoning in [55]. We are thus left with showing that χ^n satisfies (A.9). For $u, v \in L^2([0,1)^d)$ define

$$(u, v) = \frac{1}{\sqrt{h}} \sum_{i,j} \int u_i K_{ij}^h * v_j dx,$$

then by the symmetry (A.4) of the Gaussian kernel and by the symmetry of both matrices $\mathbb{A}, \mathbb{B}$ it is not hard to show that $(\cdot, \cdot)$ is symmetric. In particular we can write for any $u \in \mathcal{M}$

$$\begin{aligned}
\frac{1}{2h} d_h^2(u, \chi^{n-1}) + E_h(u) &= -E_h(u - \chi^{n-1}) + E_h(u) \\
&= -(u - \chi^{n-1}, u - \chi^{n-1}) + (u, u) \\
&= 2(\chi^{n-1}, u) - (\chi^{n-1}, \chi^{n-1}).
\end{aligned}$$

Thus (A.9) is equivalent to the fact that χ^n minimizes (χ^{n-1}, u) among all $u \in \mathcal{M}$. Since by (2)

$$(\chi^{n-1}, u) = \int \sum_i u_i \psi_i^n dx,$$

we see that χ^n minimizes the integrand pointwise, and thus it is a minimizer for the functional. □

The previous lemma allows us to apply the general theory of gradient flows in [5] to this particular problem. We record the key statement for our purposes in the following lemma, which will be applied to $(\mathcal{M}, d_h)$, where d_h is the metric (A.8).

Lemma A.1.4. *Let* $(\mathcal{M}, d)$ *be a compact metric space and* $E : \mathcal{M} \to \mathbf{R}$ *be continuous. Given* $\chi^0 \in \mathcal{M}$ *and* $h > 0$ *consider a sequence* $\{\chi^n\}_{n \in \mathbf{N}}$ *satisfying*

$$\chi^n \text{ minimizes } \frac{1}{2h} d^2(u, \chi^{n-1}) + E(u) \text{ among all } u \in \mathcal{M}.$$

Then we have for all $t \in \mathbf{N}h$

$$E(\chi(t)) + \frac{1}{2} \int_0^t \left(\frac{1}{h^2} d^2(\chi(s+h), \chi(s)) + |\partial E|^2(u(s)) \right) ds \leq E(\chi^0). \tag{A.10}$$

Here $\chi(t)$ *is the piecewise constant interpolation,* $u(t)$ *is the so-called variational interpolation, which for* $n \in \mathbf{N}$ *and* $t \in ((n-1)h, nh]$ *is defined by*

$$u(t) \in \operatorname{argmin}_{u \in \mathcal{M}} \left\{ E(u) + \frac{d^2(u, \chi^{n-1})}{2(t - (n-1)h)} \right\},$$

and $|\partial E|(u)$ *is the metric slope defined by*

$$|\partial E|(u) := \lim_{d(u,v)\to 0} \frac{(E(u)-E(v))_+}{d(u,v)} \in [0,\infty]. \tag{A.11}$$

Moreover, the variational interpolation $u(t)$ *satisfies*

$$\int_0^\infty \frac{1}{2h^2} d^2(u(t),\chi(t))dt \le E(\chi^0), \tag{A.12}$$

$$E(u(t)) \le E(\chi(t)) \text{ for all } t \ge 0. \tag{A.13}$$

A.2 Statement of results

Our main result is the convergence of the modified thresholding scheme to a weak notion of multiphase mean curvature flow. More precisely, given an initial partition $\{\Omega_1^0, ..., \Omega_N^0\}$ of $[0,1)^d$ encoded by $\chi^0 : [0,1)^d \to \{0,1\}^N$ such that $\sum_i \chi_i^0 = 1$, define $\chi^h : [0,1)^d \times \mathbf{R} \to \{0,1\}^N$ by setting

$$\begin{aligned} &\chi^h(t,x) = \chi^0(x) \text{ for } t < h, \\ &\chi^h(t,x) = \chi^n(x) \text{ for } t \in [nh,(n+1)h) \text{ for } n \in \mathbf{N}. \end{aligned} \tag{A.14}$$

If χ^0 is a function of bounded variation, we denote by $\Sigma_{ij}^0 := \partial^*\Omega_i^0 \cap \partial^*\Omega_j^0$. Our main result is contained in the following theorem.

Theorem A.2.1. *Given* $\chi^0 \in \mathcal{A}$ *and such that* $\nabla\chi^0$ *is a bounded measure and a sequence* $h \downarrow 0$*; let* χ^h *be defined by* (A.14). *Assume that there exists* $\chi : [0,1)^d \times (0,T) \to [0,1]^N$ *such that*

$$\chi^h \rightharpoonup \chi \text{ in } L^1([0,1)^d \times (0,T)). \tag{A.15}$$

Then $\chi \in \{0,1\}^N$ *almost everywhere,* $\sum_i \chi_i = 1$ *and* $\chi \in L^1((0,T), BV([0,1)^d))^N$.

If we assume that

$$\limsup_{h\downarrow 0} \int_0^T E_h(\chi^h(t))dt \le \sum_{i,j} \sigma_{ij} \int_0^T \mathcal{H}^{d-1}(\Sigma_{ij}(t))dt, \tag{A.16}$$

then χ *is a De Giorgi solution in the sense of Definition A.2.2 below.*

The convergence assumption (A.16) is motivated by a similar assumption on the implicit time discretization in the seminal paper [58] by Luckhaus and Sturzenhecker, and has also appeared in previous work in the context of the thresholding scheme [53], [54], [55]. As of now, this assumption can be verified only in particular cases, such as before the first singularity [77] or for certain types of singularities, namely mean convex ones, meaning $H > 0$. This was shown for the implicit time discretization in [22] and a proof in the case of the thresholding scheme will appear in a forthcoming work by Fuchs and the first author.

Inspired by the general framework [5] and [73], generalizing the previous two-phase version [55], we propose the following definition for weak solutions in the case of multiphase mean curvature flow.

Definition A.2.2. Given $\chi^0 \in \mathcal{A}$ and such that $\nabla\chi^0$ is a bounded measure, a map $\chi : [0,1)^d \times (0,T) \to \{0,1\}^N$ such that $\sum_i \chi_i = 1$ and $\chi \in L^1((0,T), BV([0,1)^d))^N$ is called a De Giorgi solution to the multiphase mean curvature flow with surface tensions σ_{ij} and mobilities μ_{ij} provided the following three facts hold:

1. There exist $H_{ij} \in L^2(\mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx)dt)$ which are mean curvatures in the weak sense, i.e., such that for any test vector field $\xi \in C_c^\infty([0,1)^d \times (0,T))^d$

$$\begin{aligned}\sum_{i,j} \sigma_{ij} \int_{[0,1)^d\times(0,T)} (\nabla\cdot\xi - \nu_{ij}\cdot\nabla\xi\nu_{ij})\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt \\ = -\sum_{i,j}\sigma_{ij}\int_{[0,1)^d\times(0,T)} H_{ij}\nu_{ij}\cdot\xi\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt.\end{aligned} \tag{A.17}$$

2. There exist normal velocities $V_{ij} \in L^2(\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt)$ with

$$\begin{aligned}\int_{[0,1)^d} \eta(t=0)\chi_i^0 dx + \int_{[0,1)^d\times(0,T)} \partial_t\eta\ \chi_i\ dxdt \\ + \sum_{k\neq i}\int_{[0,1)^d\times(0,T)} \eta V_{ik}\ \mathcal{H}^{d-1}_{|\Sigma_{ik}(t)}(dx)dt = 0\end{aligned}$$

   for all $\eta \in C_c^\infty([0,1)^d \times [0,T))$.

3. De Giorgi's inequality is satisfied, i.e.,

$$\begin{aligned}\limsup_{\tau\downarrow 0}\frac{1}{\tau}\sum_{i,j}\sigma_{ij}\int_{(T-\tau,T)}\mathcal{H}^{d-1}(\Sigma_{ij}(t))dt \\ +\frac{1}{2}\sum_{i,j}\int_{[0,1)^d\times(0,T)}\left(\frac{V_{ij}^2}{\mu_{ij}} + \mu_{ij}\sigma_{ij}^2 H_{ij}^2\right)\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt \le \sum_{i,j}\sigma_{ij}\mathcal{H}^{d-1}(\Sigma_{ij}^0).\end{aligned} \tag{A.18}$$

Remark A.2.3. Observe that inequality (A.18) together with the definition of the weak mean curvatures gives a notion of weak solution for the multiphase mean curvature flow incorporating both the dynamics $V_{ij} = -\sigma_{ij}\mu_{ij}H_{ij}$ and the Herring angle condition at triple junctions. Indeed if $\chi : [0,1)^d \times (0,T) \to \{0,1\}^N$ with $\sum_i \chi_i(t) = 1$ is such that the sets $\Omega_i(t) = \{\chi_i(\cdot,t) = 1\}$ meet along smooth interfaces $\Sigma_{ij} := \partial\Omega_i \cap \partial\Omega_j$ which evolve smoothly and satisfy (A.17), (A.18) then

1. The Herring angle condition at triple junctions is satisfied. Indeed by the divergence theorem on surfaces (see Theorem 11.8 and Remark 11.42 in [59]) for any $\xi \in C_c^\infty([0,1)^d)^d$

$$\begin{aligned}\int_{\Sigma_{ij}(t)} (\nabla\cdot\xi - \nu_{ij}\cdot\nabla\xi\nu_{ij})\ \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx) = -\int_{\Sigma_{ij}(t)} H_{ij}\nu_{ij}\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx) \\ + \int_{\partial\Sigma_{ij}(t)} \xi\cdot J\nu_{ij}\mathcal{H}^{d-2}(dx),\end{aligned}$$

where J denotes the rotation by ninety degrees in the normal plane to the triple junction $\partial\Sigma_{ij}(t)$. Thus (A.17) and $H_{ij} \in L^2(\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt)$ imply that

$$\begin{aligned}\sigma_{i_1i_2} &\int_{\partial\Sigma_{i_1i_2}(t)} \xi \cdot J\nu_{i_1i_2}\mathcal{H}^{d-2}(dx) \\ &+ \sigma_{i_2i_3} \int_{\partial\Sigma_{i_2i_3}(t)} \xi \cdot J\nu_{i_2i_3}\mathcal{H}^{d-2}(dx) \\ &+ \sigma_{i_3i_1} \int_{\partial\Sigma_{i_3i_1}(t)} \xi \cdot J\nu_{i_3i_1}\mathcal{H}^{d-2}(dx) = 0,\end{aligned}$$

which forces $\sigma_{i_1i_2}\nu_{i_1i_2} + \sigma_{i_2i_3}\nu_{i_2i_3} + \sigma_{i_3i_1}\nu_{i_3i_1} = 0$ at triple junctions.

2. We have $V_{ij} = -\sigma_{ij}\mu_{ij}H_{ij}$ on $\Sigma_{ij}(t)$. Indeed in the smooth case inequality (A.18) reduces to

$$\begin{aligned}&\sum_{i,j} \sigma_{ij} \int_{(0,T)} \frac{d}{dt}\mathcal{H}^{d-1}(\Sigma_{ij}(t))dt \\ &+ \frac{1}{2}\sum_{i,j} \int_{[0,1)^d\times(0,T)} \left(\frac{V_{ij}^2}{\mu_{ij}} + \mu_{ij}\sigma_{ij}^2 H_{ij}^2\right) \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt \leq 0.\end{aligned}$$

Using the Herring angle condition we have

$$\sum_{ij} \frac{d}{dt}\mathcal{H}^{d-1}(\Sigma_{ij}(t)) = \sum_{ij} \int_{[0,1)^d} V_{ij}H_{ij}\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)$$

and after completing the square we arrive at

$$\sum_{i,j} \sigma_{ij} \int_{[0,1)^d\times(0,T)} \left(\frac{V_{ij}}{\sqrt{\mu_{ij}\sigma_{ij}}} + \sqrt{\mu_{ij}\sigma_{ij}}H_{ij}\right)^2 \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt \leq 0,$$

which implies $V_{ij} = -\sigma_{ij}\mu_{ij}H_{ij}$.

The following lemma establishes, next to a compactness statement, that our convergence can be localized in the space and time variables x and t, but also in the variable z appearing in the convolution.

Lemma A.2.4. *We have the following:*

(i) Let $\{\chi^h\}_{h\downarrow 0}$ be a sequence of $\{0,1\}^N$-valued functions on $(0,T)\times[0,1)^d$ that satisfies $\chi^h \in \mathcal{A}$ for a.e. t and

$$\limsup_{h\downarrow 0} \left(\operatorname{esssup}_{t\in(0,T)} E_h(\chi^h(t)) + \int_0^T \frac{1}{2h^2} d_h^2(\chi^h(t), \chi^h(t-h))dt\right) < \infty \quad \text{(A.19)}$$

and that is piecewise constant in time in the sense of (A.14). *Such a sequence is precompact in $L^1([0,1)^d\times(0,T))^N$ and any weak limit χ is such that $\chi \in L^1((0,T), BV([0,1)^d))^N$ with*

$$\sum_{i,j} \sigma_{ij} \int_0^T \mathcal{H}^{d-1}(\Sigma_{ij}(t))dt \leq \liminf_{h\downarrow 0} \int_0^T E_h(\chi^h(t))dt. \quad \text{(A.20)}$$

(ii) *Assume that* u^h *is a sequence of* $[0,1]^N$*-valued functions with* $\sum_i u_i^h = 1$ *such that* (A.16) *holds (with* χ^h *replaced by* u^h*) and such that* $u^h \to \chi$ *in* $L^1([0,1)^d \times (0,T))^N$ *holds. Assume also that*

$$\limsup_{h\downarrow 0} \operatorname{esssup}_{t\in(0,T)} E_h(u^h(t)) < \infty. \tag{A.21}$$

Then as measures on $\mathbf{R}^d \times [0,1)^d \times (0,T)$ *we have the following weak convergences for any* $i \neq j$

$$\begin{aligned} &\frac{K_{ij}(z)}{\sqrt{h}} u_i^h(x,t) u_j^h(x-\sqrt{h}z,t)dxdtdz \\ &\qquad \rightharpoonup K_{ij}(z)(\nu_{ij}(x,t)\cdot z)_+ \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dtdz. \end{aligned} \tag{A.22}$$

$$\begin{aligned} &\frac{K_{ij}(z)}{\sqrt{h}} u_i^h(x-\sqrt{h}z,t) u_j^h(x,t)dxdtdz \\ &\qquad \rightharpoonup K_{ij}(z)(\nu_{ij}(x,t)\cdot z)_- \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dtdz. \end{aligned} \tag{A.23}$$

Here $\nu_{ij}(\cdot,t)$ *denotes the outer measure theoretic unit normal of* $\Omega_i(t)$ *restricted to the interface* $\Sigma_{ij}(t)$*. Here the convergence may be tested also with continuous functions which have polynomial growth in* $z \in \mathbf{R}^d$*.*

The next proposition is the main ingredient in the proof of Theorem A.2.1. It establishes the sharp lower bound on the distance-term.

Proposition A.2.5. *Suppose that* (A.15) *and the conclusion of Lemma A.2.4 (ii) hold. Assume also that the left hand side of* (A.24) *is finite. Then for every* $1 \leq k \leq N$ *there exists* $V_k \in L^2(|\nabla\chi_k|dt)$ *such that*

$$\partial_t \chi_k = V_k |\nabla \chi_k| dt$$

in the sense of distributions. Given $i \neq j$*, it holds that* $V_i(x,t) = -V_j(x,t)$ *on* $\Sigma_{ij}(t)$ *and if we define* $V_{ij}(x,t) := V_i(x,t)|_{\Sigma_{ij}(t)}$ *then we have*

$$\liminf_{h\downarrow 0} \int_0^T \frac{1}{h^2} d_h^2(\chi^h(t), \chi^h(t-h))dt \geq \sum_{i,j} \frac{1}{\mu_{ij}} \int_0^T \int_{\Sigma_{ij}(t)} |V_{ij}(x,t)|^2 \mathcal{H}^{d-1}(dx)dt. \tag{A.24}$$

The final ingredient is the analogous sharp lower bound for the metric slope.

Proposition A.2.6. *Suppose that the conclusion of Lemma A.2.4 (ii) holds and that* (A.15) *holds with* χ^h *replaced by* u^h*. Then for any* $i \neq j$ *there exists a mean curvature* $H_{ij} \in L^2(\mathcal{H}^{d-1}_{\Sigma_{ij}(t)}(dx)dt)$ *in the sense of* (A.17)*. Moreover the following inequality is true:*

$$\liminf_{h\downarrow 0} \int_0^T |\partial E_h|^2(u^h(t))dt \geq \sum_{i,j} \mu_{ij}\sigma_{ij}^2 \int_0^T \int_{\Sigma_{ij}(t)} |H_{ij}(x,t)|^2 \mathcal{H}^{d-1}(dx)dt. \tag{A.25}$$

We will present the proofs of Theorem A.2.1, Lemma A.2.4, Proposition A.2.5 and Proposition A.2.6 in Section A.4. Before doing that, we need a simple geometric measure theory construction.

A.3 Construction of suitable partitions of unity

In the sequel we will frequently want to localize on one of the interfaces. To do so, we need to construct a suitable family of balls on which the behavior of the flow is split into two majority phases and several minority phases. Hereafter we will ignore the time variable and consider a map $\chi : [0,1)^d \to \{0,1\}^N$ such that $\chi \in BV([0,1)^d, \mathbf{R}^N)$, $\sum_k \chi_k = 1$. Given $1 \le i < j \le N$ we denote by $\partial^*\Omega_i$ the reduced boundary of the set $\{\chi_i = 1\}$ and by $\Sigma_{ij} = \partial^*\Omega_i \cap \partial^*\Omega_j$ the interface between phase i and phase j. Given a real number $r > 0$ and a natural number $n \in \mathbf{N}$ we define

$$\mathcal{F}_n^r := \left\{ B(x, nr\sqrt{d}) : \; x \in r\mathbf{Z}^d \cap [0,1)^d \right\}$$

where the balls appearing in the definition are intended to be open. Observe that for any $n \ge 2$ and any $r > 0$ the collection of balls in $\mathcal{F}_n^r$ is a covering of $[0,1)^d$ with the property that any point $x \in [0,1)^d$ lies in at most $c(n,d)$ distinct balls belonging to $\mathcal{F}_n^r$, where $0 < c(n,d) \le (2n)^d$ is a constant that depends on n, d but not on r. Given numbers $1 \le l \ne p \le N$ we define

$$\mathcal{E}^r := \left\{ B \in \mathcal{F}_2^r : \; B \cap \Sigma_{lp} \ne \emptyset, \; \frac{\mathcal{H}^{d-1}(\Sigma_{ij} \cap 2B)}{\omega_{d-1}(4r)^{d-1}} \le \frac{1}{2^d}, \; \{i,j\} \ne \{l,p\} \right\}.$$

Here $2B$ denotes the ball with center given by the center of B and twice its radius. Given l, p as above, denote by $\{B_m^r\}$ an enumeration of $\mathcal{E}^r$ and by $\{\rho_m\}$ a smooth partition of unity subordinate to $\{B_m^r\}$. Then the following result holds true (for a proof, see the Appendix).

Lemma A.3.1. *Fix $1 \le l \ne p \le N$. With the above construction the following two properties hold.*

(i) For any $1 \le i \ne j \le N$, $\{i,j\} \ne \{l,p\}$ and any $\eta \in L^1(\mathcal{H}^{d-1}_{|\Sigma_{ij}})$

$$\lim_{r \downarrow 0} \sum_m \int_{B_m^r} \eta \mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx) = 0.$$

(ii) For any $\eta \in L^1(\mathcal{H}^{d-1}_{|\Sigma_{lp}})$

$$\lim_{r \downarrow 0} \sum_m \int \rho_m \eta \mathcal{H}^{d-1}_{|\Sigma_{lp}}(dx) = \int \eta \mathcal{H}^{d-1}_{|\Sigma_{lp}}(dx).$$

A.4 Proofs

A.4.1 Proof of Theorem A.2.1

Proof. By Lemma A.1.3, we can apply Lemma A.1.4 on the metric space $(\mathcal{M}, d_h)$ so that we get inequality (A.10) with $(E, d, \chi, u) = (E_h, d_h, \chi^h, u^h)$. Our first observation is that

$$\lim_{h \downarrow 0} E_h(\chi^0) = \sum_{i,j} \sigma_{ij} \mathcal{H}^{d-1}(\Sigma_{ij}^0),$$

which follows from the consistency, cf. Lemma A.5.2 in the Appendix. Inequality (A.10) then yields that the sequence χ^h satisfies (A.19), so that Lemma A.2.4 (i) applies to get that $\chi \in L^1((0,T), BV([0,1)^d))^N$, $\chi \in \{0,1\}^N$ a.e., $\sum_i \chi_i = 1$ and, after extracting a subsequence, $\chi^h \to \chi$ in $L^2([0,1)^d \times (0,T))^N$. We claim that this implies $u^h \to \chi$ in $L^2([0,1)^d \times (0,T))^N$. To see this, observe that (A.12) implies

$$\begin{aligned} hE_h(\chi^0) &\geq -\int_0^T E_h(u^h(t) - \chi^h(t))dt \\ &\geq C\frac{1}{\sqrt{h}}\sum_{i=1}^N \left(\int |G_\gamma^{h/2} * (u_i^h - \chi_i^h)|^2 dxdt + \int |G_\beta^{h/2} * (u_i^h - \chi_i^h)|^2 dxdt \right) \end{aligned} \tag{A.26}$$

where C is a constant which depends on $N, \mathbb{A}, \mathbb{B}$ but not on h and comes from the fact that all norms on $(1,...,1)^\perp$ are comparable. Inequality (A.26) clearly implies that $K^h * u^h - K^h * \chi^h$ converges to zero in L^2. Observe that inequality (A.13) in particular yields (A.21). Recalling (A.83) in the Appendix, we learn that $u^h - \chi^h$ converges to zero in L^2. This implies that we can apply Lemma A.2.4 (ii) both to the sequence u^h and the sequence χ^h. In particular, we may apply Proposition A.2.5 for χ^h and Proposition A.2.6 for u^h. Now the proof follows the same strategy as the one in the two-phase case in [55]. For the sake of completeness, we sketch the argument here. First of all, Lemma A.1.4 gives inequality (A.10) for (E_h, d_h, χ^h, u^h), namely for $n \in \mathbf{N}$

$$\rho(nh) \leq E_h(\chi^0), \tag{A.27}$$

where we set $\rho(t) = E_h(\chi^h(t)) + \frac{1}{2}\int_0^t \left(\frac{1}{h^2}d_h^2(\chi^h(s+h),\chi^h(s)) + |\partial E_h(u^h(s))|^2\right) ds$. Multiplying (A.27) by $\eta(nh) - \eta((n+1)h)$ for some non-increasing function $\eta \in C_c([0,T))$ we get $-\int \frac{d\eta}{dt}\rho dt \leq (\eta(0) + h\sup\left|\frac{d\eta}{dt}\right|)E_h(\chi^0)$. As test function η, we now choose $\eta(t) = \max\{\min\{\frac{T-t}{\tau}, 1\}, 0\}$ and obtain

$$\begin{aligned} &\frac{1}{\tau}\int_{T-\tau}^T E_h(\chi^h(t))dt \\ &+ \frac{1}{2}\int_0^{T-\tau} \left(\frac{1}{h^2}d_h^2(\chi^h(t),\chi^h(t-h)) + |\partial E_h(u^h(t))|^2\right) dt \leq (1+\frac{h}{\tau})E_h(\chi^0). \end{aligned} \tag{A.28}$$

Now it remains to pass to the limit as $h \downarrow 0$: to get (A.18) from inequality (A.28) one uses the lower semicontinuity (A.20) for the first left hand side term, the sharp bound (A.24) for the second left hand side term, the bound (A.25) for the last left hand side term and finally one uses the consistency Lemma A.5.2 in the Appendix to treat the right hand side term. To get (A.18) it remains to pass to the limit in $\tau \downarrow 0$. □

A.4.2 Proof of Lemma A.2.4

Proof. Argument for (i). For the compactness, the arguments in [55] adapt to this setting with minor changes. The first observation is that, by inequality (A.83) in the Appendix, one needs to prove compactness in $L^2([0,1)^d \times (0,T))^N$ of $\{K^h * \chi^h\}_{h\downarrow 0}$. For this, one just needs a modulus of continuity in time. I.e. it is sufficient to prove that there exists a constant $C > 0$ independent of h such that $I_h(s) \leq C\sqrt{s}$, where

$$I_h(s) = \int_{(s,T)\times[0,1)^d} |\chi^h(x,t) - \chi^h(x,t-s)|^2 dxdt.$$

This can be done applying word by word the argument in [55] once we show the following: for any pair $\chi, \chi' \in \mathcal{A}$, we have

$$\int |\chi - \chi'| dx \leq \frac{C}{\sqrt{h}} d_h^2(\chi, \chi') + C\sqrt{h}\left(E_h(\chi) + E_h(\chi')\right). \tag{A.29}$$

Here the constant C depends on $N, \mathbb{A}, \mathbb{B}$ but not on h.

To prove (A.29) we proceed as follows: let $\mathbb{S} \in \mathbf{R}^{N\times N}$ be a symmetric matrix which is positive definite on $(1, ..., 1)^{\perp}$. Since any two norms on a finite dimensional space are comparable, there exists a constant $C > 0$ depending on $\mathbb{S}$ and N such that

$$|\chi - \chi'| \leq |\chi - \chi'|^2 \leq C|\chi - \chi'|_{\mathbb{S}}^2$$

where $|\cdot|_{\mathbb{S}}$ denotes the norm induced by $\mathbb{S}$. For a function $u \in \mathcal{M}$ write $(\tilde{K}^h *)u^h$ for the function

$$\Big((\tilde{K}^h *)u^h\Big)_i = \sum_{j\neq i} K_{ij}^h * u_j^h.$$

Then we calculate

$$|\chi - \chi'|_{\mathbb{S}}^2 = -(\chi - \chi') \cdot (\tilde{K}^h *)(\chi - \chi') + (\chi - \chi')(\mathbb{S} + (\tilde{K}^h *))(\chi - \chi'). \tag{A.30}$$

Select $\mathbb{S} = (s_{ij})$ where $s_{ij} = -\int K_{ij}(z)dz$. Then, by our assumption (A.2) $\mathbb{S}$, is positive definite on $(1, ..., 1)^{\perp}$ and after integration on $[0, 1)^d$ identity (A.30) becomes

$$\int |\chi - \chi'|_{\mathbb{S}}^2 dx = \frac{1}{2\sqrt{h}} d_h^2(\chi, \chi') + \int (\chi - \chi')(\mathbb{S} + (\tilde{K}^h *))(\chi - \chi')dx.$$

We now proceed to estimate the integral on the right hand side. By the choice of $\mathbb{S}$ and Jensen's inequality we have

$$\int (\chi - \chi')(\mathbb{S} + (\tilde{K}^h *))(\chi - \chi')dx \leq C \int |(\mathbb{S} + (\tilde{K}^h *))(\chi - \chi')|dx$$
$$\leq C \sum_{i,j} \int K_{ij}^h(z)|(\chi_j - \chi_j')(x - z) - (\chi_j - \chi_j')(x)|dxdz.$$

Using the triangle inequality and (A.81) in the Appendix we can estimate the right hand side to obtain the following inequality

$$\begin{aligned}
&\int (\chi-\chi')(\mathbb{S}+(\tilde{K}^h*))(\chi-\chi')dx \\
&\leq C\sum_{i,j}\Bigg(\sum_{k\neq j}\int K_{ij}^h(z)\chi_j(x-z)\chi_k(x)dxdz \\
&\qquad +\sum_{k\neq j}\int K_{ij}^h(z)\chi_j(x)\chi_k(x-z)dxdz \\
&\qquad +\sum_{k\neq j}\int K_{ij}^h(z)\chi_j'(x-z)\chi_k'(x)dxdz \\
&\qquad +\sum_{k\neq j}\int K_{ij}^h(z)\chi_j'(x)\chi_k'(x-z)dxdz\Bigg).
\end{aligned}$$

Observing that there is a constant $C>0$ such that $K_{ij}\leq CK_{jk}$ we conclude that

$$\int (\chi-\chi')(\mathbb{S}+(\tilde{K}^h*))(\chi-\chi')dx \leq C\sqrt{h}\left(E_h(\chi)+E_h(\chi')\right).$$

This proves (A.29) and closes the argument for the compactness.

We also have to prove (A.20), but this follows from (A.22) with u^h replaced by χ^h once we have shown that the limit χ is such that $|\nabla\chi|$ is a bounded measure, equiintegrable in time. Indeed one can check from the proof of (A.22) that the lower bound of (A.22) does not require the extra assumption (A.16). Thus one gets that

$$\begin{aligned}
&\liminf_{h\downarrow 0}\int_0^T E_h(\chi^h(t))dt = \liminf_{h\downarrow 0}\sum_{i\neq j}\frac{1}{\sqrt{h}}\int_0^T\int_{[0,1)^d}\chi_i^h K_{ij}^h*\chi_j^h dxdt \\
&\geq \sum_{i\neq j}\liminf_{h\downarrow 0}\frac{1}{\sqrt{h}}\int_0^T\int_{[0,1)^d}\chi_i^h K_{ij}^h*\chi_j^h dxdt \\
&= \sum_{i\neq j}\liminf_{h\downarrow 0}\frac{1}{\sqrt{h}}\int_0^T\int_{[0,1)^d}\int_{\mathbf{R}^d}\chi_i^h(x,t)K_{ij}^h(z)\chi_j^h(x-z,t)dzdxdt \\
&\geq \sum_{i\neq j}\int_0^T\int_{[0,1)^d}\int_{\mathbf{R}^d}K_{ij}(z)(\nu_{ij}\cdot z)_+ dz\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt \\
&= \sum_{i\neq j}\sigma_{ij}\int_0^T\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt,
\end{aligned}$$

where in the last two lines we used (A.22) and the definition of σ_{ij}. To prove that the limit χ is such that $|\nabla\chi|$ is a bounded measure, equiintegrable in time one can proceed with an argument similar to the one used in [55] for the two-phase case. Observe that this only requires the weaker assumption (A.21).

Argument for (ii). As mentioned in the previous paragraph, we already know that the limit χ is such that $|\nabla\chi|$ is a bounded measure, equiintegrable in time. We will prove (A.22). Then (A.23) easily follows by recalling that $\nu_{ij}=-\nu_{ji}$. A standard argument (to be found in [55]) which relies on the exponential decay of the kernel yields the fact that we can test convergences (A.22) with functions with at most polynomial growth in z provided we already have the result for bounded and continuous test functions, thus we focus on this case.

Let $\xi \in C_b(\mathbf{R}^d \times [0,1)^d \times (0,T))$ be a bounded and continuous function. To show (A.22) we aim at showing that

$$\lim_{h\downarrow 0} \int \xi(z,x,t) \frac{K_{ij}(z)}{\sqrt{h}} u_i^h(x,t) u_j^h(x-\sqrt{h}z,t)dxdtdz \\ = \int \xi(z,x,t) K_{ij}(z)(\nu_{ij}(x,t)\cdot z)_+ \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dtdz. \tag{A.31}$$

Upon splitting ξ into the positive and the negative part, by linearity we may assume that $0 \le \xi \le 1$. We can split (A.31) into the local lower bound

$$\liminf_{h\downarrow 0} \int \xi(z,x,t) \frac{K_{ij}(z)}{\sqrt{h}} u_i^h(x,t) u_j^h(x-\sqrt{h}z,t)dzdxdt \\ \ge \int \xi(z,x,t) K_{ij}(z)(\nu_{ij}(x,t)\cdot z)_+ \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dtdz \tag{A.32}$$

and the global upper bound

$$\limsup_{h\downarrow 0} \int \frac{K_{ij}(z)}{\sqrt{h}} u_i^h(x,t) u_j^h(x-\sqrt{h}z,t)dzdxdt \\ \le \int K_{ij}(z)(\nu_{ij}(x,t)\cdot z)_+ \mathcal{H}^{d-1}_{\Sigma_{ij}(t)}(dx)dtdz. \tag{A.33}$$

Indeed we can recover the limsup inequality in (A.31) by splitting $\xi = 1-(1-\xi)$ and applying the local lower bound (A.32) to $1-\xi$.

We first concentrate on the local lower bounds in the case where $u^h = \chi$, namely we will show

$$\liminf_{h\downarrow 0} \int \xi(z,x,t) \frac{K_{ij}(z)}{\sqrt{h}} \chi_i(x,t)\chi_j(x-\sqrt{h}z,t)dzdxdt \\ \ge \int \xi(z,x,t) K_{ij}(z)(\nu_{ij}(x,t)\cdot z)_+ \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dtdz. \tag{A.34}$$

By Fatou's lemma the claim is reduced to showing that for a.e. point t in time and every $z \in \mathbf{R}^d$

$$\liminf_{h\downarrow 0} \int \xi(z,x,t) \frac{K_{ij}(z)}{\sqrt{h}} \chi_i(x,t)\chi_j(x-\sqrt{h}z,t)dx \\ \ge \int \xi(z,x,t) K_{ij}(z)(\nu_{ij}(x,t)\cdot z)_+ \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx).$$

Fix a point t such that $\chi(\cdot,t) \in BV([0,1)^d, \{0,1\}^N)$ and any $z \in \mathbf{R}^d$. In the sequel, we will drop those variables, so $\chi(x) = \chi(x,t)$, $\xi(x) = \xi(z,x,t)$. By approximation we may assume that $\xi \in C^\infty([0,1)^d)$. Let ρ_{mij} be a partition of unity obtained by applying the construction of Section A.3 to the function $\chi(x)$ on the interface Σ_{ij}. Let ν_i be the outer measure theoretic normal of $\Omega_i(t)$. Then by Lemma A.3.1 we have

$$\begin{aligned}
&\int \xi(x)(\nu_{ij}(x)\cdot z)_+ \mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx) \\
&= \lim_{r\downarrow 0}\left(\sum_{m\in\mathbf{N}}\int \rho_{mij}(x)\xi(x)(\nu_{ij}(x)\cdot z)_+ \mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx)\right) \\
&= \lim_{r\downarrow 0}\sum_{m\in\mathbf{N}}\left(\int \rho_{mij}(x)\xi(x)(\nu_i(x)\cdot z)_+ \mathcal{H}^{d-1}_{|\partial^*\Omega_i}(dx)\right. \\
&\qquad\qquad \left. -\sum_{k\neq i,j}\int \rho_{mij}(x)\xi(x)(\nu_{ij}(x)\cdot z)_+ \mathcal{H}^{d-1}_{|\Sigma_{ik}}(dx)\right) \\
&= \lim_{r\downarrow 0}\sum_{m\in\mathbf{N}}\int \rho_{mij}(x)\xi(x)(\nu_i(x)\cdot z)_+ \mathcal{H}^{d-1}_{|\partial^*\Omega_i}(dx).
\end{aligned} \tag{A.35}$$

We now focus on estimating the argument of the last limit. Observe that $(\nu_i(x)\cdot z)_+\mathcal{H}^{d-1}_{|\partial^*\Omega_i}(dx) = (\partial_z\chi_i)_+$, thus by definition of positive part of a measure, given $\epsilon > 0$ we can select, for any $m\in\mathbf{N}$, a function $\tilde{\xi}_m \in C^1_c(B_m)$ such that $0\le \tilde{\xi}_m \le 1$ and such that

$$\int \rho_{mij}\xi\tilde{\xi}_m\partial_z\chi_i + 2^{-m}\epsilon \ge \int \rho_{mij}\xi(\nu_i\cdot z)_+\mathcal{H}^{d-1}_{|\partial^*\Omega_i}(dx). \tag{A.36}$$

Let $\eta_m := \rho_{mij}\xi\tilde{\xi}_m \in C^1_c(B_m)$, then

$$\begin{aligned}
\int \eta_m\partial_z\chi_i &= -\int \partial_z\eta_m\chi_i dx \\
&= \lim_{h\downarrow 0}\int \frac{\eta_m(x+\sqrt{h}z)-\eta_m(x)}{\sqrt{h}}\chi_i(x)dx \\
&= \lim_{h\downarrow 0}\int \eta_m(x)\frac{\chi_i(x)-\chi_i(x-\sqrt{h}z)}{\sqrt{h}}dx.
\end{aligned}$$

Using that $\chi_i(x)-\chi_i(x-\sqrt{h}z) \le \chi_i(x)(1-\chi_i(x-\sqrt{h}z))$ (because $\chi_i\in\{0,1\}$) and that $1-\chi_i = \sum_{k\neq i}\chi_k$ we can estimate the last item by

$$\begin{aligned}
&\liminf_{h\downarrow 0}\sum_{k\neq i}\int \eta_m(x)\frac{\chi_i(x)\chi_k(x-\sqrt{h}z)}{\sqrt{h}}dx \\
&\le \liminf_{h\downarrow 0}\int \eta_m(x)\frac{\chi_i(x)\chi_j(x-\sqrt{h}z)}{\sqrt{h}}dx \\
&\qquad + \limsup_{h\downarrow 0}\sum_{k\neq i,j}\int \eta_m(x)\frac{\chi_i(x)\chi_k(x-\sqrt{h}z)}{\sqrt{h}}dx \\
&\le \liminf_{h\downarrow 0}\int \eta_m(x)\frac{\chi_i(x)\chi_j(x-\sqrt{h}z)}{\sqrt{h}}dx \\
&\qquad + \sum_{k\neq i,j}\limsup_{h\downarrow 0}\int \eta_m(x)\frac{\chi_i(x)\chi_k(x-\sqrt{h}z)}{\sqrt{h}}dx.
\end{aligned}$$

Observe that for each $m \in \mathbf{N}$, using also the consistency Lemma A.5.2

$$\begin{aligned}
&\limsup_{h\downarrow 0} \int \eta_m(x) \frac{\chi_i(x)\chi_k(x-\sqrt{h}z)}{\sqrt{h}} dx \\
&\leq \limsup_{h\downarrow 0} \int \eta_m(x) \frac{\chi_i(x)\chi_k(x-\sqrt{h}z) + \chi_i(x-\sqrt{h}z)\chi_j(x)}{\sqrt{h}} dx \\
&= \int \eta_m(x) |\nu_{ik}(x)\cdot z| \mathcal{H}^{d-1}_{|\Sigma_{ik}}(dx) \\
&\leq |z| \mathcal{H}^{d-1}(B^r_{mij} \cap \Sigma_{ik}).
\end{aligned}$$

Thus we obtain

$$\begin{aligned}
\int \eta_m \partial_z \chi_i \leq \liminf_{h\downarrow 0} \int \eta_m(x) \frac{\chi_i(x)\chi_j(x-\sqrt{h}z)}{\sqrt{h}} dx \\
+ \sum_{k\neq i,j} |z| \mathcal{H}^{d-1}(B^r_{mij} \cap \Sigma_{ik}).
\end{aligned}$$

Inserting back into (A.35), recalling also Lemma A.3.1 and the inequality (A.36), using Fatou's lemma, the fact that ρ_{mij} is a partition of unity and that $0 \leq \tilde{\xi}_m \leq 1$ we obtain that

$$\int \xi(x)(\nu_{ij}(x)\cdot z)_+ \mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx) \leq \liminf_{h\downarrow 0} \int \xi(x) \frac{\chi_i(x)\chi_j(x-\sqrt{h}z)}{\sqrt{h}} dx + \epsilon$$

and (A.34) follows letting ϵ go to zero. To derive inequality (A.32) we just apply Lemma A.5.4 in the Appendix.

To get the upper bound (A.33) we argue as follows. First of all, recall Assumption (A.16) which says

$$\limsup_{h\downarrow 0} \int_0^T E_h(u^h(t))dt \leq \int_0^T E(\chi(t))dt.$$

Now, if we define

$$e^{ij}_h(u^h) = \frac{1}{\sqrt{h}} \int_0^T \int u^h_i(t) K^h_{ij} * u^h_j(t) dx dt,$$

we have that by (A.32) $\liminf_{h\downarrow 0} e^{ij}_h(u_h) \geq e^{ij}(\chi)$, where $e^{ij}(\chi)$ is defined in the obvious way. Assume that there exists a pair i, j such that $\limsup_{h\downarrow 0} e^{ij}_h(u^h) > e^{ij}(\chi)$, then

$$\begin{aligned}
\int_0^T E(\chi(t))dt &\geq \limsup_{h\downarrow 0} \int_0^T E_h(u^h(t))dt \\
&\geq \sum_{(l,p)\neq(i,j)} \liminf_{h\downarrow 0} e^{lp}_h(u^h) + \limsup_{h\downarrow 0} e^{ij}_h(u^h) \\
&> \int_0^T E(\chi(t))dt
\end{aligned}$$

which is a contradiction. Thus we have proved (A.33). □

A.4.3 Proof of Proposition A.2.5

Proof. Since we assume that the left hand side of (A.24) is finite, in view of (A.8), upon passing to a subsequence we may assume that, in the sense of distributions, the limit

$$\lim_{h\downarrow 0} \frac{1}{h\sqrt{h}} \left(\left| G_\gamma^{h/2} * (\chi - \chi(\cdot - h)) \right|_{\mathbb{A}}^2 + \left| G_\beta^{h/2} * (\chi - \chi(\cdot - h)) \right|_{\mathbb{B}}^2 \right) = \omega$$

exists as a finite positive measure on $[0,1)^d \times (0,T)$. Here we indicated with $\chi_l^h(\cdot - h)$ the time shift of function χ_l^h. We denote by τ a small fraction of the characteristic spatial scale, namely $\tau = \alpha\sqrt{h}$ for some $\alpha > 0$, which we think as a small number. Given $1 \le l \le N$ we define

$$\delta\chi_l^h := \chi_l^h - \chi_l^h(\cdot - \tau).$$

We divide the proof into two parts: first we show that the normal velocities exist, and afterwards we prove the sharp bound. But first, let us state two distributional inequalities that will be used later. Namely

- In a distributional sense it holds that

$$\limsup_{h\downarrow 0} -\frac{1}{\sqrt{h}} \sum_{i\neq j} \delta\chi_i K_{ij}^h * \delta\chi_j \le \alpha^2 \omega. \tag{A.37}$$

- There exists a constant $C > 0$ such that for any $1 \le i \le N$ and any $\theta \in \{\gamma, \beta\}$ in a distributional sense it holds that

$$\limsup_{h\downarrow 0} \frac{1}{\sqrt{h}} (\chi_i - \chi_i(\cdot - \tau)) G_\theta^h * (\chi_i - \chi_i(\cdot - \tau)) \le C\alpha^2 \omega. \tag{A.38}$$

We observe that it suffices to prove (A.37), then (A.38) follows immediately. Indeed recall that $\mathbb{A}$ and $\mathbb{B}$ are positive definite on $(1, ..., 1)^\perp$. In particular there exists a constant $C > 0$ such that for any $v \in (1, ..., 1)^\perp$ one has $|v|_{\mathbb{A}}^2 + |v|_{\mathbb{B}}^2 \ge C|v|^2 \ge C v_i^2$ for any $i \in \{1, ..., N\}$. Applying this to the vector $v_i = G_\theta^{h/2} * \delta\chi_i$ one gets

$$|G_\theta^{h/2} * \delta\chi_i|^2 \le \frac{1}{C} |G_\theta^{h/2} * \delta\chi|_{\mathbb{A}}^2 + |G_\theta^{h/2} * \delta\chi|_{\mathbb{B}}^2.$$

The claim then follows from the definition of ω, (A.37), the symmetry (A.4) and the semigroup property (A.6). Indeed it is sufficient to check that, in the sense of distributions

$$\lim_{h\downarrow 0} \frac{1}{\sqrt{h}} \sum_{i\neq j} \delta\chi_i K_{ij}^h * \delta\chi_j + \frac{1}{\sqrt{h}} \left(|G_\gamma^{h/2} * \delta\chi|_{\mathbb{A}}^2 + |G_\beta^{h/2} * \delta\chi|_{\mathbb{B}}^2 \right) = 0.$$

To this aim, pick a test function $\eta \in C_c^\infty([0,1)^d \times (0,T))$. Spelling out the definition of the norms $|\cdot|_{\mathbb{A}}$ and $|\cdot|_{\mathbb{B}}$, the claim is proved once we show that

$$\lim_{h\downarrow 0} \frac{1}{\sqrt{h}} \sum_{i\neq j} a_{ij} \int \xi(\delta\chi_i G_\gamma^h * \delta\chi_j - G_\gamma^{h/2} * \delta\chi_i G_\gamma^{h/2} * \delta\chi_j) dx dt = 0, \tag{A.39}$$

and the same claim with a_{ij}, γ replaced by b_{ij}, β respectively.

We concentrate on (A.39). Clearly, we are done once we show that for any $i \neq j$

$$\lim_{h\downarrow 0} \frac{1}{\sqrt{h}} \int \xi(\delta\chi_i G_\gamma^h * \delta\chi_j - G_\gamma^{h/2} * \delta\chi_i G_\gamma^{h/2} * \delta\chi_j) dxdt = 0.$$

To show this, using the semigroup property (A.6) we rewrite the argument of the limit as

$$-\frac{1}{\sqrt{h}} \int [\xi, G_\gamma^{h/2} *](\delta\chi_i) G_\gamma^{h/2} * \delta\chi_j dxdt,$$

where $[\xi, G_\gamma^{h/2} *]$ denotes the commutator of multiplying by ξ and convolving with $G_\gamma^{h/2}$, i.e.

$$[\xi, G_\gamma^{h/2} *](f) = \xi G_\gamma^{h/2} * f - G_\gamma^{h/2} * (\xi f), \tag{A.40}$$

for every function f for which this expression makes sense. We observe that by the boundedness of the measures $\frac{1}{\sqrt{h}} |G_\gamma^{h/2} * \delta\chi|_{\mathbb{A}}^2$ it suffices to show

$$\lim_{h\downarrow 0} \frac{1}{\sqrt{h}} \int |[\xi, G_\gamma^{h/2} *](\delta\chi_i)|^2 dxdt = 0. \tag{A.41}$$

To prove this, spelling out the integrand, using the Cauchy-Schwarz inequality and recalling the scaling (A.5) we observe that

$$\begin{aligned}
&\int |[\xi, G_\gamma^{h/2} *](\delta\chi_i)|^2 dxdt \\
&\leq \int \left(\int |\xi(x,t) - \xi(x-z,t)|^2 G_\gamma^{h/2}(z) dz \right) G_\gamma^{h/2} * |\delta\chi_i(x,t)|^2 dxdt \\
&\leq \frac{h}{2} \sup |\nabla \xi|^2 \int G_\gamma(z) |z|^2 dz \int_0^T \int |\delta\chi_i(x,t)|^2 dxdt.
\end{aligned} \tag{A.42}$$

Observe that by the compactness of χ^h in $L^2([0,1)^d \times (0,T))$, (A.42) is of order h, thus (A.41) indeed holds true.

Now we can turn to the proof of (A.37), which is essentially already contained in the paper [55]. For the convenience of the reader we sketch the main ideas here. One reduces the claim to proving the following facts:

$$\lim_{h\downarrow 0} \frac{1}{\sqrt{h}} \sum_{ij} \delta\chi_i K_{ij}^h * \delta\chi_j - \frac{1}{\sqrt{h}} \left(\left|G_\gamma^{h/2} * \delta\chi\right|_{\mathbb{A}}^2 + \left|G_\beta^{h/2} * \delta\chi\right|_{\mathbb{B}}^2 \right) = 0, \tag{A.43}$$

$$\limsup_{h\downarrow 0} \frac{1}{\sqrt{h}} \left|G_\gamma^{h/2} * \delta\chi\right|_{\mathbb{A}}^2 - \alpha^2 \frac{1}{h\sqrt{h}} \left|G_\gamma^{h/2} * (\chi - \chi(\cdot - h))\right|_{\mathbb{A}}^2 \leq 0, \tag{A.44}$$

$$\limsup_{h\downarrow 0} \frac{1}{\sqrt{h}} \left|G_\beta^{h/2} * \delta\chi\right|_{\mathbb{B}}^2 - \alpha^2 \frac{1}{h\sqrt{h}} \left|G_\beta^{h/2} * (\chi - \chi(\cdot - h))\right|_{\mathbb{B}}^2 \leq 0. \tag{A.45}$$

Claim (A.43) was proved in the previous paragraph, while (A.44) and (A.45) are consequences of Jensen's inequality in the time variable for the convex functions $|\cdot|_{\mathbb{A}}^2$

and $|\cdot|^2_{\mathbb{B}}$ respectively. More precisely, assume without loss of generality that $\tau = Nh$ for some $N \in \mathbf{N}$, then by a telescoping argument and Jensen's inequality for $|\cdot|^2_{\mathbb{A}}$ we get

$$\begin{aligned}&\frac{1}{\sqrt{h}}|G_\gamma^{h/2} * \delta\chi|^2_{\mathbb{A}}\\ &\leq N \sum_{n=0}^{N-1} \frac{1}{\sqrt{h}}|G_\gamma^{h/2} * (\chi^h(\cdot - nh) - \chi^h(\cdot - (n+1)h))|^2_{\mathbb{A}}.\end{aligned}$$

Recalling that $N = \alpha/\sqrt{h}$ we can rewrite the right hand side as

$$\frac{\alpha^2}{N}\sum_{n=0}^{N-1} \frac{1}{h\sqrt{h}}|G_\gamma^{h/2} * (\chi^h(\cdot - nh) - \chi^h(\cdot - (n+1)h))|^2_{\mathbb{A}}.$$

This is an average of time shifts of $\alpha^2 \frac{1}{h\sqrt{h}}|G_\gamma^{h/2} * (\chi^h - \chi^h(\cdot - h))|^2_{\mathbb{A}}$. Since $Nh = o(1)$ all these time shifts are small, thus the average has the same distributional limit as $\alpha^2 \frac{1}{h\sqrt{h}}|G_\gamma^{h/2} * (\chi^h - \chi^h(\cdot - h))|^2_{\mathbb{A}}$. This proves (A.44). The argument for (A.45) is similar.

