Convolution kernels, and stability of threshold dynamics methods

Selim Esedoḡlu and Matt Jacobs

August 2, 2016

Abstract

Threshold dynamics and its extensions have proven useful in computing interfacial motions in applications as diverse as materials science and machine learning. Certain desirable properties of the algorithm, such as unconditional monotonicity in two-phase flows and gradient stability more generally, hinge on positivity properties of the convolution kernel and its Fourier transform. Recent developments in the analysis of this class of algorithms indicate that sometimes, as in the case of certain anisotropic curvature flows arising in materials science, these properties of the convolution kernel cannot be expected. Other applications, such as machine learning, would benefit from as great a level of flexibility in choosing the convolution kernel as possible. In this paper, we establish certain desirable properties of threshold dynamics, such as gamma convergence of its associated energy, for a substantially wider class of kernels than has been hitherto possible. We also present variants of the algorithm that extend some of these properties to even wider classes of convolution kernels.

1 Introduction

Originally proposed by Merriman, Bence, and Osher (MBO) in [21, 20], threshold dynamics – also known as diffusion or convolution generated motion – is a very efficient algorithm for approximating the motion by mean curvature of an interface. The algorithm generates a discrete in time approximation to mean curvature motion by alternating two very simple steps: convolution with a kernel, and pointwise thresholding. Among its benefits are 1. implicit representation of the interface as in the phase field or level set methods, allowing for graceful handling of topological changes, 2. unconditional stability, where the time step size is restricted only by accuracy considerations, and 3. very low per time step cost when implemented on uniform grids.

Motion by mean curvature arises as L^2 gradient descent, in an appropriate sense, for perimeter of sets, which in turn appears in variational models for a great variety of applications. These range from image processing and computer vision (e.g. the Mumford-Shah model [24] for image segmentation) to

materials science (e.g. Mullins' model [23] for grain boundary motion in polycrystals). More recently, such variational models and their minimization via gradient descent have also been applied in the context of machine learning and artificial intelligence (e.g. graph partitioning models for supervised clustering of data [13]). The MBO scheme, its variants, and its extensions have attracted sustained interest in the context of each one of these applications.

The elegant, streamlined nature of threshold dynamics has made it amenable to analysis and the focus of a number of theoretical investigations, see e.g. [16, 7, 10, 17, 9] and references therein. The various consistency, stability, and convergence statements contained in these contributions require various assumptions on the kernel used in the convolution step of the algorithm, such as positivity in the physical or the Fourier domain. In this paper, we present a number of new results on the original threshold dynamics algorithm and some of its variants and extensions that significantly enlarge the class of admissible kernels. We also demonstrate that some of the remaining restrictions are necessary. There are multiple reasons for seeking an extension of the theory to more general kernels. Two of these are: 1. Recent results in [9] indicate that positivity of the kernel cannot always be maintained for extensions of the MBO scheme to anisotropic (weighted) motion by mean curvature. 2. In applications such as graph partitioning, there is often little control on the properties of the convolution kernel that is typically constructed from the given edge weights of the graph.

2 Preliminaries and Notation

We will be concerned with possibly anisotropic interfacial energies defined on partitions of a domain D. D will typically be the d-dimensional annulus, i.e. a cube in $\mathbb{R}^d$ with periodic boundary conditions. By a partition of D, we mean a prescribed number N of sets $\Sigma_1, \ldots, \Sigma_N \subseteq D$ that satisfy

$$\bigcup_{i=1}^{N} \Sigma_i = D \text{ and } \Sigma_i \cap \Sigma_j = (\partial\Sigma_i) \cap (\partial\Sigma_j) \text{ for } i \neq j \tag{1}$$