Existence of the normal velocities

We now prove the existence of the normal velocities. Fix $1 \leq i \leq N$ and observe that for $w \in \{\gamma, \beta\}$ we have

$$\begin{aligned}|\chi_i - \chi_i(\cdot - \tau)| \leq &(\chi_i - \chi_i(\cdot - \tau))G_w^h * (\chi_i - \chi_i(\cdot - \tau)) + |\chi_i - G_w^h * \chi_i|\\ &+ |\chi_i(\cdot - \tau) - G_w^h * \chi_i(\cdot - \tau)|,\end{aligned} \tag{A.46}$$

which follows simply by observing that $|\chi_i - \chi_i(\cdot - \tau)| = |\chi_i - \chi_i(\cdot - \tau)|^2 = (\chi_i - \chi_i(\cdot - \tau)G_w^h * (\chi_i - \chi_i(\cdot - \tau)) + (\chi_i - \chi_i(\cdot - \tau))(\chi - G_w^h * \chi) + (\chi_i(\cdot - \tau) - \chi_i)(\chi_i(\cdot - \tau) - G_w^h * \chi_i(\cdot - \tau))$. Using Jensen's inequality and the elementary identity (A.81) in the Appendix we have

$$\begin{aligned}|\chi_i - G_w^h * \chi_i| &\leq \int G_w^h(z)|\chi_i(x) - \chi_i(x - z)|dz\\ &= \int G_w^h(z)\chi_i(x)(1 - \chi_i(x - z))dz + \int G_w^h(z)(1 - \chi_i(x))\chi_i(x - z)dz\\ &= \sum_{k \neq i} \int G_w^h(z)\chi_i(x)\chi_k(x - z)dz + \sum_{k \neq i} \int G_w^h(z)\chi_k(x)\chi_i(x - z)dz.\end{aligned}$$

Now observe that by testing (A.22) with G_w/K_{ij} (which is bounded, and thus admissible), we learn that

$$\lim_{h \downarrow 0} \frac{1}{\sqrt{h}} \int G_w^h(z)\chi_i(x)\chi_k(x - z)dz = \int G_w(z)(\nu_{ik}(x, t) \cdot z)_+ dz \mathcal{H}^{d-1}_{|\Sigma_{ik}(t)}(dx)dt.$$

Thus, if we divide (A.46) by $\sqrt{h}$ and let $h \downarrow 0$, using also (A.38) we obtain

$$\begin{aligned}\alpha|\partial_t\chi_i| &\le \liminf_{h\downarrow 0}\frac{|\delta\chi_i|}{\sqrt{h}}\\ &\le \limsup_{h\downarrow 0}\frac{|\delta\chi_i|}{\sqrt{h}}\\ &\le C\alpha^2\omega + C\mathcal{H}^{d-1}_{|\partial^*\Omega_i(t)}(dx)dt,\end{aligned} \tag{A.47}$$

where C is a constant which depends on γ, β, N, the mobilities and the surface tensions. If we divide by α and then let $\alpha \to 0$ we learn that $|\partial_t\chi_i|$ is absolutely continuous with respect to $\mathcal{H}^{d-1}_{|\partial^*\Omega_i(t)}(dx)dt$. In particular, there exists $V_i \in L^1(\mathcal{H}^{d-1}_{|\partial^*\Omega_i(t)}(dx)dt)$ which is the normal velocity of χ_i in the sense that $\partial_t\chi_i = V_i|\nabla\chi_i|$ in the sense of distributions. The optimal integrability $V_i \in L^2(\mathcal{H}^{d-1}_{|\partial^*\Omega_i(t)}(dx)dt)$ will be shown in the second part of the proof. Let us record for later use that with a similar reasoning we actually obtain that $\limsup_h \frac{|\delta\chi_i|}{\sqrt{h}}$ is absolutely continuous with respect to $\mathcal{H}^{d-1}_{|\partial^*\Omega_i(t)}(dx)dt$. Thus in particular inequality (A.47) holds with ω replaced by its absolutely continuous part with respect to $\mathcal{H}^{d-1}_{|\partial^*\Omega_i(t)}(dx)dt$; calling this ω_i^{ac}, it means

$$\limsup_{h\downarrow 0}\frac{|\delta\chi_i|}{\sqrt{h}} \le C\alpha^2\omega_i^{ac} + C\mathcal{H}^{d-1}_{|\partial^*\Omega_i(t)}(dx)dt. \tag{A.48}$$

Sharp Bound

For a given $1 \le i \le N$ we denote by $\delta\chi_i^+$ and $\delta\chi_i^-$ the positive and negative parts fo $\delta\chi_i$ respectively, i.e. we set $\delta\chi_i^+ := (\chi_i - \chi_i(\cdot - \tau))_+$ and $\delta\chi_i^- := (\chi_i - \chi_i(\cdot - \tau))_-$. Before entering into the proof of the sharp bound, we need to prove the following property. For any $i \neq j$ we have that, in a distributional sense, the following holds

$$\lim_{h\downarrow 0}\frac{1}{\sqrt{h}}\delta\chi_i^+ K_{ij}^h * \delta\chi_j^+ = 0 = \lim_{h\downarrow 0}\frac{1}{\sqrt{h}}\delta\chi_i^- K_{ij}^h * \delta\chi_j^-. \tag{A.49}$$

We focus on the first limit, the second one being analogous. The first observation is that the limit

$$\lambda := \lim_{h\downarrow 0}\frac{1}{\sqrt{h}}\delta\chi_i^+ K_{ij}^h * \delta\chi_j^+ \tag{A.50}$$

is a nonnegative bounded measure, which is absolutely continuous with respect to $\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt$. Indeed, spelling out the z-integral and using the fact that $\delta\chi_i^+ = \chi_i(1 - \chi_i(\cdot - \tau))$ we obtain

$$\begin{aligned}&\frac{1}{\sqrt{h}}\delta\chi_i^+ K_{ij}^h * \delta\chi_j^+ = \frac{1}{\sqrt{h}}\int K_{ij}^h(z)\delta\chi_i^+(x,t)\delta\chi_j^+(x-z,t)dz\\ &\le \frac{1}{\sqrt{h}}\int K_{ij}^h(z)\chi_i(x,t)\chi_j(x-z,t)dz.\end{aligned}$$

By (A.22) in Lemma A.2.4, as $h \downarrow 0$, the right hand side converges to

$$\int K_{ij}(z)(\nu_{ij}(x,t)\cdot z)_+\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt,$$

which is absolutely continuous with respect to $\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt$.

Now, given $\nu_0 \in \mathbf{S}^{d-1}$ we claim that

$$\begin{aligned}\lambda \leq &\int_{\nu_0\cdot z\leq 0} K_{ij}(z)(\nu_{ij}\cdot z)_+ \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt \\ &+ \int_{\nu_0\cdot z\geq 0} K_{ij}(z)(\nu_{ij}\cdot z)_- \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt.\end{aligned} \tag{A.51}$$

To see this, we rewrite the argument of the limit in (A.50) as

$$\frac{1}{\sqrt{h}}\int_{\mathbf{R}^d} \lambda_h(t,x,z)dz,$$

where we set $\lambda_h(t,x,z) := \chi_i(x,t)(1-\chi_i)(x,t-\tau)K^h_{ij}(z)\chi_i(x-z,t)(1-\chi_i)(x-z,t-\tau)$. Using the fact that $0\leq \chi_i \leq 1$ and $\sum_l \chi_l = 1$ we obtain the following inequalities

$$\lambda_h \leq \chi_i(x,t)K^h_{ij}(z)\chi_i(x-z,t), \tag{A.52}$$

$$\begin{aligned}\lambda_h \leq &\chi_j(x,t-\tau)K^h_{ij}(z)\chi_i(x-z,t-\tau) \\ &+ C\sum_{k\neq i,j} K^h_{ij}(z)\left(|\delta\chi_k|(x,t) + |\delta\chi_k|(x-z,t)\right).\end{aligned} \tag{A.53}$$

Here C is a constant that does not depend on h. Using inequality (A.52) on the domain $\{\nu_0\cdot z\leq 0\}$ and inequality (A.53) on the domain $\{\nu_0\cdot z \geq 0\}$ we obtain

$$\begin{aligned}\lambda \leq &\limsup_{h\downarrow 0}\frac{1}{\sqrt{h}}\int_{\nu_0\cdot z\leq 0}\chi_i(x,t)K^h_{ij}(z)\chi_i(x-z,t)dz \\ &+ \limsup_{h\downarrow 0}\frac{1}{\sqrt{h}}\int_{\nu_0\cdot z\geq 0}\chi_j(x,t-\tau)K^h_{ij}(z)\chi_i(x-z,t-\tau)dz \\ &+ C\sum_{k\neq i,j}\limsup_{h\downarrow 0}\left(\frac{1}{\sqrt{h}}\int K^h_{ij}(z)|\delta\chi_k|(x,t)dz + \frac{1}{\sqrt{h}}\int K^h_{ij}(z)|\delta\chi_k|(x-z,t)dz\right).\end{aligned}$$

Observe that for any $1\leq k\leq N$ we have

$$\limsup_{h\downarrow 0}\frac{1}{\sqrt{h}}\int K^h_{ij}(z)|\delta\chi_k|(x,t)dz = \limsup_{h\downarrow 0}\frac{1}{\sqrt{h}}\int K^h_{ij}(z)|\delta\chi_k|(x-z,t)dz.$$

This can be seen by showing that

$$\lim_{h\downarrow 0}\frac{1}{\sqrt{h}}\int K^h_{ij}(z)\left(|\delta\chi_k|(x,t) - |\delta\chi_k|(x-z,t)\right)dz = 0, \tag{A.54}$$

which can be shown to be true by testing with an admissible test function, and putting the spatial shift z on it. Thus recalling (A.22) and (A.48), we obtain that

$$\begin{aligned}\lambda \leq &\int_{\nu_0\cdot z\leq 0} K_{ij}(z)(\nu_{ij}\cdot z)_+ \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt \\ &+ \int_{\nu_0\cdot z\geq 0} K_{ij}(z)(\nu_{ij}\cdot z)_- \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt \\ &+ C\sum_{k\neq i,j}\alpha^2\omega_k^{ac} + \mathcal{H}^{d-1}_{|\partial^*\Omega_k(t)}(dx)dt.\end{aligned}$$

Since we already know that λ is absolutely continuous with respect to $\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt$, the same bound holds true if we replace the right hand side with its absolutely continuous part with respect to $\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt$. Observing that for $k \neq i, j$ by Lemma A.5.1 in the Appendix the measures $\mathcal{H}^{d-1}_{|\partial^*\Omega_k(t)}(dx)dt$ and $\mathcal{H}^{d-1}_{|\partial^*\Sigma_{ij}(t)}(dx)dt$ are mutually singular, this yields (A.51).

Writing $\lambda = \theta(x,t)\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt$ for some $L^1(\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt)$-function θ we obtain that inequality (A.51) yields

$$\begin{aligned}\theta(x,t) \leq \int_{\nu_0\cdot z\leq 0} & K_{ij}(z)(\nu_{ij}(x,t)\cdot z)_+ dz \\ &+ \int_{\nu_0\cdot z\geq 0} K_{ij}(z)(\nu_{ij}(x,t)\cdot z)_- dz\end{aligned} \tag{A.55}$$

for every $\nu_0 \in \mathbf{S}^{d-1}$ and $\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt$-a.e. $(x,t) \in [0,1)^d \times (0,T)$. By a separability argument, we see that the null set on which (A.55) does not hold can be chosen so that it is independent of the choice of ν_0. If we select $\nu_0 = \nu_{ij}(x,t)$ this yields $\theta \leq 0$ almost everywhere with respect to $\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt$. Since we already know that λ is nonnegative this gives $\lambda = 0$.

Before getting the sharp bound, we check that for any $i \neq j$ we have $V_i = -V_j$ a.e. with respect to $\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt$. To see this, we start by observing that if $\xi \in C_c^\infty([0,1)^d \times (0,T))$, thanks to the fact that $\sum_{k\neq i}\chi_k = 1-\chi_i$, we get

$$\begin{aligned}\int \xi V_i \mathcal{H}^{d-1}_{|\partial^*\Omega_i(t)}(dx)dt &= -\int \partial_t\xi\chi_i dxdt \\ &= \sum_{k\neq i}\int \partial_t\xi\chi_k dxdt \\ &= -\sum_{k\neq i}\int \xi V_k \mathcal{H}^{d-1}_{|\partial^*\Omega_k(t)}(dx)dt.\end{aligned}$$

Choosing $\xi = f(t)g(x)$ for some $f \in C_c^\infty((0,T))$ and $g \in C^\infty([0,1)^d)$, by a separability argument, we obtain that for a.e. t and every $g \in C^\infty([0,1)^d)$

$$\int gV_i\mathcal{H}^{d-1}_{|\partial^*\Omega_i(t)}(dx) = -\sum_{k\neq i}\int gV_k\mathcal{H}^{d-1}_{|\partial^*\Omega_k(t)}(dx). \tag{A.56}$$

Pick t such that (A.56) holds. Let $g \in C^\infty([0,1)^d)$ and let ρ_m be a partition of unity obtained by the construction of Section A.3 applied to the function $\chi(\cdot,t)$ on the interface $\Sigma_{ij}(t)$. Then

$$\sum_{m\in\mathbf{N}}\int \rho_m gV_i\mathcal{H}^{d-1}_{|\partial^*\Omega_i(t)}(dx) = -\sum_{m\in\mathbf{N}}\sum_{k\neq i}\int \rho_m gV_k\mathcal{H}^{d-1}_{|\partial^*\Omega_k(t)}(dx). \tag{A.57}$$

Passing to the limit $r \downarrow 0$ in (A.57) we get by Lemma A.3.1 that

$$\int gV_i\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx) = -\int gV_j\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx).$$

Since this identity holds for any $g \in C^\infty([0,1)^d)$, a density argument gives $V_i(x,t) = -V_j(x,t)$ for $\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}$-a.e. x. In other words

$$\int |V_i(x,t) + V_j(x,t)|\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx) = 0.$$

Integrating in time yields that $V_i = -V_j$ a.e. with respect to $\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt$.

We now proceed with the derivation of the sharp lower bound. Define $c_{ij} := \int K_{ij}(z)dz$. Then we have

$$\begin{aligned}
&c_{ij}(|\delta\chi_i| + |\delta\chi_j|) = c_{ij}(\delta\chi_i^+ + \delta\chi_j^- + \delta\chi_i^- + \delta\chi_j^+)\\
&= \frac{1}{2}\left(\delta\chi_i^+ K_{ij}^h * (1-\delta\chi_j^-) + (1-\delta\chi_j^-)K_{ij}^h * \delta\chi_i^+ + \delta\chi_j^- K_{ij}^h * (1-\delta\chi_i^+)\right.\\
&\quad +(1-\delta\chi_i^+)K_{ij}^h * \delta\chi_j^- + \delta\chi_i^- K_{ij}^h * (1-\delta\chi_j^+) + (1-\delta\chi_j^+)K_{ij}^h * \delta\chi_i^-\\
&\quad \left.+\delta\chi_j^+ K_{ij}^h * (1-\delta\chi_i^-) + (1-\delta\chi_i^-)K_{ij}^h * \delta\chi_j^+\right) + \left(\delta\chi_i^+ K_{ij}^h * \delta\chi_j^-\right.\\
&\quad \left.+\delta\chi_j^- K_{ij}^h * \delta\chi_i^+ + \delta\chi_i^- K_{ij}^h * \delta\chi_j^+ + \delta\chi_j^+ K_{ij}^h * \delta\chi_i^-\right).
\end{aligned}$$

Now we rewrite the terms in the second parenthesis using $-ab = a_+b_- + a_-b_+ - a_+b_+ - a_-b_-$ and then adding and subtracting the contributions of the minority phases we obtain

$$\begin{aligned}
c_{ij}(|\delta\chi_i| + |\delta\chi_j|) \le \frac{1}{2}\Bigg(&\delta\chi_i^+ K_{ij}^h * (1-\delta\chi_j^-) + (1-\delta\chi_j^-)K_{ij}^h * \delta\chi_i^+ + \delta\chi_j^- K_{ij}^h * (1-\delta\chi_i^+)\\
&+ (1-\delta\chi_i^+)K_{ij}^h * \delta\chi_j^- + \delta\chi_i^- K_{ij}^h * (1-\delta\chi_j^+) + (1-\delta\chi_j^+)K_{ij}^h * \delta\chi_i^-\\
&+ \delta\chi_j^+ K_{ij}^h * (1-\delta\chi_i^-) + (1-\delta\chi_i^-)K_{ij}^h * \delta\chi_j^+\Bigg) - \sum_{l,p} \delta\chi_l K_{lp}^h * \delta\chi_p\\
&+ \delta\chi_i^+ K_{ij}^h * \delta\chi_j^+ + \delta\chi_i^- K_{ij}^h * \delta\chi_j^- + \delta\chi_j^+ K_{ij}^h * \delta\chi_i^+\\
&+ \delta\chi_j^- K_{ij}^h * \delta\chi_i^- + \sum_{\{l,p\}\neq\{i,j\}} \delta\chi_l K_{lp}^h * \delta\chi_p.
\end{aligned} \tag{A.58}$$

Now the main idea is to split the integral of K_{ij} in the definition of c_{ij} into two parts. More precisely, by the symmetry (A.4), for any $\nu_0 \in \mathbf{S}^{d-1}$ and any $V_0 > 0$ we have

$$c_{ij} = 2\int_{0\le\nu_0\cdot z\le\alpha V_0} K_{ij}(z)dz + 2\int_{\nu_0\cdot z>\alpha V_0} K_{ij}(z)dz.$$

Substituting into (A.58) and dividing by $\sqrt{h}$ we obtain

$$
\begin{aligned}
&2\int_{0\le\nu_0\cdot z\le\alpha V_0} K_{ij}(z)dz\frac{(|\delta\chi_i|+|\delta\chi_j|)}{\sqrt{h}} \qquad\qquad \text{(A.59)}\\
&\le \frac{1}{2\sqrt{h}}\Big(\delta\chi_i^+K_{ij}^h*(1-\delta\chi_j^-)+(1-\delta\chi_j^-)K_{ij}^h*\delta\chi_i^+ +\delta\chi_j^-K_{ij}^h*(1-\delta\chi_i^+)\\
&\quad+(1-\delta\chi_i^+)K_{ij}^h*\delta\chi_j^- +\delta\chi_i^-K_{ij}^h*(1-\delta\chi_j^+)+(1-\delta\chi_j^+)K_{ij}^h*\delta\chi_i^-\\
&\quad+\delta\chi_j^+K_{ij}^h*(1-\delta\chi_i^-)+(1-\delta\chi_i^-)K_{ij}^h*\delta\chi_j^+\\
&\quad-4\int_{\nu_0\cdot z>\alpha V_0}K_{ij}(z)dz(|\delta\chi_i|+|\delta\chi_j|)\\
&\quad-2\sum_{l,p}\delta\chi_lK_{lp}^h*\delta\chi_p+2\delta\chi_i^+K_{ij}^h*\delta\chi_j^+ +2\delta\chi_i^-K_{ij}^h*\delta\chi_j^-\\
&\quad+2\delta\chi_j^+K_{ij}^h*\delta\chi_i^+ +2\delta\chi_j^-K_{ij}^h*\delta\chi_i^-\\
&\quad+2\sum_{(l,p)\ne(i,j),(l,p)\ne(j,i)}\delta\chi_lK_{lp}^h*\delta\chi_p\Big).
\end{aligned}
$$

We will be interested in bounding the lim inf of the left hand side. Observe that the distributional limit of the last five terms is non-positive. Indeed, the limit of first four terms vanish distributionally by property (A.49), while the last term is bounded from above by

$$2\sum_{(l,p)\ne(i,j),(l,p)\ne(j,i)}\delta\chi_l^+K_{lp}^h*\delta\chi_p^+ +\delta\chi_l^-K_{lp}^h*\delta\chi_p^-,$$

which vanish distributionally in the limit $h\downarrow 0$ by property (A.49). We thus obtain that the lim inf of the left hand side of (A.59) is bounded from above by

$$
\begin{aligned}
\liminf_{h\downarrow 0}\frac{1}{2\sqrt{h}}&\Big(\delta\chi_i^+K_{ij}^h*(1-\delta\chi_j^-)+(1-\delta\chi_j^-)K_{ij}^h*\delta\chi_i^+ +\delta\chi_j^-K_{ij}^h*(1-\delta\chi_i^+)\\
&+(1-\delta\chi_i^+)K_{ij}^h*\delta\chi_j^- +\delta\chi_i^-K_{ij}^h*(1-\delta\chi_j^+)+(1-\delta\chi_j^+)K_{ij}^h*\delta\chi_i^-\\
&+\delta\chi_j^+K_{ij}^h*(1-\delta\chi_i^-)+(1-\delta\chi_i^-)K_{ij}^h*\delta\chi_j^+\\
&-4\int_{\nu_0\cdot z>\alpha V_0}K_{ij}(z)dz(|\delta\chi_i|+|\delta\chi_j|)-2\sum_{l,p}\delta\chi_lK_{lp}^h*\delta\chi_p\Big).
\end{aligned}
$$

For the last term we use the sharp bound (A.37), relating this term to our dissipation measure ω. We would like to get a good bound for the other terms. This cannot be done naively as before, since we want the bound to be sharp. We claim that

$$\begin{aligned}\limsup_{h\downarrow 0} \frac{1}{\sqrt{h}} \Bigg(& \delta\chi_i^+ K_{ij}^h * (1-\delta\chi_j^-) + (1-\delta\chi_j^-) K_{ij}^h * \delta\chi_i^+ + \delta\chi_j^- K_{ij}^h * (1-\delta\chi_i^+) \\ & + (1-\delta\chi_i^+) K_{ij}^h * \delta\chi_j^- + \delta\chi_i^- K_{ij}^h * (1-\delta\chi_j^+) + (1-\delta\chi_j^+) K_{ij}^h * \delta\chi_i^- \\ & + \delta\chi_j^+ K_{ij}^h * (1-\delta\chi_i^-) + (1-\delta\chi_i^-) K_{ij}^h * \delta\chi_j^+ \\ & - 4 \int_{\nu_0\cdot z > \alpha V_0} K_{ij}(z) dz (|\delta\chi_i| + |\delta\chi_j|) \Bigg) \\ \le 8 & \int_{0\le \nu_0\cdot z\le \alpha V_0} K_{ij}(z) |\nu_{ij}(x)\cdot z| dz \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx) dt \\ & \qquad + C \sum_{k\neq i,j} (\alpha^2 \omega_k^{ac} + \mathcal{H}^{d-1}_{|\partial^*\Omega_k(t)}(dx)) dt. \end{aligned} \tag{A.60}$$

Here C is a constant that depends on $\gamma, \beta, \mathbb{A}, \mathbb{B}$, but not on h. Assume for the moment that (A.60) is true and let us conclude the argument in this case. Using (A.60) and (A.37) we obtain

$$\begin{aligned} & 2 \liminf_{h\downarrow 0} \int_{0\le \nu_0\cdot z\le \alpha V_0} K_{ij}(z) dz \frac{(|\delta\chi_i| + |\delta\chi_j|)}{\sqrt{h}} \\ & \le \alpha^2 \omega + 4 \int_{0\le \nu_0\cdot z\le \alpha V_0} K_{ij}(z) |\nu_{ij}(x)\cdot z| dz \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx) dt \\ & \qquad + C \sum_{k\neq i,j} (\alpha^2 \omega_k^{ac} + \mathcal{H}^{d-1}_{|\partial^*\Omega_k(t)}(dx) dt \end{aligned} \tag{A.61}$$

in the sense of distributions on $[0,1)^d \times (0,T)$. Observe also that the left hand side of (A.61) is an upper bound for $\int_{0\le\nu_0\cdot z\le\alpha V_0} K_{ij}(z) dz (|\partial_t\chi_i| + |\partial_t\chi_j|)$, thus the inequality still holds true if the left hand side is replaced by this term. Remember that ω_k^{ac} is absolutely continuous with respect to $\mathcal{H}^{d-1}_{|\partial^*\Omega_k(t)}(dx)dt$, thus there exist functions $W_k \in L^1(\mathcal{H}^{d-1}_{|\partial^*\Omega_k(t)}(dx)dt)$ such that $\omega_k^{ac} = W_k(x,t)\mathcal{H}^{d-1}_{|\partial^*\Omega_k(t)}(dx)dt$. We now disintegrate the measure ω, i.e. we find a Borel family $\omega_t, t\in(0,T)$, of positive measures on $[0,1)^d$ such that $\omega = \omega_t \otimes dt$. Having said this, it is not hard to see that (A.61) holds in a disintegrated version, i.e. we have for Lebesgue a.e. $t \in (0,T)$

$$\begin{aligned} & 2\alpha \int_{0\le \nu_0\cdot z\le \alpha V_0} K_{ij}(z) dz (|V_i(x,t)| \mathcal{H}^{d-1}_{|\partial^*\Omega_i(t)}(dx) + |V_j(x,t)| \mathcal{H}^{d-1}_{|\partial^*\Omega_j(t)}(dx)) \\ & \le \alpha^2 \omega_t + 4 \int_{0\le \nu_0\cdot z\le \alpha V_0} K_{ij}(z) |\nu_{ij}(x)\cdot z| dz \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx) \\ & \qquad + C \sum_{k\neq i,j} (\alpha^2 W_k(x,t) + 1) \mathcal{H}^{d-1}_{|\partial^*\Omega_k(t)}(dx). \end{aligned} \tag{A.62}$$

Here $\nu_0 \in \mathbf{S}^{d-1}$ and $V_0 \in (0,\infty)$ are arbitrary: indeed even if the set of points in time for which (A.62) holds is a priori dependent on ν_0 and V_0, a standard separability argument allows us to conclude that we can get rid of this dependence.

Fix a point t in time such that (A.62) holds. In what follows, we drop the time variable t which is fixed, so for example $V_i(x) = V_i(x,t)$, $\Sigma_{ij} = \Sigma_{ij}(t)$ and so on.

Fix $\xi \in C([0,1)^d)$, observe that by definition of V_{ij} and by using the fact that $\Sigma_{ij} \subset \partial^*\Omega_i \cap \partial^*\Omega_j$ we have

$$\begin{aligned}
&4\alpha \int_{0\le\nu_0\cdot z\le\alpha V_0} K_{ij}(z)dz \int_{[0,1)^d} \xi(x)|V_{ij}(x)|\mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx) \\
&\le \alpha^2 \int_{[0,1)^d} \xi(x)\omega_t(dx) + 4\int_{[0,1)^d}\int_{0\le\nu_0\cdot z\le\alpha V_0} K_{ij}(z)|\nu_{ij}(x)\cdot z|dz\xi(x)\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx) \\
&\qquad\qquad + C\sum_{k\ne i,j}\int_{[0,1)^d} \xi(x)(\alpha^2 W_k(x,t)+1)\mathcal{H}^{d-1}_{|\partial^*\Omega_k(t)}(dx).
\end{aligned}$$

Let us relabel ν_0, V_0 and ξ to make clear that they may depend on the pair i,j. Thus $\nu_0^{ij} \in \mathbf{S}^{d-1}$, $V_0^{ij} \in (0,\infty)$ and $\xi_{ij} \in C([0,1)^d)$ are arbitrary, and it holds

$$\begin{aligned}
&4\alpha \int_{0\le\nu_0^{ij}\cdot z\le\alpha V_0^{ij}} K_{ij}(z)dz \int_{[0,1)^d} \xi_{ij}(x)|V_{ij}(x)|\mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx) \\
&\le \alpha^2 \int_{[0,1)^d} \xi_{ij}(x)\omega_t(dx) + 4\int_{[0,1)^d}\int_{0\le\nu_0^{ij}\cdot z\le\alpha V_0^{ij}} K_{ij}(z)|\nu_{ij}(x)\cdot z|dz\xi_{ij}(x)\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx) \\
&\qquad\qquad + C\sum_{k\ne i,j}\int_{[0,1)^d} \xi_{ij}(x)(\alpha^2 W_k(x,t)+1)\mathcal{H}^{d-1}_{|\partial^*\Omega_k(t)}(dx).
\end{aligned}$$

Let $\{\rho_m\}$ be a partition of unity obtained using the construction of Section A.3 applied to the function $\chi(\cdot,t)$ on the interface $\Sigma_{ij}(t)$. Use the above inequality with ξ_{ij} replaced by $\rho_m\xi_{ij}$ and sum over m and i,j to get

$$\sum_{i<j}\sum_{m\in\mathbf{N}} \mathbf{LH}_m^{ij} \le \sum_{i<j}\sum_{m\in\mathbf{N}} (\mathbf{I}_m^{ij} + \mathbf{II}_m^{ij} + \mathbf{III}_m^{ij})$$

where we have set

$$\begin{aligned}
\mathbf{LH}_m^{ij} &= 4\alpha \int_{0\le\nu_0^{ij}\cdot z\le\alpha V_0^{ij}} K_{ij}(z)dz \int_{[0,1)^d} \rho_{mij}(x)\xi_{ij}(x)|V_{ij}(x)|\mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx), \\
\mathbf{I}_m^{ij} &= \alpha^2 \int_{[0,1)^d} \xi_{ij}(x)\rho_{mij}(x)\omega_t(dx), \\
\mathbf{II}_m^{ij} &= 4\int_{[0,1)^d}\int_{0\le\nu_0^{ij}\cdot z\le\alpha V_0^{ij}} K_{ij}(z)|\nu_{ij}(x)\cdot z|dz\rho_{mij}(x)\xi_{ij}(x)\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx), \\
\mathbf{III}_m^{ij} &= C\sum_{k\ne i,j}\int_{[0,1)^d} \rho_{mij}(x)\xi_{ij}(x)(\alpha^2 W_k(x,t)+1)\mathcal{H}^{d-1}_{|\partial^*\Omega_k(t)}(dx).
\end{aligned}$$

Observe that

$$\sum_{i<j}\sum_{m\in\mathbf{N}} \mathbf{I}_m^{ij} \le \sum_{i<j} \alpha^2 \int_{[0,1)^d} \xi_{ij}(x)\omega_t(dx)$$

because ρ_m is a partition of unity. Moreover by Lemma A.3.1 we get

$$\lim_{r\downarrow 0}\sum_{i<j}\sum_{m\in\mathbf{N}}\mathbf{LH}^{ij}_m = 4\alpha\int_{0\le\nu_0^{ij}\cdot z\le\alpha V_0^{ij}} K_{ij}(z)dz\int_{[0,1)^d}\xi_{ij}(x)|V_{ij}(x)|\mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx),$$

$$\lim_{r\downarrow 0}\sum_{i<j}\sum_{m\in\mathbf{N}}\mathbf{II}^{ij}_m = \sum_{i<j}4\int_{[0,1)^d}\int_{0\le\nu_0^{ij}\cdot z\le\alpha V_0^{ij}} K_{ij}(z)|\nu_{ij}(x)\cdot z|dz\xi_{ij}(x)\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx),$$

$$\lim_{r\downarrow 0}\sum_{i<j}\sum_{m\in\mathbf{N}}\mathbf{III}^{ij}_m = 0.$$

Putting things together we obtain that for any $\nu_0^{ij}\in\mathbf{S}^{d-1}$, any $V_0^{ij}\in(0,\infty)$, and any $\xi\in C([0,1)^d)$

$$\begin{aligned}&4\alpha\sum_{i<j}\int_{0\le\nu_0^{ij}\cdot z\le\alpha V_0^{ij}} K_{ij}(z)dz\int_{[0,1)^d}\xi_{ij}(x)|V_{ij}(x)|\mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx)\\ &\le\sum_{i<j}\alpha^2\int_{[0,1)^d}\xi_{ij}(x)\omega_t(dx)\\ &\quad+4\sum_{i<j}\int_{[0,1)^d}\int_{0\le\nu_0^{ij}\cdot z\le\alpha V_0^{ij}} K_{ij}(z)|\nu_{ij}(x)\cdot z|dz\xi_{ij}(x)\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx).\end{aligned}\tag{A.63}$$

We now claim that by approximation the above inequality is valid for any simple function $\xi_{ij}\ge 0$. To see this, it is clear that we can concentrate on $\xi_{ij}=w_{ij}\mathbf{1}_{B_{ij}}$, where $B_{ij}\subset[0,1)^d$ are Borel and $w_{ij}\ge 0$. Observe that by the dominated convergence theorem, the family

$$\mathcal{F}:=\left\{B=\prod_{i<j}B_{ij}:\ B_{ij}\in\mathcal{B}([0,1)^d)\text{ s.t. }\forall w_{ij}\ge 0\ \ \text{(A.63) holds with }\xi_{ij}=w_{ij}\mathbf{1}_{B_{ij}}\right\}$$

is a monotone class. Thus by the monotone class theorem we just need to show that it contains all the products of open sets. But this is easy because given $B_{ij}\subset[0,1)^d$ open sets, we can always find sequences η_k^{ij} of continuous functions with compact support such that $0\le\eta_k^{ij}\le\mathbf{1}_{B_{ij}}$ and such that $\eta_k^{ij}\to\mathbf{1}_{B_{ij}}$, thus the claim follows by the monotone convergence theorem.

With this in place one can use an approximation argument to replace the vector ν_0^{ij} with the $\mathcal{H}^{d-1}$-measurable vector valued function ν_{ij} obtaining the following inequality:

$$\begin{aligned}&4\alpha\sum_{i<j}\int_{[0,1)^d}\int_{0<\nu_{ij}(x)\cdot z<\alpha V_0^{ij}} K_{ij}(z)dz\xi_{ij}(x)|V_{ij}(x)|\mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx)\\ &\le\sum_{i<j}\alpha^2\int_{[0,1)^d}\xi_{ij}(x)\omega_t(dx)\\ &\quad+4\sum_{i<j}\int_{[0,1)^d}\int_{0\le\nu_{ij}\cdot z\le\alpha V_0^{ij}} K_{ij}(z)|\nu_{ij}(x)\cdot z|dz\xi_{ij}(x)\mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx).\end{aligned}\tag{A.64}$$

Now divide by α^2 and send α to zero. Record the following limits, which can be computed spelling out the definition of K_{ij}, and recalling the symmetry property (A.4) and the factorization property (A.7) for the heat kernel

$$\lim_{\alpha\downarrow 0}\frac{1}{\alpha}\int_{0<\nu_{ij}(x)\cdot z<\alpha V_0^{ij}} K_{ij}(z)dz = \frac{V_0^{ij}}{2\mu_{ij}}.$$
$$\lim_{\alpha\downarrow 0}\frac{1}{\alpha^2}\int_{0\le\nu_{ij}(x)\cdot z\le\alpha V_0^{ij}} K_{ij}(z)|\nu_{ij}(x)\cdot z|dz = \frac{(V_0^{ij})^2}{4\mu_{ij}}.$$

Then if we insert back into (A.64) we obtain

$$\sum_{i<j}\frac{2}{\mu_{ij}}\int_{[0,1)^d} V_0^{ij}\xi_{ij}(x)|V_{ij}(x)|\mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx) \tag{A.65}$$
$$\le \int_{[0,1)^d}\omega_t(dx) + \sum_{i<j}\int_{[0,1)^d}\frac{(V_0^{ij})^2}{\mu_{ij}}\xi_{ij}(x)\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx).$$

Now given $M>0$ take sequences of simple functions

$$s_m^{ij} = \sum_{k=1}^{p_m} w_k^{ij}\mathbf{1}_{B_{ij}^k}$$

such that $s_m^{ij}\to|V_{ij}|\mathbf{1}_{\{|V_{ij}|\le M\}}$ as $m\to+\infty$ monotonically almost everywhere with respect to $\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}$. We are assuming that $\{B_{ij}^k\}_{k=1,\dots,p_m}$ are disjoint and $\mathcal{H}^{d-1}$-measurable, with the property that $B_{ij}^{k_1}\cap B_{lr}^{k_2}=\emptyset$ if $\{i,j\}\ne\{l,r\}$. Choosing $V_0^{ij}=w_k^{ij}$, $\xi_{ij}=\mathbf{1}_{B_{ij}^k}$ in (A.65) and summing over k we obtain

$$\sum_{i<j}\frac{2}{\mu_{ij}}\int_{[0,1)^d} s_m^{ij}|V_{ij}(x)|\mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx)$$
$$\le \int_{[0,1)^d}\omega_t(dx) + \sum_{i<j}\int_{[0,1)^d}\frac{(s_m^{ij})^2}{\mu_{ij}}\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx).$$

Taking the limit $m\to+\infty$, using the monotone convergence theorem we obtain

$$\sum_{i<j}\frac{2}{\mu_{ij}}\int_{[0,1)^d}|V_{ij}(x)|^2\mathbf{1}_{\{|V_{ij}|\le M\}}\mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx)$$
$$\le \int_{[0,1)^d}\omega_t(dx) + \sum_{i<j}\int_{[0,1)^d}\frac{|V_{ij}(x)|^2}{\mu_{ij}}\mathbf{1}_{\{|V_{ij}|\le M\}}\mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx)$$

or, in other words,

$$\sum_{i<j}\frac{1}{\mu_{ij}}\int_{[0,1)^d}|V_{ij}(x)|^2\mathbf{1}_{\{|V_{ij}|\le M\}}\mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx)\le\int_{[0,1)^d}\omega_t(dx).$$

Recall that $\mu_{ij}=\mu_{ji}$, thus the inequality above may be rewritten as

$$\sum_{i,j}\frac{1}{2\mu_{ij}}\int_{[0,1)^d}|V_{ij}(x)|^2\mathbf{1}_{\{|V_{ij}|\le M\}}\mathcal{H}^{d-1}_{|\Sigma_{ij}}(dx)\le\int_{[0,1)^d}\omega_t(dx).$$

If we now integrate in time we learn by the monotone convergence theorem that $V_{ij}\in L^2(\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt)$ and that the sharp bound (A.24) is satisfied.

Proof of (A.60)

To prove (A.60) we proceed in several steps.

First of all, we claim that the first eight terms may be substituted by

$$2\int_{\nu_0\cdot z\geq 0} K^h_{ij}(z)\Big(|\delta\chi_i^+ - \delta\chi_j^-(-z)| + |\delta\chi_i^+(-z) - \delta\chi_j^-| \\ |\delta\chi_i^- - \delta\chi_j^+(-z)| + |\delta\chi_i^-(-z) - \delta\chi_j^+|\Big)dz. \tag{A.66}$$

To show this, observe that we may replace the implicit z-integrals in the convolution in the first eight terms by twice the integrals over the half space $\{\nu_0\cdot z\geq 0\}$ instead of $\mathbf{R}^d$. This is clearly true once we observe that, in the sense of distribution

$$\begin{aligned}&\lim_{h\downarrow 0}\frac{1}{\sqrt{h}}\Big(\delta\chi_i^+\int_{\nu_0\cdot z\geq 0} K^h_{ij}(z)(1-\delta\chi_j^-(\cdot - z))dz \\ &\qquad +(1-\delta\chi_j^-)\int_{\nu_0\cdot z\geq 0} K^h_{ij}(z)\delta\chi_i^+(\cdot - z)dz\Big) \\ =&\lim_{h\downarrow 0}\frac{1}{\sqrt{h}}\Big(\delta\chi_i^+\int_{\nu_0\cdot z\leq 0} K^h_{ij}(z)(1-\delta\chi_j^-(\cdot - z))dz \\ &\qquad +(1-\delta\chi_j^-)\int_{\nu_0\cdot z\leq 0} K^h_{ij}(z)\delta\chi_i^+(\cdot - z)dz\Big)\end{aligned} \tag{A.67}$$

and that similar identities hold exchanging the roles of i,j and $+,-$ respectively. That (A.67) holds is not difficult to show. Indeed multiplying the argument in both the limits by a test function $\xi\in C_c^\infty([0,1)^d\times(0,T))$ and integrating over space-time one observes that since the kernel is even, the argument of the second limit is just a spatial shift of z of the first one. By translation invariance the spatial shift may be put onto the test function, and thanks to the scaling of the kernel one can get the claim. We may thus substitute the first eight terms of the left hand side of (A.60) with twice the same terms with the integration with respect to z on the half space $\{\nu_0\cdot z\geq 0\}$. If we rely again on the fact that $\delta\chi_i^+\in\{0,1\}$, by identity (A.81) in the Appendix we obtain (A.66), as claimed.

Now we need two inequalities for the integrand. First note that the integrand is a mixed space-time second-order finite difference. We claim that

$$\begin{aligned}&|\delta\chi_i^+ - \delta\chi_j^-(\cdot - z)| + |\delta\chi_i^+(\cdot - z) - \delta\chi_j^-| + |\delta\chi_i^- - \delta\chi_j^+(\cdot - z)| + |\delta\chi_i^-(\cdot - z) - \delta\chi_j^+| \\ &\leq\begin{cases}|\delta\chi_i^+ - \delta\chi_i^+(\cdot - z)| + |\delta\chi_i^- - \delta\chi_i^-(\cdot - z)| + |\delta\chi_j^+ - \delta\chi_j^+(\cdot - z)| + |\delta\chi_j^- - \delta\chi_j^-(\cdot - z)| \\ +4\sum_{k\neq i,j}(|\delta\chi_k| + |\delta\chi_k(\cdot - z)|), \\ \\ |\delta\chi_i| + |\delta\chi_i(\cdot - z)| + |\delta\chi_j| + |\delta\chi_j(\cdot - z)|.\end{cases}\end{aligned} \tag{A.68}$$

The second follows from the triangle inequality. To show the first one, observe that

$$
\begin{aligned}
|\delta\chi_i^+ - \delta\chi_j^-(\cdot - z)| =&(1-\delta\chi_i^+)\delta\chi_j^-(\cdot - z) + \delta\chi_i^+(1-\delta\chi_j^-(\cdot - z))\\
\leq&(1-\delta\chi_i^+)\delta\chi_i^+(\cdot - z) + \sum_{k\neq i,j}|\delta\chi_k(\cdot - z)| + \delta\chi_j^-(1-\delta\chi_j^-(\cdot - z))\\
&+\sum_{k\neq i,j}|\delta\chi_k|
\end{aligned}
$$

and that similarly

$$
\begin{aligned}
|\delta\chi_i^+(\cdot - z) - \delta\chi_j^-| =&(1-\delta\chi_i^+(\cdot - z))\delta\chi_j^- + \delta\chi_i^+(\cdot - z)(1-\delta\chi_j^-)\\
\leq&(1-\delta\chi_i^+(\cdot - z))\delta\chi_i^+ + \sum_{k\neq i,j}|\delta\chi_k| + \delta\chi_j^-(\cdot - z)(1-\delta\chi_j^-)\\
&+\sum_{k\neq i,j}|\delta\chi_k(\cdot - z)|.
\end{aligned}
$$

Summing up the two inequalities we get

$$
\begin{aligned}
&|\delta\chi_i^+ - \delta\chi_j^-(\cdot - z)| + |\delta\chi_i^+(\cdot - z) - \delta\chi_j^-|\\
&\leq |\delta\chi_i^+ - \delta\chi_i^+(\cdot - z)| + |\delta\chi_j^- - \delta\chi_j^-(\cdot - z)| + 2\sum_{k\neq i,j}\left(|\delta\chi_k| + |\delta\chi_k(\cdot - z)|\right).
\end{aligned}
$$

Similar bounds hold for the remaining terms in (A.68).

We now split the integral (A.66) into the domains of integration $\{0 \leq \nu_0\cdot z \leq \alpha V_0\}$ and $\{\nu_0\cdot z > \alpha V_0\}$. On the first one we use the first inequality in (A.68) for the integrand. Recalling identity (A.81) and inequality (A.82) in the Appendix we obtain, and using the fact that $\sum_k \chi_k = 1$

$$
\begin{aligned}
&2\int_{0\le\nu_0\cdot z\le\alpha V_0} K_{ij}^h(z)\Big(|\delta\chi_i^+-\delta\chi_j^-(\cdot-z)|+|\delta\chi_i^+(\cdot-z)-\delta\chi_j^-|\\
&\qquad\qquad+|\delta\chi_i^--\delta\chi_j^+(\cdot-z)|+|\delta\chi_i^-(\cdot-z)-\delta\chi_j^+|\Big)dz\\
&\le 2\int_{0\le\nu_0\cdot z\le\alpha V_0} K_{ij}^h(z)\Big(|\chi_i-\chi_i(\cdot-z)|+|\chi_i(\cdot-\tau)-\chi_i(\cdot-\tau,\cdot-z)|\\
&\qquad\qquad+|\chi_j-\chi_j(\cdot-z)|+|\chi_j(\cdot-\tau)-\chi_j(\cdot-\tau,\cdot-z)|\\
&\qquad\qquad+8\sum_{k\neq i,j}|\delta\chi_k|+|\delta\chi_k(\cdot-z)|\Big)dz\\
&\le 2\int_{0\le\nu_0\cdot z\le\alpha V_0} K_{ij}^h(z)\Big(\chi_i\chi_j(\cdot-z)+\chi_i(\cdot-z)\chi_j+\sum_{k\neq i,j}\chi_i\chi_k(\cdot-z)+\chi_i(\cdot-z)\chi_k\\
&\qquad\qquad+\chi_i(\cdot-\tau)\chi_j(\cdot-\tau,\cdot-z)+\chi_i(\cdot-\tau,\cdot-z)\chi_j(\cdot-\tau)\\
&\qquad\qquad+\sum_{k\neq i,j}\chi_i(\cdot-\tau)\chi_k(\cdot-\tau,\cdot-z)+\chi_i(\cdot-\tau,\cdot-z)\chi_k(\cdot-\tau)\\
&\qquad\qquad+\chi_j\chi_i(\cdot-z)+\chi_j(\cdot-z)\chi_i+\sum_{k\neq i,j}\chi_j\chi_k(\cdot-z)+\chi_j(\cdot-z)\chi_k\\
&\qquad\qquad+\chi_j(\cdot-\tau)\chi_i(\cdot-\tau,\cdot-z)+\chi_j(\cdot-\tau,\cdot-z)\chi_i(\cdot-\tau)\\
&\qquad\qquad+\sum_{k\neq i,j}\chi_j(\cdot-\tau)\chi_k(\cdot-\tau,\cdot-z)+\chi_j(\cdot-\tau,\cdot-z)\chi_k(\cdot-\tau)\\
&\qquad\qquad+8\sum_{k\neq i,j}|\delta\chi_k|+|\delta\chi_k(\cdot-z)|\Big)\,dz.
\end{aligned}
\tag{A.69}
$$

On the set $\{\nu_0\cdot z>\alpha V_0\}$ we use the second inequality in (A.68), obtaining

$$
\begin{aligned}
&2\int_{\nu_0\cdot z>\alpha V_0} K_{ij}^h(z)\left(|\delta\chi_i^+-\delta\chi_j^-(\cdot-z)|+|\delta\chi_i^+(\cdot-z)-\delta\chi_j^-|\right.\\
&\qquad\qquad\left.+|\delta\chi_i^--\delta\chi_j^+(\cdot-z)|+|\delta\chi_i^-(\cdot-z)-\delta\chi_j^+|\right)dz\\
&\le 2\int_{\nu_0\cdot z>\alpha V_0} K_{ij}^h(z)(|\delta\chi_i|+|\delta\chi_i(\cdot-z)|+|\delta\chi_j|+|\delta\chi_j(\cdot-z)|)dz.
\end{aligned}
\tag{A.70}
$$

We now observe that for any $1\le k\le N$ we have, as we already observed in (A.54)

$$
\lim_{h\downarrow 0}\frac{1}{\sqrt{h}}\int_{0\le\nu_0\cdot z\le\alpha V_0} K_{ij}^h(z)(|\delta\chi_k(\cdot-z)|-|\delta\chi_k|)dz=0,
\tag{A.71}
$$

thus in particular

$$
\begin{aligned}
&\limsup_{h\downarrow 0}\frac{1}{\sqrt{h}}\int_{0\le\nu_0\cdot z\le\alpha V_0} K_{ij}^h(z)|\delta\chi_k(\cdot-z)|dz\\
&=\limsup_{h\downarrow 0}\frac{1}{\sqrt{h}}\int_{0\le\nu_0\cdot z\le\alpha V_0} K_{ij}^h(z)|\delta\chi_k|dz.
\end{aligned}
$$

By putting the time shift τ on the test function it is easy to check that the distributional limit of the terms of (A.69) which involve the shift τ have the same limit as the corresponding terms without the time shift. Thus recalling (A.48) and relying on (A.71) and (A.22) we obtain that inserting (A.69) and (A.70) into (A.66), the left hand side of (A.60) is bounded by

$$\begin{aligned}
&8\int_{0\le\nu_0\cdot z\le\alpha V_0} K_{ij}(z)((\nu_{ij}(x,t)\cdot z)_+ + (\nu_{ij}(x,t)\cdot z)_-)\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dtdz\\
&+C\sum_{k\neq i,j}\int_{0\le\nu_0\cdot z\le\alpha V_0} K_{ik}(z)((\nu_{ik}(x,t)\cdot z)_+ + (\nu_{ik}(x,t)\cdot z)_-)\mathcal{H}^{d-1}_{|\Sigma_{ik}(t)}(dx)dtdz\\
&+C\sum_{k\neq i,j}(\alpha^2\omega_k^{ac}+\mathcal{H}^{d-1}_{|\partial^*\Omega_k(t)}(dx)dt),
\end{aligned}$$

which clearly gives the claim once we realize that

$$\begin{aligned}
&\int_{0\le\nu_0\cdot z\le\alpha V_0} K_{ik}(z)((\nu_{ik}(x)\cdot z)_+ + (\nu_{ik}(x)\cdot z)_-)dz\\
&\le 2\int_{\mathbf{R}^d} K_{ik}(z)|z|dz\le C.
\end{aligned}$$

□

A.4.4 Proof of Proposition A.2.6

Proof. The proof is along the same lines as Proposition 2 in [55], where the claim is shown in the case of two phases. For the convenience of the reader, we outline the strategy of the full proof, providing details only for the required changes. The proof is split into several steps.

STEP 1. The first observation is that for any $h>0$, any admissible $u\in\mathcal{M}$ and any smooth vector field ξ we have the following lower bound for the metric slope, cf. (A.11)

$$\frac{1}{2}|\partial E_h|(u)\ge \delta E_h(u)_\bullet\xi-\frac{1}{2}\left(\delta d_h(\cdot,u)_\bullet\xi\right)^2.$$

Here δ denotes the first variation, which is computed considering the curve $s\to u^s$ of configurations which solve the transport equations

$$\begin{cases}\partial_s u_i^s+\xi\cdot\nabla u_i^s=0,\\ u_i^s(\cdot,0)=u_i(\cdot),\end{cases}$$

and by setting

$$\delta E_h(u)_\bullet\xi:=\frac{d}{ds}_{|s=0}E_h(u^s)\text{ and }\delta d_h(\cdot,u)_\bullet\xi:=\frac{d}{ds}_{|s=0}d(u,u^s).$$

STEP 2. The second observation is a representation formula for $\delta E_h(u)_\bullet\xi$. Namely

$$\begin{aligned}\delta E_h(u)_\bullet\xi=\sum_{i,j}\frac{1}{\sqrt{h}}\Bigg(&\int\nabla\cdot\xi u_iK_{ij}^h*u_jdx+\int\nabla\cdot\xi u_jK_{ij}^h*u_idxdt\\
&+\int[\xi,\nabla K_{ij}^h*](u_j)u_idx\Bigg).\end{aligned}$$

Here $[\xi, \nabla K_{ij}^h *]$ denotes the commutator obtained taking the convolution with ∇K_{ij}^h and multiplying by ξ, the definition is analogous to the one given in (A.40). To check this formula one starts by assuming u to be smooth and then an approximation argument gives the result for a general $u \in \mathcal{M}$.

STEP 3. Representation for $\delta d_h(\cdot, u)_\bullet \xi$. One checks that

$$\begin{aligned}
&\frac{1}{2}\left(\delta d_h(\cdot, u)_\bullet \xi\right)^2 \\
=&\frac{\sqrt{h}}{2}\sum_{i,j}\Bigg(\int u_i\xi\cdot\nabla^2 K_{ij}^h*(\xi u_j)dx+\int u_j\xi\cdot\nabla^2 K_{ij}^h*(\xi u_i)dx \\
&+\int u_i\nabla\cdot\xi\nabla K_{ij}^h*(\xi u_j)dx+\int u_j\nabla\cdot\xi\nabla K_{ij}^h*(\xi u_i)dx \\
&-\int u_i\nabla\cdot\xi K_{ij}^h*(u_j\nabla\cdot\xi)dx-\int u_j\nabla\cdot\xi K_{ij}^h*(u_i\nabla\cdot\xi)dx \\
&-\int \xi u_i\nabla K_{ij}^h*(u_j\nabla\cdot\xi)dx-\int \xi u_j\nabla K_{ij}^h*(u_i\nabla\cdot\xi)dx\Bigg).
\end{aligned}$$

Once again this formula can be easily checked when u is smooth, an approximation argument then gives the extension to the case $u \in \mathcal{M}$.

STEP 4. Passage to the limit in δE_h. We claim that

$$\lim_{h\downarrow 0}\int_0^T \delta E_h(u^h(t))_\bullet \xi dt = \sum_{i,j}\sigma_{ij}\int\left(\nabla\cdot\xi-\nu_{ij}\cdot\nabla\xi\nu_{ij}\right)\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt. \tag{A.72}$$

The proof is very similar to the two phases case, and relies on the weak convergence (A.22) and (A.23). Firstly, testing (A.22) with $\nabla\cdot\xi$ we get

$$\begin{aligned}
&\lim_{h\downarrow 0}\sum_{i,j}\frac{1}{\sqrt{h}}\int\left(\nabla\cdot\xi u_i^h K_{ij}^h*u_j^h+\nabla\cdot\xi u_j^h K_{ij}^h*u_i^h\right)dxdt \\
&=\sum_{i,j}2\sigma_{ij}\int\nabla\cdot\xi\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt.
\end{aligned}$$

For the term involving the commutator, one checks that

$$\lim_{h\downarrow 0}\left(\int[\xi,\nabla K_{ij}^h*](u_j^h)u_i^h dxdt-\int\nabla\xi z\cdot\nabla K_{ij}^h(z)u_j^h(x-z,t)u_i^h(x,t)dzdxdt\right)=0.$$

With this in place, we observe that

$$\begin{aligned}
&\lim_{h\downarrow 0}\int\nabla\xi z\cdot\nabla K_{ij}^h(z)u_j^h(x-z,t)u_i^h(x,t)dzdxdt \\
&=\int\nabla\xi(x,t)z\cdot\nabla K_{ij}(z)(\nu_{ij}(x,t)\cdot z)_+\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt
\end{aligned}$$

which can be seen by testing (A.22) with $\frac{\nabla \xi z \cdot \nabla K_{ij}(z)}{K_{ij}(z)}$, which is of polynomial growth in z. To conclude (A.72) we just need to show that for any symmetric matrix $A \in \mathbf{R}^{d\times d}$ and any unit vector ν we have

$$\int Az \cdot \nabla K_{ij}(z)(\nu \cdot z)_+ dz = -\sigma_{ij} \left(\operatorname{tr} A + \nu \cdot A\nu\right).$$

Using the definition of the kernel K_{ij} it suffices to show that

$$\int Az \cdot \nabla G_w(z)(\nu \cdot z)_+ dz = -\frac{\sqrt{w}}{\sqrt{\pi}} \left(\operatorname{tr} A + \nu \cdot A\nu\right) \ w \in \{\gamma, \beta\}.$$

STEP 5. Passage to the limit in $\delta d_h(\cdot, u)\xi$. We claim that

$$\lim_{h\downarrow 0} \frac{1}{2} \left(\delta d_h(\cdot, u^h)_{\bullet}\xi\right) = \sum_{i,j} \frac{1}{2\mu_{ij}} \int (\xi \cdot \nu_{ij})^2 \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt. \tag{A.73}$$

To prove this, we observe that the terms which do not involve the Hessian $\nabla^2 K^h_{ij}$ are all $O(\sqrt{h})$. For example, to prove that

$$\sqrt{h} \int u^h_i \nabla \cdot \xi \nabla K^h_{ij} * (\xi u^h_j) dxdt = O(\sqrt{h}), \tag{A.74}$$

spell out the integral in the convolution, use the fact that $\nabla K^h_{ij} = \frac{1}{\sqrt{h}^{d+1}} \nabla K_{ij}(\frac{z}{\sqrt{h}})$, use the fact that $\nabla \cdot \xi(x,t)\xi(x - \sqrt{h}z, t)$ is bounded and test (A.22) with $\nabla K_{ij}/K_{ij}$. The other terms can be treated similarly. For the terms involving the Hessian of the kernel, we split the claim into

$$\lim_{h\downarrow 0} \sqrt{h} \int u^h_i (\xi \cdot \nabla^2 K^h_{ij} * u_j)\xi dxdt = \frac{1}{2\mu_{ij}} \int (\xi \cdot \nu_{ij}(x,t))^2 \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt, \tag{A.75}$$

$$\sqrt{h} \int u^h_i \xi \cdot [\xi, \nabla^2 K^h_{ij} *](u^h_j) dxdt = O(\sqrt{h}). \tag{A.76}$$

The proof of (A.76) is similar to the argument for (A.74). In fact, while the additional derivative on the kernel gives an additional factor $\frac{1}{\sqrt{h}}$, we gain a factor $\sqrt{h}$ by the Lipschitz estimate

$$|\xi(x,t) - \xi(x - \sqrt{h}z, t)| \le \sqrt{h}\|\nabla \xi\|_\infty.$$

To prove identity (A.75) observe that by spelling out the z-integral, a change of variable and by testing (A.22) with $\frac{\xi(x,t)\cdot\nabla^2 K_{ij}(z)\xi(x,t)}{K_{ij}(z)}$ we obtain

$$\begin{aligned} &\lim_{h\downarrow 0} \sqrt{h} \int u^h_i (\xi \cdot \nabla^2 K^h_{ij} * u^h_j)\xi dxdt \\ &= \int \xi \cdot \nabla^2 K_{ij}(z)\xi(\nu_{ij}(x,t) \cdot z)_+ \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt. \end{aligned}$$

Now identity (A.73) follows from the following formula: For any two vectors $\xi \in \mathbf{R}^d$ and $\nu \in \mathbf{S}^{d-1}$ we have

$$\int \xi \cdot \nabla^2 K_{ij}(z)\xi(\nu \cdot z)_+ dz = \frac{1}{2\mu_{ij}}(\xi \cdot \nu)^2. \tag{A.77}$$

To check (A.77), by relying on the definition of the kernels, we just need to show that for $w \in \{\gamma, \beta\}$

$$\int \xi \cdot \nabla^2 G_w(z)\xi(\nu \cdot z)_+ dz = \frac{1}{2\sqrt{\pi w}}(\xi \cdot \nu)^2.$$

Since the kernel is isotropic, we can reduce to the case $\xi = e_1$, thus we need to prove

$$\int \partial_1^2 G_w(z)(\nu \cdot z)_+ dz = \frac{1}{2\sqrt{\pi w}}\nu_1^2.$$

This can be done after two integration by parts and observing that

$$\int_{\nu \cdot z = 0} G_w(z)\mathcal{H}^{d-1}(dz) = \frac{1}{2\sqrt{\pi w}}.$$

CONCLUSION. By STEP 1 we have

$$\frac{1}{2}\int_0^T |\partial E_h|^2(u^h)\ dt \geq \int_0^T \delta E_h(u^h)_\bullet \xi dt - \frac{1}{2}\int_0^T \left(\delta d_h(\cdot, u^h)_\bullet \xi\right)^2 dt.$$

Taking the liminf on the left hand side, using STEP 4 and STEP 5 we get that for any smooth vector field ξ

$$\begin{aligned}\liminf_{h\downarrow 0} \frac{1}{2}\int_0^T |\partial E_h|^2(u^h) dt \geq \sum_{i,j} \Bigg[\sigma_{ij} \int \left(\nabla \cdot \xi - \nu_{ij} \cdot \nabla \xi \nu_{ij}\right) \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt \\ - \frac{1}{2\mu_{ij}} \int (\xi \cdot \nu_{ij})^2\ \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt\Bigg].\end{aligned}$$

Since the left hand side is bounded, the Riesz representation theorem for L^2 yields functions $H_{ij} \in L^2(\mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt)$ such that

$$\sum_{i,j} \sigma_{ij} \int \left(\nabla \cdot \xi - \nu_{ij} \cdot \nabla \xi \nu_{ij}\right)\ \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt = -\sum_{i,j} \sigma_{ij} \int H_{ij}\nu_{ij} \cdot \xi\ \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt$$

and such that for any $\xi \in L^2(\mathcal{H}^{d-1}_{|\bigcup_{i,j}\Sigma_{ij}(t)}(dx)dt)$

$$\begin{aligned}\liminf_{h\downarrow 0} \frac{1}{2}\int_0^T |\partial E_h|(u_h)\ dt \geq \sum_{i,j} \Bigg(- \sigma_{ij} \int H_{ij}\nu_{ij} \cdot \xi \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt \\ - \frac{1}{2\mu_{ij}} \int (\xi \cdot \nu_{ij})^2 \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt\Bigg).\end{aligned}$$

Since the integration measures are mutually singular we can test with a vector field $\xi \in L^2(\mathcal{H}^{d-1}_{|\bigcup_{i,j}\Sigma_{ij}(t)}(dx)dt)$ such that $\xi_{|\Sigma_{ij}(t)} = -\mu_{ij}\sigma_{ij}H_{ij}\nu_{ij}$. This yields

$$\liminf_{h\downarrow 0} \frac{1}{2}\int_0^T |\partial E_h|^2(u_h)\ dt \geq \frac{1}{2}\sum_{i,j} \sigma_{ij}^2\mu_{ij} \int H_{ij}^2\ \mathcal{H}^{d-1}_{|\Sigma_{ij}(t)}(dx)dt. \qquad \square$$

A.5 Appendix

A.5.1 Proof of Lemma A.3.1

Before giving the proof of this result, we need a simple technical lemma.

Lemma A.5.1. *Fix $1 \leq l \neq p \leq N$. Then for any $1 \leq i \neq j \leq N$ such that $\{i,j\} \neq \{l,p\}$ the interfaces Σ_{ij} and Σ_{lp} are disjoint. In particular for $\mathcal{H}^{d-1}$-a.e. $x \in \Sigma_{lp}$ we have that*

$$\lim_{r\downarrow 0} \frac{\mathcal{H}^{d-1}(\Sigma_{ij} \cap B(x,r))}{\omega_{d-1} r^{d-1}} = 0. \tag{A.78}$$

Proof. We first show that the interfaces Σ_{ij} and Σ_{lp} are disjoint. This follows immediately once we recall that every point in the reduced boundary of a set of finite perimeter has density $1/2$ (see [59], Corollary 15.8). Assume for example that $i \neq l, p$. Thus if $y \in \Sigma_{lp}$ we have

$$\begin{aligned}
1 &\geq \limsup_{r\downarrow 0} \frac{|(\Omega_l \cup \Omega_p \cup \Omega_i) \cap B(y,r)|}{\omega_d r^d} \\
&= \lim_{r\downarrow 0} \frac{|\Omega_l \cap B(y,r)|}{\omega_d r^d} + \lim_{r\downarrow 0} \frac{|\Omega_p \cap B(y,r)|}{\omega_d r^d} + \limsup_{r\downarrow 0} \frac{|\Omega_i \cap B(y,r)|}{\omega_d r^d} \\
&= 1 + \limsup_{r\downarrow 0} \frac{|\Omega_i \cap B(y,r)|}{\omega_d r^d}
\end{aligned}$$

which says that y has density zero in Ω_i.

The fact that (A.78) holds is now a consequence of the general fact

$$\limsup_{r\downarrow 0} \frac{\mathcal{H}^{d-1}(\Sigma_{ij} \cap B(x,r))}{\omega_{d-1} r^{d-1}} = 0$$

for $\mathcal{H}^{d-1}$-a.e. $x \in (\Sigma_{ij})^c$. □

Proof of Lemma A.3.1. The argument for (i) can be found in [53] in the case of two phases and without localization, i.e. with $\eta = 1$ and $N = 2$. For the sake of completeness, we provide the proof in our case. Upon splitting into the negative and positive part, we may assume $\eta \geq 0$. Clearly the only nonzero terms in the sum are those for which $B^r_m \cap \Sigma_{ij} \neq \emptyset$. Fix such a ball: by definition there exists $y \in r\mathbf{Z}^d$ such that $B^r_m = B(y, 2r\sqrt{d})$. If $x \in \Sigma_{ij} \cap B^r_m$ then we have that $B(x, 2r\sqrt{d}) \subset B(y, 4r\sqrt{d})$, and by definition of $\mathcal{E}^r$ this yields

$$\mathcal{H}^{d-1}(B(x, 2r\sqrt{d}) \cap \Sigma_{ij}) \leq \frac{\omega_{d-1}}{2^d}(4r)^{d-1}\sqrt{d}^{\,d-1} = \frac{\omega_{d-1}}{2}(2r)^{d-1}\sqrt{d}^{\,d-1}.$$

Thus x belongs to the set of points in $\Sigma_{ij} \cap B^r_m$ such that

$$\frac{\mathcal{H}^{d-1}(B(x, 2r\sqrt{d}) \cap \Sigma_{ij})}{\omega_{d-1}(2r\sqrt{d})^{d-1}} \leq \frac{1}{2}. \tag{A.79}$$

By De Giorgi's structure theorem the approximate tangent plane exists at every point $x \in \Sigma_{ij}$, thus (A.79) cannot hold when r is small enough: moreover every point $x \in \Sigma_{ij}$ is contained in at most $c(2,d)$ balls, this means that

$$\sum_m \mathbf{1}_{\left\{z\in B_m^r\cap\Sigma_{ij}:\ \frac{\mathcal{H}^{d-1}(B(x,2r\sqrt{d})\cap\Sigma_{ij})}{\omega_{d-1}(2r\sqrt{d})^{d-1}}\leq\frac{1}{2}\right\}}(x)\eta(x)\leq c(2,d)\eta(x) \tag{A.80}$$

and that the left hand side of (A.80) converges to zero pointwise. By the dominated convergence theorem we get our claim.

Proof of (ii). Upon splitting into the negative and positive part, we may assume $\eta\geq 0$. Given a point $x\in\Sigma_{lp}$, if $y\in r\mathbf{Z}^d$ is such that $x\in B(y,2r\sqrt{d})$, then $B(y,4r\sqrt{d})\subset B(x,6r\sqrt{d})$. Thus for any $1\leq i<j\leq N$ with $(i,j)\neq(l,p)$ we have

$$\begin{aligned}\mathcal{H}^{d-1}(B(y,4r\sqrt{d})\cap\Sigma_{ij})&\leq\mathcal{H}^{d-1}(B(x,6r\sqrt{d})\cap\Sigma_{ij})\\&\leq\frac{\omega_{d-1}}{2^d}(4r)^{d-1}\sqrt{d}^{d-1}\end{aligned}$$

provided r is small enough, this follows from Lemma A.5.1. Since $\mathcal{F}_2^r$ covers $[0,1)^d$ we obtain that

$$x\in\bigcup_m B_m^r$$

for all r small enough. In other words

$$\lim_{r\downarrow 0}\sum_m\rho_m(x)\eta(x)=\eta(x)$$

pointwise on Σ_{lp}, and the argument of the limit on the right hand side is dominated by η. Thus we may once again appeal to the dominated convergence theorem and conclude the proof. ☐

A.5.2 Consistency and Monotonicity

The following results are essentially contained in [25] and [53], indeed the proofs may be adapted because we are assuming that a_{ij} and b_{ij} satisfy the triangle inequality.

Lemma A.5.2. *For every $\chi\in\mathcal{A}\cap BV([0,1)^d)^N$ we have*

$$\lim_{h\downarrow 0}E_h(\chi)=E(\chi).$$

If $\chi\in L^1((0,T),BV([0,1)^d)^N)$ such that $\chi(\cdot,t)\in\mathcal{A}$ for a.e. t. Then

$$\lim_{h\downarrow 0}\int_0^T E_h(\chi)dt=\int_0^T E(\chi)dt.$$

Even more is true: for any $g\in C^\infty([0,1)^d)$ and any pair $1\leq i\neq j\leq N$ we have

$$\begin{aligned}&\lim_{h\downarrow 0}\frac{1}{\sqrt{h}}\int_0^T\int g(x)(\chi_i(x,t)K_{ij}^h*\chi_j(x,t)+\chi_j(x,t)K_{ij}^h*\chi_i(x,t))dxdt\\&=\int g(x)K_{ij}(z)|\nu_{ij}\cdot z|dzdxdt.\end{aligned}$$

Lemma A.5.3. *For any $0<h\leq h_0$ we have*

$$E_h(u)\geq\left(\frac{\sqrt{h_0}}{\sqrt{h}+\sqrt{h_0}}\right)^{d+1}E_{h_0}(u).$$

A.5.3 Improved convergence of the energies

The following Lemma is an improvement of the convergence of the energies, the proof of this result is contained, with minor modifications, in the paper [53], Corollary 3.7.

Lemma A.5.4. *Let u^h be a sequence of $[0,1]$-valued functions such that $u^h \to \chi$ in $L^1([0,1)^d \times (0,T))$ and*

$$\lim_{h\downarrow 0} \int_0^T E_h(u^h(t))dt = \int_0^T E(\chi(t))dt.$$

Then we have that

$$\lim_{h\downarrow 0} \frac{1}{\sqrt{h}} \int G_\gamma^h(z)|f_h^\gamma(z) - f^\gamma(z)|dz = 0,$$
$$\lim_{h\downarrow 0} \frac{1}{\sqrt{h}} \int G_\beta^h(z)|f_h^\beta(z) - f^\beta(z)|dz = 0,$$

where we set

$$f_h^\gamma(z) = \sum_{i,j} a_{ij} \int u_i^h(x,t)u_j^h(x-z,t)dxdt, \quad f^\gamma(z) = \sum_{i,j} a_{ij} \int \chi_i(x,t)\chi_j(x-z,t)dxdt,$$
$$f_h^\beta(z) = \sum_{i,j} b_{ij} \int u_i^h(x,t)u_j^h(x-z,t)dxdt, \quad f^\beta(z) = \sum_{i,j} b_{ij} \int \chi_i(x,t)\chi_j(x-z,t)dxdt.$$

A.5.4 Some inequalities

Here we gather some elementary inequalities which are used frequently.

Lemma A.5.5. *Let $a, b, a', b' \in \{0,1\}$, then the following inequalities hold:*

$$|a-b| = a(1-b) + b(1-a), \tag{A.81}$$
$$|(a-a')_+ - (b-b')_+| + |(a-a')_- - (b-b')_-| \le |a-b| + |a'-b'|. \tag{A.82}$$

Proof. The first identity follows by expanding $|a-b| = |a-b|^2$. The second one is proved in [55]. For the sake of completeness, we reproduce the proof here. There are two cases. In the first one we have $(a-a')(b-b') \ge 0$ and we may assume upon replacing (a,a',b,b') with $(-a,-a',-b,-b')$ that $(a-a')$ and $(b-b')$ are non-negative. Then (A.82) reduces to

$$|(a-a') - (b-b')| \le |a-b| + |a'-b'|.$$

The second case is given by $(a-a')(b-b') \le 0$. By an argument as before we may assume $(a-a') \ge 0 \ge (b'-b)$, thus (A.82) reduces to

$$(a-a') + (b-b') \le |a-b| + |a'-b'|. \qquad \square$$

Lemma A.5.6. *There exists a constant $C > 0$ depending only on $N, \mathbb{A}, \mathbb{B}$ such that for any $v \in \mathcal{M}$*

$$\int |v - K^{h_0} * v| dx \leq C\sqrt{h_0} E_h(v) \quad \textit{for all } h_0 \geq h. \tag{A.83}$$

Proof. The proof of (A.83) is contained in the proof of Lemma 3 in [55] for the two phases case when K^h is the scaled version of the Gaussian with variance 1. The same proof may be adapted to our setting because we still have monotonicity of the energy (Lemma A.5.3) and we can prove essentially by the use of Jensen's inequality that

$$\int |v - K^h * v| dx \leq C\sqrt{h} E_h(v).$$

□

APPENDIX B

LARGE DATA LIMIT OF THE MBO SCHEME FOR DATA CLUSTERING: Γ-CONVERGENCE OF THE THRESHOLDING ENERGIES

B.1 The MBO scheme for data clustering

In this section, we provide the rigorous formulation of the MBO algorithm for data clustering originally given by Bertozzi et al. in [62], [78], [61]. Let $G = (V, E)$ be a graph with vertex set $V = \{x_1, ..., x_n\}$ and let E be the set of edges weighted by $w_{ij} = w_{ji}$, $1 \leq i, j \leq n$. We assume that $w_{ii} = 0$ for every $i = 1, \ldots, n$. For every $i = 1, \ldots, n$ we define the degree d_i as

$$d_i := \frac{1}{n} \sum_{j=1}^{n} w_{ij}.$$

We will assume that $d_i > 0$ for every $1 \leq i \leq n$. We let $\mathcal{V}$ be the space of real valued functions defined on V. We define an inner product on $\mathcal{V}$ by

$$\langle u, v \rangle_{\mathcal{V}} = \frac{1}{n} \sum_{i=1}^{n} d_i u_i v_i, \; u, v \in \mathcal{V}.$$

We let $\mathcal{E}$ be the space of antisymmetric functions on E. We define an inner product on $\mathcal{E}$ as

$$\langle F, G \rangle_{\mathcal{E}} = \frac{1}{2n^2} \sum_{i,j:\; w_{ij} \neq 0} F_{ij} G_{ij} \frac{1}{w_{ij}}, \; F, G \in \mathcal{E}.$$

Given $\epsilon > 0$ we define the derivative operator $\nabla : \mathcal{V} \to \mathcal{E}$ acting on functions $u : V \to \mathbf{R}$ as

$$(\nabla u)_{ij} = \frac{w_{ij}}{\epsilon}(u_j - u_i). \tag{B.1}$$

We denote by $\operatorname{div} : \mathcal{E} \to \mathcal{V}$ its adjoint with respect to the scalar products on $\mathcal{V}$ and $\mathcal{E}$. Explicitly, we may compute for $F \in \mathcal{E}$

$$(\operatorname{div} F)_i = \frac{1}{2\epsilon d_i} \sum_{j \neq i} (F_{ij} - F_{ji}).$$

Finally, we introduce the *random walk* graph Laplacian $\Delta := \operatorname{div} \circ \nabla : \mathcal{V} \to \mathcal{V}$. Explicitly, Δ can be identified with the matrix

$$\frac{1}{\epsilon^2}\left(\mathbb{I} - \frac{1}{n} D^{-1} W\right),$$

where $D = \operatorname{diag}(d_1, ..., d_n)$ is the diagonal matrix of degrees and $W = (w_{ij})_{i,j=1}^n$ is the matrix of weights. Given $t \in \mathbf{R}$ we let $e^{-t\Delta}$ be the exponential of the matrix $-t\Delta$. If $u \in \mathcal{V}$ then the function $v(t) = e^{-t\Delta} u$ solves the heat equation with initial value u on the graph, i.e.

$$\begin{cases} \frac{d}{dt} v(t) = -\Delta v(t), \\ v(0) = u. \end{cases}$$

We are now ready to introduce the MBO scheme for data clustering. Given a natural number $P \leq n$, a classification of the points of G into P clusters is a function $\chi : V \to \{0,1\}^P$ such that $\sum_{m=1}^P \chi_m(x) = 1$ for all $x \in V$. In other words, χ encodes a partition of the graph G into P clusters $C_m = \mathbf{1}_{\{\chi_m = 1\}}$, $m = 1, ..., P$. Let $\sigma := (\sigma_{ml})_{1 \leq m,l \leq P} \in \mathbf{R}^{P \times P}$ be a symmetric matrix with $\sigma_{ml} > 0$ for $m \neq l$ and $\sigma_{mm} = 0$. Then the MBO scheme for data clustering is as follows.

Algorithm B.1.1 (MBO scheme)**.** *Let P be the number of clusters, let $h > 0$ be the time-step size, and let N be the number of iterations to run. Let $\chi^0 : V \to \{0,1\}^P$ be a given clustering of the graph into P clusters. To obtain a clustering $\chi^N : V \to \{0,1\}^P$ using the MBO scheme, define inductively a new clustering $\chi^{q+1} : V \to \{0,1\}^P$ given the clustering $\chi^q : V \to \{0,1\}^P$ by performing the following steps for $0 \leq q < N$:*

1. ***Diffusion**. For every $m = 1, ..., P$ define*

$$u_m^q := \sum_{l \neq m} \sigma_{ml} e^{-h\Delta} \chi_l^q$$

2. ***Thresholding**. For every $m = 1, ..., P$ update the cluster by setting*

$$\{\chi_m^{q+1} = 1\} := \left\{x \in V : \; u_m^q(x) < \min_{l \neq m} u_l^q(x)\right\}.$$

We define the set $\mathcal{M}_G := \left\{u : V \to [0,1]^P : \; \sum_{m=1}^P u_m = 1 \text{ on } V\right\}$. For $u \in \mathcal{M}_G$ define the thresholding energy

$$E_G^h(u) = \frac{1}{\sqrt{h}} \sum_{i,j=1}^P \sigma_{ij} \langle u_i, e^{-h\Delta} u_j \rangle_{\mathcal{V}}.$$

The following lemma, which is essentially due to Esedoḡlu and Otto, is the main motivation for our work. We also refer to [78, Proposition 4.6].

Lemma B.1.2 ([25])**.** *Assume that the matrix σ is negative semidefinite on $(1, ..., 1)^\perp$, that means $v \cdot \sigma v \leq 0$ for all $v \in \mathbf{R}^d$ with $v \cdot (1, ..., 1) = 0$. In the setting of the previous algorithm, to obtain the new clustering $\chi^{q+1} : V \to \{0,1\}^P$ starting from $\chi^q : V \to \{0,1\}^P$ define*

$$\chi^{q+1} \in \operatorname*{argmin}_{u \in \mathcal{M}_G} \left\{ E_G^h(u) - \frac{1}{\sqrt{h}} \sum_{i,j=1}^P \sigma_{ij} \big\langle (u_i - \chi_i^q), e^{-h\Delta}(u_j - \chi_j^q) \big\rangle_{\mathcal{V}} \right\}.$$

B.2 Main results

In this section we introduce the setting of our problem and we state our main results. For the technical background and definitions we refer to Section B.5.

We assume that $M \subset \mathbf{R}^d$ is a k-dimensional compact Riemannian submanifold of $\mathbf{R}^d$. Let $\nu = \rho \,\mathrm{Vol}_M \in \mathcal{P}(M)$ be a probability measure on M, absolutely continuous with respect to the volume measure with a smooth and positive density ρ. Assume that $\{X_i\}_{i\in\mathbf{N}}$ are iid random points on M, distributed according to ν. For any $n \in \mathbf{N}$ and $\epsilon > 0$ we define the random graph $G_{n,\epsilon}$ with vertex set given by $V_{n,\epsilon} = \{X_1, ..., X_n\}$ and weights

$$w_{ij}^{(n,\epsilon)} = \frac{1}{\epsilon^k}\eta\left(\frac{|X_i - X_j|_d}{\epsilon}\right),\ i \neq j,$$

and $w_{ii}^{(n,\epsilon)} = 0$, where $\eta : [0, +\infty) \to [0, +\infty)$ is a given function and $|\cdot|_d$ denotes the standard Euclidean norm on $\mathbf{R}^d$. Here ϵ is the same as in (B.1). We stress that the graphs we constructed are random object and sometimes we will make this randomness explicit by specifying the dependence of the graph on an additional variable $\omega \in \Omega$, where Ω is the probability space on which the random objects $\{X_i\}_{i\in\mathbf{N}}$ are defined. On the function η we set the following conditions:

1. $\eta(0) > 0$ and η is continuous at 0,
2. $\eta(t) \geq 0$ for every $t > 0$, η is nonincreasing and η is $C^2((0, +\infty))$,
3. η, η', η'' have exponential decay.

Define $C_1 = \int_{\mathbf{R}^k} \eta(|y|_k)dy$ and $C_2 = \int_{\mathbf{R}^k} \eta(|y|_k)y_1^2 dy$. Assume that $P \in \mathbf{N}$ and let $\sigma \in \mathbf{R}^{P\times P}$ be a symmetric matrix, negative definite on $(1, ..., 1)^{\perp}$. We also assume that $\sigma_{ii} = 0$ for each $i = 1, ..., P$ and that $\sigma_{ij} = \sigma_{ji} > 0$ for all $i \neq j$. Finally, we assume that the coefficients of σ satisfy the *triangle inequality*, that is

$$\sigma_{ij} \leq \sigma_{il} + \sigma_{lj} \quad \forall i, j, l \in \{1, ..., P\}.$$

These assumptions are satisfied, for example, if we let σ be the matrix defined by

$$\sigma_{ij} = \begin{cases} 1 & \text{if } i \neq j, \\ 0 & \text{otherwise.} \end{cases}$$

The fact that the matrix σ is negative definite is needed in order that the MBO scheme dissipates the thresholding energy at every iteration, cf. Lemma B.1.2, while the triangle inequality ensures the lower semicontinuity of the energy. For each $n \in \mathbf{N}$ define the set

$$\mathcal{M}_n := \left\{u : V_n \to [0, 1]^P :\ \sum_{i=1}^{P} u_i = 1\right\}.$$

Given $h > 0$, define the thresholding energies on $\mathcal{M}_n$ as

$$E_{n,\epsilon}^h(u) := \frac{1}{\sqrt{h}} \sum_{i,j=1}^{P} \sigma_{ij} \langle u_i, e^{-h\Delta_{n,\epsilon}} u_j\rangle_{\nu_n}, \quad u \in \mathcal{M}_n. \tag{B.2}$$

We also define the set

$$\mathcal{M} := \left\{ u : M \to [0,1]^P \text{ measurable} : \sum_{i=1}^{P} u_i = 1 \text{ a.e.} \right\}.$$

Let $\mu = \xi \operatorname{Vol}_M$ for $\xi \in C^\infty(M)$, $\xi > 0$. Given $h > 0$ we define the thresholding energy on the weighted manifold (M, μ) as

$$E_h(u) := \frac{1}{\sqrt{h}} \sum_{i,j=1}^{P} \sigma_{ij} \int_M u_i(x) e^{-h\Delta_\xi} u_j(x) d\mu, \quad u \in \mathcal{M}.$$

Here, Δ_ξ is the weighted Laplacian on (M, μ), which is defined by its action on smooth functions $f \in C^\infty(M)$ as

$$\Delta_\xi f = -\frac{1}{\xi} \operatorname{div}(\xi \nabla f),$$

and $e^{-t\Delta_\xi}$ is the corresponding heat operator; we refer to Section B.5 for the relevant background and definitions. We are now in a position to state our main results.

Theorem B.2.1 (Discrete to nonlocal Γ-convergence)**.** *Let M be a k-dimensional compact Riemannian submanifold of $\mathbf{R}^d$. Let $\nu = \rho \operatorname{Vol}_M \in \mathcal{P}(M)$ be a probability measure, absolutely continuous with respect to the volume element with a smooth and positive density. Let $\{X_i\}_{i \in \mathbf{N}}$ be a sequence of iid random points in M, distributed according to ν. Let $\epsilon_n > 0$ be a sequence such that*

$$\lim_{n \to +\infty} \frac{\epsilon_n^{k+2} n}{\log(n)} = +\infty, \quad \lim_{n \to +\infty} \epsilon_n = 0.$$

Let G_{n,ϵ_n} be the corresponding random graphs. It holds almost surely that for any $h > 0$ if v^n converge weakly to v in $TL^2(M)$ then

$$\lim_{n \to +\infty} E^h_{n,\epsilon_n}(v^n) = \sqrt{\frac{C_1 C_2}{2}} E_{\frac{C_2 h}{2C_1}}(v),$$

where the thresholding energy on the right hand side corresponds to the weight $\xi = \rho^2$. Moreover, every sequence $v^n \in \mathcal{M}_n$ has a subsequence converging weakly in $TL^2(M)$ and every limit point lies in $\mathcal{M}$.

The main tool for proving Theorem B.2.1 is the following strong convergence of the heat operators on the graphs to the corresponding heat operator on the manifold.

Theorem B.2.2. *Let the assumptions of Theorem B.2.1 be in place. Then it holds almost surely that if $u^n \in \mathcal{V}_n$ is a sequence of functions converging weakly to $u \in L^2(M)$ in $TL^2(M)$, then for every $t > 0$ we have*

$$\lim_{n \to +\infty} e^{-t\Delta_{n,\epsilon_n}} u^n = e^{-\frac{C_2}{2C_1} t \Delta_{\rho^2}} u \text{ strongly in } TL^2(M).$$

As a corollary of Theorem B.2.2 we also obtain the following result about the consistency of one step of MBO in the large data limit, which, in the setting of random geometric graphs, answers positively to a question by Bertozzi et al. [78, Question 7.5].

Corollary B.2.3. *Let the assumptions of Theorem B.2.1 hold true. Then the following holds almost surely: Let $\chi^n : V_n \to \{0,1\}^P$ be such that the sequence $\{\chi^n\}_{n\in\mathbf{N}}$ converges weakly in $TL^2(M)$ to a function $\chi : M \to \{0,1\}^P$. Denote by $\chi^{n,h} : V_n \to \{0,1\}^P$ the outcome of one step ($N = 1$) of the MBO scheme (Algorithm B.1.1) on the n-th graph with step size $h > 0$ and initial clustering χ^n. Denote by $\chi^h : M \to \{0,1\}^P$ the outcome of one step of the MBO scheme on the manifold (Algorithm B.5.1) with initial value χ, step size $h > 0$ and diffusion parameter $\kappa = \frac{C_2}{2C_1}$. Then the sequence $\{\chi^{n,h}\}_{n\in\mathbf{N}}$ converges weakly to χ^h in $TL^2(M)$. By induction, the convergence holds for any number N of iterations.*

Remark B.2.4. Let us point out that if $\Omega \subset M$ is an open set with smooth boundary, then the functions $\chi^n : V_n \to \{0,1\}$ defined as $\chi^n := \mathbf{1}_{V_n\cap\Omega}$ are such that almost surely

$$TL^2(M) - \lim_{n\to+\infty} \chi^n = \chi := \mathbf{1}_\Omega. \tag{B.3}$$

In particular, the conclusion of Corollary B.2.3 holds true with these choices of initial values. Equation (B.3) follows by the fact that if T_n is a sequence of transport maps obtained by applying Theorem C.6.6, then there exists $\delta > 0$ and a constant C depending only on M and ρ such that if $\theta_n := \sup_{x\in M} d_M(x, T_n(x)) \leq \delta$ then

$$\int_M |\chi^n(T_n(x)) - \chi(x)| d\nu \leq C\theta_n \int_{\partial\Omega} \rho d\mathcal{H}^{k-1}. \tag{B.4}$$

The validity of (B.4) is shown in the flat case in [36, Remark 5.1], the analogous estimate on a closed manifold can be shown in a similar way.

If $u \in \mathcal{M}$ is such that $u \in BV(M, \{0,1\}^P)$ then we set $\Omega_i := \{u_i = 1\}$ and $\Sigma_{ij} := \partial^*\Omega_i \cap \partial^*\Omega_j$, the intersection of the reduced boundaries of Ω_i and Ω_j. Again, we refer to Section B.5 for the relevant background. We have the following result about the convergence of the nonlocal thresholding energy on the manifold.

Theorem B.2.5 (Nonlocal to local Γ-convergence)**.** *Let M be a k-dimensional compact Riemannian submanifold of $\mathbf{R}^d$ weighted by a measure $\mu = \xi \operatorname{Vol}_M$ with $\xi > 0$, $\xi \in C^\infty(M)$. Let $\sigma \in \mathbf{R}^{P\times P}$ be symmetric, $\sigma_{ii} = 0$ and such that σ satisfy the triangle inequality. Then on $\mathcal{M}$*

$$\Gamma(L^1(M)) - \lim_{h\downarrow 0} E_h = E,$$

where we define for $u \in \mathcal{M}$

$$E(u) = \begin{cases} \frac{1}{\sqrt{\pi}} \sum_{ij} \sigma_{ij} |D\chi_{\Omega_i}|_\xi(\Sigma_{ij}) & \text{if } u \in BV(M, \{0,1\}^P), \\ +\infty & \text{otherwise.} \end{cases}$$

Moreover, if $u \in BV(M, \{0,1\}^P) \cap \mathcal{M}$ then we have $\lim_{h\downarrow 0} E_h(u) = E(u)$. Finally, if u^h are functions in $\mathcal{M}$ such that $\sup_{h>0} E_h(u^h) < +\infty$ then the family $\{u^h\}_{h\downarrow 0}$ is precompact in $L^1(M)$ and every limit point is in $BV(M, \{0,1\}^P) \cap \mathcal{M}$.

Remark B.2.6. It should be remarked that the geometric assumptions for the previous result are not sharp. For $x \in M$ and $r > 0$ we denote by $B_r(x)$ the Riemannian ball of radius r centered at x. We also denote by p the heat kernel for the weighted Laplacian Δ_ξ (cf. (B.5.1)). For our proof to work, we need the following properties.

(i) **Doubling property**. There exists $N > 0$ such that for any $x \in M$ and any $r > 0$

$$\mu(B(x, 2r)) \leq 2^N \mu(B(x, r)). \tag{B.5}$$

(ii) **Asymptotic expansion for the heat kernel**. There exist functions $v^j \in C^\infty(M \times M), j \in \mathbf{N}$, such that for every $N > l + \frac{k}{2}$ there exists a constant $\tilde{C}_N < \infty$ such that

$$\left| \nabla^l \left(p(t, x, y) - \frac{e^{-\frac{d^2(x,y)}{4t}}}{(4\pi t)^{k/2}} \sum_{j=0}^{N} v^j(x, y) t^j \right) \right| \leq \tilde{C}_N t^{N+1-\frac{k}{2}}, \tag{B.6}$$

provided $d(x, y) \leq \frac{\operatorname{inj}(M)}{2}$, where $\operatorname{inj}(M)$ is the injectivity radius of the manifold M. Moreover we have that $v^0(x, x) = \frac{1}{\xi(x)}$.

(iii) **Gaussian bounds I**. There exists constants $Q_1, Q_2, Q_3, Q_4 > 0$ such that for every $t > 0$ and all $x, y \in M$,

$$\frac{Q_1}{\mu(B_{\sqrt{t}}(x))} e^{-\frac{d^2(x,y)}{Q_2 t}} \leq p(t, x, y) \leq \frac{Q_3}{\mu(B_{\sqrt{t}}(x))} e^{-\frac{d^2(x,y)}{Q_4 t}}. \tag{B.7}$$

(iv) **Gaussian bounds II**. There exist $\hat{C}_1, \hat{C}_2 > 0$ such that for any $x, y \in M$ and any $t > 0$

$$|\nabla_x p(t, x, y)| \leq \frac{\hat{C}_1}{\sqrt{t}\mu(B_{\sqrt{t}}(x))} \exp\left(-\frac{d^2(x, y)}{\hat{C}_2 t} \right). \tag{B.8}$$

These properties are satisfied in the case of a closed manifold as we work with. Indeed the doubling property follows from [17], the asymptotic expansion holds by construction of the heat kernel via the parametrix method, cf. [71, Chapter 3]. Finally, (C.45) and (C.46) follow from the Li–Yau inequality for weighted manifolds [75].

B.3 Semi-supervised learning

Theorem B.2.1 and Theorem B.2.5 combined together prove the *consistency* of the MBO scheme for data clustering: indeed the two Γ-convergence results prove that (local) minimizers of the energies E^h_{n,ϵ_n} converge to (local) minimizers of the weighted perimeter on the manifold if we let first $n \to +\infty$ and then $h \downarrow 0$. Of course, the only global minimizers for E^h_{n,ϵ_n} are partitions where all points are labeled in the same way, and thus the results may seem of little relevance. The full strength of Theorem B.2.1 and Theorem B.2.5 is seen in the context of *semi-supervised* learning. In semi-supervised learning one is given:

1. A dataset of n distinct points $D := \{x_1, ..., x_n\} \subset \mathbf{R}^d$.

2. A number of classes $P \in \mathbf{N}$ to split the data into.

3. A subset $\mathcal{O} \subset D$ of $L \ll n$ points and a function $u^0 : \mathcal{O} \to \{c_1, ..., c_P\}$ which assigns a label to every point in $\mathcal{O}$.

The task is to assign labels to all points in the dataset using both the known labels and the geometry of the dataset. The MBO scheme can be suitably modified to perform semi-supervised learning: The main point is to replace the heat operator in the first step of the algorithm by another differential operator with a fidelity term, see [62] for the details. This yields an algorithm which still has a minimizing movements interpretation, but the associated energy involves a different operator than the heat semigroup for the Laplacian. A different approach looks at changing the thresholding value in the thresholding step of MBO while leaving the differential operator in the diffusion step unchanged. This modified version of MBO has the advantage that the energy in the variational formulation is just the "standard" thresholding energy plus a linear term. The ideas behind the SSL MBO algorithm come from the corresponding MBO scheme for forced mean curvature flow (see [60] and [56]) and have already been adapted to data classification by Jacobs in [44]. We assume that we are given $D := \{x_1, ..., x_n\} \subset \mathbf{R}^d$ data points to be classified into P clusters. We assume that we are given a function $f : D \to \mathbf{R}^P$, the forcing term. As done in Section B.1, construct a similarity graph for the dataset. Then the MBO scheme for semi-supervised learning reads as follows.

Algorithm B.3.1 (SSL MBO)**.** *Let $h > 0$, let N the number of iterations to run. Let $\chi^0 : V \to \{0,1\}^P$ be a proposed clustering. To obtain a clustering $\chi^N : V \to \{0,1\}^P$ using the MBO scheme define inductively for $0 \le q < N$ a new clustering $\chi^{q+1} : V \to \{0,1\}^P$ starting from the clustering $\chi^q : V \to \{0,1\}^P$ by performing the following two steps:*

1. ***Diffusion**. For every $i \in \{1, ..., P\}$ define*

$$u_i = \sum_{j \neq i} \sigma_{ij} e^{-h\Delta} \chi_j^q$$

2. ***Thresholding**. Update, for every $1 \le i \le P$*

$$\left\{\chi_i^{q+1} = 1\right\} := \left\{u_i - \sqrt{h} f_i < u_j - \sqrt{h} f_j, \forall j \neq i\right\}.$$

The reason behind this approach to SSL is that the previous algorithm has a variational interpretation which adds just a linear term to the thresholding energy, namely we have the following result.

Lemma B.3.2. *Each iteration of the SSL MBO scheme decreases the energy*

$$F_h(u) = \frac{1}{\sqrt{h}} \sum_{i,j=1}^{P} \sigma_{ij} \langle u_i, e^{-h\Delta} u_j \rangle_{\mathcal{V}} - \sum_{i=1}^{P} \langle f_i, u_i \rangle_{\mathcal{V}}.$$

It is then natural to investigate the asymptotic behavior of these energies in the sense of the following theorems.

Theorem B.3.3. *Under the assumptions of Theorem B.2.1, if we additionally assume that $f^n : G_n \to \mathbf{R}^P$ are such that $f^n \to f$ in $TL^2(M)$, then it holds almost surely that for every $h > 0$ it holds that if v^n converge weakly to v in $TL^2(M)$ then*

$$\lim_{n \to +\infty} F_{n,\epsilon_n}^h(v^n) = \sqrt{\frac{C_1 C_2}{2}} F_{\frac{C_2 h}{2C_1}}(v),$$

where we set

$$F^h_{n,\epsilon_n}(v) = \frac{1}{\sqrt{h}} \sum_{i,j=1}^{P} \sigma_{ij} \langle v_i, e^{-h\Delta_{n,\epsilon_n}} v_j \rangle_{\mathcal{V}_{n,\epsilon_n}} - \sum_{i=1}^{P} \langle f_i^n, v_i \rangle_{\mathcal{V}_{n,\epsilon_n}},$$

$$F_h(u) = E_h(u) - \sqrt{\frac{2C_1}{C_2}} \sum_{i=1}^{P} \int_M f_i u_i \rho^2 d\,\mathrm{Vol}_M \quad u \in \mathcal{M},$$

where $\mathcal{M} := \left\{ u : M \to [0,1]^P \text{ measurable } : \sum_{i=1}^P u_i = 1 \text{ a.e.} \right\}$.

Theorem B.3.4. *Let M be a k-dimensional compact Riemannian submanifold of $\mathbf{R}^d$ weighted by a measure $\mu = \xi\,\mathrm{Vol}_M$ with $\xi > 0$, $\xi \in C^\infty(M)$. Let $f \in L^1(M)$. Then on $\mathcal{M}$,*

$$\Gamma(L^1(M)) - \lim_{h \downarrow 0} F_h = F,$$

where we define

$$F(u) = \begin{cases} \frac{1}{\sqrt{\pi}} \sum_{i,j=1}^P \sigma_{ij} |Du_i|_\xi(\Sigma_{ij}) - \sqrt{\frac{2C_1}{C_2}} \sum_{i=1}^P \int_M f_i u_i \rho^2 d\,\mathrm{Vol}_M & \text{if } u \in BV(M,\{0,1\})^P, \\ +\infty & \text{otherwise.} \end{cases}$$

Theorem B.3.3 and Theorem B.3.4 are easy consequences of Theorem B.2.1, Theorem B.2.5 and the stability of Γ-convergence with respect to continuous perturbations. Of course, these Theorems prove the consistency of SSL MBO once one can produce suitable forcing functions f^n which have a limit in $TL^2(M)$. In the following, let us for simplicity focus on the simple two-class setting, in which the cluster is $C = \left\{ u > \frac{1}{2} - \sqrt{h} f \right\}$. For a fixed $n \in \mathbf{N}$, if one is given a labeling function $u^0 : \mathcal{O} \to \{0,1\}$, a very intuitive choice of forcing would be

$$f = -\gamma(1 - 2u^0)\mathbf{1}_{\mathcal{O}}, \tag{B.9}$$

for some fixed constant $\gamma > 0$. Indeed, when $x \in \mathcal{O}$ and $u^0(x) = 0$, $f(x_0) = -\gamma$ so that the thresholding value for x gets higher and thus the updated set is more likely not to contain x. Observe in particular that if one chooses $\gamma > \frac{1}{\sqrt{h}}$ then the values on $\mathcal{O}$ are enforced. Similarly the case $u^0(x) = 1$ forces the updated sets to contain x. The problem with such a strategy is that this is ill-posed in the large data limit: indeed, to talk about SSL one usually assumes that

$$\lim_{n \to +\infty} \frac{|\mathcal{O}_n|}{n} = 0,$$

i.e., that the proportion of labeled points converges to zero. If we were now given labels $u^{0,n} : \mathcal{O}_n \to \{0,1\}$ then the corresponding forcings constructed according to formula (B.9) would converge to zero in $TL^1(M)$, thus the large data limit forgets the labels. To overcome this difficulty one has first to propagate the labels to the whole graph to produce a new forcing function. There are several strategies for doing this: for example, Jacobs uses a forcing fidelity term based on the graph geodesic distance, see [44]. Here we propose to construct the forcing function by means of Lipschitz learning.

Algorithm B.3.5 (SSL MBO with Lipschitz learning)**.** *Let $h > 0$, let N be the number of iterations to run. Let C be a proposed clustering. Let $u^0 : \mathcal{O} \to \{0,1\}$ given labels for a subset $\mathcal{O} \subset D$ of data points.*

1. ***Lipschitz learning - forcing construction***. *Use Lipschitz learning (eventually using a reweighted graph with self tuning weights) to propagate the given labels, i.e. find $u : D \to \mathbf{R}$ such that*

$$\begin{cases} \Delta^{(\infty)} u = 0 & \text{on } D \setminus \mathcal{O}, \\ u = u^0 & \text{on } \mathcal{O}. \end{cases}$$

   Then set $f = -\gamma(1-2u)$ for some given constant $\gamma > 0$. Here $\Delta^{(\infty)}$ denotes the infinity Laplacian on the graph.

2. ***SSL MBO***. *Perform N iterations of the SSL MBO Algorithm B.3.1 with initial clustering C and forcing f.*

The reason why Lipschitz learning is a good approach to generate the forcing is because it is a very well-posed algorithm in the large data limit. Indeed let us for simplicity consider the case when $M = \mathbb{T}^k$, the k-dimensional torus. Fix a set $\mathcal{O} \subset \{X_i\}_{i\in\mathbf{N}}$ and assume that one is given a labeling function $u^0 : \mathcal{O} \to \{0,1\}$. Denote by $\Delta_n^{(\infty)}$ the infinity Laplacian on the n-th graph, i.e. the operator which acts on $u : V_n \to \mathbf{R}$ as

$$\Delta_n^{(\infty)} u_i = \max_{1\le j\le n} w_{ij}^{(n,\epsilon_n)}(u_j - u_i) + \min_{1\le j\le n} w_{ij}^{(n,\epsilon_n)}(u_j - u_i).$$

Define $u^n : V_n \to \mathbf{R}$ as solutions of

$$\begin{cases} \Delta_n^{(\infty)} u^n = 0 & \text{on } V_n \setminus \mathcal{O}, \\ u^n = g & \text{on } \mathcal{O}. \end{cases}$$

Calder showed in [10] that u^n converges uniformly to $u \in C^{0,1}(M)$, the unique viscosity solution of the ∞-Laplace equation

$$\begin{cases} \Delta^{(\infty)} u := \sum_{l,m=1}^k \partial_l u \partial_{lm} u \partial_m u = 0 & \text{on } M \setminus \mathcal{O}, \\ u = g & \text{on } \mathcal{O}. \end{cases}$$

In particular, $f^n := -\gamma(1-2u^n)$ converges to $f := -\gamma(1-2u)$ in $TL^2(M)$ and the assumptions of Theorem B.3.3 are satisfied.

B.4 Discussion

B.4.1 Joint limit and monotonicity

We want to remark that it would be interesting to understand whether we can take the joint limit $n \to +\infty$ and $h \downarrow 0$, combing Theorem B.2.1 and Theorem B.2.5 to give that

$$\Gamma - \lim_{n\to+\infty} E_{n,\epsilon_n}^{h_n} = E, \tag{B.10}$$

where the thresholding energies $E^{h_n}_{n,\epsilon_n}$ are defined in (B.2) and E is defined in the statement of Theorem B.2.5. Here the Γ-limit has to be understood in the sense of $TL^1(M)$ convergence. At the present moment, we are not able to prove (B.10). However, let us sketch a possible approach for obtaining the Γ-lim inf inequality for (B.10). Assume that we knew that for fixed $n \in \mathbf{N}$ and for $\tilde{h} \geq h$

$$E^{\tilde{h}}_{n,\epsilon_n}(u) \leq E^{h}_{n,\epsilon_n}(u), \quad u \in \mathcal{M}_n. \tag{B.11}$$

Then the Γ-lim inf inequality in (B.10) would follow from Theorem B.2.1 and Theorem B.2.5. Indeed, assume that $u^n \in \mathcal{M}_n$ are such that $u^n \to u \in \mathcal{M}$ in $TL^1(M)$. Fix $m \in \mathbf{N}$. Since $h_n \downarrow 0$, we have that $h_n \ll h_m$ for n large enough. Thus by (B.11) and by Theorem B.2.1 we would get

$$E^{h_m}(u) = \liminf_{n\to+\infty} E^{h_m}_{n,\epsilon_n}(u^n) \leq \liminf_{n\to+\infty} E^{h_n}_{n,\epsilon_n}(u^n).$$

Now letting $m \to +\infty$ and using the consistency Theorem B.2.5 one would get

$$E(u) \leq \liminf_{n\to+\infty} E^{h_n}_{n,\epsilon_n}(u^n).$$

This means that a key ingredient for the joint limit is the monotonicity (B.11). Of course, also an approximate version of it would suffice. For example

$$E^{\tilde{h}}_{n,\epsilon_n}(u) \leq g(h)E^{h}_{n,\epsilon_n}(u) + f(\tilde{h})E^{h}_{n,\epsilon_n}(u) + z(\tilde{h})$$

where g, f, z are functions such that $\lim_{h\downarrow 0} g(h) = 1$, $\lim_{\tilde{h}\downarrow 0} f(\tilde{h}) = 0$ and $\lim_{\tilde{h}\downarrow 0} z(\tilde{h}) = 0$. The reason behind the hope for a monotonicity property for the discrete thresholding energies comes from the similar property which holds in the continuum in the Euclidean setting, see Lemma A.2 in [25]. Actually, an approximate version of this monotonicity is true also for the localized thresholding energies, see Theorem B.7.3 in the Appendix. By exploiting this result, using a suitable localization argument and the asymptotic expansion for the heat kernel (B.6) one can actually prove a similar monotonicity property for the thresholding energies on the manifold. Since the discrete thresholding energies are approximating the thresholding energy on the manifold, it is reasonable to believe that such a property holds true in some sense also at the discrete level. To support this idea, we run some numerical experiments. Quite surprisingly, it seems that the validity of this property is related to the rate $\frac{h}{\epsilon_n^2}$, in particular, it does not hold when $h \ll \epsilon_n^2$ and it seems to hold for $h \gg \epsilon_n^2$. Observe that the regime $h \gg \epsilon_n^2$ is the one which is relevant for applications, because for $h \ll \epsilon_n^2$ the MBO scheme is pinned, see [78]. It is not too difficult to show that the energies E^{h}_{n,ϵ_n} are actually *increasing* if $h \ll \epsilon_n^2$. A simple numerical experiment that we run is as follows: sample n data points from the uniform distribution on the unit sphere, see Figure B.1. Construct the similarity graph with weight functions $\eta(t) = e^{-t^2}$, randomly choose a $\{0,1\}$-valued function u which takes the value 1 on half of the data points, and then compute the thresholding energies $E^{h}_{n,\epsilon_n}(u)$ for $h \in \{2^{-5}\epsilon_n^2, ..., 2^4\epsilon_n^2\}$. The results are depicted in Figure B.2. We see that when $h \gg \epsilon^2$ the monotonicity seems to hold true: we experimented the same behavior also when using different distributions for the data points and different choices of functions u.

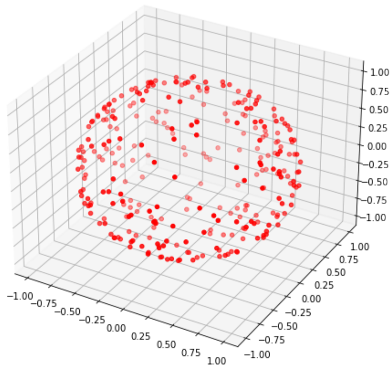

Figure B.1: Sample points from uniform distribution on the unit sphere.

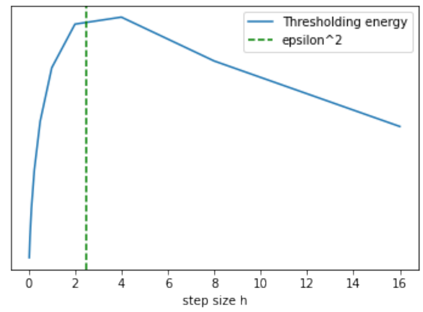

Figure B.2: The thresholding energy E^h_{n,ϵ_n} for different values of h.