Let H^s denote the s-dimensional Hausdorff measure on D. Given a non-negative, continuous, even function $\sigma : \mathbb{S}^{d-1} \to \mathbb{R}^+$ with $\sigma(x) > 0$ for $x \neq 0$, we first consider the two-phase surface energy

$$E(\Sigma, \sigma) = \int_{\partial\Sigma} \sigma(n(x))\, dH^{d-1} \tag{2}$$

where $n(x)$ denotes the outward unit normal to $\partial\Sigma$. We will also consider the multiphase extension of energy (2) to partitions. Let $N \in \mathbb{N}^+$ denote the number of phases, and define the set of distinct pairs of indices:

$$\mathcal{I}_N = \{(i, j) \in \{1, \ldots, N\} \times \{1, \ldots, N\} : i \neq j\}. \tag{3}$$

Our multiphase energy is:

$$E(\mathbf{\Sigma}, \sigma) = \sum_{(i,j)\in\mathcal{I}_N} \int_{(\partial\Sigma_i)\cap(\partial\Sigma_j)} \sigma_{i,j}(n(x))\, dH^{d-1}(x) \tag{4}$$

where we write $\mathbf{\Sigma} = (\Sigma_1, \ldots, \Sigma_N)$. It will be convenient to assume that σ has been extended to $\sigma : \mathbb{R}^d \to \mathbb{R}^+$ as

$$\sigma(x) = |x|\sigma\left(\frac{x}{|x|}\right) \text{ for } x \neq 0$$

so that it is positively 1-homogeneous. We will assume that σ is then a convex function on $\mathbb{R}^d$; this condition will ensure well-posedness of the two-phase energy (2). Define the unit ball (i.e. the *Frank diagram)* B_σ *of* σ *as*

$$B_\sigma = \{x \,:\, \sigma(x) \leq 1\}$$

which is thus a closed, convex, centrally symmetric set. We will further require B_σ to be smooth and strictly convex; this implies that we stay clear of the crystalline cases (where B_σ is a polytope) except via approximation. In two dimensions, we will also write $\sigma = \sigma(\theta)$, where θ is the angle that the unit normal makes with the positive x_1-axis. In that case, strict convexity of B_σ is equivalent to the condition

$$\sigma''(\theta) + \sigma(\theta) > 0.$$

The *Wulff shape* W_σ associated with the anisotropy σ is defined as

$$W_\sigma = \left\{ y \,:\, \sup_{x\in B_\sigma} x \cdot y \leq 1 \right\}.$$

The sets B_σ can in turn be obtained from W_σ by the formula

$$B_\sigma = \left\{ x \,:\, \sup_{y\in W_\sigma} x \cdot y \leq 1 \right\},$$

exhibiting the well known duality between B_σ and W_σ. Our assumptions on B_σ imply that W_σ is also strictly convex and has smooth boundary.

For $d = 2$ or 3, we will study approximations for L^2 gradient flow of energies (2) and (4), which are known as weighted mean curvature flow (of an interface and a network). The normal speed of an interface in three dimensions under this flow is given by

$$v_\perp(x) = \mu(n(x)) \left(\left(\partial_{s_1}^2 \sigma(n(x)) + \sigma(n(x)) \right) \kappa_1(x) + \left(\partial_{s_2}^2 \sigma(n(x)) + \sigma(n(x)) \right) \kappa_2 \right) \tag{5}$$

where κ_1 and κ_2 are the two principal curvatures, and ∂_{s_i} denotes differentiation along the great circle on $\mathbb{S}^2$ that passes through $n(x)$ and has as its tangent the i-th principal curvature direction. In two dimensions, the expression simplifies to

$$v_\perp(x) = \mu(n(x))\sigma(n(x)) \left(\kappa_1(x) + \kappa_2(x)\right). \tag{6}$$