B.4.2 Extensions

Here, we summarize the necessary changes to extend our results to other choices of graph Laplacians and to data dependent weights. With the notation and in the setting of Section B.2, given $\lambda \in \mathbf{R}$ we define

$$w_{ij}^{(n,\epsilon,\lambda)} = \frac{w_{ij}^{(n,\epsilon)}}{\left(d_i^{(n,\epsilon)} d_j^{(n,\epsilon)}\right)^{\lambda}}, \tag{B.12}$$

and

$$d_i^{(n,\epsilon,\lambda)} = \frac{1}{n}\sum_{j=1}^{n} w_{ij}^{(n,\epsilon,\lambda)}. \tag{B.13}$$

We can then consider the weighted graph $G_{n,\epsilon,\lambda}$ where the vertex set is given by $V_n := \{X_1, ..., X_n\}$ and the weights are given by (B.12). Let $W^{(n,\epsilon,\lambda)}$ be the matrix of weights, and $D^{(n,\epsilon,\lambda)}$ the diagonal matrix of the degrees (B.13). We then consider the following operators on $\mathcal{V}_{n,\epsilon,\lambda}$:

$$\Delta^{rw}_{n,\epsilon,\lambda} u = \frac{1}{\epsilon^2}\left(\mathbb{I} - \frac{1}{n}(D^{(n,\epsilon,\lambda)})^{-1} W^{(n,\epsilon,\lambda)}\right) u, \qquad u \in \mathcal{V}_{(n,\epsilon,\lambda)},$$
$$\Delta^{un}_{n,\epsilon,\lambda} u = \frac{1}{\epsilon^2}\left(D^{(n,\epsilon,\lambda)} - \frac{1}{n} W^{(n,\epsilon,\lambda)}\right) u, \qquad u \in \mathcal{V}_{(n,\epsilon,\lambda)}.$$

We define inner products on $\mathcal{V}_{(n,\epsilon,\lambda)}$ as

$$\langle u, v\rangle_{\mathcal{V}_{(n,\epsilon,\lambda)},rw} = \frac{1}{n}\sum_{i=1}^{n} d_i^{(n,\epsilon,\lambda)} u_i v_i, \qquad u, v \in \mathcal{V}_{(n,\epsilon,\lambda)},$$
$$\langle u, v\rangle_{\mathcal{V}_{(n,\epsilon,\lambda)},un} = \frac{1}{n}\sum_{i=1}^{n} u_i v_i, \qquad u, v \in \mathcal{V}_{(n,\epsilon,\lambda)}.$$

We also define inner products on $\mathcal{E}_{(n,\epsilon,\lambda)}$ as

$$\langle F,G\rangle_{\mathcal{E}_{(n,\epsilon,\lambda)}} = \frac{1}{2n^2}\sum_{i,j:\ w_{ij}^{(n,\epsilon,\lambda)}\neq 0} F_{ij}G_{ij}\frac{1}{w_{ij}^{(n,\epsilon,\lambda)}}, \quad F,G\in\mathcal{E}_{(n,\epsilon,\lambda)}.$$

We define the Dirichlet energies

$$E_{(n,\epsilon,\lambda)}(u) = \frac{1}{2}|\nabla u|^2_{\mathcal{E}_{(n,\epsilon,\lambda)}}, \quad u\in\mathcal{V}_{(n,\epsilon,\lambda)}. \tag{B.14}$$

Finally, we define the thresholding energies

$$E^h_{(n,\epsilon,\lambda,rw)}(u) = \frac{1}{\sqrt{h}}\sum_{i,j=1}^{P}\sigma_{ij}\langle u_i, e^{-h\Delta^{rw}_{(n,\epsilon,\lambda)}}u_j\rangle_{\mathcal{V}_{(n,\epsilon,\lambda,rw)}}, \quad u\in\mathcal{M}_n,$$
$$E^h_{(n,\epsilon,\lambda,un)}(u) = \frac{1}{\sqrt{h}}\sum_{i,j=1}^{P}\sigma_{ij}\langle u_i, e^{-h\Delta^{un}_{(n,\epsilon,\lambda)}}u_j\rangle_{\mathcal{V}_{(n,\epsilon,\lambda,un)}}, \quad u\in\mathcal{M}_n.$$

We then have the following results.

Theorem B.4.1. *Let $\lambda\in\mathbf{R}$. Under the assumptions of Theorem B.2.1 it holds almost surely that if $u^n\in\mathcal{V}_{(n,\epsilon,\lambda)}$ is a sequence of functions converging weakly to $u\in L^2(M)$ in TL^2, then for every $t>0$ we have*

$$\lim_{n\to+\infty} e^{-t\Delta^{rw}_{(n,\epsilon_n,\lambda)}}u^n = e^{-\frac{C_2}{2C_1}t\Delta_{\rho^s}}u \text{ strongly in } TL^2,$$
$$\lim_{n\to+\infty} e^{-t\Delta^{un}_{(n,\epsilon_n,\lambda)}}u^n = e^{-\frac{C_2}{2C_1^{2\lambda}}t\rho^{1-2\lambda}\Delta_{\rho^s}}u \text{ strongly in } TL^2,$$

where $s=2(1-\lambda)$.

Theorem B.4.2. *Under the assumptions of Theorem B.2.1, for every $\lambda\in\mathbf{R}$ it holds almost surely that for each fixed $h>0$, if v^n converges weakly to v in $TL^2(M)$,*

$$\lim_{n\to+\infty} E^h_{(n,\epsilon_n,\lambda,rw)}(v^n) = C_1^{\frac{1}{2}-2\lambda}\sqrt{\frac{C_2}{2}}E^{rw}_{\frac{C_2h}{2C_1},\lambda}(v),$$
$$\lim_{n\to+\infty} E^h_{(n,\epsilon_n,\lambda,un)}(v^n) = \sqrt{\frac{C_2}{2C_1^{2\lambda}}}E^{un}_{\frac{C_2}{2C_1^{2\lambda}},\lambda}(v),$$

where we define, for $v\in\mathcal{M}$,

$$E^{rw}_{h,\lambda}(v) := \frac{1}{\sqrt{h}}\sum_{i,j=1}^{P}\sigma_{ij}\int_M u_i e^{-h\Delta_{\rho^s}}u_j\rho^s d\,\mathrm{Vol}_M,$$
$$E^{un}_{h,\lambda}(v) := \frac{1}{\sqrt{h}}\sum_{i,j=1}^{P}\sigma_{ij}\int_M u_i e^{-h\rho^{1-2\lambda}\Delta_{\rho^s}}u_j\rho d\,\mathrm{Vol}_M,$$

with $s=2(1-\lambda)$.

The proofs of Theorem B.4.1 and Theorem B.4.2 are completely analogous to the proofs of Theorem B.2.2 and Theorem B.2.1 (which consider the random walk Laplacian with $\lambda = 0$). The only needed changes are:

1. Replace the use of Theorem B.5.14 with the convergence of the corresponding Laplacian (see [39, Theorem 30]).

2. Replace the use of Theorem B.5.16 with the analogous statement for the Dirichlet energies (B.14). It seems that this is not written down anywhere in the literature, but the proof of García Trillos and Slepčev [35] should be easily adapted to this setting.

Using Theorem B.4.2 one can clearly extend also the analogous statement for the semi-supervised MBO scheme.

B.5 Preliminaries

B.5.1 Weighted manifolds

Hereafter, $M = (M, g, \mu)$ will be a compact Riemannian manifold with $\partial M = \emptyset$. We will assume that $\mu = \xi \operatorname{Vol}_M$ for some $\xi \in C^\infty(M)$ such that $\xi > 0$.

For every $x \in M$, we denote by $\langle \cdot, \cdot \rangle_x$ the inner product on $T_x M$ induced by the metric g, i.e., for any $v, w \in T_x M$ we have $\langle v, w \rangle_x = g_x(v, w)$. Let $f : M \to \mathbf{R}$ be a smooth function. Then the gradient $\nabla f(x) \in T_x M$ is defined uniquely by the relation

$$\langle \nabla f(x), Y \rangle_x = d_x f(Y) \ \forall Y \in T_x M.$$

Let $\Gamma(TM)$ be the space of smooth vector fields on M. We can define the (weighted) divergence operator $\operatorname{div}_\xi : \Gamma(TM) \to C^\infty(M)$ by the requirement that for any $f \in C^\infty(M)$ and $Y \in \Gamma(TM)$

$$\int_M \langle \nabla f(x), Y(x) \rangle_x d\mu(x) = -\int_M f(x) \operatorname{div}_\xi Y(x) d\mu(x).$$

It is easy to check that the divergence can be expressed in local coordinates as

$$\operatorname{div}_\xi Y = \frac{1}{\xi \sqrt{\det(g)}} \sum_{i=1}^{k} \partial_i \left(\xi \sqrt{\det(g)} Y_i \right).$$

We also define the weighted Laplacian $\Delta_\xi : C^\infty(M) \to C^\infty(M)$ as $\Delta_\xi = -\operatorname{div}_\xi \circ \nabla$. A distribution on M is a continuous linear functional $T : C^\infty(M) \to \mathbf{R}$. We denote by $\mathcal{D}'(M)$ the space of distributions on M. We follow the terminology of [38] and say that a distributional vector field is a continuous linear functional $V : \Gamma(TM) \to \mathbf{R}$. If V is a distributional vector field, we define its divergence as the distribution $\operatorname{div}_\xi V \in \mathcal{D}'(M)$ such that $\operatorname{div}_\xi V(f) = -\langle V, \nabla f \rangle$. We define the Sobolev space

$$W^{1,2}(M) := \left\{ u \in L^2(M, \mu) : \ \nabla u \in L^2(TM, \mu) \right\},$$

which is a Hilbert space when endowed with the inner product

$$(u, v)_{W^{1,2}(M)} = (u, v)_{L^2(M,\mu)} + (\nabla u, \nabla v)_{L^2(M,\mu)}.$$

We also denote $W^{1,2}(M)$ by $H^1(M)$. We denote by $H^{-1}(M)$ its dual. If T is a distribution, we define $\Delta_\xi T \in \mathcal{D}'(M)$ by requiring $\Delta_\xi T(f) = T(\Delta_\xi(f))$. In particular if $u \in L^2(M,\mu)$, then $\Delta_\xi u \in \mathcal{D}'(M)$. Define

$$\mathcal{W}^{2,2}(M) = \left\{u \in W^{1,2}(M):\ \Delta_\xi u \in L^2(M,\mu)\right\}.$$

It is a standard result that Δ_ξ can be extended uniquely to a self-adjoint operator on $\mathcal{W}^{2,2}(M)$, see for instance [38, Theorem 4.6]. It can be shown that Δ_ξ is a nonnegative self-adjoint operator in $L^2(M)$ and $\operatorname{spec}(\Delta_\xi) \subset [0,+\infty)$. For $u \in L^2(M,\mu)$ we denote by $T(t)u = v(t,\cdot)$ the solution to the Cauchy problem

$$\begin{cases} \partial_t v = -\Delta_\xi v & \text{in } (0,+\infty)\times M, \\ v(0,x) = u(x) & \text{on } M. \end{cases}$$

More precisely, the map $t \in (0,+\infty) \mapsto T(t)u = v(t,\cdot) \in L^2(M)$ is characterized by the following properties:

- It is strongly differentiable in $L^2(M)$.
- For every $t > 0$ we have $T(t)u \in \operatorname{dom}(\Delta_\xi)$ and
$$\frac{dT(t)u}{dt} = -\Delta_\xi T(t)u.$$
- $T(t)u \to u$ in $L^2(M)$ as $t \downarrow 0$.

One way of constructing $T(t)$ is by means of the spectral resolution of Δ_ξ. I.e., one defines linear operators $T(t) : L^2(M) \to L^2(M)$ by

$$T(t) := \int_0^\infty e^{-t\gamma} dE_\gamma,$$

where E_γ is the spectral resolution of Δ_ξ. We refer to [38, Chapter 7] for the details. Furthermore, one can show that there exists a smooth map $p : (0,+\infty)\times M \times M \to \mathbf{R}$ such that for any $u \in L^2(M)$ and every $t > 0$

$$e^{-t\Delta_\xi}u(x) := T(t)u(x) = \int_M p(t,x,y)u(y)d\mu(y).$$

We call p the heat kernel for Δ_ξ.

Another more constructive way to prove the existence of the heat kernel is by the so-called *parametrix method*. This has the advantage of giving immediately the asymptotic expansion (B.6). However, the construction is technical and we think it is not worth sketching it here. The reader is referred to [71, Chapter 3], where this construction is carried out in detail for the case of constant density $\xi = 1$.

B.5.2 The MBO scheme on weighted manifolds

In this subsection, we recall the MBO scheme on weighted manifolds, which can be used to approximate the evolution by multiphase (weighted) mean curvature flow. Hereafter M is a k-dimensional closed Riemannian manifold endowed with a weight $\xi \in C^\infty(M)$, $\xi > 0$.

Algorithm B.5.1 (MBO scheme on manifolds)**.** *Let P be the number of phases, let $h > 0$ be a time-step size and let $\kappa > 0$ be a diffusion coefficient. Let $\chi^0 : M \to \{0,1\}^P$ be a partition of M into P phases. To obtain an approximation of the evolution of χ^0 by multiphase mean curvature flow define inductively a new partition $\chi^{n+1} : M \to \{0,1\}^P$ starting from $\chi^n : M \to \{0,1\}^P$ by performing the following steps:*

1. ***Diffusion***. *For every $m = 1, ..., P$ define*
$$u_m^n := \sum_{l \neq m} \sigma_{ml} e^{-\kappa h \Delta_\xi} \chi_l^n.$$

2. ***Thresholding***. *Define a new partition $\chi^{n+1} : V \to \{0,1\}^P$ by defining, for every $m = 1, ..., P$*
$$\left\{\chi_m^{n+1} = 1\right\} := \left\{x \in M : \ u_m^n(x) < \min_{l \neq m} u_l^n(x)\right\}.$$

We then have the following minimizing movements interpretation for the previous algorithm.

Lemma B.5.2 ([25])**.** *Assume that σ is negative semidefinite on $(1, ..., 1)^\perp$, which means that $v \cdot \sigma v \leq 0$ for all $v \in \mathbf{R}^d$ with $v \cdot (1, ..., 1)^\perp = 0$. Given a step-size $h > 0$ and a diffusion parameter $\kappa > 0$, to obtain the new partition $\chi^{n+1} : M \to \{0,1\}^P$ starting from $\chi^n : M \to \{0,1\}^P$ one can define*
$$\chi^{n+1} \in \operatorname{argmin}_{u \in \mathcal{M}} \left\{ \sqrt{\kappa} E_{\kappa h}(u) - \frac{1}{\sqrt{h}} \sum_{i \neq j} \sigma_{ij} \int_M (u_i - \chi_i^n) e^{-\kappa h \Delta_\xi} (u_j - \chi_j^n) \xi d \operatorname{Vol}_M \right\}.$$

B.5.3 BV functions on weighted manifolds

We begin this section introducing the total variation $|Du|_\xi$ of a function $u \in L^1(M)$. We define
$$|Du|_\xi(M) = \sup \left\{ \int_M u \operatorname{div}_\xi Y d\mu : \ Y \in \Gamma(TM), |Y| \leq 1 \right\}.$$
We say that $u \in L^1(M)$ is in $BV(M)$ provided $|Du|_\xi(M) < +\infty$. One can prove the following result.

Theorem B.5.3. *Let $u \in BV(M)$, then there exist a Radon measure $|Du|_\xi \in \mathcal{M}_+(M)$ and a $|Du|_\xi$-measurable vector field σ_u such that $|\sigma_u| = 1$ $|Du|_\xi$-almost everywhere and such that*
$$\int_M u \operatorname{div}_\xi X d\mu = - \int_M \langle \sigma_u, X \rangle d|Du|_\xi \ \forall X \in \Gamma(TM). \tag{B.15}$$

The proof of the theorem is an adaptation of the classical Riesz representation theorem: first one works locally on an open set $V \subset M$ using an orthonormal frame $\{E_1, ..., E_k\}$. Following the same lines of the proof of the Riesz representation theorem one can check that there exist a Radon measure $\gamma_V \in \mathcal{M}_+(V)$ and a γ_V measurable vector field σ_u^V such that $|\sigma_u^V| = 1$ γ_V-a.e. and such that (B.15) holds true for all

$X \in \Gamma(TV)$. Then one checks that if V_1, V_2 are two open subsets of M, the construction is consistent on $V_1 \cap V_2$. One can then take a covering $\{V_i\}_i$ and apply the construction on each element of the covering. Taking a partition of unity $\{\rho_i\}$ subordinate to the covering one defines

- $|Du|_\xi(W) := \sum_i \int_{W \cap V_i} \rho_i d\gamma_{V_i}$.
- $\sigma|_{V_i} := \sigma_{V_i}$.

A subset $E \subset M$ is said to be of *finite perimeter* if $\chi_E \in BV(M)$. If E is a set of finite perimeter, we denote by $\mathrm{Per}_\xi(E) := |D\chi_E|_\xi(M)$ its perimeter. We will make use of the following elementary lemma, which follows easily from Theorem B.5.3.

Lemma B.5.4. *Let $u \in L^1(M)$. Then $u \in BV(M)$ if and only if for every chart (V, ψ) the map $u \circ \psi^{-1}$ is in $BV(\psi(V))$. In that case we have that for any chart (V, ψ)*

$$\psi_\# |Du|_\xi = \gamma |D(u \circ \psi^{-1})|,$$

where $\gamma = \xi \circ \psi^{-1} \sqrt{\det g^{ij}}$.

Remark B.5.5. For a set of finite perimeter $E \subset M$, we define its reduced boundary $\partial^*_M E$ as follows:

$$\partial^*_M E = \{x \in M : \exists (V, \psi) \text{ chart of } M \text{ s.t. } \psi(x) \in \partial^* \psi(E)\},$$

where $\partial^* \psi(E)$ is the reduced boundary of the set $\psi(E)$ in the usual Euclidean setting. From now on, we will also denote by $\partial^* E$ the set $\partial^*_M E$. One can check that:

- The definition is well posed.
- $|D\chi_E|_\xi$ is concentrated on $\partial^* E$. In particular $|D\chi_E|_\xi$-a.e. point x is in the reduced boundary of E.
- If $x \in \partial^* E$, then in normal coordinates (V, ψ) centered around x we have that $\sigma_{\chi_E}(x) = \nu_{\psi(E)}(\underline{o})$, where $\nu_{\psi(E)}$ is the measure theoretic inner unit normal for $\psi(E)$ and $\underline{o}$ are the coordinates for the center of the chart x.
- If $E, F \subset M$ are sets of finite perimeter, then it holds that for $|D\chi_F|_\xi$-a.e. point $x \in \partial^* E \cap \partial^* F$ we have $\sigma_E(x) = \langle \sigma_E(x), \sigma_F(x) \rangle_x \sigma_F(x)$.

We also record the following elementary lemma, which can be proved by using Lemma B.5.4 and the analogous statement in the Euclidean setting.

Lemma B.5.6. *Let u^n be a sequence of functions in $BV(M)$ such that*

$$\sup_{n \in \mathbf{N}} \int_M |Du^n|_\xi < +\infty.$$

Then $\{u^n\}$ is precompact in $L^1(M)$ and every limit point is in $BV(M)$.

B.5.4 Transportation distance

Here we recall the definition of TL^p-convergence introduced in [36] and we introduce the notion of weak TL^p-convergence. Let (M, g) be a k-dimensional compact Riemannian manifold. For a fixed $1 \leq p < \infty$ let $\mu, \nu \in \mathcal{P}(M)$, $u \in L^p(\mu), v \in L^p(\nu)$, we set

$$d_{TL^p}((\mu, u), (\nu, v)) := \inf_{\pi \in \Gamma(\mu,\nu)} \left\{ \left(\int_{M\times M} d_M^p(x, y) + |u(x) - u(y)|^p d\pi \right)^{1/p} \right\},$$

where the infimum is taken over the space of couplings between μ and ν, which we denote by $\Gamma(\mu, \nu)$. For $p = \infty$, $\mu, \nu \in \mathcal{P}(M)$, $u \in L^p(\mu), v \in L^p(\nu)$ we set

$$d_{TL^\infty}((\mu, u), (\nu, v)) := \inf_{\pi \in \Gamma(\mu,\nu)} \left\{ \operatorname{esssup}_{x,y\in M} \left(d_M(x, y) + |u(x) - u(y)| \right) \right\}.$$

We call d_{TL^p} the TL^p-metric. It can be shown that d_{TL^p} is a metric on

$$\mathcal{L}^p := \{(\mu, u) : \ \mu \in \mathcal{P}(M), \ u \in L^p(\mu)\},$$

this is done in[36] for the Euclidean case, the case of a compact manifold is analogous. Let $\{\pi_n\}_n \subset \Gamma(\mu, \nu)$ be a sequence of transport plans between μ and ν, we say that the these are *p-stagnating* if

$$\lim_{n\to+\infty} \int_{M\times M} d_M^p(x, y) d\pi_n = 0.$$

Transport maps T_n between $\mu, \nu \in \mathcal{P}(M)$ are said to be *p-stagnating* if the corresponding transport plans $(Id \times T_n)_\# \mu$ are p-stagnating. The following propositions are straightforward generalizations of [36, Proposition 3.12] and [35, Proposition 2.6].

Proposition B.5.7. *Let $(\mu^n, u^n), (\mu, u) \in \mathcal{L}^p$, $n \in \mathbf{N}$, $1 \leq p < +\infty$. Assume that μ is absolutely continuous with respect to* Vol_M. *Then the following are equivalent:*

(i) $(\mu^n, u^n) \to (\mu, u)$ *in the TL^p sense.*

(ii) For every sequence of p-stagnating transport maps T_n we have

$$\lim_{n\to+\infty} \int_M |u^n(T_n(x)) - u(x)|^p d\mu(x) = 0.$$

(iii) There exists a sequence of p-stagnating transport maps T_n such that

$$\lim_{n\to+\infty} \int_M |u^n(T_n(x)) - u(x)|^p d\mu(x) = 0.$$

Proposition B.5.8. *Suppose that $(\mu^n, u^n) \to (\mu, u)$ in $TL^2(M)$ and $(\mu^n, v^n) \to (\mu, v)$ in $TL^2(M)$. Then*

$$\lim_{n\to\infty} \langle u^n, v^n \rangle_{L^2(\mu^n)} = \langle u, v \rangle_{L^2(\mu)}.$$

We will also make use of the following result, which can easily be derived from [33, Theorem 2].

Theorem B.5.9. *Let M be a k-dimensional compact Riemannian submanifold of $\mathbf{R}^d$. Let $\rho \in C^\infty(M)$, $\rho > 0$ such that $\nu := \rho \operatorname{Vol}_M \in \mathcal{P}(M)$. Let $\{X_i\}_{i\in\mathbf{N}}$ be iid random points in M distributed according to ν and let $\nu^n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ be the associated empirical measures. Then there is a constant $C > 0$ such that almost surely there exist transport maps T_n such that $(T_n)_\# \nu = \nu^n$ and*

$$\begin{cases} \limsup_{n\to+\infty} \frac{n^{1/2}\sup_{x\in M} d_M(x,T_n(x))}{log^{3/4}(n)} \leq C if\ k=2 \\ \limsup_{n\to+\infty} \frac{n^{1/k}\sup_{x\in M} d_M(x,T_n(x))}{log^{1/k}(n)} \leq C if\ k\geq 3 \end{cases} \tag{B.16}$$

The correct notion of convergence for obtaining Theorem B.2.1 is weak TL^2-convergence, because this is the topology in which we get Γ-compactness. More generally, let us introduce the notion of *weak TL^p-convergence*.

Definition B.5.10. Let μ be a probability measure on M which is absolutely continuous with respect to the volume measure Vol_M, and let $u \in L^p(\mu)$. A sequence $(\mu^n, u^n) \in \mathcal{L}^p$ is said to *converge weakly to* (μ, u) *in* TL^p if there exists a sequence of q-stagnating transport maps T_n between μ and μ^n such that the functions $u^n \circ T_n$ converge weakly to u in $L^p(\mu)$. Here, $q = \frac{p}{p-1}$ is the conjugate exponent for p.

We record the following useful result, which says that the previous definition is independent of the sequence of q-stagnating transport maps.

Proposition B.5.11. *Let $1 < p < +\infty$ and $q = \frac{p}{p-1}$. Let μ be a probability measure on M which is absolutely continuous with respect to the volume measure and let $u \in L^p(\mu)$. Assume that $(u^n, \mu^n) \in \mathcal{L}^p$ is a sequence converging weakly to (u, μ) in TL^p. Then for every sequence S_n of q-stagnating transport maps between μ and μ^n the functions $u^n \circ S_n$ converge weakly to u in $L^p(\mu)$.*

Proof. We let T_n the sequence of q-stagnating transport maps as in Definition B.5.10. Let S_n be an arbitrary sequence of q-stagnating transport maps between μ and μ^n. Observe that $\|u^n \circ S_n\|_{L^p(\mu)} = \|u^n\|_{L^p(\mu_n)} = \|u^n \circ T_n\|_{L^p(\mu)}$. In particular, the sequence $u^n \circ S_n$ is bounded in $L^p(\mu)$ and thus, up to extracting a subsequence, we may assume that it converges weakly to a limit $v \in L^p(\mu)$. We need to show that $v = u$. To do so, pick $\varphi \in C^\infty(M)$ and observe that by Hölder's inequality

$$\begin{aligned} &\left| \int_M u^n \circ T_n \varphi d\mu - \int_M u^n \circ T_n \varphi_{|V_n} \circ T_n d\mu \right| \\ &\leq \left(\int_M |u^n \circ T_n|^p d\mu \right)^{1/p} \left(\int_M (\operatorname{Lip}\varphi)^q d_M(x, T_n(x))^q d\mu \right)^{1/q}, \end{aligned}$$

a similar estimate holds true with T_n replaced by S_n. This clearly implies that

$$\begin{aligned} &\left| \int_M u^n \circ T_n \varphi d\mu - \int_M u^n \circ S_n \varphi d\mu \right| \\ &\leq C \left(\int_M d_M(x, T_n(x))^q d\mu \right)^{1/q} + C \left(\int_M d_M(x, S_n(x))^q d\mu \right)^{1/q} \end{aligned}$$

The right hand side converges to zero as $n \to +\infty$ because the sequences of transport maps are q-stagnating. □

Finally, we have the following natural improvement of Proposition B.5.8.

Proposition B.5.12. *Let $1 \leq p < +\infty$. Let μ be a probability measure on M which is absolutely continuous with respect to the volume element. Assume that $(u^n, \mu^n) \in \mathcal{L}^p$ converges weakly in TL^p to (u, μ) and that (v^n, μ^n) is a sequence in $\mathcal{L}^q$ converging strongly in TL^q to (v, μ), where q is the conjugate exponent of p. Then*

$$\lim_{n \to +\infty} \int_M u^n v^n d\mu^n = \int_M uv d\mu.$$

B.5.5 Asymptotics for the graph Laplacian and for the degrees

Here we recall the results about the convergence of the graph Laplacian contained in [19] and [39]. To do so, we need to introduce some definitions. We write $\eta_\epsilon(t) := \frac{1}{\epsilon^k}\eta(\frac{t}{\epsilon})$, and we introduce the function $k_\epsilon : M \times M \to \mathbf{R}$ defined as $k_\epsilon(x, x) = 0, \ x \in M$ and

$$k_\epsilon(x, y) = \eta_\epsilon(|x - y|_d),$$

where $|\cdot|_d$ denotes the standard Euclidean norm in the ambient space $\mathbf{R}^d$. Observe that

$$\lim_{\epsilon \downarrow 0} \int_M k_\epsilon(x, y) d\nu(y) = C_1 \rho(x) \quad \text{uniformly in } x \in M. \tag{B.17}$$

Define also

$$d^{(n,\epsilon)}(x) = \frac{1}{n} \sum_{i=1}^{n} k_\epsilon(x, X_i), \ x \in M,$$

so that the weights and the degrees on the graph G_{n,ϵ_n} can be expressed via

$$w_{ij}^{(n,\epsilon)} = k_\epsilon(X_i, X_j), \ d_i^{(n,\epsilon)} = d^{(n,\epsilon)}(X_i).$$

In this way the random walk graph Laplacian may be extended to an operator acting on functions $f \in C^\infty(M)$ as

$$\Delta_{n,\epsilon} f(x) = \frac{1}{\epsilon^2} \left(f(x) - \sum_{j=1}^{n} \frac{k_\epsilon(x, X_j) f(X_j)}{n d^{(n,\epsilon)}(x)} \right).$$

Finally, for any $\epsilon > 0$ and any $f \in C^\infty(M)$ we define

$$\Delta_\epsilon f(x) = \frac{1}{\epsilon^2} \left(f(x) - \frac{\int_M k_\epsilon(x, y) f(y) d\nu(y)}{\int_M k_\epsilon(x, y) d\nu(y)} \right).$$

The following theorem is contained in Coifman and Lafon [19].

Theorem B.5.13. *Let the assumptions on M in Theorem B.2.1 be in place. For $K \in \mathbf{R}$ define*

$$E_K = \{f \in C^\infty(M) : \ \|f\|_{C^3} \leq K\}.$$

Then uniformly in x and uniformly on E_K we have

$$\Delta_\epsilon f(x) = \frac{C_2}{2C_1} \Delta_{\rho^2} f(x) + o(\epsilon).$$

One can then apply the previous theorem to obtain the following slight modification of Theorem 28 in [39].

Theorem B.5.14. *Let the assumptions of Theorem B.2.1 be satisfied. Then with probability one, for all* $f \in C^\infty(M)$ *we have that* $\Delta_{n,\epsilon_n} f \to \frac{C_2}{2C_1}\Delta_{\rho^2} f$ *in* TL^2.

In the following, we need also the following simple result about the convergence of the degrees.

Lemma B.5.15. *Let the assumptions of Theorem B.2.1 be satisfied. Then it holds with probability one that if* T_n *is a sequence of transport maps such that*

$$\lim_{n\to+\infty} \sup_{x\in M} d_M(x, T_n(x)) = 0,$$

then

$$\lim_{n\to+\infty} \|d^{(n,\epsilon_n)} \circ T_n - C_1\rho\|_{L^\infty(\nu)} = 0. \tag{B.18}$$

Moreover, $\lim_{n\to+\infty} \|\frac{1}{(d^{(n,\epsilon_n)}\circ T_n)^{1/2}} - \frac{1}{(C_1\rho)^{1/2}}\|_{L^\infty(\nu)} = 0.$

Proof. The second assertion follows from (B.18) and the fact that $\rho > 0$ on the compact manifold M. In the following we write d^n for $d^{(n,\epsilon_n)}$. To prove (B.18) we first observe that

$$\begin{aligned}\|d^n \circ T_n - C_1\rho\|_{L^\infty(\nu)} \le \|d^n \circ T_n - C_1\rho\circ T_n\|_{L^\infty(\nu)} \\ + \operatorname{Lip}(\rho)\sup_{x\in M} d_M(x, T_n(x)).\end{aligned}$$

Thus we only need to prove that $\|d^n \circ T_n - C_1\rho\circ T_n\|_{L^\infty(\nu)}$ converges to zero. To this aim, observe that we may write

$$\|d^n \circ T_n - C_1\rho\circ T_n\|_{L^\infty(\nu)} = \max_{i=1,\dots,n} |d^n(X_i) - C_1\rho(X_i)|,$$

and that by the fact that X_i's are identically distributed, if $\gamma > 0$, then

$$\begin{aligned}\mathbb{P}\Big(\max_{i=1,\dots,n} |d^n(X_i) - C_1\rho(X_i)| \ge \gamma\Big) &\le n\mathbb{P}\Big(|d^n(X_1) - C_1\rho(X_1)| \ge \gamma\Big) \\ &= n\int_M \mathbb{P}\Big(|d^n(x) - C_1\rho(x)| \ge \gamma\Big) d\nu(x).\end{aligned}$$

Fix $x \in M$,then

$$\begin{aligned}\mathbb{P}\Big(|d^n(x) - C_1\rho(x)| \ge \gamma\Big) \le\ & \mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^n k_{\epsilon_n}(x, X_j) - \int_M k_{\epsilon_n}(x,y)d\nu(y)\right| \ge \frac{\gamma}{2}\right) \\ &+ \mathbb{P}\left(\left|\int_M k_{\epsilon_n}(x,y)d\nu(y) - C_1\rho(x)\right| \ge \frac{\gamma}{2}\right).\end{aligned}$$

The second term on the right hand side is zero for n sufficiently large because of (B.17). Thus for n large enough depending on γ,

$$\mathbb{P}\Big(|d^n(x) - C_1\rho(x)| \geq \gamma\Big) \leq \mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} k_{\epsilon_n}(x, X_j) - \int_M k_{\epsilon_n}(x, y)d\nu(y)\right| \geq \frac{\gamma}{2}\right).$$

We now proceed at estimating the right hand side. Observe that $Y_j := k_{\epsilon_n}(x, X_j)$ are iid random variables with

$$\mathbb{E}[Y_j] = \int_M k_{\epsilon_n}(x, y)d\nu(y), \ |Y_j| \leq \frac{\|\eta\|_\infty}{\epsilon_n^k}, \ \mathrm{Var}(Y_j) \leq C_1 c_0 \frac{\|\eta\|_\infty}{\epsilon_n^k}.$$

Thus we can apply Bernstein's inequality to get the following bound

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} k_{\epsilon_n}(x, X_j) - \int_M k_{\epsilon_n}(x, y)d\nu(y)\right| \geq \frac{\gamma}{2}\right)$$
$$\leq 2\exp\left(\frac{-n\frac{\gamma^2}{4}\epsilon_n^k}{2\|\eta\|_\infty C_1 c_0 + \frac{2}{3}\|\eta\|_\infty \frac{\gamma}{2}}\right)$$

Putting things together and summing over n we have that, for some $n(\gamma) \in \mathbf{N}$ depending on γ

$$\sum_{n\in\mathbf{N}} \mathbb{P}\Big(\|d^n \circ T_n - C_1\rho \circ T_n\|_{L^\infty(\nu)} \geq \gamma\Big)$$
$$\leq n(\gamma) + \sum_{n=n(\gamma)+1} 2n\exp\left(\frac{-n\frac{\gamma^2}{4}\epsilon_n^k}{2\|\eta\|_\infty C_1 c_0 + \frac{2}{3}\|\eta\|_\infty \frac{\gamma}{2}}\right).$$

The latter sum is finite if $\frac{n\epsilon_n^k}{\log(n)} \to +\infty$. We conclude by Borel-Cantelli's lemma that almost surely $\|d^n \circ T_n - C_1\rho \circ T_n\|_{L^\infty(\nu)} \to 0$, which concludes the proof. □

B.5.6 Γ-convergence of the Dirichlet energies

In the setting of Section B.2, we define for each random graph G_{n,ϵ_n} the Dirichlet energy functional E_n as:

$$E_n(u) = \frac{1}{2}|\nabla u|^2_{\mathcal{E}_{n,\epsilon_n}}, \quad u \in \mathcal{V}_n. \tag{B.19}$$

In the proof of Theorem B.2.2 we will need the following result about the Γ-convergence of the Dirichlet energies defined on the graphs to the Dirichlet energy on the manifold.

Theorem B.5.16. *Let M be a k-dimensional compact Riemannian manifold embedded in $\mathbf{R}^d$, $k \geq 2$. Let $\rho > 0$ be a smooth function on M such that $\nu = \rho\,\mathrm{Vol}_M \in \mathcal{P}(M)$. If $k \geq 2$ and $\frac{\epsilon_n^k n}{\log(n)} \to +\infty$ as $n \to +\infty$ then*

$$E_n \xrightarrow{\Gamma - TL^2} \frac{C_2}{2}E,$$

where the energy E is defined on $L^2(M)$ as

$$E(u) = \begin{cases} \frac{1}{2}\int_M |\nabla u|^2 \rho^2 d\,\mathrm{Vol}_M & \text{if } u \in H^1(M) \\ +\infty & \text{otherwise.} \end{cases}$$

Moreover if $u \in C^\infty(M)$, then $\limsup_{n\to+\infty} E_n(u) \le \frac{C_2}{2}E(u)$. Finally, we have the following compactness property: if u^n are such that

$$\sup_{n\in\mathbf{N}} E_n(u^n) < +\infty, \ \sup_{n\in\mathbf{N}} \|u^n\|_{L^2(M)} < +\infty$$

then the sequence u^n is precompact in TL^2.

The proof of Theorem 8 is a straightforward adaptation of the argument given in Theorem 1.4 in [35] for the flat case. For the sake of completeness, we include a proof in the Appendix.

B.5.7 The optimal energy dissipation inequality

Let us recall the following notion of weak solution of the heat equation on the weighted manifold M.

Definition B.5.17. Let $(M, g, \mu = \xi\,\mathrm{Vol}_M)$ be a weighted k-dimensional compact Riemannian submanifold of $\mathbf{R}^d$, with $\xi > 0$ smooth. Let $u^0 \in L^2(\mu)$ and $c > 0$. A weak solution for the diffusion equation

$$\begin{cases} \partial_t u = -c\Delta_\xi u \\ u(x,0) = u^0 \end{cases} \tag{B.20}$$

is a function $u \in L^2_{loc}([0,+\infty), H^1(M))$ such that $u' \in L^2_{loc}([0,+\infty), H^{-1}(M))$ for which

1. $u(0) = u^0$,
2. $(\partial_t u, \xi w)_{H^1;H^{-1}} + c\int_M g_x(\nabla u, \nabla w)\,\xi d\,\mathrm{Vol}_M = 0$ for a.e. t and all $w \in H^1(M)$.

The following lemma is a well-known result, which says that the equation (B.20) is completely characterized by the energy dissipation inequality (B.21).

Lemma B.5.18. *Let $u \in L^2_{loc}([0,+\infty), H^1(M))$ be such that $u' \in L^2_{loc}([0,+\infty), L^2(M))$. Let $u^0 \in C^\infty(M)$. Then u is a weak solution of* (B.20) *if and only if $u(0) = u^0$ and u satisfies the optimal energy dissipation inequality for a.e. $t \in [0,+\infty)$, i.e.,*

$$cE[u(t)] + \frac{1}{2}\int_0^t \int_M c^2|\Delta_\xi u|^2 \xi d\,\mathrm{Vol}_M\,ds + \frac{1}{2}\int_0^t \int_M |u'|^2 \xi d\,\mathrm{Vol}_M\,ds \le cE[u^0] \tag{B.21}$$

where we define, for $v \in H^1(M)$,

$$E[v] = \frac{1}{2}\int_M |\nabla v|^2 \xi d\,\mathrm{Vol}_M\,.$$

Proof. We first observe that whenever $u \in L^2_{loc}([0,+\infty), H^1(M))$ is a function such that $u' \in L^2_{loc}([0,+\infty), L^2(M))$ and $\Delta_\xi u \in L^2_{loc}([0,+\infty), L^2(M))$ we have

$$\frac{d}{dt} cE[u(t)] = c \int_M \Delta_\xi u \partial_t u \; \xi d \operatorname{Vol}_M . \tag{B.22}$$

Now, assume first that u is a weak solution of (B.20). Then by parabolic regularity, u is smooth and (B.22) is thus true. Using the equation for $\partial_t u$ we obtain that

$$\begin{aligned} & c \int_M \Delta_\xi u \partial_t u \xi d \operatorname{Vol}_M \\ & = -\frac{c^2}{2} \int_M |\Delta_\xi u|^2 \xi d \operatorname{Vol}_M - \frac{1}{2} \int_M |\partial_t u|^2 \xi d \operatorname{Vol}_M . \end{aligned}$$

Thus, integrating (B.22) in time we get the required inequality (actually, equality).

Conversely, assume that (B.21) is satisfied. Then we clearly infer that $\Delta_\xi u \in L^2([0,+\infty), L^2(M))$, we can use (B.22) in (B.21) to get, after completing the square,

$$\frac{1}{2} \int_0^t \int_M (c\Delta_\xi u + u')^2 \, \xi d \operatorname{Vol}_M ds \leq 0.$$

which forces $\partial_t u = -c\Delta_\xi u$, thus u is a weak solution of (B.20). □

In a similar way, one can prove that solutions of the heat equation on a graph also satisfy an energy dissipation inequality. Namely, we have the following result.

Lemma B.5.19. *Let $G_{n,\epsilon}$ be a graph as constructed in Section B.1. Let $u^0 \in \mathcal{V}_{n,\epsilon}$. Let $v(x,t) = e^{-t\Delta_{n,\epsilon}} u^0(x)$. Then for all $t \in [0,+\infty)$ the optimal energy dissipation inequality is satisfied, i.e.*

$$E_n[v(t)] + \frac{1}{2} \int_0^t |\Delta_{n,\epsilon} v(s)|^2_{\mathcal{V}_{n,\epsilon}} ds + \frac{1}{2} \int_0^t |\frac{d}{ds} v(s)|^2_{\mathcal{V}_{n,\epsilon}} ds \leq E_n[u^0],$$

where E_n is the Dirichlet energy defined in (B.19) *with ϵ_n replaced by ϵ.*

B.6 Proofs

B.6.1 Convergence of the heat operators

Proof of Theorem B.2.2. We first prove the following result: Assume that $u^0 \in C^\infty(M)$, then for every $t > 0$ we have

$$e^{-t\Delta_{n,\epsilon_n}} (u^0|_{V_n}) \to e^{-\frac{C_2}{2C_1} t \Delta_{\rho^2}} u^0 \text{ in } TL^2. \tag{B.23}$$

To this aim, observe that it suffices to prove the result for the case $C_1 = 1$; the general case follows by rescaling the weight functions. Define, for $n \in \mathbf{N}, t \geq 0$ and $x \in \mathcal{V}_n(\omega)$,

$$v^n(\omega, x, t) := \left(e^{-t\Delta_{n,\epsilon_n}} u^0|_{V_n(\omega)}\right)(x).$$

Here ω is a sample point from some probability space $(\Omega, \mathbb{P})$ on which the random variables $\{X_i\}_{i\in\mathbf{N}}$ are defined. We want to prove that $\mathbb{P}$-a.s. for every sequence of

2-stagnating transport maps T_n from ν to $\nu^n(\omega) := \frac{1}{n}\sum_{i=1}^n \delta_{X_i(\omega)}$ we have that, for any $t > 0$

$$\lim_{n\to+\infty} \int_M |v^n(t, T_n(x)) - u(t,x)|^2 d\nu = 0, \tag{B.24}$$

where $u(t,\cdot) = e^{-\frac{C_2}{2}t\Delta_{\rho^2}}u$. To this aim, pick $\omega \in \Omega$ such that the following conditions are satisfied:

(i) There exists a sequence of 2-stagnating transport maps T_n such that (C.53) is satisfied.

(ii) $E_n \xrightarrow{\Gamma-TL^2} \frac{C_2}{2}E$.

(iii) $\Delta_{n,\epsilon_n} f \xrightarrow{TL^2} \frac{C_2}{2C_1}\Delta_{\rho^2} f$ for every $f \in C^\infty(M)$.

(iv) $\|d^{(n,\epsilon_n)} \circ T_n - C_1\rho\|_{L^\infty(\nu)} \to 0$.

Observe that by Theorem C.6.6, Theorem B.5.16, Theorem B.5.14 and Lemma B.5.15 these conditions hold for $\mathbb{P}$-a.e. $\omega \in \Omega$. Thus if we prove (B.24) for such an ω we are done. From now on, we will assume ω to be fixed, so we drop this variable for ease of notation. Recalling Proposition B.5.7, we just need to show (B.24) for the sequence of transport maps in (i). Define $\tilde{v}^n(t,x) := v^n(t, T_n(x))$, for $t \in [0,+\infty), x \in M$. We also write d^n for $d^{(n,\epsilon_n)}$. By condition (iv) and assuming that n is sufficiently large we have that for all $x \in V_n$,

$$\frac{C_1}{2c} \le d^n(x) \le 2C_1 c.$$

Here $c > 0$ is a constant such that

$$\frac{1}{c} \le \rho \le c \text{ on } M.$$

In particular we have that there exists a constant $C > 0$ such that for any $w \in \mathcal{V}_{n,\epsilon_n}$, if $\tilde{w} := w \circ T_n$

$$\frac{1}{C}|w|_{\mathcal{V}_{n,\epsilon_n}} \le \|\tilde{w}\|_{L^2(\nu)} \le C|w|_{\mathcal{V}_{n,\epsilon_n}}. \tag{B.25}$$

Step 1. We claim that given $T > 0$ there exists a constant $C_T < \infty$ for which

$$\sup_{n\in\mathbf{N}} \|\tilde{v}^n\|_{L^\infty([0,T],L^2(M))} \le C_T, \tag{B.26}$$

$$\sup_{n\in\mathbf{N}} \|\frac{d\tilde{v}^n}{dt}\|_{L^\infty([0,T],L^2(M))} \le C_T. \tag{B.27}$$

We start by proving (B.26). This is easy: Using the equation we obtain

$$\begin{aligned}\frac{d}{dt}\frac{1}{2}|v^n(t)|^2_{\mathcal{V}_{n,\epsilon_n}} &= -\langle v^n(t), \Delta_{n,\epsilon_n} v^n(t)\rangle_{\mathcal{V}_{n,\epsilon_n}} \\ &= -\langle \nabla_{n,\epsilon_n} v^n(t), \nabla_{n,\epsilon_n} v^n(t)\rangle_{\mathcal{V}_{n,\epsilon_n}} \le 0.\end{aligned}$$

Thus after integrating in time and recalling (B.25) we easily obtain (B.26), once we recall that $u^0 \in C^\infty(M)$.
To show (B.27) we notice that $q^n(x,t) := \frac{d}{dt}v^n(x,t)$ is the unique solution to

$$\frac{d}{dt}q^n = -e^{-t\Delta_{n,\epsilon_n}}(\Delta_{n,\epsilon_n}u^0|_{G_n})$$

In particular, arguing as in the proof of (B.26), we get an L^∞-bound on the L^2-norm of q^n. Namely, for any $T>0$ and a possibly different constant C_T

$$\sup_{n\in\mathbf{N}} \|\tilde{q}^n\|_{L^\infty([0,T],L^2(M))} \le C_T.$$

From this, it is not hard to show that the map $t \to \tilde{v}^n(t)$ is weakly differentiable with derivative given by $t \to \tilde{q}^n(t)$. In particular (B.27) follows at once.

Step 2. Compactness.
We may apply Lemma B.5.19 to obtain that for any $n \in \mathbf{N}$ and any $t > 0$

$$E_n[v^n(t)] + \frac{1}{2}\int_0^t |\Delta_{n,\epsilon_n}v^n(s)|^2_{\mathcal{V}_{n,\epsilon_n}} ds + \frac{1}{2}\int_0^t |\frac{d}{ds}v^n(s)|^2_{\mathcal{V}_{n,\epsilon_n}} ds \le E_n[u^0|_{G_n}]. \quad \text{(B.28)}$$

Fix a time horizon $T>0$ and a countable dense subset $\{t_p\}_{p\in\mathbf{N}} \subset [0,T]$. From (B.28) with $t = t_j$ and the compactness property in Theorem B.5.16 we have that, for each $p \in \mathbf{N}$, the sequence $\tilde{v}^n(t_p)$ is precompact in $L^2(M)$. By a diagonal argument we can thus find a subsequence n_j and functions $u_{t_p} \in L^2(M)$ such that

$$L^2(\nu) - \lim \tilde{v}^{n_j}(t_p) = u_{t_p}\ \forall p \in \mathbf{N}.$$

We claim that $v^{n_j}(t)$ is a Cauchy sequence in $L^2(\nu)$ for any $t \in [0,T)$. Indeed for $p \in \mathbf{N}$, $j,l \in \mathbf{N}$ using the triangle inequality and (B.27) we have

$$\begin{aligned}
&\|\tilde{v}^{n_{j+l}}(t) - \tilde{v}^{n_j}(t)\|_{L^2(\nu)} \\
&\le \|\tilde{v}^{n_{j+l}}(t) - \tilde{v}^{n_{j+l}}(t_p)\|_{L^2(\nu)} + \|\tilde{v}^{n_{j+l}}(t_p) - \tilde{v}^{n_j}(t_p)\|_{L^2(\nu)} + \|\tilde{v}^{n_j}(t_p) - \tilde{v}^{n_j}(t)\|_{L^2(\nu)} \\
&\le 2C_T|t-t_p| + \|\tilde{v}^{n_{j+l}}(t_p) - \tilde{v}^{n_j}(t_p)\|_{L^2(\nu)}.
\end{aligned}$$

Now given $\gamma > 0$ select $p \in \mathbf{N}$ such that $|t - t_p| \le \frac{\gamma}{4C_T}$ and $j \in \mathbf{N}$ such that $\|\tilde{v}^{n_{j+l}}(t_p) - \tilde{v}^{n_j}(t_p)\|_{L^2(\nu)} \le \frac{\gamma}{2}$ for any $l \in \mathbf{N}$, then

$$\|\tilde{v}^{n_{j+l}}(t) - \tilde{v}^{n_j}(t)\|_{L^2(\nu)} \le \gamma$$

whenever $l \in \mathbf{N}$; thus $\tilde{v}^{n_j}(t) \to u_t \in L^2(M)$. Define $u(t) = u_t,\ t \in [0,T]$. We have just proved that

$$L^2(\nu) - \lim_{j\to+\infty} \tilde{v}^{n_j}(t) = u(t)\ \forall t \in [0,T].$$

We need to show that u is characterized by (B.20), this will be shown in Step 4 where we will pass to the limit into (B.28). To this aim, we will need to be able to pass to the limit in the second and third terms on the left hand side of (B.28). By (B.26) and (B.27) over a further, non-relabeled subsequence we have that there exists $v \in L^2([0,T],L^2(\nu))$ with $v' \in L^2([0,T],L^2(\nu))$ such that

$$\tilde{v}^{n_j} \xrightarrow{L^2(L^2)} v,\ \frac{d}{dt}\tilde{v}^{n_j} \xrightarrow{L^2(L^2)} \frac{d}{dt}v,$$

and by uniqueness of the limit this implies $u(t) = v(t)$. For later, we also record that

$$\liminf_{j\to+\infty} \int_0^t |\frac{d}{ds} v^{n_j}|^2_{\mathcal{V}_{n_j,\epsilon_{n_j}}} ds \geq \int_0^t \int_M |\frac{d}{ds} u(s)|^2 \rho^2 d\nu. \tag{B.29}$$

This easily follows by the weak lower semicontinuity of the $L^2((0,t), L^2(M))$ norm once we observe that $d^{n_j} \circ T_{n_j} \frac{d}{dt} \tilde{v}^{n_j}$ converges weakly to $\rho u'$ in this space.

Step 3. We claim that $\Delta_{\rho^2} u \in L^2_{loc}((0,+\infty), L^2(\nu))$ and that for every $T > 0$

$$\liminf_{j\to+\infty} \int_0^T |\Delta_{n_j,\epsilon_{n_j}} v^{n_j}|^2_{\mathcal{V}_{n_j,\epsilon_{n_j}}} dt \geq \int_0^T c^2 |\Delta_{\rho^2} u|^2 \rho^2 dx dt, \tag{B.30}$$

where $c := \frac{C_2}{2}$.
To show this we observe that from (B.28) we obtain that up to taking a further subsequence, the functions $\widetilde{\Delta_{n_j,\epsilon_{n_j}}} v^{n_j} := \Delta_{n_j,\epsilon_{n_j}} v^{n_j} \circ T_{n_j}$ converge weakly in $L^2([0,T], L^2(\nu))$ to a function $w \in L^2([0,T], L^2(\nu))$. We claim that

$$\frac{1}{c} w = \Delta_{\rho^2} u \text{ in the sense of distributions.} \tag{B.31}$$

If (B.31) is true, then (B.30) follows by the lower semicontinuity of the L^2-norm. To show (B.31) we take $f \in C_c^\infty((0,+\infty))$ and $g \in C^\infty(M)$. Then using Theorem B.5.14, Lemma B.5.15, the convergence of the functions $\tilde{v}^{n_j}$ and using the fact that the Laplacian on the graph is self-adjoint we have

$$\begin{aligned}
&\int_0^{+\infty} f(t) \int_M u(t) \, (\Delta_{\rho^2} g) \, \rho^2 d \operatorname{Vol}_M dt \\
&= \lim_{j\to+\infty} \frac{1}{c} \int_0^{+\infty} f(t) \int_M \tilde{v}^{n_j}(t) \widetilde{\Delta_{n_j,\epsilon_{n_j}}} g \tilde{d}^{n_j} d\nu dt \\
&= \lim_{j\to+\infty} \frac{1}{c} \int_0^{+\infty} f(t) \langle v^{n_j}, \Delta_{n_j,\epsilon_j} g \rangle_{\mathcal{V}_{n_j,\epsilon_{n_j}}} dt \\
&= \lim_{j\to+\infty} \frac{1}{c} \int_0^{+\infty} f(t) \langle \Delta_{n_j,\epsilon_{n_j}} v^{n_j}, g \rangle_{\mathcal{V}_{n_j,\epsilon_{n_j}}} dt \\
&= \lim_{j\to+\infty} \frac{1}{c} \int_0^{+\infty} f(t) \int_M \widetilde{\Delta_{n_j,\epsilon_{n_j}}} v^{n_j} \tilde{g}^{n_j} \tilde{d}^{n_j} d\nu dt.
\end{aligned}$$

Observing that $\tilde{g}^{n_j} := g \circ T_{n_j}$ converges uniformly to g we get that the last limit equals

$$\frac{1}{c} \int_0^{+\infty} f(t) \int_M w(t) g \rho^2 d \operatorname{Vol}_M dt.$$

In particular for every $f \in C_c^\infty((0,+\infty))$ and $g \in C^\infty(M)$ we have

$$\int_0^{+\infty} f(t) \int_M u(t) \Delta_{\rho^2} g \rho^2 d \operatorname{Vol}_M dt = \frac{1}{c} \int_0^{+\infty} f(t) \int_M w(t) g \rho^2 d \operatorname{Vol}_M dt,$$

which clearly gives (B.31).

Step 4. Proof of (B.23).
Using Theorem B.5.16, the weak lower semicontinuity (B.29) and the lower bound obtained in Step 3 we can pass to the limit in (B.28) to get that for all times $t > 0$

$$\frac{C_2}{2}E[u(t)] + \frac{1}{2}\int_0^t \int_M \left(\frac{C_2}{2}\right)^2 |\Delta_{\rho^2} u|^2 \rho^2 dx ds + \frac{1}{2}\int_0^t \int_M |u'|^2 \rho^2 dx ds \leq \frac{C_2}{2}E[u^0]$$

In particular, applying Lemma B.5.18 with $c = C_2/2$ we see that u is the unique solution to

$$\begin{cases} \partial_t u = -\frac{C_2}{2}\Delta_{\rho^2} u \\ u(0) = u^0 \end{cases}$$

which, in particular, implies that the limit is independent of the chosen subsequence, thus the whole sequence converges to u, as claimed.

Step 5. Conclusion.
Let T_n be a sequence of transportation maps obtained by applying Theorem C.6.6. By Proposition B.5.7 we just need to show that for any $t > 0$ the functions $e^{-t\Delta_{n,\epsilon_n}} u^n \circ T_n$ converge strongly to $e^{-t\frac{C_2}{2C_1}\Delta_{\rho^2}} u$ in $L^2(M)$. For $s > 0$ define $v^n(x, s) = e^{-s\Delta_{n,\epsilon_n}} u^n$. By differentiating the norm we have

$$\frac{d}{ds}|v^n|^2_{\mathcal{V}_{n,\epsilon_n}} = -|\nabla_n v^n|^2_{\mathcal{E}_{n,\epsilon_n}}.$$

Thus after integrating in s and by using Fatou's Lemma we have that for a fixed $t > 0$

$$\int_0^t \liminf_{n\to+\infty} |\nabla_n v^n(s)|^2_{\mathcal{E}_{n,\epsilon_n}} ds \leq C\|u\|^2_{L^2(\nu)}.$$

In particular there exist $0 < s < t$ and a subsequence n_j such that

$$\sup_{j\in\mathbf{N}} E_{n_j}[v^{n_j}(s)] < +\infty.$$

By Lemma B.5.19 applied to $w^{n_j}(r) := e^{-r\Delta_{n_j,\epsilon_{n_j}}} v^{n_j}(s)$ we infer that

$$\sup_{j\in\mathbf{N}} E_{n_j}[v^{n_j}(t)] = \sup_{j\in\mathbf{N}} E_{n_j}[w^{n_j}(t-s)] < +\infty.$$

In particular, by the compactness statement of Theorem B.5.16 we obtain that, upon taking a further subsequence, the functions $v^{n_j}(t)$ converge strongly in $TL^2(M)$ to a function $v(t) \in L^2(M)$. We claim that $v = e^{-t\frac{C_2}{2C_1}\Delta_{\rho^2}} u$. To see this, let $g \in C^\infty(M)$. Setting $\tilde{v}^{n_j}(t) = v^{n_j}(t) \circ T_{n_j}$ and $\tilde{g}^{n_j} = g \circ T_{n_j}$, using the fact that $\tilde{g}^{n_j}$ converges uniformly to g and the fact that $e^{-t\Delta_{n,\epsilon_n}}$ are self-adjoint operators we get

$$\begin{aligned}
\int_M vg\rho^2 d\,\mathrm{Vol}_M &= \lim_{j\to+\infty} \int_M \tilde{v}^{n_j}(t)\tilde{g}\tilde{d}^{n_j} d\nu \\
&= \lim_{j\to+\infty} \langle v^{n_j}(t), g\rangle_{\mathcal{V}_{n_j,\epsilon_{n_j}}} \\
&= \lim_{j\to+\infty} \langle u^{n_j}, e^{-t\Delta_{n_j,\epsilon_{n_j}}} g\rangle_{\mathcal{V}_{n_j,\epsilon_{n_j}}} \\
&= \lim_{j\to+\infty} \int_M \tilde{u}^{n_j} \widetilde{e^{-t\Delta_{n_j,\epsilon_{n_j}}} g}\,\tilde{d}^{n_j} d\nu \\
&= \int_M u e^{-t\frac{C_2}{2C_1}\Delta_{\rho^2}} g\rho^2 d\,\mathrm{Vol}_M,
\end{aligned}$$

where in the last step we used (B.23). Using the self-adjointness of the heat semigroup on M we infer that for any smooth function $g \in C^\infty(M)$,

$$\int_M vg\rho^2 d\,\mathrm{Vol}_M = \int_M e^{-t\frac{C_2}{2C_1}\Delta_{\rho^2}} ug\rho^2 d\,\mathrm{Vol}_M \,.$$

Thus $v = e^{-t\frac{C_2}{2C_1}\Delta_{\rho^2}} u$. In particular, the limit does not depend on the chosen subsequence, thus we obtain the claim. □

B.6.2 Discrete-to-nonlocal

Proof of Theorem B.2.1. The proof follows from Theorem B.2.2, Proposition B.5.11 and the convergence of the degrees in Lemma B.5.15. The precompactness statement is a consequence of the general fact that bounded sets in L^2 are weakly precompact. □

B.6.3 Bertozzi's question

Proof of Corollary B.2.3. By Theorem B.2.1 we know that almost surely, for each $h > 0$,

$$\Gamma(TL^2(M) - \text{weak}) - \lim_{n\to+\infty} E^h_{n,\epsilon_n} = \sqrt{\frac{C_1C_2}{2}} E_{\frac{C_2 h}{2C_1}}.$$

By the same argument used in the proof of Theorem B.2.1 we have that almost surely, for every $h > 0$, the sequence of energies

$$D^h_{n,\epsilon_n}(u) := \frac{1}{\sqrt{h}} \sum_{i\neq j} \sigma_{ij} \langle u_i - \chi^n_i, e^{-h\Delta_{n,\epsilon_n}}(u_j - \chi^n_j)\rangle_{\nu_n}, \quad u \in \mathcal{M}_n,$$

Γ-converges in the weak-$TL^2(M)$ topology to the energy

$$D_h(u) = \sqrt{\frac{C_1C_2}{2}} \frac{1}{\sqrt{\frac{C_2 h}{2C_1}}} \sum_{i\neq j} \int_M (u_i - \chi_j) e^{-h\frac{C_2}{2C_1}\Delta_{\rho^2}} (u_j - \chi_j)\rho^2 d\,\mathrm{Vol}_M, \quad u \in \mathcal{M}.$$

In particular for every $h > 0$ we have

$$\Gamma(TL^2(M) - \text{weak}) - \lim_{n\to+\infty} (E^h_{n,\epsilon_n} - D^h_{n,\epsilon_n}) = \sqrt{\frac{C_1C_2}{2}} E_{\frac{C_2 h}{2C_1}} - D_h.$$

This yields that the minimizers of $E^h_{n,\epsilon_n} - D^h_{n,\epsilon_n}$ converge weakly in $TL^2(M)$ to minimizers of $\sqrt{\frac{C_1C_2}{2}} E_{\frac{C_2 h}{2C_1}} - D_h$. The conclusion is then a consequence of the minimizing movements interpretations in Lemma B.1.2 and Lemma B.5.2. □

B.6.4 Nonlocal-to-local

Proof of Theorem B.2.5. Γ-lim sup. The Γ-lim sup inequality is a consequence of the consistency part of the theorem, namely that for every $u \in BV(M, \{0,1\}^P) \cap \mathcal{M}$

$$\lim_{h\downarrow 0} E_h(u) = E(u). \tag{B.32}$$

It is clear that (B.32) is a consequence of the following claim: Assume that $E, F \subset M$ are sets of finite perimeter, then

$$\lim_{h\downarrow 0} \frac{1}{\sqrt{h}} \int_M \chi_F \left(\chi_E - e^{-h\Delta_\xi}\chi_E\right) d\mu = \frac{1}{\sqrt{\pi}} \int_{\partial^* E \cap \partial^* F} \langle \sigma_E(x), \sigma_F(x)\rangle_x |D\chi_F|_\xi(x). \quad \text{(B.33)}$$

Indeed, simply apply (B.33) to $E = \{u_i = 1\}$, $F = \{u_j = 1\}$, multiply by σ_{ij} and sum over all pairs $i \neq j$, since then $\langle \sigma_E(x), \sigma_F(x)\rangle_x = -1$ on $\partial^* E \cap \partial^* F$. We now prove (B.33) in four steps.

Step 1. Recalling the notation (B.5.1), we can rewrite

$$T(h)\chi_E(x) = \chi_E(x) + \int_0^h \frac{d}{dt} T(t)\chi_E(x)dt.$$

Using Theorem B.5.3 we obtain that the argument of the limit in (B.33) is equal to

$$\begin{aligned} &\frac{1}{\sqrt{h}} \int_M \chi_F \int_0^h \Delta_\xi T(t)\chi_E(x)dtd\mu(x) \\ &= \frac{1}{\sqrt{h}} \int_{\partial^* F} \langle \sigma_{\chi_F}(x), \int_0^h \nabla T(t)\chi_E(x)dt\rangle d|D\chi_F|_\xi(x). \end{aligned}$$

Thus by the discussion in Remark B.5.5 it suffices to show that for every $x \in \partial^* E \cap \partial^* F$ such that $\sigma_E(x) = \langle \sigma_E(x), \sigma_F(x)\rangle_x \sigma_F(x)$ we have

$$\frac{1}{\sqrt{\pi}} \langle \sigma_F(x), \sigma_E(x)\rangle = \lim_{h\downarrow 0} \left\langle \sigma_F(x), \frac{1}{\sqrt{h}} \int_0^h \nabla T(t)\chi_E(x)dt \right\rangle. \quad \text{(B.34)}$$

Step 2. We claim that for $s < 1/2$, equation (B.34) is equivalent to

$$\frac{1}{\sqrt{\pi}} \langle \sigma_F(x), \sigma_E(x)\rangle = \lim_{h\downarrow 0} \left\langle \sigma_F(x), \frac{1}{\sqrt{h}} \int_0^h \nabla T(t)(\chi_{E\cap B_{h^s}(x)}(\cdot))(x)dt \right\rangle. \quad \text{(B.35)}$$

To prove this equivalence, we fix $s < 1/2$ and use (C.46) to show

$$\lim_{h\downarrow 0} \frac{1}{\sqrt{h}} \int_0^h \int_{M\setminus B_{h^s}(x)} |\nabla_x p(t,x,y)| d\mu(y) = 0. \quad \text{(B.36)}$$

Clearly (B.36) then implies the equivalence between (B.34) and (B.35). For $j \in \mathbf{N}$ and $t < h$ we denote by B_j the ball $B_{2^j t^s}(x)$. Observe that $M \setminus B_{h^s}(x) \subset M \setminus B_{t^s}(x)$. To prove (B.36) we use the Gaussian upper bound (C.46) to estimate

$$\begin{aligned} &\frac{1}{\sqrt{h}} \int_0^h \int_{M\setminus B_{h^s}(x)} |\nabla p(t,x,y)| d\mu(y)dt \\ &\leq \sum_{j=0}^{[\mathrm{diam}(M)]} \frac{1}{\sqrt{h}} \int_0^h \int_{B_{j+1}\setminus B_j} \frac{\hat{C}_1}{\sqrt{t}\mu(B_{\sqrt{t}}(x))} \exp\left(\frac{-d^2(x,y)}{\hat{C}_2 t}\right) d\mu(y)dt \\ &\leq \sum_{j=0}^{[\mathrm{diam}(M)]} \frac{\hat{C}_1}{\sqrt{h}} \int_0^h \frac{\mu(B_{j+1})}{\sqrt{t}\mu(B_{\sqrt{t}}(x))} \exp\left(-\frac{2^{2j}}{\hat{C}_2 t^{1-2s}}\right) dt. \quad \text{(B.37)} \end{aligned}$$

Observe that the doubling property (B.5) gives $\frac{\mu(B_{j+1})}{\mu(B_{\sqrt{t}}(x))} \leq 3\frac{2^{jN}t^{sN}}{t^{N/2}}$. Thus (B.37) is estimated by

$$\begin{aligned}
&\hat{C}_1 \sum_{j=0}^{[\mathrm{diam}(M)]} \frac{1}{\sqrt{h}} \int_0^h \frac{2^{jN}t^{Ns}}{\sqrt{t}t^{N/2}} \exp\left(-\frac{2^{2j}}{\hat{C}_2 t^{1-2s}}\right) dt \\
&= \hat{C}_1 \sum_{j=0}^{[\mathrm{diam}(M)]} 2^{jN} \int_0^h t^{-N/2+sN-1} \exp\left(-\frac{2^{2j}}{\hat{C}_2 t^{1-2s}}\right) dt \\
&\leq \hat{C}_1 \sum_{j=0}^{[\mathrm{diam}(M)]} 2^{jN} \exp\left(-\frac{2^{2j}}{2\hat{C}_2 h^{1-2s}}\right) \int_0^h t^{-N/2+sN-1} \exp\left(-\frac{1}{2\hat{C}_2 t^{1-2s}}\right) dt,
\end{aligned}$$

which converges to zero as $h \downarrow 0$, since the integrand is uniformly bounded and the prefactor converges to zero as $h \downarrow 0$.

Step 3. We claim that

$$\begin{aligned}
&\lim_{h\downarrow 0} \frac{1}{\sqrt{h}} \left\langle \sigma_F(x), \frac{1}{\sqrt{h}} \int_0^h \int_M \nabla_x p(t,x,y) \chi_{E\cap B_{h^s}(x)}(y) d\mu(y) \right\rangle \qquad \text{(B.38)} \\
&= \lim_{h\downarrow 0} \frac{1}{\sqrt{h}} \left\langle \sigma_F(x), \int_0^h \int_M \nabla_x \left(\frac{e^{-\frac{d(x,y)^2}{4t}}}{(4\pi t)^{k/2}} v^0(x,y) \right) \chi_{E\cap B_{h^s}(x)}(y) d\mu(y) dt \right\rangle,
\end{aligned}$$

where v^0 is the coefficient in the asymptotic expansion (B.6).

To see this, observe that (B.6) applied with $l = 1$ and some $N > \frac{k}{2} + l$ yields

$$\begin{aligned}
&\lim_{h\downarrow 0} \frac{1}{\sqrt{h}} \left\langle \sigma_F(x), \frac{1}{\sqrt{h}} \int_0^h \int_M \nabla_x p(t,x,y) \chi_{E\cap B_{h^s}(x)}(y) d\mu(y) \right\rangle \\
&= \lim_{h\downarrow 0} \frac{1}{\sqrt{h}} \sum_{j=0}^{N} \left\langle \sigma_F(x), \int_0^h \int_M \nabla_x \left(\frac{e^{-\frac{d(x,y)^2}{4t}}}{(4\pi t)^{k/2}} v^j(x,y) t^j \right) \chi_{E\cap B_{h^s}(x)}(y) d\mu(y) dt \right\rangle.
\end{aligned}$$

Thus, all we need to show is that the limit as $h \downarrow 0$ of the terms on the right hand side corresponding to $j \geq 1$ vanishes, i.e., that for $j \geq 1$

$$\lim_{h\downarrow 0} \frac{1}{\sqrt{h}} \left\langle \sigma_F(x), \int_0^h \int_M \nabla_x \left(\frac{e^{-\frac{d(x,y)^2}{4t}}}{(4\pi t)^{k/2}} v^j(x,y) t^j \right) \chi_{E\cap B_{h^s}(x)}(y) d\mu(y) \right\rangle = 0. \qquad \text{(B.39)}$$

To verify (B.39), we compute the argument in the limit in normal coordinates around x. Let $\Psi : B_R(x) \to B_R(\underline{o})$ be normal coordinates around x. Then $g_{ij}(\underline{o}) = \delta_{ij}$ is the identity matrix and $d^2(x,y) = |\Psi(y)|_k$. Writing ν_F for the vector of coordinates of $\sigma_F(x)$, the argument of the limit may be written as

$$\frac{1}{\sqrt{h}} \int_0^h \int_{B_{h^s}(x)} \nu_F \cdot \left(\frac{-z}{2t} v^j(x,\Phi(z)) + D v^j(x,\Phi(z)) \right) \frac{e^{-\frac{|z|^2}{4t}}}{(4\pi t)^{k/2}} t^j \chi_{\Psi(E)}(z) \gamma(z) dz dt,$$

where we set $\gamma(z) := \sqrt{\det(g)(z)}\rho(\Phi(z))$, with $\Phi = \Psi^{-1}$. By the smoothness of the coefficients v^j, by the compactness of the manifold and by the fact that $j \geq 1$, if $h < 1$ we can bound this integral by

$$Ch^{j-1}\frac{1}{\sqrt{h}}\int_0^h \int_{\mathbf{R}^k} \left(\frac{1}{2}+1\right)\frac{e^{-\frac{|z|^2}{4t}}}{(4\pi t)^{k/2}}dzdt \leq C\sqrt{h}, \tag{B.40}$$

where C is a constant depending on v^j, M and ξ. Thus we have (B.39).

Step 4. Conclusion. We now compute the limit on the right-hand side of (B.38). As before, we work in normal coordinates centered at x. With the same notation as in Step 3, the argument of the limit may be rewritten as

$$\begin{aligned}
&\frac{1}{\sqrt{h}}\left\langle \sigma_F(x), \int_0^h \int_M \nabla_x \left(\frac{e^{-\frac{d(x,y)^2}{4t}}}{(4\pi t)^{k/2}}v^0(x,y)\right)\chi_{E\cap B_{h^s}(x)}(y)d\mu(y)dt\right\rangle \\
&= \frac{1}{\sqrt{h}}\int_0^h \int_{B_{h^s}(x)} \nu_F \cdot \left(\frac{-z}{2t}v^0(x,\Phi(z)) + Dv^0(x,\Phi(z))\right)\frac{e^{-\frac{|z|^2}{4t}}}{(4\pi t)^{k/2}}\chi_{\Psi(E)}(z)\gamma(z)dzdt.
\end{aligned}$$

As in Step 3, one can show that

$$\lim_{h\downarrow 0}\frac{1}{\sqrt{h}}\int_0^h \int_{B_{h^s}(x)} \nu_F \cdot Dv^0(x,\Phi(z))\frac{e^{-\frac{|z|^2}{4t}}}{(4\pi t)^{k/2}}\chi_{\Psi(E)}(z)\gamma(z)dzdt = 0.$$

Thus, all we need to show is that

$$\lim_{h\downarrow 0} -\frac{1}{\sqrt{h}}\int_0^h \int_{B_{h^s}(o)} \nu_F \cdot \frac{z}{2t}v^0(x,\Phi(z))\frac{e^{-\frac{|z|^2}{4t}}}{(4\pi t)^{k/2}}\chi_{\Psi(E)}(z)\gamma(z)dzdt = \frac{1}{\sqrt{\pi}}\nu_E \cdot \nu_F.$$

This is essentially already done in [67]. We sketch the short argument for completeness. After a change of variables in space and time, the argument of the limit may be written as

$$-\int_0^1 \frac{1}{\sqrt{t}}\int_{B_{\frac{h^s}{\sqrt{ht}}}(o)} \nu_F \cdot \frac{z}{2}v^0(x,\Phi(\sqrt{ht}z))\frac{e^{-\frac{|z|^2}{4}}}{(4\pi)^{k/2}}\chi_{\frac{\Psi(E)}{\sqrt{ht}}}\gamma(\sqrt{ht}z)dzdt.$$

By De Giorgi's structure theorem we have

$$L^1_{loc} - \lim_{h\downarrow 0}\chi_{\Psi(E)}(\sqrt{ht}) = \chi_{H_{\nu_E}} \ \forall t \in [0,1],$$

where H_{ν_E} is the half space given by

$$H_{\nu_E} := \left\{z \in \mathbf{R}^k : \ \nu_E \cdot z \leq 0\right\}.$$

By an application of the dominated convergence theorem we infer that on any compact set $K \subset \mathbf{R}^k$,

$$\lim_{h\downarrow 0}\int_0^1 \int_K |\chi_{\frac{\Psi(E)}{\sqrt{ht}}} - \chi_{H_{\nu_E}}|dzdt = 0.$$

In particular, upon taking a subsequence, we may assume that

$$\lim_{h\downarrow 0}\chi_{\frac{\Psi(E)}{\sqrt{ht}}}(z)=\chi_{H_{\nu_E}}(z) \text{ for a.e. } (t,z)\in[0,1]\times\mathbf{R}^k.$$

Moreover we have that

$$\lim_{h\downarrow 0} v^0(x,\Phi(\sqrt{ht}z))\gamma(\sqrt{ht}z)=1, \text{ uniformly in } t\in[0,1].$$

Thus by an application of the dominated convergence theorem we get

$$\begin{aligned}
&\lim_{h\downarrow 0} -\int_0^1 \frac{1}{\sqrt{t}}\int_{B_{\frac{h^s}{\sqrt{ht}}}(o)} \nu_F\cdot\frac{z}{2}v^0(x,\Phi(\sqrt{ht}z))\frac{e^{-\frac{|z|^2}{4}}}{(4\pi)^{k/2}}\chi_{\frac{\Psi(E)}{\sqrt{ht}}}\gamma(\sqrt{ht}z)dzdt\\
&=-\int_0^1\frac{1}{2\sqrt{t}}\int_{\nu_E\cdot y\le 0}(y\cdot\nu_F)G_1(z)dz\\
&=\int_0^1\frac{1}{2\sqrt{t}}\int_{\nu_E\cdot y\le 0}(\nu_F\cdot\nu_E)(y\cdot\nu_E)_-G_1(z)dz\\
&=\frac{1}{\sqrt{\pi}}(\nu_F\cdot\nu_E).
\end{aligned}$$

Γ-lim inf. To prove the Γ-lim inf inequality we use the blow-up method of Fonseca and Müller [30] (see also [1] and [4]).
Given $u_h\in\mathcal{M}$ such that $u^h\to u\in\mathcal{M}$ in $L^1(M)$, we want to prove that for every sequence $h_n\downarrow 0$

$$\liminf_{n\to+\infty} E_{h_n}(u^{h_n})\ge E(u). \tag{B.41}$$

Clearly, we may without loss of generality assume that the left hand side of (B.41) is finite.

Step 1. $u\in BV(M,\{0,1\}^P)$.

By Lemma B.5.4 we just need to show that $u\circ\psi$ is in $BV(\psi(V))$ for every chart (V,ψ) of M. It is clear that one can restrict to the case when $V=B_r(x_0)$, $r\le R<\frac{\text{inj}(M)}{2}$, $x_0\in M$ for some fixed R and $\psi=\exp_{x_0}^{-1}$. The statement for a general chart then follows by compactness. So we fix $V=B_r(x_0)$ and $\psi=\exp_{x_0}^{-1}$. We observe that if $N\ge\frac{k}{2}$, by the asymptotic expansion for the heat kernel (B.6) with $l=0$ and $t=h$ we get

$$\begin{aligned}
E_{h_n}(u^{h_n}) &\ge \frac{1}{\sqrt{h_n}}\sum_{i,j}\sigma_{ij}\int_{B_r(x_0)} u_i^{h_n}e^{-h_n\Delta_\xi}u_j^{h_n}d\mu\\
&\ge\frac{1}{\sqrt{h_n}}\sum_{i,j}\sigma_{ij}\int_{B_r(x_0)}u_i^{h_n}(x)\int_{B_r(x_0)}p(h_n,x,y)u_j^{h_n}(y)d\mu(y)d\mu(x)\\
&\ge\sum_{l=0}^N\frac{1}{\sqrt{h_n}}\sum_{i,j}\sigma_{ij}\int_{B_r(x_0)}u_i^{h_n}(x)\int_{B_r(x_0)}\frac{e^{\frac{-d^2(x,y)}{4h_n}}}{(4\pi h_n)^{k/2}}v^l(x,y)h_n^l u_j^{h_n}(y)d\mu(y)d\mu(x)\\
&\quad-C_N\sqrt{h_n}
\end{aligned}$$

If $l\ge 1$, with an estimate similar to the one used in (B.40) we obtain that

$$\left|\frac{1}{\sqrt{h_n}}\sum_{i,j}\sigma_{ij}\int_{B_r(x_0)}u_i^{h_n}(x)\int_{B_r(x_0)}\frac{e^{\frac{-d^2(x,y)}{4h_n}}}{(4\pi h_n)^{k/2}}v^l(x,y)h_n^l u_j^{h_n}(y)d\mu(y)d\mu(x)\right|\le C\sqrt{h_n},$$

where C depends on σ, v_l, M and ξ. Thus we have

$$\begin{aligned}E_{h_n}(u^{h_n}) \geq &\frac{1}{\sqrt{h_n}}\sum_{i,j}\sigma_{ij}\int_{B_r(x_0)} u_i^{h_n}(x)\int_{B_r(x_0)}\frac{e^{\frac{-d^2(x,y)}{4h_n}}}{(4\pi h_n)^{k/2}}v^0(x,y)u_j^{h_n}(y)d\mu(y)d\mu(x)\\ &- C\sqrt{h_n}.\end{aligned}$$

The first term on the right hand side may be rewritten in local coordinates as

$$\frac{1}{\sqrt{h_n}}\sum_{i,j}\sigma_{ij}\int_{B_r(\underline{o})}\tilde{u}_i^{h_n}(x)\int_{B_r(\underline{o})}\frac{e^{\frac{-d^2(\psi^{-1}(x),\psi^{-1}(y))}{4h_n}}}{(4\pi h_n)^{k/2}}\tilde{v}^0(x,y)\tilde{u}_j^{h_n}(y)\gamma(y)dy\ \gamma(x)dx, \quad \text{(B.42)}$$

with $\gamma(x) = \sqrt{\det(g)}\xi(\psi^{-1}(x))$, $\tilde{v}^0(x,y) = v_0(\psi^{-1}(x),\psi^{-1}(y))$ and $\tilde{u} = u\circ\psi^{-1}$. Let L be such that $d(\psi^{-1}(x),\psi^{-1}(y)) \leq L|x-y|_k$. Then (B.42) may be bounded from below by

$$\frac{\inf_{x,y\in B_r(\underline{o})}\{\tilde{v}^0(x,y)\gamma(y)\gamma(x)\}}{L^{k+1}}E^{euclid}_{\frac{h_n}{L^2}}(1_{B_r(\underline{o})}\tilde{u}^{h_n}),$$

where we set

$$E^{euclid}_{\frac{h_n}{L^2}}(1_{B_r(\underline{o})}\tilde{u}^{h_n}) = \frac{1}{\sqrt{\frac{h_n}{L^2}}}\sum_{i,j}\sigma_{ij}\int_{\mathbf{R}^k}\tilde{u}_i^{h_n}(x)\int_{\mathbf{R}^k}\frac{e^{\frac{-L^2|x-y|^2}{4h_n}}}{(4\pi\frac{h_n}{L^2})^{k/2}}\tilde{u}_j^{h_n}dy\ dx.$$

In particular we obtain that

$$+\infty > \liminf_{n\to+\infty} E^{euclid}_{\frac{h_n}{L^2}}(\mathbf{1}_{B_r(\underline{o})}\tilde{u}^{h_n}),$$

which says that $\mathbf{1}_{B_r(\underline{o})}\tilde{u}$ is in $BV(\mathbf{R}^k,\{0,1\}^P)$ by an application of, for example, Lemma A.4 in [25].

Step 2. We now turn to (B.41). By Step 1 we know that $u \in BV(M,\{0,1\}^P)$. We set $\Omega_i := \{u_i = 1\}$. Passing to a subsequence if necessary, we may assume that

$$\lim_{n\to+\infty} E_{h_n}(u^{h_n}) = \liminf_{n\to+\infty} E_{h_n}(u^{h_n}) < +\infty. \quad \text{(B.43)}$$

We define the Radon measures $\lambda_{ij}^{h_n}$ by setting

$$\lambda_{ij}^{h_n}(W) := \frac{1}{\sqrt{h_n}}\sigma_{ij}\int_W u_i^{h_n}e^{-h\Delta_\xi}u_j^{h_n}d\mu,\ W\in\mathcal{B}(M).$$

Then by (B.43), upon passing to a further subsequence, we may assume that there exist Radon measures λ_{ij} such that

$$\lim_{n\to+\infty}\lambda_{ij}^{h_n} = \lambda_{ij} \text{ weakly-}* \text{ in the sense of Radon measures.} \quad \text{(B.44)}$$

In particular we obtain that

$$\liminf_{n\to+\infty} E_{h_n}(u^{h_n}) = \liminf_{n\to+\infty}\sum_{i,j}\sigma_{ij}\lambda_{ij}^{h_n}(M) \geq \sum_{i,j}\sigma_{ij}\lambda_{ij}(M).$$

Thus to conclude the proof of the Γ-lim inf inequality it suffices to show the following:
It holds that if $\underline{x} \in \Sigma_{ij}$ then

$$\sum_m \sigma_{mq} \frac{d\lambda_{mq}}{d|Du_i|_\xi}(\underline{x}) \geq \frac{2\sigma_{ij}}{\sqrt{\pi}}. \tag{B.45}$$

Indeed, if (B.45) is true, then using the fact that the interfaces $\Sigma_{ij} = \partial^*\Omega_i \cap \partial^*\Omega_j$ are disjoint

$$\begin{aligned}
\left(\sum_{m,q} \sigma_{mq}\lambda_{mq}\right)(M) &\geq \sum_{i<j} \left(\sum_{m,q} \sigma_{mq}\lambda_{mq}\right)(\Sigma_{ij}) \\
&\geq \sum_{i<j} \int_{\Sigma_{ij}} \sum_{m,q} \sigma_{mq} \frac{d\lambda_{mq}}{d|Du_i|_\xi} d|Du^i|_\xi \\
&\geq \sum_{i<j} \frac{2\sigma_{ij}}{\sqrt{\pi}} \int_{\Sigma_{ij}} d|Du_i|_\xi \\
&= \frac{1}{\sqrt{\pi}} \sum_{i,j} \sigma_{ij} |Du_i|_\xi(\Sigma_{ij}).
\end{aligned}$$

We now prove (B.45). Fix $\delta > 0$, then there exists $R < \frac{\mathrm{inj}(M)}{2}$ such that for any $x \in M$

$$y, z \in B_{\frac{R}{2}}(x) \Rightarrow d(y,z) \leq (1+\delta)|\exp_x^{-1}(y) - \exp_x^{-1}(z)|. \tag{B.46}$$

Fix $i, j \in \{1, ..., P\}$, with $i \neq j$ and $\underline{x} \in \Sigma_{ij}$. For every $m, q \in \{1, ..., P\}$ with $m \neq q$ we have that

$$\frac{d\lambda_{mq}}{d|Du_i|_\xi}(\underline{x}) = \lim_{r\downarrow 0} \frac{\lambda_{mq}(B_r(\underline{x}))}{|Du_i|_\xi(B_r(\underline{x}))}.$$

Observe also that, using Lemma B.5.4 applied with $V = B_r(\underline{x})$ and $\psi(y) = \exp_{\underline{x}}^{-1}(y)$,

$$\lim_{r\downarrow 0} \frac{|Du_i|_\xi(B_r(\underline{x}))}{\omega_{k-1}r^{k-1}\gamma(\underline{o})} = \lim_{r\downarrow 0} \frac{\int_{B_r(\underline{o})} \gamma d\mathcal{H}^{k-1}}{\omega_{k-1}r^{k-1}\gamma(\underline{o})} = 1.$$

In particular

$$\frac{d\lambda_{mq}}{d|Du_i|_\xi}(\underline{x}) = \lim_{r\downarrow 0} \frac{\lambda_{mq}(B_r(\underline{x}))}{\omega_{k-1}r^{k-1}\gamma(\underline{o})}.$$

Observe that there exists an at most countable set $Q \subset \mathbf{R}$ such that if $r \notin Q$

$$\lambda_{mq}(\partial B_r(\underline{x})) = 0.$$

Thus, by the weak convergence (B.44) of the $\lambda_{mq}^{h_n}$ we have

$$\frac{d\lambda_{mq}}{d|Du_i|_\xi}(\underline{x}) = \lim_{r\downarrow 0, r\notin Q} \lim_{n\to+\infty} \frac{\lambda_{mq}^{h_n}(B_r(\underline{x}))}{\gamma(\underline{o})\omega_{k-1}r^{k-1}}.$$

We now set $\tilde{u}^{h_n} = u^{h_n} \circ \exp_{\underline{x}}$. Given a measurable function f defined on $B_r(\underline{o})$ we define the blow-up at scale r as $R_r f(y) := f(ry),\ y \in B_1$. By De Giorgi's structure theorem we know that

$$\lim_{r\downarrow 0} R_r\tilde{u}_i = \chi_{H_{\nu^{(i)}}} \text{ in } L^1(B_1), \tag{B.47}$$

$$\lim_{r\downarrow 0} R_r\tilde{u}_j = \chi_{H_{\nu^{(j)}}} \text{ in } L^1(B_1) \tag{B.48}$$

where we define

$$H_{\nu^{(m)}} := \{z \in \mathbf{R}^k : \ z\cdot\nu^{(m)} \le 0\}, \ m \in \{1,...,P\}.$$

Here $\nu^{(m)}$ is the outer unit normal of $\exp_{\underline{x}}^{-1}(\Omega_m) \subset \mathbf{R}^k$ at $\underline{o}$. Observe furthermore that for $q \neq i,j$ it holds that

$$\lim_{r\downarrow 0} R_r\tilde{u}_q = 0 \text{ in } L^1(B_1).$$

Indeed, this follows by the constraint $\sum_m R_r u_m^{h_n} = 1$ and (B.47), (B.48). Upon selecting a subsequence, we may thus choose a sequence r_n of radii such that

$$\begin{aligned}
&\lim_{n\to+\infty} r_n = \lim_{n\to+\infty} \frac{h_n}{r_n^2} = 0,\\
&\lim_{n\to+\infty} \frac{\lambda_{mq}^{h_n}(B_{r_n}(\underline{x}))}{\omega_{k-1}r_n^{k-1}\gamma(\underline{o})} = \frac{d\lambda_{mq}}{d|Du_i|}(\underline{x}),\\
&\lim_{n\to+\infty} R_{r_n}\tilde{u}_i^{h_n} = \chi_{H_{\nu^{(i)}}} \text{ in } L^1(B_1),\\
&\lim_{n\to+\infty} R_{r_n}\tilde{u}_j^{h_n} = \chi_{H_{\nu^{(j)}}} \text{ in } L^1(B_1),\\
&\lim_{n\to+\infty} R_{r_n}\tilde{u}_m^{h_n} = 0 \text{ in } L^1(B_1) \text{ for } m \neq i,j.
\end{aligned}$$

We now use once more the expansion (B.6) with some $N \ge \frac{k}{2}$ and observe that

$$\begin{aligned}
&\left|\lambda_{mq}^{h_n}(B_{r_n}(\underline{x})) - \sum_{l=0}^{N} \frac{1}{\sqrt{h_n}} \int_{B_{r_n}(\underline{x})} u_m^{h_n} \int_{B_{r_n}(\underline{x})} \frac{e^{\frac{-d^2(x,y)}{4h_n}}}{(4\pi h_n)^{k/2}} v^l(x,y) h_n^l u_q^{h_n}(y)d\mu(y)d\mu(x)\right|\\
&\le C\sqrt{h_n} r_n^{2k}.
\end{aligned}$$

Moreover, similarly as for (B.40) we get that for $l \ge 1$

$$\left|\frac{1}{\sqrt{h_n}} \int_{B_{r_n}(\underline{x})} u_m^{h_n} \int_{B_{r_n}(\underline{x})} \frac{e^{\frac{-d^2(x,y)}{4h_n}}}{(4\pi h_n)^{k/2}} v^l(x,y) h_n^l u_q^{h_n}(y)d\mu(y)d\mu(x)\right| \le C\sqrt{h_n} r_n^k.$$

From these two estimates we conclude that

$$\begin{aligned}
&\sum_{m,q} \sigma_{mq} \frac{d\lambda_{mq}}{d|Du_i|_\xi}(\underline{x})\\
&= \lim_{n\to+\infty} \sum_{m,q} \sigma_{mq} \frac{1}{\omega_{k-1}r_n^{k-1}\gamma(\underline{o})\sqrt{h_n}} \times\\
&\qquad \int_{B_{r_n}(\underline{x})} u_m^{h_n} \int_{B_{r_n}(\underline{x})} \frac{e^{\frac{-d^2(x,y)}{4h_n}}}{(4\pi h_n)^{k/2}} v^0(x,y) u_q^{h_n}(y)d\mu(y)d\mu(x).
\end{aligned}$$

By (B.46), for n large enough the previous limit may be estimated from below by

$$\liminf_{n\to+\infty} \frac{c_n}{\gamma(\underline{o})\omega_{k-1}r_n^{k-1}\sqrt{h_n}} \sum_{m,q} \sigma_{mq} \int_{B_{r_n}(\underline{o})} \tilde{u}_m^{h_n} \int_{B_{r_n}(\underline{o})} \frac{e^{\frac{-(1+\delta)^2|x-y|^2}{4h_n}}}{(4\pi h_n)^{k/2}} \tilde{u}_q^{h_n} dydx, \tag{B.49}$$

where $\tilde{u} := u \circ \exp_{\underline{x}}$ and

$$c_n := \inf_{x,y\in B_{r_n}(\underline{o})} \left\{ v^0(\exp_{\underline{x}}(x), \exp_{\underline{x}}(y))\gamma(x)\gamma(y) \right\}.$$

Observe that $c_n \to \gamma(\underline{o})$ as $n \to +\infty$. In particular (B.49) equals

$$\liminf_{n\to+\infty} \frac{1}{\omega_{k-1}r_n^{k-1}\sqrt{h_n}} \sum_{m,q} \sigma_{mq} \int_{B_{r_n}(\underline{o})} \tilde{u}_m^{h_n} \int_{B_{r_n}(\underline{o})} \frac{e^{\frac{-(1+\delta)^2|x-y|^2)}{4h_n}}}{(4\pi h_n)^{k/2}} \tilde{u}_q^{h_n} dydx.$$

We now perform the changes of variables $x \mapsto r_n x$ and $y \mapsto r_n y$, so that the previous quantity is equal to

$$\liminf_{n\to+\infty} \frac{1}{\omega_{k-1}(1+\delta)^{k+1}} E^{B_1}_{\frac{h_n}{r_n^2(1+\delta)^2}}(R_{r_n}\tilde{u}^{h_n}\mathbf{1}_{B_1}), \tag{B.50}$$

where we define for $t > 0$ and $f \in \mathcal{A}_{B_1} := \left\{ f : B_1 \to [0,1]^P \text{ such that } \sum_m f_m = 1 \right\}$

$$E_t^{B_1}(f) := \sum_{mq} \sigma_{mq} \frac{1}{\sqrt{t}} \int_{B_1} f_m G_t * f_q dx.$$

Here G_t denotes the standard k-dimensional Euclidean heat kernel at time t. Let $\beta \in C_c^\infty(B_1)$, $0 \le \beta \le 1$, then for $f \in \mathcal{A}_{B_1}$

$$E_t^{B_1}(f) \ge E_t^{B_1}(f,\beta) := \sum_{m,q} \sigma_{mq} \frac{1}{\sqrt{t}} \int_{B_1} \beta f_m G_t * f_q dx. \tag{B.51}$$

We record the following result, a proof of which is given in the Appendix.

Theorem B.6.1. *If $\sigma \in \mathbf{R}^{k\times k}$ is symmetric, $\sigma_{mm} = 0$, σ satisfy the triangle inequality and $\beta \in C_c^\infty(B_1)$ with $\beta \ge 0$, then on $\mathcal{A}_{B_1}$*

$$\Gamma - \lim_{t\downarrow 0} E_t^{B_1}(\cdot,\beta) = E(\cdot,\beta) \text{ in } L^1(B_1),$$

where we define, for $f \in \mathcal{A}_{B_1}$,

$$E(u,\beta) := \begin{cases} \frac{1}{\sqrt{\pi}} \sum_{m,q} \sigma_{mq} \int_{S_{mq}} \beta(x) d\mathcal{H}^{d-1}(x) & \text{if } f \in BV(B_1,\{0,1\}^P), \\ +\infty & \text{otherwise.} \end{cases}$$

Here, for $f \in BV(B_1,\{0,1\}^P)$, we set $S_{mp} := \partial^*\{f_m = 1\} \cap \partial^*\{f_q = 1\}$.

In particular, we may use the Γ-lim inf part of Theorem B.6.1 in (B.50) to obtain that for any $\beta \in C_c^\infty(B_1), 0 \le \beta \le 1$ we have

$$\sum_{m,q} \sigma_{mq} \frac{d\lambda_{mq}}{d|Du_i|_\xi(\underline{x})} \ge \frac{1}{\sqrt{\pi}\omega_{k-1}(1+\delta)^{k+1}} \sigma_{ij} \Bigg(\int_{\{\nu^{(i)}\cdot x=0\}} \beta(x) d\mathcal{H}^{d-1}(x) + \int_{\{\nu^{(j)}\cdot x=0\}} \beta(x) d\mathcal{H}^{d-1}(x) \Bigg).$$

Taking the supremum over all such β gives

$$\sum_{m,q} \sigma_{mq} \frac{d\lambda_{mq}}{d|Du_i|_\xi(\underline{x})} \geq \frac{2\sigma_{ij}}{\sqrt{\pi}\omega_{k-1}(1+\delta)^{k+1}} \mathcal{H}^{k-1}(B_1^{(k-1)}) = \frac{2\sigma_{ij}}{\sqrt{\pi}(1+\delta)^{k+1}}.$$

The previous inequality holds for every $\delta > 0$, thus if we let $\delta \downarrow 0$ we recover (B.45) and the proof of the Γ-$\limsup$ inequality is completed.

Compactness. To prove the last item of the theorem we proceed adapting the ideas of [25] for the flat case. Fix $i \in \{1, ..., P\}$ and define $m := \min_{j \neq i} \sigma_{ij}$. Then if $u \in \mathcal{M}$ we have that

$$\begin{aligned} E_h(u) &= \sum_{i,j} \sigma_{ij} \int_M u_j e^{-h\Delta_\xi} u_i d\mu \\ &\geq m \int_M (1-u_i) e^{-h\Delta_\xi} u_i d\mu \\ &= \frac{m}{2} \int_{M\times M} p(h,x,y)((1-u_i(x))u_i(y) + u_i(x)(1-u_i(y))d\mu(y)d\mu(x) \qquad \text{(B.52)} \\ &\geq \frac{m}{2} \int_{M\times M} p(h,x,y)|u_i(y) - u_i(x)|d\mu(y)d\mu(x). \end{aligned}$$

We now fix $C > 0$ to be determined later. By Stokes theorem $\int_M \nabla_x p(h,x,y)d\mu(y) = 0$, thus using the Gaussian upper bound (C.46) and the Gaussian lower bound (C.45) we observe

$$\begin{aligned} \int_M |De^{-Ch\Delta_\xi} u_i|_\xi &= \int_M |\nabla e^{-Ch\Delta_\xi} u_i|_x d\mu(x) \\ &= \int_M \left| \int_M \nabla_x p(h,x,y) u_i(y) \right|_x d\mu(x) \\ &\leq \int_{M\times M} |\nabla_x p(h,x,y)|_x |u_i(y) - u_i(x)| d\mu(y)d\mu(x) \\ &\leq \frac{C_1}{Q_1} \int_{M\times M} \frac{\mu(B_{\sqrt{C_2 Ch}}(x))}{\mu(B_{\sqrt{Ch}})} p\left(\frac{C_2 Ch}{Q_2}, x, y\right) |u_i(y) - u_i(x)| d\mu(y)d\mu(x). \end{aligned}$$

If we take $C = \frac{Q_2}{C_2}$, using the doubling property (B.5) and the bound (B.52) we end up with

$$\begin{aligned} \int_M |De^{-Ch\Delta_\xi} u_i|_\xi &\leq \frac{C_1}{Q_1} \int_{M\times M} \frac{\mu(B_{\sqrt{Q_2 h}}(x))}{\mu(B_{\sqrt{\frac{Q_2 h}{C_2}}}(x))} p(h,x,y)|u_i(y) - u_i(x)| d\mu(y)d\mu(x) \\ &\leq \tilde{C} E_h(u). \end{aligned}$$

In particular, using this with $u = u_h$ we see that for every $i \in \{1, ..., P\}$

$$\sup_{h>0} \int_M |De^{-Ch\Delta_\xi} u_i^h|_\xi < +\infty.$$

By Lemma B.5.6 we know that up to extracting a subsequence, for every $i \in \{1, ..., P\}$ there exists $v_i \in BV(M)$ such that

$$\lim_{h \downarrow 0} \|e^{-Ch\Delta_\xi} u_i^h - v_i\|_{L^1(M)} = 0.$$

The result now follows by observing that

$$\lim_{h \downarrow 0} \|e^{-Ch\Delta_\xi} u_i^h - u_i^h\|_{L^1(M)} = 0.$$ □

B.7 Appendix

Proof of Theorem B.5.16

The proof is a slight modification of [35, Theorem 1.4]. The idea is still to reduce the problem to a nonlocal Γ-convergence result, namely to reduce it to the following statement.

Theorem B.7.1. *Let M be a k-dimensional compact Riemannian submanifold of $\mathbf{R}^d$. Assume that η is as in Section B.2, let $\xi > 0$ be a smooth function on M. Given $\epsilon > 0$ and $u \in L^2(M)$ define*

$$G_\epsilon(u) := \frac{1}{\epsilon^2} \int_{M \times M} \frac{1}{\epsilon^k} \eta\left(\frac{|x-y|_d}{\epsilon}\right) (u(x) - u(y))^2 \xi(x)\xi(y) d\,\mathrm{Vol}_M(x) d\,\mathrm{Vol}_M(y),$$

where $|\cdot|_d$ denotes the Euclidean distance in $\mathbf{R}^d$. Then

$$\Gamma - \lim_{\epsilon \downarrow 0} G_\epsilon = 2C_2 E,$$

where E is the Dirichlet energy (B.5.16). *Moreover, for any $u \in C^\infty(M)$ we have that $\limsup_{\epsilon \downarrow 0} G_\epsilon(u) \leq 2C_2 E(u)$. Finally, we have the following compactness property: if $\epsilon_n \downarrow 0$ and u^n are such that*

$$\sup_{n \in \mathbf{N}} G_{\epsilon_n}(u^n) < +\infty, \ \sup_{n \in \mathbf{N}} \|u^n\|_{L^2(M)} < +\infty,$$

then the sequence u^n is precompact in $L^2(M)$.

To reduce the proof of Theorem B.5.16 to Theorem B.7.1 one proceeds along the same lines of the proof of [35, Theorem 1.4]. Let T_n be the optimal transport maps obtained by applying Theorem C.6.6. To follow the proof of [35, Theorem 1.4] the only additional observation is that, since M is a Riemannian submanifold of $\mathbf{R}^d$, for any $x, y \in M$ we have

$$|x - y|_d \leq d_M(x, y).$$

In particular this yields

$$\|I - T_n\|_\infty \leq \sup_{x \in M} d_M(x, T_n(x)).$$

This will give the reduction to Theorem B.5.16. In the case $k = 2$ the previous argument works as long as one assumes

$$\frac{\epsilon_n n^{1/2}}{\log^{3/4}(n)} \gg 1.$$

The extra logarithmic factor may be removed by using [13, Proposition 2.11].

We are left with proving Theorem B.7.1. This can in turn be deduced from the corresponding result in the Euclidean case, namely the following.

Theorem B.7.2. *Let $D \subset \mathbf{R}^k$ be a bounded open set with smooth boundary, let $\tilde{\xi} : \overline{D} \to (0, +\infty)$ be a smooth function. Then*

$$\Gamma - \lim_{\epsilon \downarrow 0} \tilde{G}_\epsilon^{D,\tilde{\xi}} = 2C_2 \tilde{E}^{D,\tilde{\xi}} \text{ in } L^2(D),$$

and $\lim_{\epsilon\downarrow 0} \tilde{G}_\epsilon^{D,\tilde{\xi}}(u) = 2C_2\tilde{E}^{D,\tilde{\xi}}(u)$ whenever $u \in L^2(\overline{D})$, where we set for $u \in L^2(D)$

$$\tilde{G}_\epsilon^{D,\tilde{\xi}}(u) = \frac{1}{\epsilon^{k+2}} \int_{D\times D} \eta\left(\frac{|x-y|_k}{\epsilon}\right) |u(x)-u(y)|^2 \tilde{\xi}(x)\tilde{\xi}(y) dx dy,$$

and

$$\tilde{E}^{D,\tilde{\xi}}(u) = \begin{cases} \frac{1}{2}\int_D |Du|^2 \tilde{\xi}^2 dx & \text{if } u \in H^1(D) \\ +\infty & \text{otherwise.} \end{cases}$$

Finally, the following compactness property holds true: if $\epsilon_n \downarrow 0$ and u^n are such that

$$\sup_{n\in\mathbf{N}} \tilde{G}_{\epsilon_n}^{D,\tilde{\xi}}(u^n) < +\infty, \ \sup_{n\in\mathbf{N}} \|u^n\|_{L^2(D)} < +\infty$$

then the sequence u^n is precompact in $L^2(D)$.

With this result, we can prove Theorem B.7.1.

Proof of Theorem B.7.1. Γ-lim inf. Let $u^n \to u$ in $L^2(M)$. We want to prove that for any sequence $\epsilon_n \downarrow 0$ we have

$$\liminf_{n\to+\infty} G_{\epsilon_n}(u^n) \geq 2C_2 E(u). \tag{B.53}$$

To this aim, we may assume as before that the left hand side of (B.53) is finite. For any given $R > 0$ the family

$$\mathcal{F}_R := \left\{ \overline{B_r(x)} \subset M : \ r \leq R, \ B_r(x) \subset\subset \Psi(U), \ \Psi \ 1-\text{Lipschitz chart} \right\}$$

is a Vitali covering. Since the manifold is compact, or more precisely by the doubling property (B.5), we can select countably many disjoint balls B_i in the above family, such that

$$\mathrm{Vol}_M\left(M \setminus \bigcup_{i=1}^{\infty} \overline{B_i}\right) = 0. \tag{B.54}$$

By construction, each ball B_i is contained in $\Psi_i(U_i)$ for some 1-Lipschitz local parametrization Ψ_i. Here 1-Lipschitz is understood between Euclidean spaces, i.e.

$$|\Psi_i(y_1) - \Psi_i(y_2)|_d \leq |y_1 - y_2|_k, \ y_1, y_2 \in U_i.$$

In particular we have that since η is non-increasing

$$\eta\left(\frac{|\Psi_i(y_1) - \Psi_i(y_2)|_d}{\epsilon_n}\right) \geq \eta\left(\frac{|y_1 - y_2|_k}{\epsilon_n}\right). \tag{B.55}$$

By (B.55) and since the balls are disjoint we obtain

$$\begin{aligned}
&G_{\epsilon_n}(u^n) \\
&\geq \sum_{i\in\mathbf{N}} \frac{1}{\epsilon_n^2} \int_{U_i\times U_i} \frac{1}{\epsilon_n^k} \eta\left(\frac{|x-y|_k}{\epsilon}\right) (u^n \circ \Psi_i(x) - u^n \circ \Psi(y))^2 \tilde{\xi}_i(x)\tilde{\xi}_i(y) dxdy \\
&= \sum_{i\in\mathbf{N}} \tilde{G}_{\epsilon_n}^{U_i,\tilde{\xi}_i}(u^n \circ \Psi_i),
\end{aligned}$$

where $\tilde{\xi}_i(y) = \xi(\Psi_i(y))\sqrt{\det(g)}$. Now an application of Fatou's Lemma, Theorem B.7.2 and (B.54) give the lim inf inequality.