In addition to (5), a condition known as the *Herring angle condition* [15] holds along triple junctions: For $d = 3$, at a junction formed by the meeting of the three phases Σ_i, Σ_j , and Σ_k , this condition reads

$$(\ell \times n_{i,j})\sigma_{i,j}(n_{i,j}) + (\ell \times n_{j,k})\sigma_{j,k}(n_{j,k}) + (\ell \times n_{k,i})\sigma_{k,i}(n_{k,i}) \\ + n_{j,i}\sigma'_{i,j}(n_{i,j}) + n_{k,j}\sigma'_{j,k}(n_{j,k}) + n_{i,k}\sigma'_{k,i}(n_{k,i}) = 0 \quad (7)$$

where $n_{i,j}$ is the unit normal vector to the interface $(\partial\Sigma_i) \cap (\partial\Sigma_j)$ pointing in the Σ_i to Σ_j direction, $\ell = n_{j,k} \times n_{i,j}$ is a unit vector tangent to the triple junction, and $\sigma'_{i,j}(n_{i,j})$ denotes derivative of $\sigma_{i,j}$ taken on $\mathbb{S}^2$ in the direction of the vector $\ell \times n_{i,j}$. In the isotropic setting, (7) simplifies to the following more familiar form, known as Young's law:

$$\sigma_{i,j}n_{i,j} + \sigma_{j,k}n_{j,k} + \sigma_{k,i}n_{k,i} = 0. \quad (8)$$

Finally, we note that well-posedness (lower semi-continuity) of the multi-phase energy (4) in its full generality is a complicated matter [3]. At the very least, the surface tensions $\sigma_{i,j} : \mathbb{R}^d \to \mathbb{R}^+$ need to be convex and satisfy a pointwise triangle inequality

$$\sigma_{i,j}(n) + \sigma_{j,k}(n) \geq \sigma_{i,k}(n) \quad (9)$$

for all distinct i, j, and k, and all $n \in \mathbb{S}^{d-1}$. In case the $\sigma_{i,j}$ are positive constants, (9) is known to be also sufficient for well-posedness of model (4).

3 Previous Work

In its simplest form, the two-phase MBO algorithm as presented in the original paper [20] generates a discrete in time, continuous in space approximation to the motion by mean curvature of an interface $\partial\Sigma^0$ (given as the boundary of an initial set Σ^0) as follows:

Algorithm 1: (MBO'92): Given a time step size $\delta t > 0$, alternate the following steps:

1. Convolution:

$$\psi^k = \frac{1}{(\delta t)^{\frac{d}{2}}} K\left(\frac{x}{\sqrt{\delta t}}\right) * \mathbf{1}_{\Sigma^k} \quad (10)$$

2. Thresholding:

$$\Sigma^{k+1} = \left\{x \,:\, \psi^k(x) \geq \frac{1}{2}\right\}. \quad (11)$$

where Σ^k denotes the approximate solution at time $t = k\delta t$, and the convolution kernel $K \in L^1(\mathbb{R}^d)$ satisfies

$$K(x) \in L^1(\mathbb{R}^d) \;\; , \; xK(x) \in L^1(\mathbb{R}^d), \text{ and } K(x) = K(-x) \quad (12)$$

together with

$$\int_{\mathbb{R}^d} K(x)\,dx = 1. \tag{13}$$

For convenience, we will write

$$K_\varepsilon(x) = \frac{1}{\varepsilon^d} K\left(\frac{x}{\varepsilon}\right)$$

for the rescaled versions of a given convolution kernel K. In the original papers [21, 20], the kernel K is taken to be the Gaussian:

$$G(x) = \frac{1}{(4\pi)^{\frac{d}{2}}} \exp\left(-\frac{|x|^2}{4}\right)$$

although the possibility of choosing it to be not necessarily radially symmetric for anisotropic curvature motions is also mentioned.

There have been multiple studies devoted to the question of convergence for Algorithm 1. In [19], [26], [25], consistency of the scheme is studied via Taylor expansion after one step of the algorithm is applied on a set with a smooth boundary. Rigorous convergence results were first given in [11] and [4]. [16] studies the algorithm with fairly general convolution kernels K, and establishes its convergence to the viscosity solution [12, 8] of certain anisotropic curvature flows provided that K satisfies certain conditions, chief among which is positivity. Positivity of K implies that the scheme preserves a comparison principle known to hold for the evolution (5) and is crucial in the viscosity solutions approach.