Γ-lim sup. By a diagonal argument, we may reduce to proving the Γ-lim sup inequality in the case when $u \in H^1(M) \cap C^\infty(M)$. For such a function and any sequence $\epsilon_n \downarrow 0$ we claim that

$$\limsup_{n\to+\infty} G_{\epsilon_n}(u) \leq 2C_2 E(u).$$

We will in a second moment construct a suitable covering $\{W_1, ..., W_N\}$ which satisfies

$$M \subset \bigcup_{i=1}^{N} W_i,$$

where $N \in \mathbf{N}$ and $W_i = \Psi_i(U_i)$ with Ψ_i local parametrizations defined on a bounded domain $U_i \subset \mathbf{R}^k$ with Lipschitz boundary such that

$$|\Psi_i(y_1) - \Psi_i(y_2)|_d \geq |y_1 - y_2|_k \text{ for } y_1, y_2 \in U_i. \tag{B.56}$$

Let δ be the Lebesgue number of the given covering. Define the set

$$F_\delta := \{(x, y) \in M \times M : \ d(x, y) \geq \delta\}.$$

It is clear that F_δ is compact, thus by continuity we have that

$$|x - y|_d \geq c_\delta > 0, \ x, y \in F_\delta.$$

In particular by the exponential decay of η we get that there exists two positive constants c_1, c_2 such that

$$\eta\left(\frac{|x-y|_d}{\epsilon}\right) \leq c_1 \exp\left(\frac{-c_2 c_\delta}{\epsilon}\right), \ x, y \in F_\delta. \tag{B.57}$$

Now observe that

$$
\begin{aligned}
G_{\epsilon_n}(u) =& \frac{1}{\epsilon_n^{2+k}} \int_{F_\delta} \eta \left(\frac{|x-y|_d}{\epsilon_n} \right) |u(x)-u(y)|^2 \xi(x)\xi(y) d \operatorname{Vol}_M(x) d \operatorname{Vol}_M(y) \\
&+ \frac{1}{\epsilon_n^{2+k}} \int_{M \times M \setminus F_\delta} \eta \left(\frac{|x-y|_d}{\epsilon_n} \right) |u(x)-u(y)|^2 \xi(x)\xi(y) d \operatorname{Vol}_M(x) d \operatorname{Vol}_M(y).
\end{aligned}
$$

Recalling (B.57) we may observe that the first term on the right hand side converges to zero as $n \to +\infty$. We claim that the lim sup of the second right hand side term is bounded above by

$$
\sum_{i=1}^N \int_{W_i} |\nabla u|^2 \xi d \operatorname{Vol}_M .
$$

To show this, observe that if $(x,y) \in M \times M \setminus F_\delta$ then by definition there exists $1 \le i \le N$ such that $(x,y) \in W_i \times W_i$, in particular

$$
\begin{aligned}
&\frac{1}{\epsilon_n^{2+k}} \int_{M \times M \setminus F_\delta} \eta \left(\frac{|x-y|_d}{\epsilon_n} \right) |u(x)-u(y)|^2 \xi(x)\xi(y) d \operatorname{Vol}_M(x) d \operatorname{Vol}_M(y) \\
&\le \sum_{i=1}^N \frac{1}{\epsilon_n^{2+k}} \int_{W_i \times W_i} \eta \left(\frac{|x-y|_d}{\epsilon_n} \right) |u(x)-u(y)|^2 \xi(x)\xi(y) d \operatorname{Vol}_M(x) d \operatorname{Vol}_M(y) \\
&= \sum_{i=1}^N \tilde{G}_{\epsilon_n}^{U_i, \tilde{\xi}_i}(u \circ \Psi_i),
\end{aligned}
$$

where $\tilde{\xi}_i(y) = \xi(\Psi_i(y))\sqrt{\det(g)}$. Recalling Theorem B.7.2, if we let $n \to +\infty$ we obtain

$$
\limsup_{n \to +\infty} G_{\epsilon_n}(u) \le C_2 \sum_{i=1}^N \int_{W_i} |\nabla u|^2 \xi d \operatorname{Vol}_M .
$$

We now claim that given any $\alpha > 0$ we can find $N \in \mathbf{N}$ and a covering $W_1, ..., W_N$ as before such that

$$
\sum_{i=1}^N \int_{W_i} |\nabla u|^2 \xi d \operatorname{Vol}_M - \int_M |\nabla u|^2 \xi d \operatorname{Vol}_M < \alpha. \tag{B.58}
$$

This can be done as follows. Given any point $x \in M$ we can find a smooth function $\gamma : \mathbf{R}^k \to \mathbf{R}^{d-k}$ and a number $R > 0$ such that, upon translating and rotating the axes, the map

$$
\Psi(y) = (y, \gamma(y)), \; y \in Q(0,R) := (0,R)^k
$$

is a local parametrization around x. Clearly we have that (B.56) is true. Define $V_x := \Psi(Q(0, \frac{R}{2}))$. Since the manifold is compact, we can find $N \in \mathbf{N}$ and points $x_1, ..., x_N$ such that the sets $V_i := V_{x_i}$, $i = 1, ..., N$, cover M. Now define $\tilde{V}_1 = V_1$, $\tilde{V}_{i+1} = V_{i+1} \setminus \cup_{j=1}^i V_i$. Then $\{\tilde{V}_i\}$ is a partition of M. Define $A_i := \Psi_i^{-1}(V_i)$. Then the sets $A_i \subset Q(0, \frac{R_i}{2})$ have Lipschitz boundary. Given $\theta > 0$ sufficiently small define, for any $1 \le i \le N$

$$
\begin{aligned}
&A_i^\theta := \{y \in Q(0,R_i) : \ d(y,A_i) < \theta\}\,, \\
&W_i := \Psi_i(A_i^\theta).
\end{aligned}
$$

Clearly, $\{W_i, ..., W_N\}$ is an open covering satisfying (B.56). We now check that it satisfies (B.58) provided θ is small enough. Observe that there exists a constant $C > 0$ such that for any $1 \le i \le N$

$$
\begin{aligned}
\mathrm{Vol}_M(W_i \setminus \tilde{V}_i) &\le C\mathcal{L}^k\left(A_i^\theta \setminus A_i\right) \\
&\le C\mathcal{L}^k\left(\{y \in Q(0,R_i) : \ |y - \partial A_i|_k \le \theta\}\right).
\end{aligned}
$$

Recall that, since ∂A_i is $(k-1)$-rectifiable, we have

$$
\mathcal{H}^{k-1}(\partial A_i) = \lim_{\theta \downarrow 0} \frac{\mathcal{L}^k\left(\{y \in Q(0,R_i) : \ |y - \partial A_i|_k \le \theta\}\right)}{\theta}.
$$

The right-hand side is the Minkowski content, cf. [28, Theorem 3.2.39]. In particular for a given $\tilde{\alpha} > 0$, we can choose θ so small that

$$
\mathrm{Vol}_M(W_i \setminus \tilde{V}_i) \le C\tilde{\alpha}.
$$

Now observe that since $\tilde{V}_i \subset W_i$

$$
\begin{aligned}
&\sum_{i=1}^N \int_{W_i} |\nabla u|^2 \xi d\,\mathrm{Vol}_M - \int_M |\nabla u|^2 \xi d\,\mathrm{Vol}_M \\
&= \sum_{i=1}^N \int_{W_i} |\nabla u|^2 \xi d\,\mathrm{Vol}_M - \int_{\tilde{V}_i} |\nabla u|^2 \xi d\,\mathrm{Vol}_M \\
&\le \tilde{C} N \tilde{\alpha}.
\end{aligned}
$$

Choosing $\tilde{\alpha} = \frac{\alpha}{\tilde{C}N}$ we get (B.58). In particular

$$
\limsup_{n \to +\infty} G_{\epsilon_n}(u) \le 2C_2 E(u) + 2C_2\alpha,
$$

and letting $\alpha \downarrow 0$ we get the lim sup inequality.

The compactness property follows easily from Theorem B.7.2. □

Γ-convergence of the localized thresholding energies

Here we sketch the proof of Theorem B.6.1. The upper bound in the Γ-convergence is obtained by using Lemma 3.6 in [53]. For the lower bound, one just needs the following approximate monotonicity, which was proved by Otto and one of the authors in the first version of the preprint preceeding [53], but did not appear in the published version.

Theorem B.7.3. *Let $\sigma \in \mathbf{R}^{P \times P}$ be a symmetric matrix such that σ_{ij} satisfy the triangle inequality. Let $\beta \in C_c^\infty(B_1)$, where $B_1 \subset \mathbf{R}^k$ is the unit ball. For $t > 0$ define*

$E_t^{B_1}(\cdot,\beta)$ as in (B.51). Let $k_t(z)=\frac{1}{\sqrt{t}^k}k_1(\frac{z}{\sqrt{t}})$, with $k_1(z)=|z|G_1(z)$. Then, defining for $u:B_1\to[0,1]^P$ with $\sum_m u_m=1$,

$$\tilde{E}_t(u)=\frac{1}{\sqrt{t}}\sum_{i,j}\sigma_{ij}\int u_ik_t*u_jdx,$$

we have that for all such u and all $0<h\le h_0$

$$E_{h_0}^{B_1}(u,\beta)\le\left(\frac{\sqrt{h_0}+\sqrt{h}}{\sqrt{h_0}}\right)^{k+1}E_h^{B_1}(u,\beta)+C\|D\beta\|_{L^\infty}\tilde{E}_h(u)\sqrt{h_0}. \tag{B.59}$$

Here C is a constant that does not depend on h nor on h_0.

The original proof was based on the ideas used for proving the monotonicity of the non-localized thresholding energies in [25]. For the convenience of the reader, we include a proof for the simpler two phase setting. In that case one has to prove (B.59) with

$$\tilde{E}_t^{B_1}(u,\beta)=\frac{1}{\sqrt{t}}\int_{B_1}\beta(1-u)G_t*udx,\ u:B_1\to[0,1]$$

and

$$\tilde{E}_t(u)=\frac{1}{\sqrt{t}}\int_{B_1}k_t*(1-u)udx,\ u:B_1\to[0,1].$$

Proof of Theorem B.7.3 in the two phase setting. Clearly, statement (B.59) is a consequence of the following two items.

$$\sqrt{h_1}^{k+1}E_{h_1}^{B_1}(u,\beta)\le\sqrt{h_2}^{k+1}E_{h_2}^{B_1}(u,\beta)\qquad\forall 0<h_1\le h_2, \tag{B.60}$$

$$E_{N^2h}^{B_1}(u,\beta)\le E_h^{B_1}(u,\beta)+C(N-1)\sqrt{h}\|D\beta\|_\infty\tilde{E}_h(u)\qquad\forall N\in\mathbf{N},\ \forall h>0. \tag{B.61}$$

To see this, let $0<h\le h_0$. Let $N\in\mathbf{N}$ be such that

$$(N-1)\sqrt{h}\le\sqrt{h_0}<N\sqrt{h}.$$

Then we have

$$\begin{aligned}E_{h_0}^{B_1}(u,\beta)&\overset{(B.60)}{\le}\left(\frac{N\sqrt{h}}{\sqrt{h_0}}\right)^{k+1}E_{N^2h}^{B_1}(u,\beta)\\&\overset{(B.61)}{\le}\left(\frac{N\sqrt{h}}{\sqrt{h_0}}\right)^{k+1}\left(E_h^{B_1}(u,\beta)+C(N-1)\sqrt{h}\|D\beta\|_\infty\tilde{E}_h(u)\right)\\&\le\left(\frac{\sqrt{h}+\sqrt{h_0}}{\sqrt{h_0}}\right)^{k+1}E_h^{B_1}(u,\beta)+C\sqrt{h_0}\|D\beta\|_\infty\tilde{E}_h(u).\end{aligned}$$

We are thus left with proving (B.60) and (B.61).
Item (B.60). This follows by showing that

$$\frac{d}{d\sqrt{h}}\left(\sqrt{h}^{k+1}E_h^{B_1}(u,\beta)\right)\ge 0,$$

which follows by differentiation. Indeed,

$$\frac{d}{d\sqrt{h}}\left(\sqrt{h}^{k+1}E_h^{B_1}(u,\beta)\right) = \frac{d}{d\sqrt{h}}\int_{B_1}\beta(1-u)G_1\left(\frac{z}{\sqrt{h}}\right)udx$$
$$= -\frac{1}{\sqrt{h}}\int_{B_1}\beta(1-u)\nabla G_1\left(\frac{z}{\sqrt{h}}\right)\cdot\frac{z}{\sqrt{h}}udx \geq 0.$$

Item (B.61). We let $0 < h_0$ be such that $\sqrt{h_0} = \sqrt{h_1} + \sqrt{h}$. Then we observe that

$$\begin{aligned}\sqrt{h_0}E_{h_0}^{B_1}(u,\beta) &= \int_{B_1}\int_{\mathbf{R}^k}\beta(x)(1-u)(x)G_1(z)u(x-\sqrt{h_1}z-\sqrt{h}z)dzdx \\ &\leq \int_{B_1}\int_{\mathbf{R}^k}\beta(x)(1-u)(x-\sqrt{h_1}z)G_1(z)u(x-\sqrt{h_1}z-\sqrt{h}z)dzdx \\ &\quad + \int_{B_1}\int_{\mathbf{R}^k}\beta(x)(1-u)(x)G_1(z)u(x-\sqrt{h_1}z)dzdx,\end{aligned}$$

where we used the inequality

$$(1-u)u'' \leq (1-u')u'' + (1-u)u', \ \forall u,u',u'' \in [0,1],$$

applied to $u = u(x), u' = u(x-\sqrt{h_1}z), u'' = u(x-\sqrt{h}z-\sqrt{h_1}z)$. We record that the second term on the right hand side is equal to

$$\sqrt{h_1}E_{h_1}^{B_1}(u,\beta).$$

The other term is estimated as follows. First we change variable in x, and then we estimate $|\beta(x)-\beta(x-\sqrt{h_1}z)| \leq \|D\beta\|_\infty\sqrt{h_1}|z|$ to get

$$\begin{aligned}&\int_{\mathbf{R}^k}\int_{B_1-\sqrt{h_1}z}\beta(x+\sqrt{h}z)(1-u)(x)G_1(z)u(x-\sqrt{h}z)dzdx \\ &\leq \int_{\mathbf{R}^k}\int_{B_1-\sqrt{h_1}z}\beta(x)(1-u)(x)G_1(z)u(x-\sqrt{h}z)dxdz \\ &\quad + \sqrt{h_1}\|D\beta\|_\infty\int_{\mathbf{R}^k}\int_{B_1-\sqrt{h_1}z}(1-u)(x)|z|G_1(z)u(x-\sqrt{h}z)dzdx \\ &= \int_{\mathbf{R}^k}\int_{B_1}\beta(x)(1-u)(x)G_1(z)u(x-\sqrt{h}z)dxdz \\ &\quad + \sqrt{h_1}\|D\beta\|_\infty\int_{\mathbf{R}^k}\int_{B_1-\sqrt{h_1}z}(1-u)(x)k_1(z)u(x-\sqrt{h}z)dzdx,\end{aligned}$$

where in the last equality we used the fact that β is supported in B_1. Observe that

$$\begin{aligned}
&\int_{\mathbf{R}^k}\int_{B_1-\sqrt{h_1}z}(1-u)(x)k_1(z)u(x-\sqrt{h}z)dzdx \\
&\leq \int_{\mathbf{R}^k}\int_{\mathbf{R}^k}(1-u)(x)k_1(z)u(x-\sqrt{h}z)dzdx \\
&= \int_{\mathbf{R}^k}\int_{\mathbf{R}^k}(1-u)(x)k_h(z)u(x-z)dzdx \\
&= \int_{\mathbf{R}^k}\int_{\mathbf{R}^k}(1-u)(x)k_h(z-x)u(z)dzdx \\
&= \int_{\mathbf{R}^k}\int_{B_1}(1-u)(x)k_h(z-x)u(z)dzdx \\
&= \int_{B_1}k_h*(1-u)(z)u(z)dz = \sqrt{h}\tilde{E}_h(u).
\end{aligned}$$

Here we used that u is supported in B_1. Putting things together we obtain that

$$\begin{aligned}
\sqrt{h_0}E_{h_0}^{B_1}(u,\beta) \leq &\sqrt{h_1}E_{h_1}^{B_1}(u,\beta) + \sqrt{h}E_h^{B_1}(u,\beta) \\
&+ \sqrt{h_1}\sqrt{h}\|D\beta\|_\infty\tilde{E}_h(u).
\end{aligned} \tag{B.62}$$

If we now apply inductively (B.62) with $h_1 = (N-1)^2h$ and $h_0 = N^2h$ one gets

$$\begin{aligned}
N\sqrt{h}E_{N^2h}^{B_1}(u,\beta) &\leq N\sqrt{h}E_h^{B_1}(u,\beta) + \sum_{i=1}^{N-1} ih\|D\beta\|_\infty\tilde{E}_h(u) \\
&= N\sqrt{h}E_h^{B_1}(u,\beta) + \frac{(N-1)N}{2}h\|D\beta\|_\infty\tilde{E}_h(u).
\end{aligned}$$

Dividing by $N\sqrt{h}$ yields (B.61). □

Data Availability

The datasets generated during and/or analysed during the current study are available in the GitHub repository `https://github.com/jonalelmi/Data-th-en`.

APPENDIX C

LARGE DATA LIMIT OF THE MBO SCHEME FOR DATA CLUSTERING: CONVERGENCE OF THE DYNAMICS

Notation. In the present work, we make extensive use of the Landau symbols o, O. To explain these, we let $\{a_\omega\}_{\omega\in\Omega}$, $\{b_\omega\}_{\omega\in\Omega}$ be two families of real numbers, with $b_\omega > 0$, indexed by $\omega \in \Omega \subset \mathbf{R}$. Let $\omega_0 \in \mathbf{R} \cup \{-\infty, +\infty\}$ be a limit point for the set Ω, which will be clear from the context. We say that $a_\omega = O(b_\omega)$ if

$$\limsup_{\omega\to\omega_0} \frac{a_\omega}{b_\omega} < +\infty.$$

We say that $a_\omega = o(b_\omega)$ if

$$\lim_{\omega\to\omega_0} \frac{a_\omega}{b_\omega} = 0.$$

We also alternatively write $a_\omega \lesssim b_\omega$ for $a_\omega = O(b_\omega)$ and $a_\omega \ll b_\omega$ for $a_\omega = o(b_\omega)$. In the following, usually (Ω, ω_0) will be $(\mathbf{N}, +\infty)$ or $(\mathbf{R}^+, 0)$, and this will be clear from the context.

C.1 The MBO scheme on graphs

In this section, we describe the MBO algorithm on graphs originally given by Bertozzi et al. in [62, 78, 61]. We refer to [50] for more information about its use in data clustering. We consider a weighted connected graph $G = (V, W)$ with n vertices, with $W_{ii} = 0$ for every $i = 1, ..., n$. For each vertex $x_i \in V, i \in \{1, ..., n\}$, we can define

$$d(x_i) = \frac{1}{n}\sum_{j=1}^{n} w_{ij}.$$

We define $D := \mathrm{diag}(d(x_1), ..., d(x_n))$. We let $\mathcal{V} := \{u | u : V \to \mathbf{R}\}$, the set of functions defined on V, which we endow this with the inner product

$$\langle u, v\rangle_{\mathcal{V}} := \frac{1}{n}\sum_{i=1}^{n} d(x_i)u(x_i)v(x_i).$$

We define the random walk Laplacian $\Delta : \mathcal{V} \to \mathcal{V}$ as the operator induced by the matrix

$$\Delta := \left(I - \frac{1}{n}D^{-1}W\right).$$

One can check that Δ is non-negative and self-adjoint with respect to $\langle\cdot,\cdot\rangle_{\mathcal{V}}$, in particular, it has n eigenvalues (counted with multiplicity) which we order in the following way

$$0 = \lambda^1 \leq ... \leq \lambda^n.$$

We denote by $\{v^l\}_{1\leq l\leq n}$ a basis of corresponding eigenvectors, orthonormal with respect to $\langle\cdot,\cdot\rangle_{\mathcal{V}}$. For $0 < K \leq n$ we define a kernel $H^K : (0,+\infty) \times V \times V \to \mathbf{R}$ via

$$H^K(t,x,y) := \sum_{l=1}^{K} e^{-t\lambda^l} v^l(x) v^l(y) \frac{d(y)}{n}.$$

The choice $K = n$ corresponds to the heat kernel associated to Δ, which is the unique function $H : (0,+\infty) \times V \times V \to \mathbf{R}$ with the property that for every $u_0 \in \mathcal{V}$, the function

$$u(t,x) := e^{-t\Delta} u_0(x) := \sum_{y\in V} H(t,x,y) u_0(y),\ x \in V,\ t > 0$$

satisfies

$$\begin{cases} \partial_t u = -\Delta u & \text{on } (0,+\infty) \times V, \\ \lim_{t\downarrow 0} u(t,x) = u_0(x) & \text{on } V. \end{cases}$$

We are now ready to introduce the MBO scheme on graphs.

Algorithm C.1.1 (MBO scheme)**.** *Fix a time-step size $h > 0$ and initial conditions $\chi^0 : V \to \{0,1\}$. For each $l \in \mathbf{N}$ define inductively $\chi^{l+1} : V \to \{0,1\}$ as follows:*

1. ***Diffusion**. Define*
$$u^l := e^{-h\Delta} \chi^l.$$

2. ***Thresholding**. Define χ^{l+1} by*
$$\{\chi^{l+1} = 1\} = \left\{u^l \geq \frac{1}{2}\right\}.$$

We then define the piecewise constant in time, right-continuous interpolation

$$\chi^{h,G}(t,x) = \chi^l(x) \text{ for } t \in [lh,(l+1)h) \text{ and } x \in V.$$

We are interested in understanding whether this approximation is consistent at the level of the evolution by mean curvature flow on the manifold.

In practice, computing the exact diffusion in the first step of the algorithm may be computationally intractable. For this reason, one usually implements the MBO scheme by considering only a smaller number of eigenvectors of the Laplacian, say K. In other words, one uses the following more efficient variant of MBO.

Algorithm C.1.2 (Approximate MBO scheme)**.** *Fix a time-step size $h > 0$ and initial conditions $\chi^0 : V \to \{0,1\}$. For each $l \in \mathbf{N}$ define inductively $\chi^{l+1} : V \to \{0,1\}$ as follows:*

1. ***Diffusion**. Define*
$$u^l(x) := \sum_{y\in V} H^K(h,x,y) \chi^l(y).$$

2. ***Thresholding***. *Define* χ^{l+1} *by*

$$\{\chi^{l+1} = 1\} = \left\{u^l \geq \frac{1}{2}\right\}.$$

Again, we then define the piecewise constant in time, right-continuous interpolation

$$\chi^{h,G,K}(t,x) = \chi^l(x) \text{ for } t \in [lh,(l+1)h) \text{ and } x \in V.$$

At present, the choice of h and the exact value of K to pick in order to get a good approximation of the MBO scheme is obtained by trial and error. In this work, under the standard *manifold assumption*, we rigorously justify that an admissible regime to get a consistent result in the large-data limit is $K \geq (\log(n))^q$, $h \gg (\log(n))^{-\alpha}$ for some $q, \alpha > 0$ (see Theorem C.2.4 for the precise choices of q, α).

C.2 Main results

Hereafter $M \subset \mathbf{R}^d$ is a k-dimensional closed Riemannian submanifold. We denote by $\{x_i\}_{i=1}^{+\infty}$ a sequence of points on M, and for each $n \in \mathbf{N}$ we define weighted graphs $G_n = (V_n, W_n)$ where the vertex set V_n is given by $\{x_1, ..., x_n\}$ and the adjacency matrix $W_n = (w_{ij}^{(n,\epsilon_n)})_{1\leq i,j\leq n}$ is given by

$$\begin{aligned} w_{ii}^{(n,\epsilon_n)} &= 0 \text{ for } 1 \leq i \leq n, \\ w_{ij}^{(n,\epsilon_n)} &= \frac{1}{\epsilon_n^k}\eta\left(\frac{\|x_i - x_j\|_d}{\epsilon_n}\right) \text{ for } 1 \leq i,j \leq n,\ i \neq j. \end{aligned}$$

Here $\epsilon_n > 0$ are given length scales and $\eta : [0,+\infty) \to [0,+\infty)$ is a non-increasing function with support on the interval $[0,1]$, whose restriction to the interval $[0,1]$ is Lipschitz continuous. We define

$$C_1 := \int_{\mathbf{R}^k} \eta(|y|_k)dy, \quad C_2 := \int_{\mathbf{R}^k} \eta(|y|_k)y_1^2 dy, \quad \kappa(\eta) := \frac{C_2}{2C_1}.$$

We also define, for every $x \in M$ and every $n \in \mathbf{N}$

$$d_n(x) := \frac{1}{n}\sum_{j=1}^{n} \frac{1}{\epsilon_n^k}\eta\left(\frac{\|x - x_j\|_d}{\epsilon_n}\right)\mathbf{1}_{\{x\neq x_j\}}.$$

Note that, when $x = x_i$ for some $1 \leq i \leq n$, then $d_n(x)$ is the degree of the i-th node. We denote by $D_n := \operatorname{diag}(d_n(x_1), ..., d_n(x_n))$ the diagonal matrix of the degrees. The random walk Laplacian Δ_n is the linear operator induced by the $(n \times n)$-matrix given by

$$\Delta_n := \frac{1}{\epsilon_n^2}\left(I - \frac{1}{n}D_n^{-1}W_n\right).$$

We denote by $\{v_n^l\}_{1\leq l\leq n}$ an orthonormal basis (with respect to the inner product $\langle\cdot,\cdot\rangle_{\mathcal{V}_n}$) made of eigenvectors for the Laplacian Δ_n corresponding to the eigenvalues $\{\lambda_n^l\}_{1\leq l\leq n}$, which are ordered in the following way

$$0 = \lambda_n^1 \leq \lambda_n^2 \leq ... \leq \lambda_n^n.$$

Like in Section C.1, for every $0 < K \leq n$ we define

$$H_n^K(t, x, y) = \sum_{l=1}^{K} e^{-t\lambda_n^l} v_n^l(x) v_n^l(y) \frac{d_n(y)}{n},$$

and we set $H_n = H_n^n$ when $K = n$. Assume that we are given a sequence of operators $S_n : (0, +\infty) \times \mathcal{V}_n \to \mathcal{V}_n$ which are linear in the second variable, we then consider the following abstract version of the MBO scheme on the n-th graph.

Algorithm C.2.1 (Abstract MBO scheme)**.** *Fix a time-step size $h_n > 0$ and initial conditions $\chi^{0,G_n} : V_n \to \{0, 1\}$. For each $l \in \mathbf{N}$ define inductively $\chi^{l+1,G_n} : V_n \to \{0, 1\}$ as follows:*

1. ***Diffusion**. Define*
$$u_n^l := S_n(h_n, \chi^{l,G_n}).$$
2. ***Thresholding**. Define χ^{l+1,G_n} by*
$$\left\{\chi^{l+1,G_n} = 1\right\} = \left\{u_n^l \geq \frac{1}{2}\right\}.$$

We then define $\chi^{h_n,G_n} : [0, +\infty) \times V_n \to \{0, 1\}$ by

$$\chi^{h_n,G_n}(t, x) := \chi^{l,G_n}(x), \ x \in V_n, \ t \in [lh_n, (l + 1)h_n).$$

For convenience, we will mostly work with the $\{-1, 1\}$-valued functions

$$u^{h_n,G_n}(t, x) := 2\chi^{h_n,G_n}(t, x) - 1.$$

We also define the upper and lower limits of the family $\{u^{h_n,G_n}\}_{n\in\mathbf{N}}$ as

$$u^*(t, x) := \sup \left\{ \limsup_{n\to+\infty} u^{h_n,G_n}(t_n, x_n) \middle| \ t_n > 0, \ \lim_{n\to+\infty} t_n = t, \ x_n \in G_n, \ \lim_{n\to+\infty} x_n = x \right\}, \tag{C.1}$$

$$u_*(t, x) := \inf \left\{ \liminf_{n\to+\infty} u^{h_n,G_n}(t_n, x_n) \middle| \ t_n > 0, \ \lim_{n\to+\infty} t_n = t, \ x_n \in G_n, \ \lim_{n\to+\infty} x_n = x \right\}. \tag{C.2}$$

Let $\xi > 0$ be a smooth function on the manifold M. Let $\Omega \subset M$ be an open set with smooth boundary Γ_0. We let $u : [0, +\infty) \times M \to \mathbf{R}$ be the unique viscosity solution of the level set formulation of the mean curvature flow with density ξ (see Section C.3 for the details) with initial value $sd(\cdot, \Gamma_0) = d_M(x, \Omega^c) - d_M(x, \Omega)$, the signed distance function from Γ_0. For any $t > 0$ we also define

$$\Omega_t := \{x \in M \mid u(t, x) > 0\}, \ \Gamma_t = \{x \in M \mid u(t, x) = 0\}. \tag{C.3}$$

Let us denote by Δ_ξ the weighted Laplacian on M with weight $\mu := \xi\,\mathrm{Vol}_M$, i.e.,

$$\Delta_\xi f = -\frac{1}{\xi}\,\mathrm{div}\,(\xi\nabla f) \quad \text{for } f \in C^\infty(M).$$

Let $H : (0,+\infty) \times M \times M \to \mathbf{R}$ denote the corresponding heat kernel.

Our first main result is the following conditional convergence of the abstract formulation of the MBO scheme.

Theorem C.2.2. *Assume that:*

(i) The operators S_n satisfy the maximum principle up to errors $h_n^{3/2}$, i.e., for n large enough and for each $u, v \in \mathcal{V}_n$ it holds

$$u \le v \Rightarrow S_n(h_n, u) \le S_n(h_n, v) + \left(\max_{V_n}|u| + \max_{V_n}|v|\right) O(h_n^{3/2}).$$

(ii) The operators S_n approximate the heat operator on the manifold, i.e. there exists a constant $\kappa > 0$ such that for every function $f \in C^\infty(M)$ we have

$$\max_{x\in V_n}\left|S(h_n, f)(x) - e^{-h\kappa\Delta_\xi} f(x)\right| = (\sup|f|)\, o(\sqrt{h_n}) + \mathrm{Lip}(f)\, O(h_n^{3/2}). \tag{C.4}$$

where the functions $o(\sqrt{h_n}), O(h_n^{3/2})$ are independent of f.

(iii) The operators S_n almost preserve the total mass in the sense that

$$\max_{x\in V_n}|S_n(h_n, \mathbf{1}_{G_n})(x) - 1| = O(h_n^{3/2}).$$

Then u^* and u_* defined in (C.1) *and, respectively,* (C.2) *satisfy*

$$u_*(x,t) = 1 \quad \textit{if } x \in \Omega_t, \tag{C.5}$$
$$u^*(x,t) = -1 \; \textit{if } x \in (\Omega_t \cup \Gamma_t)^c. \tag{C.6}$$

Here Ω_t and Γ_t are defined as in (C.3).

Remark C.2.3. Let us compare Theorem C.2.2 with the work [68], where the authors prove convergence of the dynamics of the graph MBO scheme to a viscosity solution to mean curvature flow in the case of regular, two-dimensional grids. More precisely, they work in the following setting: the manifold M is the standard Euclidean plane $\mathbf{R}^2$, the sequence of graphs G_n are given by $G_n := \epsilon_n\mathbf{Z}^2$ for a sequence of localization parameters $\epsilon_n \downarrow 0$ and for $(i,j),(l,m) \in \mathbf{Z}^2$ one sets

$$w_{(i,j),(l,s)} = \begin{cases} 1 & \text{if } |i-l| + |m-j| = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{C.7}$$

In this way we can define an infinite dimensional weight matrix W_n whose entries are indexed by $\mathbf{Z}^2 \times \mathbf{Z}^2$ and are defined as (C.7). To put ourselves in a setting that is precisely the one we are working in we could actually work with $M = \mathbf{T}^2$, the 2-dimensional torus, and the sequence of graphs $G_n \cap \mathbf{T}^2$, but to keep the discussion simple we prefer to continue this discussion in the precise setting of [68]. Let $v : \epsilon_n \mathbf{Z}^2 \to \mathbf{R}$ be a function which is zero outside a compact subset of $\mathbf{R}^2$. We denote by $S_n(t,v) : [0,+\infty) \times \epsilon_n \mathbf{Z}^2 \to \mathbf{R}$ the solution to the heat equation on G_n with initial value v, i.e., $u := S_n(t,v)$ solves

$$\begin{cases} \dfrac{d}{dt} u(t,(i,j)) = \dfrac{1}{\epsilon_n^2}\Big[u(t,(i+1,j)) + u(t,(i-1,j)) \\ \qquad\qquad + u(t,(i,j+1)) + u(t,(i,j-1)) \qquad \text{for } (i,j) \in \epsilon_n \mathbf{Z}^2, \\ \qquad\qquad - 4u(t,(i,j))\Big] \\ u(0,(i,j)) = v((i,j)) \qquad\qquad \text{for } (i,j) \in \epsilon_n \mathbf{Z}^2. \end{cases}$$

In other words, $S_n(\cdot, v)$ is the heat operator on G_n applied to v. By using Fourier analysis methods, it can be shown that for every $h > 0$ and every $(x_1, x_2) \in \epsilon_n \mathbf{Z}^2$

$$S_n(h,v)((x_1,x_2)) = \sum_{(i,j)\in \epsilon_n \mathbf{Z}^2} Q_{i-x_1}\left(\frac{2h}{\epsilon_n^2}\right) Q_{j-x_2}\left(\frac{2h}{\epsilon_n^2}\right) v((i,j)),$$

where

$$Q_l(\alpha) := \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(l\xi) e^{\alpha(\cos(\xi)-1)} d\xi. \tag{C.8}$$

Using the asymptotic expansions [68, Proposition 3] for (C.8) it is not hard to prove that for any smooth, compactly supported function $f \in C_c^\infty(\mathbf{R}^2)$

$$\sup_{(i,j)\in\epsilon_n\mathbf{Z}^2} \left| S_n(h,f)((i,j)) - G_h^{\mathbf{R}^2} * f((i,j)) \right| = \mathrm{Lip}(f) o(\epsilon_n) + \sup|f| O\left(\frac{\epsilon_n^2}{h}\right) + \sup|f| O\left(\frac{\epsilon_n}{\sqrt{h}} \log\left(\frac{\epsilon_n}{\sqrt{h}}\right)\right), \tag{C.9}$$

where $G_h^{\mathbf{R}^2}$ denotes the heat kernel in the Euclidean plane at time h. In particular, when $\epsilon_n = h_n^\alpha$ for $\alpha \geq \frac{3}{2}$, we see that (C.9) implies (C.4). This allows us to use Theorem C.2.2 to recover the results of [68] when $\alpha \geq \frac{3}{2}$. Actually, an inspection of the proof of Theorem C.2.2 shows that to check that u^* and u_* are, respectively, a viscosity subsolution and a viscosity supersolution to mean curvature flow, the estimate (C.4) can be replaced by

$$\max_{x\in V_n} \left| S(h_n, f)(x) - e^{-h\kappa\Delta_\xi} f(x) \right| = (\sup|f|)\, o(\sqrt{h_n}) + \mathrm{Lip}(f)\, O(h_n^\gamma),$$

for some $\gamma > 1$. In particular, we see that in the setting of the two-dimensional regular grid this is satisfied whenever $\epsilon_n = h_n^\gamma$. This allows to recover the full parameter range $\gamma > 1$ of [68]. We need the slightly sharper assumption $\gamma = \frac{3}{2}$ for checking the initial conditions for u^* and u_*.

C.2.1 Results on the MBO scheme and on the approximate MBO scheme

The MBO scheme as stated in Algorithm C.1.1 corresponds to the choices $S_n(t,\cdot) = e^{-t\Delta_n}(\cdot)$, the heat semigroup on the n-th graph, which acts on functions $u \in \mathcal{V}_n$ by

$$e^{-t\Delta_n}(u)(x) = \sum_{y\in V_n} H_n(t,x,y)u(y).$$

Let $0 < K_n \leq n$ be a sequence of numbers converging to $+\infty$, then the approximate MBO scheme as stated in Algorithm C.1.2 corresponds to the choices $S_n = P_n$, where the operators P_n act on functions $u \in \mathcal{V}_n$ by

$$P_n(t,u)(x) := \sum_{y\in V_n} H_n^{K_n}(t,x,y)u(y). \tag{C.10}$$

Our second main result states that on random geometric graphs the operators $e^{-t\Delta_n}(\cdot)$ and P_n satisfy the assumptions of Theorem C.2.2 with high probability.

Theorem C.2.4. *Let us assume that $\nu := \rho\,\mathrm{Vol}_M$ is a probability measure with a smooth and positive density ρ. Assume that the points $\{x_i\}_{i=1}^{+\infty}$ in the above construction are i.i.d. random points sampled from M, distributed according to ν. Assume that $q > 0, \frac{2}{k} > s > 0$ are such that:*

(i) $q > \frac{1}{\frac{2}{k}-s}$,

(ii) We have that $\inf_{i\in\mathbf{N}}(\lambda_{i+1}-\lambda_i) > 0$.

(iii) $K_n \geq (\log(n))^q$,

(iv) $h_n \gg (\log(n))^{-\alpha}$, *with* $\alpha = -1 + \frac{2q}{k} - sq \geq 0$,

(v) $\epsilon_n \ll (\log(n))^{-\beta}$, *with* $\beta = -\frac{1}{2} + 4q + \frac{13q}{k} - \frac{sq}{2} \geq 0$,

(vi) We have

$$\epsilon_n \gtrsim \begin{cases} \left(\frac{\log(n)}{n}\right)^{\frac{1}{k}} & \text{if } k \geq 3,\\ \left(\frac{\log(n)}{n}\right)^{\frac{1}{8}} & \text{if } k = 2, \end{cases}$$

Then the operators $e^{-t\Delta_n}(\cdot)$ *and* P_n *satisfy conditions* (i), (ii) *and* (iii) *in Theorem C.2.2 (with $\xi = \rho^2$ and $\kappa = \kappa(\eta)$) on G_n with probability greater than*

$$1 - C\epsilon_n^{-6k}\exp(-\frac{n\epsilon_n^{k+4}}{C}) - Cn\exp(-\frac{n}{C(\log(n))^{2q}}).$$

Remark C.2.5. Let us comment on this second result.

(i) For each $k \geq 2$, the space of admissible parameters (s,q) in Theorem C.2.4 is quite *large*. To see this, we plot the space of admissible parameters. The shaded region represents the space of admissible pairs (s,q).

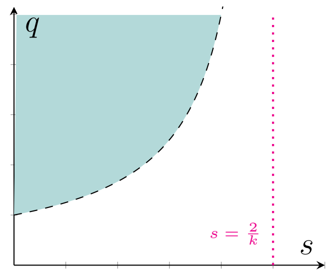

Figure C.1: Parameter space.

(ii) Condition (ii) in Theorem C.2.4 concerns the geometry of the manifold M. It implies in particular that the eigenvalues of the Laplacian Δ_{ρ^2} are simple. Condition (ii) in Theorem C.2.4 is for example satisfied by the k-torus and by the k-sphere with standard unit density, see [16, Chapter II, Section 2] and [16, Chapter II, Section 4].

(iii) Let us observe that conditions (v) and (vi) in Theorem C.2.4 are compatible, indeed the right-hand side of (v) in Theorem C.2.4 is a rational function of $\log(n)$, while the lower bound in condition (vi) in Theorem C.2.4 converges to zero as a power of n, up to a logarithmic factor. We also remark that items (iv) and (v) of Theorem C.2.4 imply

$$\epsilon_n \lesssim h_n^{3/2},$$

while we expect that the convergence of the scheme should be true up to the critical scaling

$$\epsilon_n \ll h_n.$$

Observe furthermore that condition (iv) in Theorem C.2.4 gives a lower bound for h_n of the form

$$h_n \gg (\log(\delta_n))^{\alpha},$$

where $\delta_n = (\frac{1}{n})^{1/k}$ is the characteristic distance between the nodes of the graph. This is perhaps not too surprising because the diffusion needs some time to smear out the fine details in the graph that appear at its characteristic length scale. A similar condition already appeared in [24].

(iv) In the proof of Theorem C.2.4 we will assume, for simplicity, that $K_n = \log(n)^q \in \mathbf{N}$. In this setting we will use condition (v) of Theorem C.2.4 in the form

$$\epsilon_n \ll \frac{\sqrt{\log(n)}}{K_n^{1+\frac{1}{k}-\frac{s}{2}} \left(\lambda_{K_n}^{\frac{2}{k}+1} + 1\right)^2 \left(\lambda_{K_n}^{4+\frac{k}{2}} + 1\right)}. \tag{C.11}$$

Observe that condition (v) of Theorem C.2.4 implies (C.11) because by Weyl's law we have $\lambda_{K_n} \sim K_n^{2/k}$.

Corollary C.2.6. *In the setting of Theorem C.2.4, if we additionally assume that*

$$\epsilon_n \gg \left(\frac{\log(n)}{n}\right)^{\frac{1}{k+4}}, \tag{C.12}$$

then the conclusion of Theorem C.2.2 holds almost surely both for the MBO scheme, Algorithm C.1.1, and the approximate MBO scheme, Algorithm C.1.2.

An important ingredient for the proof of Theorem C.2.4 is the following lemma, which gives an estimate of the distance between the approximate heat kernel on the graph and the heat kernel on the manifold in a uniform sense. Such heat kernel estimates are of independent interest, for example, one should compare with [24, Theorem 3], where the authors obtain a similar estimate when the frequency cut-off K_n and the time-scale h_n are fixed. In Lemma C.2.7 we improve their result by showing how to choose K_n in terms of n as $n \to +\infty$.

Lemma C.2.7. *In the setting of Theorem C.2.4, there exist constants* $a_1, a_2, a_3, a_4 > 0$ *such that if* n *is large enough, then, with probability greater than* $1 - a_1 \epsilon_n^{-6k} \exp(-a_2 n \epsilon_n^{k+4}) - a_3 n \exp(-a_4 \frac{n}{(\log(n))^{2q}})$*, we have*

$$\max_{x,y \in V_n} \left| H_{\epsilon_n}^{K_n}(h_n, x, y) - \frac{\rho(y)}{n} H(\kappa(\eta) h_n, x, y) \right| = o\left(\frac{\sqrt{h_n}}{n} \right). \tag{C.13}$$

C.3 The level set equation for MCF on a weighted manifold

In this section, we provide the basic framework for viscosity solutions to mean curvature flow in weighted Riemannian manifolds.

Hereafter (M, g) is a k-dimensional closed Riemannian manifold, and $\xi > 0$ is a smooth function on M. Recall that the evolution of a smooth open set Ω_0 by mean curvature follows the trajectory of steepest descent of the area functional, which is defined as

$$\Omega \mapsto \int_{\partial \Omega} dS,$$

where Ω ranges over all open sets in M with a smooth boundary. When we consider a weight ξ on the manifold, the correct functional to consider is the weighted-area functional, defined as

$$\Omega \mapsto \int_{\partial \Omega} \xi dS,$$

where Ω ranges over all open sets in M with smooth boundary. We define the evolution of mean curvature flow with density ξ - hereafter denoted as MCF_ξ - as the trajectory of steepest descent of this functional. To derive an equation for MCF_ξ we consider a family $\{\Omega(t)\}_{0 \le t < T}$ of smooth open sets evolving smoothly in time with normal velocity vector V. Denote by $\nu(t)$ a suitable extension of the outer unit normal of $\partial\Omega(t)$. We then have by Gauss' Theorem

$$\begin{aligned} \frac{d}{dt} \int_{\partial\Omega(t)} \xi dS &= \frac{d}{dt} \int_{\Omega(t)} \frac{1}{\xi} \operatorname{div}(\xi \nu(t)) \xi d \operatorname{Vol}_M \\ &= \int_{\partial\Omega(t)} \frac{1}{\xi} \operatorname{div}(\xi \nu(t)) g(V(t), \nu(t)) \xi dS. \end{aligned}$$

We thus see that the trajectory of steepest descent is given by

$$g(V, \nu) = -\frac{1}{\xi} \operatorname{div}(\xi \nu).$$

We are thus led to the following definition.

Definition C.3.1. Let (M, g) be a smooth k-dimensional closed Riemannian manifold. Let $\xi > 0$ be a smooth function on M. A family $\{\Omega_t\}_{t\geq 0}$ of smooth open subsets of M is said to evolve by MCF_ξ if

$$g(V, \nu) = -\frac{1}{\xi}\operatorname{div}(\xi\nu). \tag{C.14}$$

where V is the velocity vector field of the evolution and ν is the outer unit normal field.

Remark C.3.2. Using that the mean curvature $H(t)$ of $\partial\Omega(t)$ satisfies $\operatorname{div}(\nu(t)) = H(t)$, equation (C.14) can be rewritten as

$$g(V, \nu) = -H - g\left(\frac{\nabla\xi}{\xi}, \nu\right), \tag{C.15}$$

which yields the following interpretation for (C.14): the evolution by MCF_ξ as defined in Definition C.3.1 is driven by the minimization of two quantities, area and density. The first term on the right-hand side of (C.15) forces the evolution to follow a trajectory which decreases as much as possible the area of $\partial\Omega(t)$, whereas the second term on the right-hand side forces the evolution to move towards regions where the density ξ is low.

We now derive the corresponding level set formulation for the above evolution in the spirit of [27, 18]. Let $u : [0, +\infty) \times M \to \mathbf{R}$ be a smooth function, assume for this heuristic discussion that $Du \neq 0$ everywhere. For any $s \in \mathbf{R}$ define $\Omega_t^s := \{x \in M : u(t, x) > s\}$ and assume that $\{\Omega_t^s\}_{t\geq 0}$ evolves by MCF_ξ defined in Definition C.3.1. Let $s \in \mathbf{R}$ and let $x : (0, T) \to M$ a smooth curve such that $x(t) \in \partial\Omega_t^s$ for every time $0 < t < T$. Then

$$\begin{aligned} 0 &= \frac{d}{dt}u(t, x(t)) \\ &= (\partial_t u)(t, x(t)) + g(\nabla u(t, x(t)), \dot{x}(t)). \end{aligned}$$

Using the fact that the outer normal to the super level set Ω_t^s is given by $\nu(t, x) = -\frac{\nabla u(t,x)}{|\nabla u(t,x)|}$ and plugging in (C.14) we obtain

$$\begin{aligned} (\partial_t u)(t, x(t)) &= |\nabla u(t, x(t))| g(\nu(t, x(t)), V(t, x(t))) \\ &= -|\nabla u(t, x(t))| \frac{1}{\xi(x(t))}\operatorname{div}(\xi\nu)(t, x(t)). \end{aligned}$$

Using the product rule for the divergence and recalling that $\nu = -\frac{\nabla u}{|\nabla u|}$ we observe that u solves

$$\partial_t u = \left\langle g - \frac{Du \otimes Du}{|Du|^2}, D^2 u\right\rangle + g\left(\frac{\nabla\xi}{\xi}, \nabla u\right), \tag{C.16}$$

where we denoted by $\langle\cdot,\cdot\rangle$ the extension of g to the linear bundle of $T^*M \otimes T^*M$, i.e. for A, B sections of $T^*M \otimes T^*M$ we have in local coordinates

$$\langle A, B\rangle := \sum_{i,j,k,l=1}^{k} A_{ij} g^{jk} g^{kl} B_{li}.$$

From (C.16) we are led to the following definition.

Definition C.3.3. Let $u : (0,T)\times M \to \mathbf{R}$ be a smooth function with $Du \neq 0$ everywhere. Then u is said to solve the level set formulation of MCF$_\xi$ if (C.16) holds on $(0,T)\times M$.

Remark C.3.4. Another way of deriving directly equation (C.16) without relying on (C.14) is by computing the steepest descent of the total variation functional $\int_M |\nabla u|\xi d\,\mathrm{Vol}_M$ with respect to the metric

$$(\delta u, \delta u) = \int_M \left(\frac{\delta u}{|\nabla u|}\right)^2 |\nabla u|\xi d\,\mathrm{Vol}_M\,.$$

Indeed, consider a smooth function $u : (0,T)\times M \to \mathbf{R}$ with $Du \neq 0$, we then compute

$$\begin{aligned}\frac{d}{dt}\int_M |\nabla u(x,t)|\xi(x) d\,\mathrm{Vol}_M &= \int_M g\left(\frac{\nabla u(x,t)}{|\nabla u(x,t)|}, \nabla\partial_t u(x,t)\right)\xi(x) d\,\mathrm{Vol}_M \\ &= -\int_M \mathrm{div}\left(\xi\frac{\nabla u}{|\nabla u|}\right)(t,x)\partial_t u(t,x) d\,\mathrm{Vol}_M\,.\end{aligned}$$

Thus the steepest descent of the total variation functional with respect to the metric defined above is given by requiring

$$\partial_t u = |\nabla u|\frac{1}{\xi}\,\mathrm{div}\left(\xi\frac{\nabla u}{|\nabla u|}\right),$$

which is equivalent to (C.16).

We are now ready to introduce a weak notion of solution for (C.16) based on the notion of viscosity solution. In the context of mean curvature flow with constant density $\xi = 1$ it was introduced in [27] and [18] in the Euclidean case, and in [41] on curved manifolds. If $U \subset (0,T)\times M$ is an open set, $(t_0, x_0) \in U$ and if $u : (0,T)\times M \to \mathbf{R}$ is an upper (lower) semi-continuous function, a smooth function $\varphi : U \to \mathbf{R}$ is said to be tangent to u at (t_0, x_0) from above (below), if $u - \varphi$ has a local maximum (minimum) at (t_0, x_0).

Definition C.3.5. An upper (lower) semi-continuous function $u : (0,T)\times M \to \mathbf{R}$ is said to be a viscosity subsolution (supersolution) for (C.16) if for every $(t_0, x_0) \in (0,T)\times M$ and every smooth function φ tanget to u from above (below):

(i) If $D\varphi(t_0, x_0) \neq 0$ then

$$\partial_t\varphi \leq \left\langle g - \frac{D\varphi\otimes D\varphi}{|D\varphi|^2}, D^2\varphi\right\rangle + g\left(\frac{\nabla\xi}{\xi}, \nabla\varphi\right)\ (\geq)\quad \text{at } (t_0, x_0)$$

(ii) Otherwise there exists $\nu \in T^*_{x_0}M$ with $|\nu| \leq 1$ such that

$$\partial_t\varphi \leq \langle g - \nu\otimes\nu, D^2\varphi\rangle\ (\geq)\quad \text{at } (t_0, x_0)$$

We say that u is a viscosity solution if it is both a subsolution and a supersolution.

In [41] the author introduces the notion of viscosity subsolution/supersolution to mean curvature flow on a manifold (which corresponds to choosing the constant density $\xi = 1$) requiring continuity of the function u. We need to work with this slightly more general definition because the functions u_* and u^* in Theorem C.2.2 are not continuous. We recall the following useful characterization of Definition C.3.5, which says that we need to check condition (ii) only when also $D^2\varphi(t_0, x_0) = 0$.

Proposition C.3.6. *Let $u : (0, T) \times M \to \mathbf{R}$ be an upper (lower) semicontinuous function. Then u is a viscosity subsolution (supersolution) of the level set formulation of MCF$_\xi$ if and only if whenever φ is tangent to u at (t_0, x_0) from above (below),* (i) *is satisfied and if $D\varphi(t_0, x_0) = 0$ and $D^2\varphi(t_0, x_0) = 0$, then*

$$\partial_t\varphi(t_0, x_0) \leq 0 \ (\geq).$$

Proposition C.3.6 is proved in the Euclidean case in [6, Proposition 2.2]. On a manifold, the proof is analogous. We recall the following comparison principle.

Theorem C.3.7. *Let M be a closed k-dimensional Riemannian manifold. Let $\xi > 0$ be a smooth function on M. Let u be a subsolution of* (C.16) *on $(0, T] \times M$ and let v be a viscosity supersolution of* (C.16) *on $(0, T] \times M$. Define*

$$u^*(x) := \limsup_{y \to x,\, t \to 0} u(t, y), \ v_*(x) := \liminf_{y \to x,\, t \to 0} v(t, y).$$

Assume that $u^* \leq v_*$ and that either u^* or v_* is continuous. Then for every $t \in (0, T]$

$$u(t, \cdot) \leq v(t, \cdot).$$

Theorem C.3.7 is proved when $\xi = 1$ is the constant density and the functions u, v are assumed to be continuous in [41]. A careful look at the proof reveals that the same argument goes trough with the above assumptions. When $M = \mathbf{R}^k$ is the flat Euclidean space, an even more general version of Theorem C.3.7 can be found in [3, Theorem 18]. We also recall the following result concerning the existence of viscosity solutions, which can be again found in [41] for the case of a constant density $\xi = 1$.

Theorem C.3.8. *Let M be a k-dimensional closed Riemannian manifold, and let $\xi > 0$ be a smooth function on M. Let $u_0 : M \to \mathbf{R}$ be continuous. Then there exists a unique viscosity solution $u : [0, T) \times M \to \mathbf{R}$ to* (C.16) *such that $u(0) = u_0$.*

Finally, we recall the following *relabeling property*, which is proved in [41] in the case of a constant density $\xi = 1$.

Lemma C.3.9. *Let M be a k-dimensional closed Riemannian manifold, and let $\xi > 0$ be a smooth function on M. Let $u : [0, T) \times M \to \mathbf{R}$ be a viscosity solution to* (C.16)*. Then for every continuous map $\Psi : \mathbf{R} \to \mathbf{R}$, the function $v := \Psi \circ u$ is a viscosity solution to* (C.16)*.*

C.4 MBO scheme on manifolds

As in the previous section, M will denote a k-dimensional closed Riemannian manifold and $\xi > 0$ will denote a smooth function on M. The following algorithm can be used to approximate the evolution of an open set $\Omega_0 \subset M$ with smooth boundary by MCF_ξ.

Algorithm C.4.1 (MBO scheme on manifolds)**.** *Fix a time-step size $h > 0$, a diffusion coefficient $\kappa > 0$ and a (bounded) drift $f : M \to \mathbf{R}$. Let $\Omega_0 \subset M$ be an open set with a smooth boundary. For each $n \in \mathbf{N}$ define inductively Ω_{l+1} as follows.*

1. ***Diffusion**. Define*
$$u_l := e^{-h\kappa\Delta_\xi}\mathbf{1}_{\Omega_l}.$$
2. ***Thresholding**. Define Ω_{n+1} by*
$$\Omega_{l+1} = \left\{ u_l \geq \frac{1}{2} + f\sqrt{h} \right\}.$$

We then have the following result for one step of MBO.

Theorem C.4.2. *Let M, ξ be as above. Let Ω_0 be a smooth open set such that* $\mathrm{diam}(\Omega_0) < \frac{\mathrm{inj}(M)}{2}$*. Let Ω_1 be obtained by applying one step of MBO with a bounded drift $f : M \to \mathbf{R}$ to Ω_0 with a given step size $h > 0$ and a given diffusion coefficient $\kappa > 0$. Let $x \in \partial\Omega_0$. Let $\nu(x) \in T_xM$ be the outer unit normal to $\partial\Omega_0$ at x and define*

$$z(x) := \begin{cases} \sup\{s \in \mathbf{R}^- \mid \ \exp_x(s\nu(x)) \in \Omega_1\} & \text{if } x \notin \Omega_1, \\ \inf\ \{s \in \mathbf{R}^+ \mid \ \exp_x(s\nu(x)) \notin \Omega_1\} & \text{if } x \in \Omega_1. \end{cases}$$

Then we have

$$|z(x)| \leq Vh,$$

where the constant V depends only on κ, the L^∞-norm of f, the ambient manifold M, and the C^0-norm of the second fundamental form of $\partial\Omega_0$.

Corollary C.4.3. *Let $x_0 \in M$ and $R < \frac{\mathrm{inj}(M)}{4}$ be fixed. Then there is a constant $C_R < +\infty$ such that if $\frac{R}{2} < r \leq R$ and, in the above theorem, $\Omega_0 = B_r(x_0)$, then*

$$|z(x)| \leq C_R h$$

for every $x \in \partial B_r(x_0)$.

Finally, we have the following consistency result, which will be crucial in proving Theorem C.2.2.

Theorem C.4.4. *Let h_n be a sequence of positive real numbers converging to zero. Assume that $\psi_{h_n} : (0, +\infty) \times M \to \mathbf{R}$ are $C^{1,2}((0, +\infty) \times M)$ functions converging in $C^{1,2}((0, +\infty) \times M)$ to a function $\psi : (0, +\infty) \times M \to \mathbf{R}$. Assume that $(s_{h_n}, z_{h_n}) \in (0, +\infty) \times M$ are converging to a point $(s, z) \in [0, +\infty) \times M$. Assume also that $\delta_n := \psi_{h_n}(s_{h_n}, z_{h_n})$ are such that*

$$\lim_{n\to+\infty} \frac{\delta_n}{\sqrt{h_n}} = 0. \tag{C.17}$$

Then we have that:

(i) If $D\psi(s,z) \neq 0$ then

$$\liminf_{n\to+\infty} \frac{1}{\sqrt{\kappa h_n}}\left(\frac{1}{2} - \int_{\{\psi_{h_n}(t_{h_n}-h_n,\cdot)\geq 0\}} H(\kappa h_n, z_{h_n}, y)\xi(y)d\,\mathrm{Vol}_M\right)$$
$$\geq \frac{1}{2\sqrt{\pi}|D\psi(s,z)|}\left(\partial_t\psi - \left\langle g - \frac{D\psi\otimes D\psi}{|D\psi|^2}, D^2\psi\right\rangle - g\left(\frac{\nabla\xi}{\xi}, \nabla\psi\right)\right)(s,z). \tag{C.18}$$

(ii) Otherwise if $D\psi(s,z) = 0, D^2\psi(s,z) = 0$ and

$$\frac{1}{2} - \int_{\{\psi_{h_n}(t_{h_n}-h_n,\cdot)\geq 0\}} H(\kappa h_n, z_{h_n}, y)\xi(y)d\,\mathrm{Vol}_M \leq o(\sqrt{h_n}),$$

then

$$\partial_t\psi(s,z) \leq 0.$$

C.5 Proofs

C.5.1 Conditional convergence: Proof of Theorem C.2.2

The purpose of this section is the proof of Theorem C.2.2, which is inspired by the works [6] and [68].

Proof of Theorem C.2.2. Let u be the unique viscosity solution to MCF_ξ from Theorem C.3.8 with $\xi = \rho^2$, starting from $u(0,\cdot) = sd(\cdot,\Gamma_0) := d_M(x,\Omega_0^c) - d_M(x,\Omega_0)$. We will show later that u^* and u_* are, respectively, a viscosity subsolution and a viscosity supersolution of the level set formulation of MCF_ξ according to Definition C.3.5. We furthermore claim that for every $x \in M$,

$$u^*(0,x) \leq \mathrm{sign}^*(u(0,x)), \tag{C.19}$$
$$u_*(0,x) \geq \mathrm{sign}_*(u(0,x)), \tag{C.20}$$

where sign^* and sign_* are, respectively, the upper semi-continuous envelope and the lower semi-continuous envelope of the sign function.

Once these facts are proved, it follows from Theorem C.3.7 that for every $x \in M$ and every $t \geq 0$,

$$u^*(t,x) \leq \mathrm{sign}^*(u(t,x)), \tag{C.21}$$
$$u_*(t,x) \geq \mathrm{sign}_*(u(t,x)). \tag{C.22}$$

To see this, we observe that if $\Psi : \mathbf{R} \to \mathbf{R}$ is a continuous function such that $\Psi \geq \mathrm{sign}^*$, then the relabeling property in Lemma C.3.9 implies that $\Psi \circ u$ is a continuous solution to (C.16) with $u^*(0,x) \leq \mathrm{sign}^*(u(0,x)) \leq \Psi(u(0,x))$ for every $x \in M$, thus Theorem C.3.7 implies that for every $0 \leq t \leq T$ and every $x \in M$

$$u^*(t,x) \leq \inf_{\Psi\in C(\mathbf{R}),\Psi\geq\mathrm{sign}^*} \Psi(u(t,x)) = \mathrm{sign}^*(u(t,x)).$$

A similar argument gives (C.22). Let us now conclude the proof of the theorem assuming that (C.21) and (C.22) hold. If $x \in \Omega_t$, then $u(t,x) > 0$, thus (C.22) yields $u_*(t,x) = 1$. In a similar way (C.21) implies that $u^*(t,x) = -1$ on $(\Omega_t \cup \Gamma_t)^c$. We are thus left with proving that u^* is a subsolution, that u_* is a supersolution and with verifying the initial conditions (C.19) and (C.20).

We now show that indeed u^* is a viscosity subsolution. Pick a test functions φ tangent to u^* at $(t_0, x_0) \in (0, +\infty) \times M$ from above. We may assume without loss of generality that

$$\lim_{t \to +\infty} \max_M \varphi(t, \cdot) = +\infty, \tag{C.23}$$

and that $u^* - \varphi$ has a strict global maximum at (t_0, x_0). Thanks to Proposition C.3.6, we only need to check that

1. Either $D\varphi(t_0, x_0) \neq 0$ and

$$\partial_t \varphi \leq \left\langle g - \frac{D\varphi \otimes D\varphi}{|D\varphi|^2}, D^2\varphi \right\rangle + g\left(\frac{\nabla \xi}{\xi}, \nabla\varphi\right) \text{ at } (t_0, x_0).$$

2. Or $D\varphi(t_0, x_0) = 0$, $D^2\varphi(t_0, x_0) = 0$ and

$$\partial_t \varphi(t_0, x_0) \leq 0.$$

If $(t_0, x_0) \in \{u^* = -1\}$ or $(t_0, x_0) \in \operatorname{Int}\{u^* = 1\}$ the claim is trivial, because in that case u^* is constant in a neighborhood of (t_0, x_0). We thus assume that $(t_0, x_0) \in \partial\{u^* = 1\}$. By definition, there exists a sequence (t_{n_j}, z_{n_j}) such that $z_{n_j} \in G_{n_j}$ for every $j \in \mathbf{N}$ and, as $j \to +\infty$,

$$\begin{aligned} n_j &\to +\infty, \\ z_{n_j} &\to x_0, \\ t_{n_j} &\to t_0, \\ u^{n_j, G_{n_j}}(t_{n_j}, z_{n_j}) &\to u^*(t_0, x_0). \end{aligned}$$

For every $j \in \mathbf{N}$, pick

$$(s_j, x_j) \in \operatorname{argmax}_{x \in G_{n_j}, s \in (0,+\infty)} \left\{ u^{n_j, G_{n_j}}(s, x) - \varphi(s, x) \right\}. \tag{C.24}$$

We observe that, up to extracting a subsequence, $(s_j, x_j) \to (t_0, x_0)$ as $j \to +\infty$. Indeed by the compactness of M and the assumption (C.23), we may assume that the sequence (s_j, x_j) converges to some limit point $(\underline{s}, \underline{x})$. Then by definition of u^*, by the choice (C.24) and by the properties of the points (t_{n_j}, z_{n_j}) we must have

$$\begin{aligned} (u^* - \varphi)(\underline{s}, \underline{x}) &\geq \limsup_{j \to +\infty} \, (u^{n_j, G_{n_j}} - \varphi)(s_j, x_j) \\ &\geq \limsup_{j \to +\infty} \, (u^{n_j, G_{n_j}} - \varphi)(t_{n_j}, z_{n_j}) \\ &= (u^* - \varphi)(t_0, x_0). \end{aligned}$$

This forces $(t_0, x_0) = (\underline{s}, \underline{x})$, because (t_0, x_0) is a strict global maximum for $u^* - \varphi$. It is also easy to check that $u^{n_j, G_{n_j}}(s_j, x_j) = 1$ for j large enough. We now pick a sequence

$\delta_j \downarrow 0$ to be determined later, and we define $\theta_j : \mathbf{R} \to [-1,1]$ to be a smooth function such that

$$\theta_j(t) = \operatorname{sign}(t) \text{ for } |t| \geq \delta_j,$$
$$\|\theta_j'\|_\infty \leq \frac{2}{\delta_j}.$$

We claim that

$$u^{n_j,G_{n_j}}(s,z) \leq \theta_j(\varphi(s,z) - \varphi(s_j,x_j) + \delta_j) \tag{C.25}$$

for every j large enough, $z \in G_{n_j}$ and $s \in (0,+\infty)$. Indeed, inequality (C.25) holds trivially if $u^{n_j,G_{n_j}}(s,z) = -1$. If instead $u^{n_j,G_{n_j}}(s,z) = 1$, probing (C.24) with (s,z), we have

$$\begin{aligned} 1 = u^{n_j,G_{n_j}}(s,z) &\leq u^{n_j,G_{n_j}}(s_j,x_j) - \varphi(s_j,x_j) + \varphi(s,z) \\ &= 1 - \varphi(s_j,x_j) + \varphi(s,z), \end{aligned}$$

where we used that $u^{n_j,G_{n_j}}(s_j,x_j) = 1$ for j large enough. In particular

$$0 \leq -\varphi(s_j,x_j) + \varphi(s,z),$$

which, by definition of θ_j, yields (C.25).

We now choose $s = s_j - h_{n_j}$ in (C.25), we apply $S_{n_j}(h_{n_j},\cdot)$ to both sides of the inequality and we evaluate at x_j. Recalling assumption (i) of Theorem C.2.2 we get

$$\begin{aligned} &S_{n_j}(h_{n_j}, u^{n_j,G_{n_j}}(s_j - h_{n_j},\cdot))(x_j) \\ &\leq S_{n_j}(h_{n_j}, \theta_j(\varphi(s_j - h_{n_j},\cdot) - \varphi(s_j,x_j) + \delta_j))(x_j) + O\left(h_{n_j}^{3/2}\right). \end{aligned}$$

We now apply sign^* to both sides of the inequality to get

$$1 = u^{n_j,G_{n_j}}(s_j,x_j) \leq \operatorname{sign}^*\left(S_{n_j}(h_{n_j}, \theta_j(\varphi(s_j - h_{n_j},\cdot) - \varphi(s_j,x_j) + \delta_j))(x_j) + O\left(h_{n_j}^{3/2}\right)\right),$$

which, by definition of the function sign^*, implies

$$0 \leq S_{n_j}(h_{n_j}, \theta_j(\varphi(s_j - h_{n_j},\cdot) - \varphi(s_j,x_j) + \delta_j))(x_j) + O\left(h_{n_j}^{3/2}\right).$$

We now divide both sides of the previous inequality by 2 and we add $1/2$ to both sides of the inequality. Using assumption (iii) of Theorem C.2.2 and the linearity of S_n in the second variable yields

$$\frac{1}{2} \leq S_{n_j}\left(h_{n_j}, \left(\frac{1+\theta_j}{2}\right)\left(\varphi(s_j - h_{n_j},\cdot) - \varphi(s_j,x_j) + \delta_j\right)\right)(x_j) + O\left(h_{n_j}^{3/2}\right).$$

Define

$$f_j(z) := \left(\frac{1+\theta_j}{2}\right)\left(\varphi(s_j - h_{n_j},z) - \varphi(s_j,x_j) + \delta_j\right).$$

Then by applying the estimate (C.4) in assumption (ii) in Theorem C.2.2 we obtain

$$\frac{1}{2} \leq (e^{-h_{n_j}\kappa\Delta_\xi} f_j)(x_j) + o(h_{n_j}^{1/2}) + \frac{2}{\delta_j} O(h_{n_j}^{3/2}).$$

In other words, we have

$$o\left(h_{n_j}^{1/2}\right)+\frac{2}{\delta_j}O\left(h_{n_j}^{3/2}\right)\geq\frac{1}{2}-\int_M H(h_{n_j}\kappa,x_j,y)f_j(y)\xi(y)d\,\mathrm{Vol}_M(y)$$
$$\geq\frac{1}{2}-\int_{\{\varphi(s_j-h_{n_j},\cdot)-\varphi(s_j,x_j)+\delta_j\geq 0\}}H(h_{n_j}\kappa,x_j,y)\xi(y)d\,\mathrm{Vol}_M(y).$$

We divide the previous inequality by $\sqrt{h_{n_j}\kappa}$, and we choose $\delta_j=h_{n_j}^{2/3}$ so that on the one hand $\frac{h_{n_j}}{\delta_j}\to 0$ and on the other hand we can apply Theorem C.4.4. If $D\varphi(t_0,x_0)\neq 0$, then by (i) in Theorem C.4.4,

$$0\geq\frac{1}{2\sqrt{\pi}|D\psi(s,z)|}\left(\partial_t\psi-\left\langle g-\frac{D\psi\otimes D\psi}{|D\psi|^2},D^2\psi\right\rangle-g\left(\frac{D\xi}{\xi},D\psi\right)\right)(t_0,x_0),$$

which gives (i) in Definition C.3.5. If $D\varphi(t_0,x_0)=0$ and $D^2\varphi(t_0,x_0)=0$ then we can apply (ii) in Theorem C.4.4 to get the second item in the equivalent description of viscosity subsolution in Proposition C.3.6. Thus u^* is a viscosity subsolution. In a similar way one can prove that u_* is a supersolution.

We are left with checking the initial conditions for u^* and u_*. Again, we focus on the inequality (C.19) for u^*, since the argument for u_* is similar. Observe that

$$\mathrm{sign}^*(u(0,x))=\begin{cases}1 & \text{if } x\in\overline{\Omega_0}\\ -1 & \text{if } x\in M\setminus\overline{\Omega_0}\end{cases}$$

and since $u^*\in\{-1,1\}$, we just have to show that $u^*(0,x)=-1$ for $x\in M\setminus\overline{\Omega_0}$. To this aim, pick a sequence $(t_n,z_n)\in(0,+\infty)\times G_n$ such that $t_n\to 0$ and $z_n\to x$ as $n\to+\infty$. We have to show that $u^{n,G_n}(t_n,z_n)=-1$ for n large enough. For $q\in\mathbf{R}$, denote by $T^{q,G_n}(h_n)(\Omega_0)$ the outcome of the abstract thresholding scheme with thresholding value given by q and step size h_n on the graph G_n with initial value $\Omega_0\cap V_n$. For $m\in\mathbf{N}$ we also write $(T^{q,G_n}(h_n))^m$ for $T^{q,G_n}(h_n)\circ\ldots\circ T^{q,G_n}(h_n)$. Since $x\in M\setminus\overline{\Omega_0}$ there exists $R>0$ such that $B_R(x)\subset M\setminus\overline{\Omega_0}$. We denote by $w_n:V_n\to[0,+\infty)$ a sequence of non-negative functions which, for n large enough and for every $u,v\in\mathcal{V}_n$, $|u|\leq 1$, $|v|\leq 1$, satisfy

$$u\leq v\Rightarrow S(h_n,u)\leq S(h_n,v)+w_n, \tag{C.26}$$
$$a_n:=\|w_n\|_{L^\infty(G_n)}=O(h_n^{3/2}),$$
$$\max_{x\in V_n}|S(h_n,\mathbf{1}_{G_n})(x)-1|<a_n. \tag{C.27}$$

Such functions exist by assumptions (i) and (iii) in Theorem C.2.2. We now check that

$$V_n\setminus\left(T^{1/2,G_n}(h_n)\right)^m(\Omega_0)\supset\left(T^{1/2+2ma_n,G_n}(h_n)\right)^m(B_R(x)). \tag{C.28}$$

To see this, we proceed by induction over m. We treat just the base case $m=1$, the inductive step being analogous. To prove (C.28) for $m=1$, we show

$$V_n\setminus T^{1/2,G_n}(h_n)(\Omega_0)\supset T^{1/2+a_n,G_n}(h_n)(M\setminus\Omega_0)\supset T^{1/2+2a_n,G_n}(h_n)(B_R(x)). \tag{C.29}$$

To see this, let $y\in T^{1/2+a_n,G_n}(h_n)(M\setminus\Omega_0)$, observe that by (C.27) we have

$$S(h_n,\mathbf{1}_{\Omega_0})(y)+\frac{1}{2}+a_n\leq S(h_n,\mathbf{1}_{\Omega_0})(y)+S_n(h_n,\mathbf{1}_{M\setminus\Omega_0})(y)<1+a_n,$$

in particular, we have that $y \in V_n \setminus T^{1/2,G_n}(h_n)(\Omega_0)$. Thus $V_n \setminus T^{1/2,G_n}(h_n)(\Omega_0) \supset T^{1/2+a_n,G_n}(M \setminus \Omega_0)$. We now observe that since $\mathbf{1}_{B_R(x)} \leq \mathbf{1}_{M\setminus\Omega_0}$, (C.26) yields that for $y \in T^{1/2+2a_n,G_n}(h_n)(B_R(x))$

$$\frac{1}{2} + 2a_n \leq S(h_n, \mathbf{1}_{B_R(x)})(y) \leq S(h_n, \mathbf{1}_{M\setminus\Omega_0})(y) + a_n,$$

which yields (C.29).

We will show that there is a constant $C < +\infty$ such that

$$\left(T^{1/2+2\left[\frac{t_n}{h_n}\right]a_n,G_n}(h_n)\right)^{\left[\frac{t_n}{h_n}\right]}(B_R(x)) \supset B_{R-Ct_n}(x) \cap V_n. \tag{C.30}$$

Once this is proved, we have that using also (C.28), since $t_n \downarrow 0$,

$$M \setminus \left(T^{1/2,G_n}(h_n)\right)^{\left[\frac{t_n}{h_n}\right]}(\Omega_0) \supset B_{\frac{R}{2}}(x)$$

when n is large enough. In particular, since z_n is converging to x, we must have that $u^{n,G_n}(t_n, z_n) = -1$ for n large enough. Finally, to show (C.30) we argue as follows. Let C_R be the constant in Corollary C.4.3. Let $f \in C_c^\infty(B_R(x))$ such that $\mathbf{1}_{B_{R-C_R h_n}(x)} \leq f \leq \mathbf{1}_{B_R(x)}$ with $\mathrm{Lip}(f) \leq c/h_n$, using assumptions (i) and (ii) in Theorem C.2.2 we have for $y \in M \cap V_n$

$$\begin{aligned} S_n(h_n, \mathbf{1}_{B_R(x)})(y) &\geq S_n(h_n, f)(y) + O(h_n^{3/2}) \\ &\geq e^{-h_n\kappa\Delta_\xi} f(y) + O(h_n^{1/2}) \\ &\geq e^{-h_n\kappa\Delta_\xi} \mathbf{1}_{B_{R-C_R h_n}(x)}(y) + O(h_n^{1/2}). \end{aligned}$$

Observe that $\frac{1}{2} + 2\left[\frac{t_n}{h_n}\right] a_n = \frac{1}{2} + O(h_n^{1/2})$, in particular, we can apply Corollary C.4.3 to obtain, for n large enough, whenever $y \in B_{R-2C_R h_n}(x) \cap V_n$

$$e^{-h_n\kappa\Delta_\xi} \mathbf{1}_{B_{R-C_R h_n}(x)}(y) + O(h_n^{1/2}) \geq \frac{1}{2} + 2\left[\frac{t_n}{h_n}\right] a_n.$$

By an induction argument we get (C.30). □

C.5.2 Heat kernel estimate in random geometric graphs: Proof of Theorem C.2.4

The main purpose of this subsection is the proof of Theorem C.2.4. We first introduce some notation. We denote by $\{\lambda_l\}_{l=1}^{+\infty}$ the eigenvalues of the weighted Laplacian Δ_{ρ^2} on the manifold (M, g), which are ordered in the following way (recall that we are assuming that the eigenvalues are simple)

$$0 = \lambda_1 < \lambda_2 < \lambda_3 < \ldots$$

We denote by $\{f_l\}_{l=1}^{+\infty}$ an orthonormal basis (with respect to the $L^2(\rho^2\,\mathrm{Vol}_M)$-inner product on M) made of the corresponding eigenvectors. Then, for $x, y \in M$, the heat kernel on M can be written as

$$H(t, x, y) = \sum_{l=1}^{+\infty} e^{-t\lambda_l} f_l(x) f_l(y). \tag{C.31}$$

Proof of Theorem C.2.4. As we pointed out in Remark C.2.5, in the present proof we will for simplicity assume that $K_n = \log(n)^q \in \mathbf{N}$. We will indicate by γ the quantity $\gamma := \inf_{i\in\mathbf{N}}(\lambda_{i+1} - \lambda_i)$, which is positive by Item (ii) in Theorem C.2.4.

Observe that items (i) and (iii) in Theorem C.2.2 hold exactly (i.e. without error) for the choice $S_n(t,\cdot) = e^{-t\Delta_n}(\cdot)$. To show that these hold true with high probability also for the choice $S_n = P_n$ defined in (C.10) we take $w \in \mathcal{V}_n$ and we consider, for $x \in V_n$, the difference

$$\begin{aligned}\left|e^{-h_n\Delta_n}w(x) - P_n(h_n,w)(x)\right| &= \left|\sum_{y\in V_n}\sum_{l=K_n+1}^{n} e^{-h_n\lambda_n^l}v_n^l(x)v_n^l(y)\frac{d_n(y)}{n}w(y)\right| \\ &\leq n\max_{z\in V_n}|w(z)|\max_{z\in V_n}|d_n(z)|\frac{1}{n}\max_{z\in V_n}\sum_{l=K_n+1}^{n} e^{-h_n\lambda_n^l}(v_n^l(z))^2,\end{aligned}$$

where in the last line we used the Cauchy–Schwarz inequality. To get items (i) and (iii) in Theorem C.2.4 for P_n, it thus suffices to show that

$$\mathcal{R}_n := \max_{z\in V_n} d_n(z)\max_{z\in V_n}\frac{1}{n}\sum_{l=K_n+1}^{n} e^{-h_n\lambda_n^l}(v_n^l(z))^2 = O\left(\frac{h_n^{3/2}}{n}\right).$$

To show this, we start by observing that for every $n \in \mathbf{N}$, every $z \in V_n$ and $1 \leq l \leq n$

$$1 = \langle v_n^l, v_n^l\rangle_{\mathcal{V}_n} \geq \frac{d_n(z)}{n}(v_n^l(z))^2. \tag{C.32}$$

By applying Theorem C.6.5 we can also choose n so large that, with probability greater than $1 - Q_6\epsilon_n^{-k}\exp(-Q_7 n\epsilon_n^{k+2})$, we have

$$\max_{z\in V_n}|d_n(z) - C_1\rho(z)| \leq Q_8\epsilon_n,$$

and we can clearly assume that n is so large that

$$C_1\frac{\min\rho}{2} \leq d_n \leq 2C_1\max\rho.$$

Using (C.32) and the ordering $\lambda_n^l \geq \lambda_n^{K_n}$ for $n \geq l \geq K_n$ we get

$$\begin{aligned}\mathcal{R}_n &\leq \frac{C}{n}\left(n^2 e^{-\lambda_n^{K_n}h_n}\right) \\ &= \frac{C}{n}\left(n^2 e^{-\kappa(\eta)\lambda_{K_n}h_n} e^{-\left(\lambda_n^{K_n}-\kappa(\eta)\lambda_{K_n}\right)h_n}\right).\end{aligned}$$

We now use Theorem C.6.4 and Theorem C.6.1 to infer that with probability greater than $1 - Q_1\epsilon_n^{-6k}\exp(-Q_2 n\epsilon_n^{k+4}) - Q_3 n\exp(-Q_4 n\left(\lambda_{K_n}+1\right)^{-k})$ we have

$$\mathcal{R}_n \leq \frac{C}{n}\left(n^2 e^{-\kappa(\eta)\lambda_{K_n}h_n} e^{\frac{C\epsilon_n}{\gamma}\left(\lambda_{K_n}^{4+\frac{k}{2}}+1\right)h_n}\right).$$

By Weyl's law we have that $\lambda_{K_n} \sim K_n^{2/k}$, thus

$$\mathcal{R}_n \leq \frac{C}{n}\left(n^2 e^{-cK_n^{2/k}h_n} e^{\frac{\tilde{C}\epsilon_n}{\gamma}K_n^{\frac{8}{k}+1}}\right).$$

Recalling the conditions (iv), (v) and (ii) in Theorem C.2.4, as well as the scaling $K_n = (\log(n))^q$ we get

$$\begin{aligned}\mathcal{R}_n &\leq \frac{C}{n}\left(n^2 e^{-c(\log(n))^{\frac{2q}{k}-\alpha}}\right)\\ &= \frac{Ch_n^{3/2}}{n}\left(\frac{n^{2-c(\log(n))^{\frac{2q}{k}-1-\alpha}}}{h_n^{3/2}}\right)\\ &\leq \frac{Ch_n^{3/2}}{n}\left(n^{2-c(\log(n))^{\frac{2q}{k}-1-\alpha}}(\log(n))^{\frac{3\alpha}{2}}\right).\end{aligned}$$

So $\mathcal{R}_n = O\left(\frac{h_n^{3/2}}{n}\right)$ because by the definition of α in (iv) in Theorem C.2.4 we have $\frac{2q}{k} - 1 - \alpha > 0$.
We are left with proving item (ii) in Theorem C.2.2 for both $e^{-t\Delta_n}(\cdot)$ and P_n. We prove it for $e^{-t\Delta_n}(\cdot)$, the proof for P_n being analogous. The proof is divided into three steps.

Step 1. We claim that with probability greater than $1 - a_1\epsilon_n^{-6k}\exp(-a_2 n\epsilon_n^{k+4}) - a_3 n\exp(-a_4 n(\lambda_{K_n}+1)^{-k})$

$$\max_{x,y\in V_n}\left|H^n_{\epsilon_n}(h_n,x,y) - \frac{\rho(y)}{n}H(\kappa(\eta)h_n,x,y)\right| = o\left(\frac{\sqrt{h_n}}{n}\right). \tag{C.33}$$

To show (C.33) we pick two points $x, y \in V_n$ and we compute

$$\begin{aligned}\left|H^n_{\epsilon_n}(h_n,x,y) - \frac{\rho(y)}{n}H(\kappa(\eta)h_n,x,y)\right| \leq &\left|H^{K_n}_{\epsilon_n}(h_n,x,y) - \frac{\rho(y)}{n}H(\kappa(\eta)h_n,x,y)\right|\\ &+\left|\sum_{l=K_n+1}^{n} e^{-h_n\lambda_n^l}v_n^l(x)v_n^l(y)\frac{d_n(y)}{n}\right|.\end{aligned}$$

From Lemma C.2.7 we get that the first term on the right-hand side is $o\left(\frac{\sqrt{h_n}}{n}\right)$ with probability greater than $1 - a_1\epsilon_n^{-6k}\exp(-a_2 n\epsilon_n^{k+4}) - a_3 n\exp(-a_4 n(\lambda_{K_n}+1)^{-k})$, while the second term is estimated in the same way as the term $\mathcal{R}_n$ in the previous part of the proof.

Step 2. We choose an optimal transport map

$$T_n \in \operatorname*{argmin}_{T_\# \nu=\nu_n}\sup_{x\in M} d_M(x,T(x)),\ \theta_n := \sup_{x\in M} d_M(x,T_n(x)).$$

We claim that, with probability greater than $1 - a_1\epsilon_n^{-6k}\exp(-a_2 n\epsilon_n^{k+4}) - a_3 n\exp(-a_4 n(\lambda_{K_n}+1)^{-k})$, we have for every $f \in C^\infty(M)$,

$$\begin{aligned}\max_{x\in V_n}\left|e^{-h_n\Delta_n}f(x) - e^{-\kappa(\eta)h_n\Delta_{\rho^2}}f(x)\right| \leq &L_1\sup_M|f|\frac{\theta_n}{\sqrt{h_n}}e^{\frac{2\theta_n\operatorname{diam}(M)}{h_n}}.\\ &+\sup_M|f|o(\sqrt{h_n}) + L_2\left(\sup_M|f| + \operatorname{Lip}(f)\right)(C.34)\theta_n,\end{aligned}$$

where the constants L_1, L_2 and the function in $o(\sqrt{h_n})$ depend only on M.

To show (C.34), we work under the assumption that we are in the event in which the estimate of Step 1 holds true; this happens with probability greater than

$$1 - a_1\epsilon_n^{-6k}\exp(-a_2 n\epsilon_n^{k+4}) - a_3 n\exp(-a_4 n\,(\lambda_{K_n}+1)^{-k}).$$

We take $f \in C^\infty(M)$ and $x \in V_n$. Then by using the triangle inequality

$$\begin{aligned}
&|e^{-h_n\Delta_n}f(x) - e^{-\kappa(\eta)h_n\Delta_{\rho^2}}f(x)|\\
&= \left|\sum_{y\in V_n} H^n_{\epsilon_n}(h_n,x,y)f(y) - \int_M H(\kappa(\eta)h_n,x,y)f(y)\rho^2(y)d\,\mathrm{Vol}_M(y)\right|\\
&\le \sum_{y\in V_n}\left|H^n_{\epsilon_n}(h_n,x,y)f(y) - \frac{\rho(y)}{n}H(\kappa(\eta)h_n,x,y)f(y)\right|\\
&\quad + \left|\sum_{y\in V_n}\frac{\rho(y)}{n}H(\kappa(\eta)h_n,x,y)f(y) - \int_M H(\kappa(\eta)h_n,x,y)f(y)\rho^2(y)d\,\mathrm{Vol}_M(y)\right|.
\end{aligned}$$

For the first term on the right-hand side, we use the estimate in Step 1 to infer

$$\begin{aligned}
\sum_{y\in V_n}\left|H^n_{\epsilon_n}(h_n,x,y)f(y) - \frac{\rho(y)}{n}H(\kappa(\eta)h_n,x,y)f(y)\right| &\le n\sup_M|f|o\left(\frac{\sqrt{h_n}}{n}\right)\\
&= \sup_M|f|o(\sqrt{h_n}).
\end{aligned}$$

For the second term, we recall that $(T_n)_\#\nu = \nu_n$, thus

$$\begin{aligned}
&\left|\sum_{y\in V_n}\frac{\rho(y)}{n}H(\kappa(\eta)h_n,x,y)f(y) - \int_M H(\kappa(\eta)h_n,x,y)f(y)\rho^2(y)d\,\mathrm{Vol}_M(y)\right|\\
&= \left|\int_M H(\kappa(\eta)h_n,x,T_n(y))f(T_n(y))\rho(T_n(y))d\nu(y) - \int_M H(\kappa(\eta)h_n,x,y)f(y)\rho(y)d\nu(y)\right|.
\end{aligned}$$

By the smoothness of ρ and f, we observe that

$$\left|\int_M H(\kappa(\eta)h_n,x,y)\left(f(T_n(y))\rho(T_n(y)) - f(y)\rho(y)\right)d\nu(y)\right| \le L_2\left(\sup_M|f| + \mathrm{Lip}(f)\right)\theta_n,$$

so we are left with showing that

$$\begin{aligned}
&\left|\int_M (H(h_n,x,T_n(y)) - H(h_n,x,y))f(T_n(y))\rho(T_n(y))d\nu(y)\right|\\
&\le L_1\sup_M|f|\frac{\theta_n}{\sqrt{h_n}}e^{\frac{\theta_n\,\mathrm{diam}(M)}{h_n}}. \qquad (C.35)
\end{aligned}$$

To prove (C.35) we fix $x, y \in M$ and we consider the length minimizing constant-speed geodesic $\sigma_{n,y} : [0,1] \to M$ from y to $T_n(y)$, i.e.,

$$\mathrm{Length}(\sigma_{n,y}|_{[0,s]}) = d_M(y, \sigma_{n,y}(s)).$$

By the fundamental theorem of calculus, the Cauchy–Schwarz inequality and the boundedness of ρ we obtain

$$
\begin{aligned}
&\left|\int_M (H(h_n, x, T_n(y)) - H(h_n, x, y)) f(T_n(y)) \rho(T_n(y)) d\nu(y)\right| \\
&\leq C \sup_M |f| \int_0^1 \int_M |\nabla H(h_n, x, \sigma_{n,y}(s))| |\dot{\sigma}_{n,y}(s)| d\nu(y) ds \\
&\leq C\theta_n \sup_M |f| \int_0^1 \int_M \frac{\hat{Q}_1}{\sqrt{h_n} \mu(B_{\sqrt{h_n}}(x))} \exp\left(-\frac{d_M^2(x, \sigma_{n,y}(s))}{\hat{Q}_2 h_n}\right) d\nu(y) ds, \quad \text{(C.36)}
\end{aligned}
$$

where in the last line we used the fact that the speed of the constant-speed geodesic $\sigma_{n,y}$ is equal to its length – which can be bounded by $C\theta_n$ by definition of θ_n – and we estimated the gradient of the heat kernel by an application of Theorem C.6.2. We now observe that by the reverse triangle inequality

$$
\begin{aligned}
|d_M^2(x, \sigma_{n,y}(s)) - d_M^2(x, y)| &= (d_M(x, y) - d_M(x, \sigma_{n,y}(s)))(d_M(x, \sigma_{n,y}(s)) + d_M(x, y)) \\
&\leq 2\theta_n d_M(x, y).
\end{aligned}
$$

Inserting this estimate into (C.36) and using the Gaussian lower bound for the heat kernel from Theorem C.6.2 yields

$$
\begin{aligned}
&\left|\int_M (H(h_n, x, T_n(y)) - H(h_n, x, y)) f(T_n(y)) \rho(T_n(y)) d\nu\right| \\
&\leq C \frac{\theta_n}{\sqrt{h_n}} e^{\frac{2\theta_n \operatorname{diam}(M)}{h_n}} \sup_M |f| \int_M H(\tilde{Q} h_n, x, y) d\nu(y) \\
&\leq L_1 \sup_M |f| \frac{\theta_n}{\sqrt{h_n}} e^{\frac{2\theta_n \operatorname{diam}(M)}{h_n}}.
\end{aligned}
$$

Step 2. Conclusion. To conclude the proof of the theorem from (C.34) one clearly just needs to prove that

$$
\limsup_{n \to +\infty} \frac{\theta_n}{h_n^{3/2}} < +\infty.
$$

We first treat the case $k \geq 3$. Observe that, by Theorem C.6.6

$$
\limsup_{n \to +\infty} \frac{n^{1/k} \theta_n}{\log^{1/k}(n)} < +\infty.
$$

In particular, using also assumption (vi)

$$
\limsup_{n \to +\infty} \frac{\theta_n}{h_n^{3/2}} = \limsup_{n \to +\infty} \left(\frac{n^{1/k} \theta_n}{\log^{1/k}(n)} \frac{\log^{1/k}(n)}{\epsilon_n n^{1/k}} \frac{\epsilon_n}{h_n^{3/2}} \right) < +\infty,
$$

provided

$$
\limsup_{n \to +\infty} \frac{\epsilon_n}{h_n^{3/2}} < +\infty. \quad \text{(C.37)}
$$

To check that (C.37) is satisfied, we observe that by the assumptions (iv) and (v) in Theorem C.2.4 we get

$$
\limsup_{n \to +\infty} \frac{\epsilon_n}{h_n^{3/2}} \leq \limsup_{n \to +\infty} (\log(n))^{\frac{3}{2}\alpha - \beta},
$$

the right-hand side of which is finite since assumption (i) in Theorem C.2.4 implies $\frac{3}{2}\alpha - \beta \leq 0$.
For the case $k = 2$ we proceed analogously. Recall that by Theorem C.6.6

$$\limsup_{n\to+\infty} \frac{n^{1/2}\theta_n}{\log^{3/4}(n)} < +\infty.$$

In particular, using also assumption (vi) in Theorem C.2.4 we obtain

$$\limsup_{n\to+\infty} \frac{\theta_n}{h_n^{3/2}} = \limsup_{n\to+\infty} \left(\frac{\theta_n n^{1/2}}{\log^{3/4}(n)} \left(\frac{\log(n)}{\epsilon_n^8 n}\right)^{1/2} \frac{\epsilon_n^4 \log^{1/4}(n)}{h_n^{3/2}} \right) < +\infty,$$

provided

$$\limsup_{n\to+\infty} \frac{\epsilon_n^4 \log^{1/4}(n)}{h_n^{3/2}} < +\infty.$$

To show this, we estimate ϵ_n using assumption (v) in Theorem C.2.4 and esimate h_n using assumption (iv) in Theorem C.2.4

$$\limsup_{n\to+\infty} \frac{\epsilon_n^4 \log^{1/4}(n)}{h_n^{3/2}} \leq \limsup_{n\to+\infty} (\log(n))^{\frac{1}{4}+\frac{3}{2}\alpha-4\beta} < +\infty,$$

which follows from (i) in Theorem C.2.4. □

Proof of Lemma C.2.7. As in the proof of Theorem C.2.4, we will for simplicity assume that $K_n = \log(n)^q \in \mathbf{N}$. We will indicate by γ the quantity $\gamma := \inf_{i\in\mathbf{N}}(\lambda_{i+1} - \lambda_i)$, which is positive by Item (ii) in Theorem C.2.4.

To show (C.13), fix two points $x, y \in V_n$. By using the expansion (C.31) and the triangle inequality we have

$$\left| H_{\epsilon_n}^{K_n}(h_n, x, y) - \frac{\rho(y)}{n} H(\kappa(\eta)h_n, x, y) \right| \leq \mathbf{I}_n + \mathbf{II}_n,$$

where we define

$$\mathbf{I}_n = \left| \sum_{l=1}^{K_n-1} e^{-h_n\lambda_n^l} v_n^l(x) v_n^l(y) \frac{d_n(y)}{n} - e^{-h_n\kappa(\eta)\lambda^l} f_l(x) f_l(y) \frac{\rho(y)}{n} \right|,$$
$$\mathbf{II}_n = \left| \sum_{l=K_n}^{+\infty} e^{-h_n\kappa(\eta)\lambda^l} f_l(x) f_l(y) \frac{\rho(y)}{n} \right|.$$

We now proceed to show that these two terms are both of order $o\left(\frac{\sqrt{h_n}}{n}\right)$.
To control term $\mathbf{II}_n$ we follow the ideas in [24] and [7]. By the Cauchy–Schwarz inequality and by the fact that ρ is bounded we get

$$\mathbf{II}_n \leq \frac{C}{n} \max_{z\in M} \sum_{l=K_n}^{+\infty} e^{-h_n\kappa(\eta)\lambda_l} f_l^2(z).$$

To control the right hand side, fix $z \in M$. We define a measure ω_z on $\mathbf{R}$ by

$$\omega_z := \sum_{l=K_n}^{+\infty} f_l^2(z)\delta_{\lambda_l}(d\lambda).$$

Then an integration by parts yields

$$\begin{aligned}\sum_{l=K_n}^{+\infty} e^{-h_n\kappa(\eta)\lambda_l} f_l^2(z) &= \int_{\mathbf{R}} e^{-\kappa(\eta)h_n\lambda} d\omega_z(d\lambda)\\ &= \left[e^{-\kappa(\eta)h_n\lambda}\omega_z([0,\lambda])\right]_{\lambda=0}^{+\infty} + \int_{\mathbf{R}} \kappa(\eta)h_n e^{-\kappa(\eta)h_n\lambda}\omega_z([0,\lambda])d\lambda\\ &\leq \limsup_{\lambda\to+\infty}\left(e^{-h_n\kappa(\eta)\lambda}\sum_{\lambda_{K_n}\leq\lambda_l\leq\lambda} f_l^2(z)\right)\\ &\quad + \int_{\lambda_{K_n}}^{+\infty} h_n\kappa(\eta)e^{-h_n\kappa(\eta)\lambda}\omega_z([0,\lambda])d\lambda.\end{aligned}$$

Now we use Theorem C.6.2 to show that the first term on the right hand side vanishes. Recalling the notation $\mu := \xi\,\mathrm{Vol}_M$, and using the Gaussian upper bounds in Theorem C.6.2 we get in particular

$$\begin{aligned}\sum_{\lambda_{K_n}\leq\lambda_l\leq\lambda} f_l^2(z) \leq e\sum_{0\leq\lambda_l\leq\lambda} e^{-\frac{\lambda_l}{\lambda}} f_l^2(z) \leq eH\left(\frac{1}{\lambda}, z, z\right)\\ \leq \frac{C}{\mu(B_{\lambda^{-1/2}}(x)))} \leq C\lambda^{\frac{k}{2}},\end{aligned} \tag{C.38}$$

so that indeed

$$\limsup_{\lambda\to+\infty} e^{-h_n\frac{\kappa(\eta)}{2}\lambda}\sum_{\lambda_{K_n}\leq\lambda_l\leq\lambda} f_l^2(z) \leq \limsup_{\lambda\to+\infty} e^{-h_n\frac{\kappa(\eta)}{2}\lambda} C\lambda^{\frac{k}{2}} = 0.$$

We thus obtain, using (C.38) once more with λ_{K_n} replaced by zero,

$$\begin{aligned}\mathbf{II}_n &\leq \frac{C}{n}\int_{\lambda_{K_n}}^{+\infty} h_n\kappa(\eta)e^{-h_n\kappa(\eta)\lambda}\lambda^{k/2}d\lambda\\ &= \frac{C}{n}\left(h_n\kappa(\eta)\right)^{-\frac{k}{2}}\int_{\kappa(\eta)h_n\lambda_{K_n}}^{+\infty} e^{-\lambda}\lambda^{k/2}d\lambda\\ &\leq \frac{C}{n}h_n^{-\frac{k}{2}}\int_{ch_nK_n^{2/k}}^{+\infty} e^{-\lambda}\lambda^{k/2}d\lambda,\end{aligned}$$

where we used Weyl's law in the last step. If $ch_nK_n^{\frac{2}{k}} - \frac{k}{2} \geq 1$, we can estimate the right hand side by

$$\frac{C}{n}h_n^{-\frac{k}{2}}\left(ch_nK_n^{\frac{2}{k}}\right)^{\frac{k}{2}+1} e^{-ch_nK_n^{\frac{2}{k}}} = \frac{\tilde{C}}{n}K_ne^{-A}A,$$

where $A = ch_nK_n^{\frac{2}{k}}$. Now we follow the reasoning in the proof of [24, Theorem 3] to obtain $K_nAe^{-A} \leq \frac{1}{K_n}e^{-\frac{A}{2}}$ provided $A \geq 8\log(K_n)$, which is satisfied because of our assumption (iv) in Theorem C.2.4. Thus, using again our assumptions on h_n

$$\begin{aligned}\mathbf{II}_n &\leq \frac{\tilde{C}\sqrt{h_n}}{n}\left(\frac{e^{-c(\log(n))^{\frac{2q}{k}-\alpha}}}{(\log(n))^q\sqrt{h_n}}\right)\\ &\leq \frac{\tilde{C}\sqrt{h_n}}{n}\left(e^{-c(\log(n))^{\frac{2q}{k}-\alpha}}(\log(n))^{\frac{\alpha}{2}-q}\right).\end{aligned}$$

Thus we obtain that $\mathbf{II}_n = o\left(\frac{\sqrt{h_n}}{n}\right)$ because of the definition of α.

Regarding the term $\mathbf{I}_n$, we use the triangle inequality, to decompose this into four terms

$$\mathbf{I}_n \leq \mathbf{I}_n^a + \mathbf{I}_n^b + \mathbf{I}_n^c + \mathbf{I}_n^d,$$

where

$$\mathbf{I}_n^a = \left| \sum_{l=1}^{K_n-1} \left(e^{-h_n \lambda_n^l} - e^{-\kappa(\eta) h_n \lambda_l} \right) \frac{\rho(y)}{n} f_l(x) f_l(y) \right|,$$
$$\mathbf{I}_n^b = \left| \sum_{l=1}^{K_n-1} e^{-h_n \lambda_n^l} \left(C_1 \frac{\rho(y)}{n} - \frac{d_n(y)}{n} \right) \frac{f_l(x)}{C_1^{1/2}} \frac{f_l(y)}{C_1^{1/2}} \right|,$$
$$\mathbf{I}_n^c = \left| \sum_{l=1}^{K_n-1} e^{-h_n \lambda_n^l} \frac{d_n(y)}{n} \left(\frac{f_l(x)}{C_1^{1/2}} - v_n^l(x) \right) \frac{f_l(y)}{C_1^{1/2}} \right|,$$
$$\mathbf{I}_n^d = \left| \sum_{l=1}^{K_n-1} e^{-h_n \lambda_n^l} \frac{d_n(y)}{n} v_n^l(x) \left(\frac{f_l(y)}{C_1^{1/2}} - v_n^l(y) \right) \right|.$$

We now proceed at estimating these four terms.
Term $\mathbf{I}_n^a$. We observe that $\lambda_n^1 = \lambda_1 = 0$, thus in the sum we can neglect the term corresponding to $l = 1$, i.e.

$$\mathbf{I}_n^a \leq \frac{C}{n} \sum_{l=2}^{K_n-1} \left| e^{-h_n \lambda_n^l} - e^{-h_n \kappa(\eta) \lambda_l} \right| \|f_l\|_{C^0(M)}^2.$$

Since $s \mapsto e^{-s}$ is 1-Lipschitz continuous on $[0, +\infty)$, for every $2 \leq l \leq K_n - 1$ we have

$$\left| e^{-h_n \lambda_n^l} - e^{-\kappa(\eta) h_n \lambda_l} \right| \leq |\lambda_n^l - \kappa(\eta) \lambda_l| h_n \leq Q_5 \frac{\|f_l\|_{C^3(M)}}{\gamma} \epsilon_n h_n,$$

where the last inequality holds with probability greater than $1 - Q_1 \epsilon_n^{-6k} \exp(-Q_2 n \epsilon_n^{k+4}) - Q_3 n \exp(-Q_4 n (\lambda_{\bar{l}} + 1)^{-k})$ because of Theorem C.6.4. In particular using also Theorem C.6.1 to control the C^0 and C^3 norm of the eigenfunctions and using the fact that for $l \leq K_n$ we have $\lambda_l \leq \lambda_{K_n}$ we can bound

$$\mathbf{I}_n^a \leq \frac{C h_n}{n} \left(\frac{K_n \left(\lambda_{K_n}^{1+\frac{k}{2}} + 1 \right)^2 \left(\lambda_{K_n}^{4+\frac{k}{2}} + 1 \right) \epsilon_n}{\gamma} \right).$$

From this, we obtain that $\mathbf{I}_n^a = o\left(\frac{\sqrt{h_n}}{n}\right)$, because by our assumptions on ϵ_n in (v) of Theorem C.2.4 and our assumptions on the spectral gap in (ii) of Theorem C.2.4 we clearly have

$$\left(\frac{K_n \left(\lambda_{K_n}^{1+\frac{k}{2}} + 1 \right)^2 \left(\lambda_{K_n}^{4+\frac{k}{2}} + 1 \right) \epsilon_n}{\gamma} \right) = o(1).$$

Term $\mathbf{I}_n^b$. Using Theorem C.6.4, Theorem C.6.5 and Theorem C.6.1 we have that with probability greater than $1 - Q_1\epsilon_n^{-6k}\exp(-Q_2 n\epsilon_n^{k+4}) - Q_3 n\exp(-Q_4 n\,(\lambda_{\bar{l}}+1)^{-k})$ $- Q_6\epsilon_n^{-k}\exp(-Q_7 n\epsilon_n^{k+2})$, for each $1 \le l \le K_n - 1$ we can estimate

$$\begin{aligned}&\left| e^{-h_n\lambda_n^l}\left(C_1\frac{\rho(y)}{n} - \frac{d_n(y)}{n}\right)\frac{f_l(x)}{C_1^{1/2}}\frac{f_l(y)}{C_1^{1/2}}\right| \\ &\le \frac{C}{n}e^{-h_n\kappa(\eta)\lambda_l}e^{-h_n\left(\lambda_n^l-\kappa(\eta)\lambda_l\right)}\|C_1\rho - d_n\|_{L^\infty(G_n)}\|f_l\|^2_{L^\infty(M)} \\ &\le \frac{C}{n}e^{Ch_n\frac{\left(\lambda_{K_n}^{4+\frac{k}{2}}+1\right)\epsilon_n}{\gamma}}\left(\lambda_{K_n}^{1+\frac{k}{2}}+1\right)^2\epsilon_n.\end{aligned}$$

In particular, multiplying and dividing by $\sqrt{h_n}$ and summing over $l = 1, ..., K_n$, we obtain

$$\mathbf{I}_n^b \le \frac{C\sqrt{h_n}}{n}\left(\frac{K_n}{\sqrt{h_n}}e^{ch_n\frac{\left(\lambda_{K_n}^{4+\frac{k}{2}}+1\right)\epsilon_n}{\gamma}}\left(\lambda_{K_n}^{1+\frac{k}{2}}+1\right)^2\epsilon_n\right).$$

By Weyl's law and our by assumptions (v), (iv) and (ii) in Theorem C.2.4, this is again an $o\left(\frac{\sqrt{h_n}}{n}\right)$ term.

The terms $\mathbf{I}_n^c, \mathbf{I}_n^d$ are treated similarly. In particular $\mathbf{I}_n = o\left(\frac{\sqrt{h_n}}{n}\right)$ provided we are in the event in which Theorem C.6.4 and Theorem C.6.5 apply. This happens with probability greater than

$$\begin{aligned}&1 - Q_1\epsilon_n^{-6k}\exp(-Q_2 n\epsilon_n^{k+4}) - Q_3 n\exp(-Q_4 n\,(\lambda_{\bar{l}}+1)^{-k}) - Q_6\epsilon_n^{-k}\exp(-Q_7 n\epsilon_n^{k+2}) \\ &\ge 1 - (Q_1+Q_6)\epsilon_n^{-6k}\exp(-\min(Q_2,Q_7)n\epsilon_n^{k+4}) - Q_3 n\exp(-Q_4 n\,(\lambda_{\bar{l}}+1)^{-k}) \\ &= 1 - a_1\epsilon_n^{-6k}\exp(-a_2 n\epsilon_n^{k+4}) - a_3 n\exp(-a_4 n\,(\lambda_{K_n}+1)^{-k}),\end{aligned}$$

provided n is large enough, this concludes our argument for (C.13). □

Proof of Corollary C.2.6. We know from Theorem C.2.4 that for n large enough, assumptions (i), (ii), (iii) of Theorem C.2.2 hold for both the choices of the operators $e^{-t\Delta_n}$ and P_n on the graph G_n on an event A_n such that

$$\mathbb{P}(A_n) \ge 1 - C\epsilon_n^{-6k}\exp(-\frac{1}{C}n\epsilon_n^{k+4}) - Cn\exp(-\frac{n}{C(\log(n))^{2q}}).$$

For $\overline{n} \in \mathbf{N}$ large enough we consider the set

$$C_{\overline{n}} := \bigcap_{n\ge\overline{n}} A_n.$$

Theorem C.2.2 says that, in the event $C_{\overline{n}}$, for both the choices of the operators $e^{-t\Delta_n}$ and P_n we have that (C.5) and (C.6) hold true. Observe that

$$\mathbb{P}(C_{\overline{n}}) \ge 1 - \sum_{n\ge\overline{n}} C\epsilon_n^{-6k}\exp(-\frac{1}{C}n\epsilon_n^{k+4}) - Cn\exp(-\frac{n}{C(\log(n))^{2q}}),$$

In particular, we have that

$$\begin{aligned}\mathbb{P}\left(\{u^* \text{ and } u_* \text{ satisfy (C.5) and (C.6)}\}\right) \geq \mathbb{P}\left(\bigcup_{\overline{n}\in\mathbf{N}} C_{\overline{n}}\right) &= \lim_{\overline{n}\to+\infty} \mathbb{P}(C_{\overline{n}}) \\ &\geq 1 - \lim_{\overline{n}\to+\infty} \sum_{n\geq\overline{n}} \Bigg(C\epsilon_n^{-6k} \exp(-\frac{1}{C} n\epsilon_n^{k+4}) \\ &\qquad - Cn\exp(-\frac{n}{C(\log(n))^{2q}})\Bigg).\end{aligned} \tag{C.39}$$

We thus just need to show that

$$\lim_{\overline{n}\to+\infty} \sum_{n\geq\overline{n}} \left(C\epsilon_n^{-6k} \exp(-\frac{1}{C} n\epsilon_n^{k+4}) - Cn\exp(-\frac{n}{C(\log(n))^{2q}})\right) = 0,$$

in other words, we need to prove that the series is convergent. To this end, observe that

$$\begin{aligned}C\epsilon_n^{-6k} \exp(-\frac{1}{C} n\epsilon_n^{k+4}) &= C\exp\left(-6k\log(\epsilon_n) - \frac{1}{C} n\epsilon_n^{k+4}\right) \\ &= C\exp\left(\log(n)\left(-6k\frac{\log(\epsilon_n)}{\log(n)} - \frac{1}{C}\frac{n\epsilon_n^{k+4}}{\log(n)}\right)\right) \\ &= Cn^{\left(-6k\frac{\log(\epsilon_n)}{\log(n)} - \frac{1}{C}\frac{n\epsilon_n^{k+4}}{\log(n)}\right)}.\end{aligned}$$

In a similar way, we have

$$Cn\exp(-\frac{n}{C(\log(n))^{2q}}) = Cn^{\left(1-\frac{1}{C}\frac{n}{(\log(n))^{2q+1}}\right)}.$$

To prove the convergence of the series appearing in (C.39) it is sufficient to show

$$\lim_{n\to+\infty}\left(-6k\frac{\log(\epsilon_n)}{\log(n)} - \frac{1}{C}\frac{n\epsilon_n^{k+4}}{\log(n)}\right) = \lim_{n\to+\infty}\left(1 - \frac{1}{C}\frac{n}{(\log(n))^{2q+1}}\right) = -\infty.$$

The second limit is easily treated. To treat the first limit, observe that by assumption (C.12) in Corollary C.2.6 we have

$$\lim_{n\to+\infty}\frac{n\epsilon_n^{k+4}}{\log(n)} = +\infty.$$

To conclude the proof, we show that

$$\inf_{n\in\mathbf{N}}\frac{\log(\epsilon_n)}{\log(n)} > -\infty. \tag{C.40}$$

Indeed, we have

$$\frac{\log(\epsilon_n)}{\log(n)} = \frac{\log\left(\frac{\epsilon_n n^{\frac{1}{k+4}}}{\log^{\frac{1}{k+4}}(n)}\right)}{\log(n)} - \frac{1}{k+4} + \frac{\log\log(n)}{\log(n)},$$

The first term is bounded from below because it is asymptotically nonnegative by (C.12) . The last term converges to zero as $n \to +\infty$. Thus (C.40) holds and the proof is complete. □

C.5.3 MBO on manifolds

Proof of Theorem C.4.2. We let $\hat{x} := \exp_x(z(x)\nu(x))$. Then we have

$$\frac{1}{2} + \omega_1\sqrt{h} = \int_{\Omega_0} H(\kappa h, \hat{x}, y)\rho^2(y) d\,\mathrm{Vol}_M$$

By the Gaussian upper bounds on the heat kernel in Theorem C.6.2, we have that $d_M(\hat{x}, \partial\Omega_0) \leq \tilde{C}\sqrt{h}$, for a fixed constant $\tilde{C}$, independent of Ω_0. In particular, we infer from the asymptotic expansion of the heat kernel in Theorem C.6.3 that

$$\frac{1}{2} + \omega_1\sqrt{h} = \int_{\Omega_0} \frac{e^{-\frac{d_M^2(\hat{x},y)}{4\kappa h}}}{(4\pi\kappa h)^{k/2}} v_0(\hat{x}, y)\rho^2(y) d\,\mathrm{Vol}_M + O(h). \tag{C.41}$$

Since $d(\hat{x}, \partial\Omega_0) \leq \tilde{C}h$, and $\mathrm{diam}(\Omega_0) \leq \frac{\mathrm{inj}(M)}{2}$, we can rewrite the integral in (C.41) in exponential coordinates around $\hat{x}$, i.e.

$$\frac{1}{2} + \omega_1 \circ \exp_{\hat{x}} \sqrt{h} = \int_{\tilde{\Omega}_0} \frac{e^{-\frac{|y|^2}{4\kappa h}}}{(4\pi\kappa h)^{k/2}} v_0(\hat{x}, \exp_{\hat{x}}(y))\rho^2(\exp_{\hat{x}}(y)) dy + O(h),$$

where $\tilde{\Omega}_0 := \exp_{\hat{x}}^{-1}(\Omega_0)$. Recalling that $v_0(\hat{x}, \hat{x}) = \frac{1}{\rho^2(\hat{x})}$, a Taylor expansion of the function $y \mapsto v_0(\hat{x}, \exp_{\hat{x}}(y))\rho^2(\exp_{\hat{x}}(y))$ around zero reveals that

$$\frac{1}{2} + \omega_1 \circ \exp_{\hat{x}} \sqrt{h} = \int_{\tilde{\Omega}_0} \frac{e^{-\frac{|y|^2}{4\kappa h}}}{(4\pi\kappa h)^{k/2}} dy + O(\sqrt{h}).$$

In other words, there exists a bounded function ω_2 on $\mathbf{R}^k$ such that

$$\frac{1}{2} + \omega_2\sqrt{h} = \int_{\tilde{\Omega}_0} \frac{e^{-\frac{|y|^2}{4\kappa h}}}{(4\pi\kappa h)^{k/2}} dy.$$

In other words, we have that $0 \in \partial E$, where

$$E = \left\{ v \in \mathbf{R}^k |\ \frac{1}{2} + \omega_2(v)\sqrt{h} \leq \int_{\tilde{\Omega}_0} \frac{e^{-\frac{|v-y|^2}{4\kappa h}}}{(4\pi\kappa h)^{k/2}} dy \right\},$$

and thus the normal distance $z(x)$ coincides with the normal distance of $\partial\tilde{\Omega}_0$ and E at the point $\exp_{\hat{x}}^{-1}(x) \in \partial\tilde{\Omega}_0$. The conclusion of the proof is then obtained by applying the following result. □

Proposition C.5.1. *Let $\Omega \subset \mathbf{R}^k$ be a smooth open set. Let E be obtained by applying one step of MBO with diffusion coefficient $\kappa > 0$, bounded drift $\omega : \mathbf{R}^k \to \mathbf{R}$ and step size $h > 0$. Let $x \in \partial\Omega$. Let $\nu(x)$ be the outer unit normal to $\partial\Omega$ at x, define*

$$z(x) := \begin{cases} \sup\{l \in \mathbf{R}^- |\ x + l\nu(x) \in E\} & \text{if } x \notin E \\ \inf\{l \in \mathbf{R}^+ |\ x + l\nu(x) \notin E\} & \text{if } x \in E. \end{cases}$$

Then we have

$$|z(x)| \leq \tilde{C}h,$$

where the constant $\tilde{C}$ depends only on k, κ and the C^0-norm of the second fundamental form of $\partial\Omega$.

Proposition C.5.1 is a weaker version of [31, Theorem 4.1], which makes rigorous the original ideas in [60]. In those works, the authors identify the exact first order coefficient of the expansion of $z(x)$ in h. Since we do not need this, we present a proof of our weaker statement.

Proof of Proposition C.5.1. For ease of notation, we assume that $\kappa = 1$. We treat the case when $z(x) > 0$, the other case is similar. First of all, we observe that $z(x) \leq \tilde{C}_k\sqrt{h}$, for a constant $\tilde{C}_k$ depending just on the dimension k. We now choose a coordinate system in which $x = 0$ and $\nu(x) = e_k$. We may find an open set U containing the origin and a smooth function $\zeta : \mathbf{R}^{k-1} \to \mathbf{R}$ such that $\zeta(0) = 0, D\zeta(0) = 0$ and

$$U \cap \Omega = \left\{v \in \mathbf{R}^k |\ v_k < \zeta(v_1, ..., v_{k-1})\right\}.$$

Using the fact that $z(x) = O(\sqrt{h})$ and the exponential decay of the heat kernel, we have that there exists a bounded function $\omega : \mathbf{R}^k \to \mathbf{R}$ such that

$$\frac{1}{2} + \omega((0, z(x)))\sqrt{h} = \int_{\mathbf{R}^{k-1}} \int_{-\infty}^{\zeta(y)+z(x)} \frac{e^{-\frac{|y|^2+|s|^2}{4h}}}{(4\pi h)^{k/2}} dsdy. \tag{C.42}$$

Recalling that the Gaussian integrates to $1/2$ over half-spaces, we get that (C.42) reads

$$\omega((0, z(x)))\sqrt{h} = \int_{\mathbf{R}^{k-1}} \int_{0}^{\zeta(y)+z(x)} \frac{e^{-\frac{|y|^2+|s|^2}{4h}}}{(4\pi h)^{k/2}} dsdy.$$

Since $\zeta(0) = 0$ and $D\zeta(0) = y$, there exists a bounded function ζ_1 such that $\zeta(v) = \zeta_1(v)|v|^2$. We also observe that

$$e^{-t} \geq 1 - t,\ \ t \geq 0.$$

In particular

$$\begin{aligned}\omega((0, z(x)))\sqrt{h} &\geq \frac{1}{(4\pi h)^{k/2}} \int_{\mathbf{R}^{k-1}} e^{-\frac{|y|^2}{4h}} \int_{0}^{\zeta(y)+z(x)} \left(1 - \frac{s^2}{4h}\right) dsdy \\ &= \frac{1}{(4\pi h)^{k/2}} \int_{\mathbf{R}^{k-1}} e^{-\frac{|y|^2}{4h}} \left(\zeta_1(y)|y|^2 + z(x) - \frac{1}{12h}\left(\zeta_1(y)|y|^2 + z(x)\right)^3\right) dy.\end{aligned}$$

By using the change of variable $y \to \sqrt{h}y$ we obtain

$$\omega((0, z(x))\sqrt{h} \geq \frac{1}{h^{1/2}}\left(z(x) + \frac{q_1}{h}z(x)^3 + q_2 h + q_3 h^2 + q_4 z(x)^2\right),$$

where q_1, q_2, q_3, q_4 are coefficients depending on the first six moments of the function $y \mapsto e^{-|y|^2}$. By multiplying both sides by $\sqrt{h}$ we get

$$\omega((0, z(x))h - \left(q_2 h + q_3 h^2 + q_4 z(x)^2\right) \geq z(x) + \frac{q_1}{h}z(x)^3.$$

By applying [31, Lemma 6.1] (which holds true even if we additionally consider a bounded drift ω), we have that $z(x) = O(h^{3/2})$. In particular, for h small enough

$$\frac{1}{2} < 1 - \frac{q_1}{h}z(x)^2,$$

in other words

$$2\omega((0, z(x))h - 2\left(q_2 h + q_3 h^2 + q_4 z(x)^2\right) \geq z(x) \geq 0,$$

from which we conclude that $z(x) = O(h)$. □

Proof of Corollary C.4.3. Denote by $\tilde{C}_{r,x_0}$ the constant obtained by applying Theorem C.4.2 to $\Omega_0 = B_r(x_0)$. Since $\tilde{C}_{r,x_0}$ depends on Ω_0 only through the C^0 norm of the second fundamental form S_{r,x_0} of $\partial B_r(x_0)$, it is sufficient to show that this can be bounded independently of $\frac{R}{2} \leq r \leq R$ and $x_0 \in M$. We clearly have that

$$(0, \operatorname{diam}(M)) \times M \ni (r, x_0) \mapsto \|S_{r,x_0}\|_{C^0}$$

is a continuous function. It is thus bounded on the compact set

$$W := \left\{(r,x) \in (0,+\infty) \times M : \frac{R}{2} \leq r \leq R, x \in M\right\}.$$

□

Proof of Theorem C.4.4. For ease of notation, let us assume that $\kappa = 1$. We start by observing that

$$\int_{M \setminus B_{h_n^{\frac{1}{4}}}(z_{h_n})} H(h_n, z_{h_n}, y)\xi(y) d\operatorname{Vol}_M(y) = o\left(\sqrt{h_n}\right).$$

This is proved by using the Gaussian bounds from Theorem C.6.2, as we did in [50, Theorem 3, Step 2]. In particular, both in (i) and in (ii) of Theorem C.4.4 we can replace the domain of integration with

$$\{\psi_{h_n}(s_{h_n} - h_n, \cdot) \geq 0\} \cap B_{h_n^{\frac{1}{4}}}(z_{h_n}).$$

In this way, the sequence of integrals can be computed in normal coordinates around z_{h_n}, i.e.,

$$\begin{aligned}
&\int_{\{\psi_{h_n}(s_{h_n} - h_n,\cdot)\geq 0\} \cap B_{h_n^{\frac{1}{4}}}(z_{h_n})} H(h_n, z_{h_n}, y)\xi(y) d\operatorname{Vol}_M(y) \\
&= \int_{\left\{\tilde{\psi}_{h_n}(s_{h_n} - h_n,\cdot)\geq 0\right\} \cap B_{h_n^{\frac{1}{4}}}(0)} H(h_n, z_{h_n}, \exp_{z_{n_n}}(y))\xi(\exp_{z_{n_n}}(y))\sqrt{\det(g)}dy,
\end{aligned}$$

where we set

$$\tilde{\psi}_{h_n}(t,y) := \psi_{h_n}(t, \exp_{z_{h_n}}(y)), \ y \in B_{\frac{\operatorname{inj}(M)}{2}}(0).$$

Using the asymptotic expansion for the heat kernel in Theorem C.6.3, it is easy to see that

$$\begin{aligned}
&\int_{\left\{\tilde{\psi}_{h_n}(s_{h_n} - h_n,\cdot)\geq 0\right\} \cap B_{h_n^{\frac{1}{4}}}(0)} H(h_n, z_{h_n}, \exp_{z_{n_n}}(y))\xi(\exp_{z_{n_n}}(y))\sqrt{\det(g)}dy \\
&= \int_{\left\{\tilde{\psi}_{h_n}(s_{h_n} - h_n,\cdot)\geq 0\right\} \cap B_{h_n^{\frac{1}{4}}}(0)} \frac{e^{-\frac{|y|^2}{4h_n}}}{(4\pi h_n)^{k/2}} v_0(z_{h_n}, \exp_{z_{h_n}}(y))\xi(\exp_{z_{n_n}}(y))\sqrt{\det(g)}dy \\
&\quad + o(\sqrt{h_n}).
\end{aligned}$$

In particular, in both (i) and (ii) in Theorem C.4.4 the integrals may be substituted with

$$\int_{\left\{\tilde{\psi}_{h_n}(s_{h_n} - h_n,\cdot)\geq 0\right\} \cap B_{h_n^{\frac{1}{4}}}(0)} \frac{e^{-\frac{|y|^2}{4h_n}}}{(4\pi h_n)^{k/2}} v_0(z_{h_n}, \exp_{z_{h_n}}(y))\xi(\exp_{z_{n_n}}(y))\sqrt{\det(g)}dy.$$

These integrals may be furthermore decomposed into the sums $\mathbb{I}_n + \mathbb{III}_n$,

$$\mathbb{I}_n := \int_{\left\{\tilde{\psi}_{h_n}(s_{h_n}-h_n,\cdot)\geq 0\right\}\cap B_{h_n^{\frac{1}{4}}}(0)} \frac{e^{-\frac{|y|^2}{4h_n}}}{(4\pi h_n)^{k/2}}dy,$$

$$\mathbb{III}_n := \int_{\left\{\tilde{\psi}_{h_n}(s_{h_n}-h_n,\cdot)\geq 0\right\}\cap B_{h_n^{\frac{1}{4}}}(0)} \frac{e^{-\frac{|y|^2}{4h_n}}}{(4\pi h_n)^{k/2}}(w_n(y)-1)dy,$$

where we define

$$w_n(y) := v_0(z_{h_n}, \exp_{z_{h_n}}(y))\xi(\exp_{z_{n_n}}(y))\sqrt{\det(g)}.$$

We claim that

$$\lim_{n\to+\infty} \mathbb{III}_n = \begin{cases} 0 & \text{if } \nabla\psi(s,z)=0, \\ \frac{1}{2\sqrt{\pi}|\nabla\psi(s,z)|}\langle \frac{\nabla\xi}{\xi}(z), \nabla\psi(s,z)\rangle & \text{otherwise.} \end{cases} \tag{C.43}$$

Using (C.48) we see that

$$w_n(y) = \sqrt{\frac{\xi(\exp_{z_{n_n}}(y))\det(g)}{\xi(z_{h_n})\det(d_{\exp^{-1}_{z_{h_n}}(y)}(\exp_{z_{h_n}}))}}.$$

In particular, denoting $\tilde{\xi}_n = \xi\circ\exp_{z_{h_n}}$ and $D_n := \det(d_{\exp^{-1}_{z_{h_n}}(y)}(\exp_{z_{h_n}}))$ we get

$$Dw_n = \frac{1}{2w_n(y)}\frac{\left((D_y\tilde{\xi}_n)\det(g)+\tilde{\xi}_nD_y\det(g))\tilde{\xi}_n(0)D_n - \tilde{\xi}_n\det(g)\tilde{\xi}_n(0)D_yD_n\right)}{\tilde{\xi}_n(0)^2D_n^2}.$$

We now recall that, in normal coordinates $g(z_{h_n}) = Id$, $Dg(z_{h_n}) = 0$, in particular

$$Dw_n(z_{h_n}) = \frac{1}{2}\frac{D\tilde{\xi}_n}{\tilde{\xi}_n}(0),$$

and by a Taylor expansion

$$Dw_n(y) = 1 + \frac{1}{2}\frac{D\tilde{\xi}_n}{\tilde{\xi}_n}(0)\cdot y + O(|y|^2);$$

in particular, we infer that

$$\mathbb{III}_n = \frac{1}{2}\frac{D\tilde{\xi}_n}{\tilde{\xi}_n}(0)\cdot\int_{\left\{\tilde{\psi}_{h_n}(s_{h_n}-h_n,\cdot)\geq 0\right\}\cap B_{h_n^{\frac{1}{4}}}(0)} \frac{e^{-\frac{|y|^2}{4h_n}}}{(4\pi h_n)^{k/2}}ydy + O(h_n).$$

Now we claim that

$$\lim_{n\to+\infty}\frac{1}{2\sqrt{h_n}}\frac{D\tilde{\xi}_n}{\tilde{\xi}_n}(0)\cdot\int_{\left\{\tilde{\psi}_{h_n}(s_{h_n}-h_n,\cdot)\geq 0\right\}\cap B_{h_n^{\frac{1}{4}}}(0)} \frac{e^{-\frac{|y|^2}{4h_n}}}{(4\pi h_n)^{k/2}}ydy \tag{C.44}$$

$$= \frac{1}{2\sqrt{\pi}|\nabla\psi(s,z)|}\frac{D\tilde{\xi}}{\tilde{\xi}}(0)\cdot D\tilde{\psi}(s,0),$$

where $\tilde{\xi} = \xi \circ \exp_z$. Of course (C.44) gives (C.43).

To see that (C.44) holds, we start by changing variable in the integral by setting $y = \frac{y}{\sqrt{h_n}}$, which gives that the argument in the limit equals

$$\frac{1}{2}\frac{D\tilde{\xi}_n}{\tilde{\xi}_n}(0) \cdot \int_{\left\{y|\ \tilde{\psi}_{h_n}(s_{h_n}-h_n,\sqrt{h_n}y)\geq 0\right\}\cap B_{h_n^{-\frac{1}{4}}}(0)} \frac{e^{-\frac{|y|^2}{4}}}{(4\pi)^{k/2}} y dy.$$

We now let R_n be a sequence of orthogonal matrices such that $R_n^T e_1 = \frac{D\tilde{\xi}_n(0)}{|D\tilde{\xi}_n(0)|}$ and without loss of generality we assume that the sequence converges to an orthogonal matrix R. We change variable by setting $y = R_n^T y$ and we get that the argument of the limit becomes

$$\frac{|D\tilde{\xi}_n(0)|}{2} \int_{\mathcal{C}_n \cap B_{h_n^{-\frac{1}{4}}}(0)} \frac{e^{-\frac{|y|^2}{4}}}{(4\pi)^{k/2}} y_1 dy,$$

where we define

$$\mathcal{C}_n := \left\{ y \in \mathbf{R}^k |\ \tilde{\psi}_{h_n}(s_{h_n} - h_n, R_n\sqrt{h_n} y) \geq 0 \right\}.$$

We now observe that, by Taylor expanding $\tilde{\psi}_{h_n}(t_{h_n} - \cdot, \cdot)$ around $(0,0)$

$$\begin{aligned}\tilde{\psi}_{h_n}(s_{h_n} - h_n, R_n\sqrt{h_n}y) =& \delta_{h_n} + \sqrt{h_n} R_n^T D\tilde{\psi}_{h_n}(s_{h_n}, 0) \cdot y \\ & - h_n \partial_s \tilde{\psi}_{h_n}(s_{h_n}, 0) + o(|y|^2 + h_n^2),\end{aligned}$$

thus

$$\begin{aligned}\mathcal{C}_n = \Bigg\{ y \in \mathbf{R}^k |\ & \frac{\delta_{h_n}}{\sqrt{h_n}} + R_n^T D\tilde{\psi}_{h_n}(s_{h_n}, 0) \cdot y \\ & - \sqrt{h_n}\partial_s \tilde{\psi}_{h_n}(s_{h_n}, 0) + o(\sqrt{h_n}|y|^2 + h_n^{\frac{3}{2}}) \geq 0 \Bigg\}.\end{aligned}$$

Recalling assumption (C.17) this re-reads

$$\mathcal{C}_n = \left\{ y \in \mathbf{R}^k |\ R_n^T D\tilde{\psi}_{h_n}(s_{h_n}, 0) \cdot y + o(1) \geq 0 \right\}.$$

Observe also that

$$\begin{aligned}R_n D\tilde{\xi}_n(0) &= |D\tilde{\xi}_n(0)| e_1 \\ &= \sqrt{\langle \nabla\xi(z_{h_n}), \nabla\xi(z_{h_n})\rangle} e_1 \underset{n\to+\infty}{\to} \sqrt{\langle \nabla\xi(z), \nabla\xi(z)\rangle} e_1,\end{aligned}$$

but also

$$R_n D\tilde{\xi}_n(0) = D\tilde{\xi} \circ R_n^T(0) = D\xi \circ \exp_{z_{h_n}} \circ R_n^T(0)) \underset{n\to+\infty}{\to} RD(\xi \circ \exp_z)(0).$$

In other words we must have $D\tilde{\xi}(0) = |D\tilde{\xi}(0)| R^T e_1$. In particular

$$\lim_{n\to+\infty}\frac{|D\tilde{\xi}_n(0)|}{2}\int_{\mathcal{C}_n\cap B_{h_n^{-\frac{1}{4}}}(0)}\frac{e^{-\frac{|y|^2}{4}}}{(4\pi)^{k/2}}y_1dy=\frac{|D\tilde{\xi}(0)|}{2}\int_{\left\{y|\ R^TD\tilde{\psi}(s,0)\cdot y\geq 0\right\}}\frac{e^{-\frac{|y|^2}{4}}}{(4\pi)^{k/2}}y_1dy$$
$$=\frac{|D\tilde{\xi}(0)|}{2}\int_{\left\{y|\ D\tilde{\psi}(s,0)\cdot y\geq 0\right\}}\frac{e^{-\frac{|y|^2}{4}}}{(4\pi)^{k/2}}Ry\cdot e_1dy$$
$$=\frac{1}{2}\frac{D\tilde{\xi}}{\tilde{\xi}}(0)\cdot\int_{\left\{y|\ D\tilde{\psi}(s,0)\cdot y\geq 0\right\}}\frac{e^{-\frac{|y|^2}{4}}}{(4\pi)^{k/2}}ydy.$$

If $\nabla\psi(t,z)=0$, then the last integral is zero, being component-wise the integral over the whole space of on odd-function. Otherwise we change variable according to $y=O^Ty$, where O is an orthogonal matrix such that $OD\tilde{\psi}(s,0)=|D\tilde{\psi}(s,0)|e_1$, which gives that the last integral equals

$$\frac{1}{2}\frac{OD\tilde{\xi}}{\tilde{\xi}}(0)\cdot\int_{\{y|\ y_1\geq 0\}}\frac{e^{-\frac{|y|^2}{4}}}{(4\pi)^{k/2}}ydy=\frac{1}{2}\frac{OD\tilde{\xi}}{\tilde{\xi}}(0)\cdot e_1\frac{1}{\sqrt{\pi}}$$
$$=\frac{1}{2\sqrt{\pi}|D\tilde{\psi}(s,0)|}\frac{D\tilde{\xi}}{\tilde{\xi}}(0)\cdot D\tilde{\psi}(s,0).$$

We are now in a position to prove (i) and (ii) in Theorem C.4.4.

Item (i). By the discussion above, the left hand side of (C.18) may be substituted with

$$\liminf_{n\to+\infty}\frac{1}{\sqrt{h_n}}\left(\frac{1}{2}-\mathbb{I}_n-\mathbb{III}_n\right)$$
$$\geq\liminf_{n\to+\infty}\frac{1}{\sqrt{h_n}}\left(\frac{1}{2}-\mathbb{I}_n\right)-\frac{1}{2\sqrt{\pi}|\nabla\psi(s,z)|}\langle\frac{\nabla\xi}{\xi}(z),\nabla\psi(s,z)\rangle,$$

where we used (C.43) in the second line. To estimate

$$\liminf_{n\to+\infty}\frac{1}{\sqrt{h_n}}\left(\frac{1}{2}-\mathbb{I}_n\right)$$

we can use [6, Proposition 4.1] applied with

$$(t_h,x_h)=(s_h,0),$$
$$(t,x)=(s,0),$$
$$\phi_h(t,\cdot)=\tilde{\psi}_h(t,\cdot).$$

The only difference is that here we do not assume that $\phi(t_h,x_h)=0$, but $\phi(t_h,x_h)=o(\sqrt{h})$ - one can check that the result holds true also with this modification by the same proof of [6, Proposition 4.1]. In particular, we get

$$\liminf_{n\to+\infty}\frac{1}{\sqrt{h_n}}\left(\frac{1}{2}-\int_{\{\psi_{h_n}(t_{h_n}-h_n,\cdot)\geq 0\}}H(h_n,z_{h_n},y)\xi(y)d\,\mathrm{Vol}_M\right)$$
$$\geq\frac{1}{2\sqrt{\pi}|D\tilde{\psi}(s,0)|}\left(\partial_t\tilde{\psi}+\Delta\tilde{\psi}-\frac{D\tilde{\psi}\cdot D^2\tilde{\psi}D\tilde{\psi}}{|D\tilde{\psi}|^2}-\frac{D\tilde{\xi}}{\tilde{\xi}}\cdot D\tilde{\psi}\right),$$

which is equal to the right hand side of (C.18) because we are using exponential coordinates around z (recall our convention $\Delta = -\sum_{i=1}^k \partial_{ii}^2$).

Item (ii). Once again, by the above discussion, we can assume that

$$\frac{1}{2} - \mathbb{I}_n \leq o(\sqrt{h_n}),$$

and the result follows by applying [6, Proposition 4.1] with

$$\begin{aligned}(t_h, x_h) &= (s_h, 0),\\ (t, x) &= (s, 0),\\ \phi_h(t, \cdot) &= \tilde{\psi}_h(t, \cdot).\end{aligned}$$

In this case, there are two differences from the original version [6, Proposition 4.1]. First of all, we again do not assume that $\phi_h(t_h, x_h) = 0$, but we assume $\phi_h(t_h, x_h) = o(\sqrt{h})$. Then, we assume that $\frac{1}{2} - \mathbb{I}_n \leq o(\sqrt{h_n})$ and not the stronger $\frac{1}{2} - \mathbb{I}_n \leq 0$. But a quick inspection of the proof of [6, Proposition 4.1] reveals that these changes are irrelevant for the argument to work. □

C.6 Appendix

C.6.1 Results on weighted manifolds

Hereafter we collect some results about weighted Laplacians and heat kernels on closed manifolds. Let (M, g) be a k-dimensional, compact Riemannian manifold endowed with a measure $\mu := \xi \operatorname{Vol}_M$, with $\xi \in C^\infty(M)$, $\xi > 0$. We denote by Δ_ξ the associated Laplacian, which is defined on $f \in C^\infty(M)$ as

$$\Delta_\xi f := -\frac{1}{\xi} \operatorname{div}(\xi \nabla f).$$

We denote by H the corresponding heat kernel, i.e., H is a real valued function defined on $(0, +\infty) \times M \times M$ such that for any $u \in L^2(M)$ the function

$$e^{-t\Delta_\xi} u(x) := T(t)u(x) = \int_M H(t, x, y) u(y) d\mu(y),$$

defined for $(t, x) \in (0, +\infty) \times M$, is the unique solution to the Cauchy problem

$$\begin{cases} \partial_t v = -\Delta_\xi v & \text{in } (0, +\infty) \times M, \\ v(0, x) = u(x) & \text{on } M, \end{cases}$$

where the initial value at $t = 0$ is attained in the sense that

$$\lim_{t \downarrow 0} e^{-t\Delta_\xi} u = u \text{ in } L^2(M).$$

We will use the following results.

Theorem C.6.1. *Let M, ξ be as above. Let f be an $L^2(\xi)$-normalized eigenfunction of Δ_ξ corresponding to the eigenvalue λ, then for each integer $m \geq 0$*

$$\|f\|_{C^m(M)} \leq C_{M,m} \left(\lambda^{m+1+\frac{k}{2}} + 1\right).$$

Theorem C.6.2. *Let M, ξ be as above. There exists constants $Q_1, Q_2, Q_3, Q_4, \hat{Q}_1, \hat{Q}_2 > 0$ such that for every $t > 0$ and all $x, y \in M$,*

$$\frac{Q_1}{\mu(B_{\sqrt{t}}(x))} e^{-\frac{d^2_M(x,y)}{Q_2 t}} \leq H(t,x,y) \leq \frac{Q_3}{\mu(B_{\sqrt{t}}(x))} e^{-\frac{d^2_M(x,y)}{Q_4 t}}. \tag{C.45}$$

$$|\nabla_x H(t,x,y)| \leq \frac{\hat{Q}_1}{\sqrt{t}\mu(B_{\sqrt{t}}(x))} \exp\left(-\frac{d^2_M(x,y)}{\hat{Q}_2 t}\right). \tag{C.46}$$

Theorem C.6.3. *Let M, ξ be as above. There exist functions $v_j \in C^\infty(M \times M)$, $j \in \mathbf{N}$, such that for every $N > l + \frac{k}{2}$ there exists a constant $\tilde{C}_N < \infty$ such that*

$$\left| \nabla^l \left(H(t,x,y) - \frac{e^{-\frac{d^2_M(x,y)}{4t}}}{(4\pi t)^{k/2}} \sum_{j=0}^{N} v_j(x,y) t^j \right) \right| \leq \tilde{C}_N t^{N+1-\frac{k}{2}}, \tag{C.47}$$

provided $d(x,y) \leq \frac{\mathrm{inj}(M)}{2}$. Moreover we have

$$v_0(x,y) = \frac{1}{\sqrt{\xi(x)\xi(y)\det(d_{\exp_x^{-1}(y)} \exp_x)}}. \tag{C.48}$$

Theorem C.6.1 follows by the Sobolev embedding theorem and the L^2-regularity theory for elliptic equations on manifolds. Theorem C.6.2 follows from the Li–Yau inequality for weighted manifolds [75]. The asymptotic expansion in Theorem C.6.3 follows by constructing the heat kernel by means of the *parametrix* method: this construction is technical and we refer to [71], where this is carried out for the case $\xi = 1$. Here we just sketch the first part of the construction for a general density ξ, which gives (C.48). The idea is that when x, y are close enough, say $d(x,y) < \frac{\mathrm{inj}(M)}{2}$, a good approximation for the heat kernel should be given by

$$H_N(t,x,y) := G_t(x,y)\left(v_0(x,y) + ... + t^N v_N(x,y)\right), \tag{C.49}$$

for smooth functions v_j and $t > 0$. Here

$$G_t(x,y) := \frac{e^{-\frac{d^2_M(x,y)}{4t}}}{(4\pi t)^{k/2}}.$$

Since the Ansatz (C.49) should be an approximation of the heat kernel, we would like to have

$$0 = \partial_t H_N + \Delta_\xi H_N, \tag{C.50}$$

where Δ_ξ denotes the weighted Laplacian with respect to the y-variable. We now compute the right hand side of the above equation by using exponential coordinates around x: we denote them by $(r,\theta) \in [0,R) \times \mathbb{S}^{k-1}$. Observe that

$$\begin{aligned} \partial_t H_N &= \partial_t G_t(v_0 + ... + t^N v_N) + G_t(v_1 + ... + N t^{N-1} v_N) \\ &= \left(\frac{r^2}{4t^2} - \frac{k}{2t}\right) G_t(v_0 + ... + t^N v_N) + G_t(v_1 + ... + N t^{N-1} v_N). \end{aligned}$$

Furthermore

$$\begin{aligned}\Delta_\xi H_N &= G_t\left(\Delta_\xi v_0 + ... + t^N \Delta_\xi v_N\right)\\ &\quad + \Delta_\xi G_t(v_0 + ... + t^N v_N) - 2\langle \nabla G_t, \left(\nabla v_0 + ... + t^N \nabla v_N\right)\rangle.\end{aligned}$$

Using Gauss' Lemma and the fact that G_t is independent of θ we get

$$\begin{aligned}2\langle \nabla G_t, \left(\nabla v_0 + ... + t^N \nabla v_N\right)\rangle &= 2\partial_r G_t(\partial_r v_0 + ... + t^N \partial_r v_N)\\ &= -\frac{r}{t} G_t(\partial_r v_0 + ... + t^N \partial_r v_N).\end{aligned}$$

We also observe that by definition of Δ_ξ and by using again Gauss' Lemma and the independence of G_t from θ

$$\Delta_\xi G_t = \Delta G_t - \langle \frac{\nabla \xi}{\xi}, \nabla G_t\rangle = \Delta G_t + \frac{r}{2t}\frac{\partial_r \xi}{\xi} G_t.$$

We define

$$D(y) := \det(d_{\exp_x^{-1}(y)} \exp_x).$$

Using the expression of the Laplacian in spherical coordinates and the invariance of G_t with respect to θ we get

$$\Delta G_t = -\frac{\partial^2 G_t}{\partial r^2} - \partial_r G_t \left(\frac{\partial_r D}{D} + \frac{k-1}{r}\right) = -\left(\frac{r^2}{4t^2} - \frac{k}{2t}\right) G_t + \frac{r}{2t}\frac{\partial_r D}{D} G_t.$$

Putting things together we have

$$\begin{aligned}\partial_t H_N + \Delta_\xi H_N = G_t\bigg(&(v_1 + ... + N t^{N-1} v_N) - (\Delta_\xi v_0 + ... t^N \Delta_\xi v_N)\\ &+ \frac{r}{2t}\partial_r \log(D\xi)(v_0 + ... + t^N v_N) + \frac{r}{t}(\partial_r v_0 + ... + \partial_r v_N)\bigg).\end{aligned}$$

Although we cannot get (C.50) exactly, we can choose v_j in such a way that

$$\partial_t H_N + \Delta_\xi H_N = G_t t^N \Delta_\xi v_N.$$

In other words, we choose the coefficients in such a way that

$$\frac{r}{2t}\partial_r \log(D\xi) v_0 + \frac{r}{t}\partial_r v_0 = 0, \tag{C.51}$$

$$j t^{j-1} v_j - t^{j-1}\Delta_\xi v_{j-1} + t^{j-1}\frac{r}{2}\partial_r \log(D\xi) v_j + r t^{j-1}\partial_r v_j = 0, \text{ for } 1 \le j \le N. \tag{C.52}$$

Once one solves (C.51), one can show inductively that (C.52) admits a smooth solution v_j. It is easily seen that (C.51) can be solved explicitly to give

$$v_0(x,y) = \frac{1}{\sqrt{\xi(x)\xi(y)\det(d_{\exp_x^{-1}(y)} \exp_x)}}.$$

From here, the construction of the heat kernel and the estimate (C.47) follow verbatim as in [71].

C.6.2 Results on random geometric graphs

In this subsection we use the setting and the notation of Section C.2, with the points $\{x_i\}_{i=1}^{+\infty}$ being given by i.i.d. random points on M, distributed according to a probability distribution $\nu = \rho \operatorname{Vol}_M \in \mathcal{P}(M)$, with $\rho \in C^\infty(M)$, $\rho > 0$. The following two results are proved in [13, 14] for the unnormalized graph Laplacian, but the proof of the statements extends when we work with the random walk Laplacian. Hereafter, given $l \in \mathbf{N}$, we set

$$\gamma_l := \inf_{j<l, j\in\mathbf{N}} (\lambda_{j+1} - \lambda_j).$$

Theorem C.6.4. *In the above-mentioned setting, if additionally, the eigenvalues of Δ_{ρ^2} are simple, then for every $\bar{l} \in \mathbf{N}$ we have that with probability greater than*

$$1 - Q_1 \epsilon_n^{-6k} \exp(-Q_2 n \epsilon_n^{k+4}) - Q_3 n \exp(-Q_4 n \, (\lambda_{\bar{l}} + 1)^{-k})$$

we have for every $l \leq \bar{l}$

$$|\lambda_n^l - \kappa(\eta)\lambda_l| + \max_{z\in V_n} \left| v_n^l(z) - \frac{f_l(z)}{C_1^{1/2}} \right| \leq Q_5 \frac{\|f_l\|_{C^3(M)}}{\gamma_l} \epsilon_n.$$

Theorem C.6.5. *In the above-mentioned setting, if n is large enough, with probability greater than $1 - Q_6 \epsilon_n^{-k} \exp(-Q_7 n \epsilon_n^{k+2})$, we have that*

$$\max_{z\in V_n} |d_{n,\epsilon_n}(z) - C_1 \rho(z)| \leq Q_8 \epsilon_n.$$

We also recall the following result, which may be easily derived from [33, Theorem 2].

Theorem C.6.6. *Let (M, g) be a k-dimensional closed Riemannian manifold. Let $\rho \in C^\infty(M)$, $\rho > 0$ such that $\nu := \rho \operatorname{Vol}_M \in \mathcal{P}(M)$. Let $\{X_i\}_{i\in\mathbf{N}}$ be i.i.d. random points in M distributed according to ν and let $\nu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ be the associated empirical measures. Then there is a constant $C > 0$ such that almost surely there exist transport maps T_n such that $(T_n)_\# \nu = \nu_n$ and*

$$\begin{cases} \limsup_{n\to+\infty} \frac{n^{1/2} \sup_{x\in M} d_M(x, T_n(x))}{log^{3/4}(n)} \leq C & \text{if } k = 2, \\ \limsup_{n\to+\infty} \frac{n^{1/k} \sup_{x\in M} d_M(x, T_n(x))}{log^{1/k}(n)} \leq C & \text{if } k \geq 3. \end{cases} \tag{C.53}$$

BIBLIOGRAPHY

[1] Giovanni Alberti and Giovanni Bellettini. A non-local anisotropic model for phase transitions: asymptotic behaviour of rescaled energies. *Eur. J. Appl. Math.*, 9(3):261–284, 1998.

[2] Fred Almgren, Jean E. Taylor, and Lihe Wang. Curvature-driven flows: a variational approach. *SIAM J. Control. Optim.*, 31(2):387–438, 1993.

[3] Luigi Ambrosio and Norman Dancer. *Calculus of variations and partial differential equations*. Springer-Verlag, Berlin, 2000. Topics on geometrical evolution problems and degree theory, Papers from the Summer School held in Pisa, September 1996, Edited by G. Buttazzo, A. Marino and M. K. V. Murthy.

[4] Luigi Ambrosio, Guido De Philippis, and Luca Martinazzi. Gamma-convergence of nonlocal perimeter functionals. *Manuscripta Math.*, 134(3-4):377–403, 2011.

[5] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005.

[6] Guy Barles and Christine Georgelin. A simple proof of convergence for an approximation scheme for computing motions by mean curvature. *SIAM J. Numer. Anal.*, 32(2):484–500, 1995.

[7] Pierre Bérard, Gerard Besson, and Sylvestre Gallot. Embedding Riemannian manifolds by their heat kernel. *Geom. Funct. Anal.*, 4(4):373–398, 1994.

[8] Kenneth A. Brakke. *The motion of a surface by its mean curvature*, volume 20 of *Mathematical Notes*. Princeton University Press, Princeton, N.J., 1978.

[9] Leon Bungert, Jeff Calder, and Tim Roith. Uniform convergence rates for lipschitz learning on graphs. *IMA J. Numer. Anal.*, 09 2022.

[10] Jeff Calder. Consistency of Lipschitz learning with infinite unlabeled data and finite labeled data. *SIAM J. Math. Data Sci.*, 1(4):780–812, 2019.

[11] Jeff Calder. The game theoretic p-Laplacian and semi-supervised learning with few labels. *Nonlinearity*, 32(1):301–330, 2019.

[12] Jeff Calder, Brendan Cook, Matthew Thorpe, and Dejan Slepcev. Poisson learning: Graph based semi-supervised learning at very low label rates. *CoRR*, abs/2006.11184, 2020.

[13] Jeff Calder and Nicolás García Trillos. Improved spectral convergence rates for graph Laplacians on ε-graphs and k-NN graphs. *Appl. Comput. Harmon. Anal.*, 60:123–175, 2022.

[14] Jeff Calder, Nicolás García Trillos, and Marta Lewicka. Lipschitz regularity of graph Laplacians on random data clouds. *SIAM J. Math. Anal.*, 54(1):1169–1222, 2022.

[15] Jeff Calder, Dejan Slepčev, and Matthew Thorpe. Rates of Convergence for Laplacian Semi-Supervised Learning with Low Labeling Rates. *arXiv e-prints*, page arXiv:2006.02765, June 2020.

[16] Isaac Chavel. *Eigenvalues in Riemannian geometry*, volume 115 of *Pure and Applied Mathematics*. Academic Press, Inc., Orlando, FL, 1984.

[17] Jeff Cheeger, Mikhail Gromov, and Michael Taylor. Finite propagation speed, kernel estimates for functions of the Laplace operator, and the geometry of complete Riemannian manifolds. *J. Differ. Geom.*, 17(1):15–53, 1982.

[18] Yun Gang Chen, Yoshikazu Giga, and Shun'ichi Goto. Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations. *J. Differ. Geom.*, 33(3):749–786, 1991.

[19] Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Appl. Comput. Harmon. Anal.*, 21(1):5–30, 2006.

[20] Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)*, 27(1):1–67, 1992.

[21] Ennio De Giorgi and Tullio Franzoni. Su un tipo di convergenza variazionale. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei*, 58(6):842–850, 1975.

[22] Guido De Philippis and Tim Laux. Implicit time discretization for the mean curvature flow of mean convex sets. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5), 21:911-930*, 2020.

[23] Matthew M. Dunlop, Dejan Slepčev, Andrew M. Stuart, and Matthew Thorpe. Large data and zero noise limits of graph-based semi-supervised learning algorithms. *Appl. Comput. Harmon. Anal.*, 49(2):655–697, 2020.

[24] David B. Dunson, Hau-Tieng Wu, and Nan Wu. Spectral convergence of graph Laplacian and heat kernel reconstruction in L^∞ from random samples. *Appl. Comput. Harmon. Anal.*, 55:282–336, 2021.

[25] Selim Esedoḡlu and Felix Otto. Threshold dynamics for networks with arbitrary surface tensions. *Commun. Pur. Appl. Math.*, 68(5):808–864, 2015.

[26] Lawrence C. Evans. Convergence of an algorithm for mean curvature motion. *Indiana Univ. Math. J.*, 42(2):533–557, 1993.

[27] Lawrence C. Evans and Joel Spruck. Motion of level sets by mean curvature. I. *J. of Differ. Geom.*, 33(3):635–681, 1991.

[28] Herbert Federer. *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

[29] Julian Fischer, Sebastian Hensel, Tim Laux, and Theresa Simon. The local structure of the energy landscape in multiphase mean curvature flow: Weak-strong uniqueness and stability of evolutions. *arXiv: 2003.05478*, 2020.

[30] Irene Fonseca and Stefan Müller. Relaxation of quasiconvex functionals in $\mathrm{BV}(\Omega, \mathbf{R}^p)$ for integrands $f(x, u, \nabla u)$. *Arch. Ration. Mech. Anal.*, 123(1):1–49, 1993.

[31] Jakob Fuchs and Tim Laux. Strong convergence of the thresholding scheme for the mean curvature flow of mean convex sets. *arXiv: 2201.00413*, 2022. To appear in *Adv. Calc. Var.*

[32] Michael Gage and Richard S. Hamilton. The heat equation shrinking convex plane curves. *J. Differential Geom.*, 23(1):69–96, 1986.

[33] Nicolás García Trillos, Moritz Gerlach, Matthias Hein, and Dejan Slepčev. Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace-Beltrami operator. *Found. Comput. Math.*, 20(4):827–887, 2020.

[34] Nicolás García Trillos and Dejan Slepčev. Continuum limit of total variation on point clouds. *Arch. Ration. Mech. Anal.*, 220(1):193–241, 2016.

[35] Nicolás García Trillos and Dejan Slepčev. A variational approach to the consistency of spectral clustering. *Appl. Comput. Harmon. Anal.*, 45(2):239–281, 2018.

[36] Nicolás García Trillos, Dejan Slepčev, James von Brecht, Thomas Laurent, and Xavier Bresson. Consistency of Cheeger and ratio graph cuts. *J. Mach. Learn. Res.*, 17:Paper No. 181, 46, 2016.

[37] Matthew A. Grayson. A short note on the evolution of a surface by its mean curvature. *Duke Math. J.*, 58(3):555–558, 1989.

[38] Alexander Grigor'yan. *Heat kernel and analysis on manifolds*, volume 47 of *AMS/IP Studies in Advanced Mathematics*. American Mathematical Society, Providence, RI; International Press, Boston, MA, 2009.

[39] Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg. Graph Laplacians and their convergence on random neighborhood graphs. *J. Mach. Learn. Res.*, 8:1325–1368, 2007.

[40] Sebastian Hensel and Tim Laux. A new varifold solution concept for mean curvature flow: Convergence of the allen-cahn equation and weak-strong uniqueness. *arXiv: 2109.04233*, 2021.

[41] Tom Ilmanen. Generalized flow of sets by mean curvature on a manifold. *Indiana Univ. Math. J.*, 41(3):671–705, 1992.

[42] Tom Ilmanen. Convergence of the Allen-Cahn equation to Brakke's motion by mean curvature. *J. Differential Geom.*, 38(2):417–461, 1993.

[43] Hitoshi Ishii, Gabriel E. Pires, and Panagiotis E. Souganidis. Threshold dynamics type approximation schemes for propagating fronts. *J. Math. Soc. Japan*, 51(2):267–308, 1999.

[44] Matt Jacobs. A fast MBO scheme for multiclass data classification. *Sixth International Conference on Scale Space and Variational Methods in Computer Vision*, 2016.

[45] Matt Jacobs, Inwon Kim, and Alpár R. Mészáros. Weak Solutions to the Muskat Problem with Surface Tension Via Optimal Transport. *Arch. Ration. Mech. Anal.*, 239(1):389–430, 2021.

[46] Matt Jacobs, Ekaterina Merkurjev, and Selim Esedoḡlu. Auction dynamics: a volume constrained MBO scheme. *J. Comput. Phys.*, 354:288–310, 2018.

[47] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[48] David Kinderlehrer and Chun Liu. Evolution of grain boundaries. *Math. Models Methods Appl. Sci.*, 11(4):713–729, 2001.

[49] Tim Laux and Jona Lelmi. Large data limit of the MBO scheme for data clustering. To appear in *Proc. Appl. Math. Mech.*

[50] Tim Laux and Jona Lelmi. Large data limit of the MBO scheme for data clustering: Γ-convergence of the thresholding energies. *arXiv: 2112.06737*, 2021.

[51] Tim Laux and Jona Lelmi. De Giorgi's inequality for the thresholding scheme with arbitrary mobilities and surface tensions. *Calc. Var. Partial Differ. Equ.*, 61(1):Paper No. 35, 42, 2022.

[52] Tim Laux and Jona Lelmi. Large data limit of the MBO scheme for data clustering: convergence of the dynamics. *arXiv: 2209.05837*, 2022.

[53] Tim Laux and Felix Otto. Convergence of the thresholding scheme for multi-phase mean-curvature flow. *Calc. Var. Partial Diff. Equ.*, 55(5):Paper No .129 pp. 74, 2016.

[54] Tim Laux and Felix Otto. Brakke's inequality for the thresholding scheme. *Calc. Var. Partial Diff. Equ.*, 59(1):Paper No. 39, pp. 26, 2020.

[55] Tim Laux and Felix Otto. The thresholding scheme for mean curvature flow and De Giorgi's ideas for minimizing movements. *Adv. Stud. Pure Math. 85*, pages 63–93, 2020.

[56] Tim Laux and Drew Swartz. Convergence of thresholding schemes incorporating bulk effects. *Interfaces Free Bound.*, 19(2):273–304, 2017.

[57] Tim Laux and Nung Kwan Yip. Analysis of diffusion generated motion for mean curvature flow in codimension two: a gradient-flow approach. *Arch. Ration. Mech. Anal.*, 232(2):1113–1163, 2019.

[58] Stephan Luckhaus and Thomas Sturzenhecker. Implicit time discretization for the mean curvature flow equation. *Calc. Var. Partial Differ. Equ.*, 3(2):253–271, 1995.

[59] Francesco Maggi. *Sets of finite perimeter and geometric variational problems*, volume 135 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2012.

[60] Pierre Mascarenhas. Diffusion generated motion by mean curvature. *CAM Reports, Department of Mathematics, University of California, Los Angeles*, 1992.

[61] Ekaterina Merkurjev, Cristina Garcia-Cardona, Andrea L. Bertozzi, Arjuna Flenner, and Allon G. Percus. Diffuse interface methods for multiclass segmentation of high-dimensional data. *Appl. Math. Lett.*, 33:29–34, 2014.

[62] Ekaterina Merkurjev, Tijana Kostić, and Andrea L. Bertozzi. An MBO scheme on graphs for classification and image processing. *SIAM J. Imaging Sci.*, 6(4):1903–1930, 2013.

[63] Barry Merriman, James K. Bence, and Stanley J. Osher. Diffusion generated motion by mean curvature. *CAM Reports, Department of Mathematics, University of California, Los Angeles*, 1992.

[64] Barry Merriman, James K. Bence, and Stanley J. Osher. Motion of multiple functions: a level set approach. *J. of Comput. Phys.*, 112(2):334–363, 1994.

[65] Peter W. Michor and David Mumford. Riemannian geometries on spaces of plane curves. *J. Eur. Math. Soc.*, 8(1):1–48, 2006.

[66] Alexander Mielke. On evolutionary Γ-convergence for gradient systems. In *Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity*, volume 3 of *Lect. Notes Appl. Math. Mech.*, pages 187–249. Springer, 2016.

[67] Michele Miranda, Jr., Diego Pallara, Fabio Paronetto, and Marc Preunkert. Short-time heat flow and functions of bounded variation in $\mathbf{R}^N$. *Ann. Fac. Sci. Toulouse Math.*, 16(1):125–145, 2007.

[68] Oleksandr Misiats and Nung Kwan Yip. Convergence of space-time discrete threshold dynamics to anisotropic motion by mean curvature. *Discrete Contin. Dyn. Syst.-A*, 36(11):6379–6411, 2016.

[69] William W. Mullins. Two-dimensional motion of idealized grain boundaries. *J. Appl. Phys.*, 27(8):900–904, 1956.

[70] Boaz Nadler, Nathan Srebro, and Xueyuan Zhou. Semi-supervised learning with the graph laplacian: the limit of infinite unlabelled data. In *NIPS 2009*, 2009.

[71] Steven Rosenberg. *The Laplacian on a Riemannian manifold*, volume 31 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1997.

[72] Tiago Salvador and Selim Esedoḡlu. A simplified threshold dynamics algorithm for isotropic surface energies. *J. Sci. Comput.*, 79(1):648–669, 2019.

[73] Etienne Sandier and Sylvia Serfaty. Gamma-convergence of gradient flows with applications to Ginzburg-Landau. *Comm. Pure Appl. Math.*, 57(12):1627–1672, 2004.

[74] Isaac Jacob Schoenberg. Metric spaces and positive definite functions. *Trans. Am. Math. Soc.*, 44(3):522–536, 1938.

[75] Alberto G. Setti. Gaussian estimates for the heat kernel of the weighted Laplacian and fractal measures. *Can. J. Math.*, 44(5):1061–1078, 1992.

[76] Dejan Slepčev and Matthew Thorpe. Analysis of p-Laplacian regularization in semisupervised learning. *SIAM J. Math. Anal.*, 51(3):2085–2120, 2019.

[77] Drew Swartz and Nung Kwan Yip. Convergence of diffusion generated motion to motion by mean curvature. *Commun. Partial. Differ. Equ.*, 42(10):1598–1643, 2017.

[78] Yves van Gennip, Nestor Guillen, Braxton Osting, and Andrea L. Bertozzi. Mean curvature, threshold dynamics, and phase field theory on finite graphs. *Milan J. Math.*, 82(1):3–65, 2014.