In [10], a variational formulation for the original MBO scheme (10) & (11) was given. In particular, it was shown that the following functional defined on sets, with kernel K chosen to be the Gaussian G, which had previously been established [1, 22] to be a non-local approximation to (isotropic) perimeter, is dissipated by the MBO scheme at every step, regardless of time step size:

$$E_{\sqrt{\delta t}}(\Sigma, K_{\sqrt{\delta t}}) = \frac{1}{\sqrt{\delta t}} \int_{\Sigma^c} K_{\sqrt{\delta t}} * \mathbf{1}_\Sigma \, dx. \tag{14}$$

Thus, (14) is a Lyapunov functional for Algorithm 1, establishing its unconditional gradient stability. Moreover, the following *minimizing movements* [2, 18] interpretation involving (14) for Algorithm 1 was given in [10]:

$$\Sigma^{k+1} = \arg\min_{\Sigma} E_{\sqrt{\delta t}}(\Sigma, K_{\sqrt{\delta t}}) + \frac{1}{\sqrt{\delta t}} \int (\mathbf{1}_\Sigma - \mathbf{1}_{\Sigma^k}) K_{\sqrt{\delta t}} * (\mathbf{1}_\Sigma - \mathbf{1}_{\Sigma^k})\,dx \tag{15}$$

where the kernel K was again taken to be the Gaussian. In [10], variational formulation (14) & (15) was also extended to the *multiphase* setting.

Let us recall the following fact from [10] that ensures (14) is a Lyapunov functional for Algorithm 1, establishing the connection between the variational problem (14) and threshold dynamics, and underlining the significance of $\widehat{K}$:

Proposition 3.1. *(from [10]) Let K satisfy (12) and (13). If $\widehat{K} \geq 0$, threshold dynamics algorithm (10) & (11) decreases energy (14) at every time step, regardless of the time step size.*

In [10], the variational formulation (15) was then extended to the multiphase energy (4) in case the surface tensions $\sigma_{i,j}$ are constant but possibly distinct:

$$E(\mathbf{\Sigma}, \sigma) = \sum_{(i,j)\in\mathcal{I}_N} \sigma_{i,j} H^{d-1}(\partial\Sigma_i \cap \partial\Sigma_j) \tag{16}$$

in which case the Lyapunov functional becomes

$$E_{\sqrt{\delta t}}(\mathbf{\Sigma}, K_{\sqrt{\delta t}}) = \frac{1}{\sqrt{\delta t}} \sum_{(i,j)\in\mathcal{I}_N} \sigma_{i,j} \int_{\Sigma_j} K_{\sqrt{\delta t}} * \mathbf{1}_{\Sigma_i}\, dx. \tag{17}$$

We also consider a relaxation of (17):

$$E_{\sqrt{\delta t}}(\mathbf{u}, K_{\sqrt{\delta t}}) = \frac{1}{\sqrt{\delta t}} \sum_{(i,j)\in\mathcal{I}_N} \sigma_{i,j} \int_{D} u_j K_{\sqrt{\delta t}} * \mathbf{1}_{\Sigma_i}\, dx \tag{18}$$

over the following convex set of functions satisfying a box constraint:

$$\mathcal{K} = \left\{ \mathbf{u} \in L^1(D, [0,1]^N) \,:\, \sum_{i=1}^{N} u_i(x) = 1 \text{ a.e. } x \in D \right\}. \tag{19}$$

There is a corresponding minimizing movements scheme that can be derived from (51) that leads to the following extension of threshold dynamics, Algorithm 2, to the constant but possibly unequal surface tension multiphase energy (16).

Algorithm 2: ([10]): Given a time step size $\delta t > 0$, alternate the following steps:

1. Convolution:

$$\psi_i^k = K_{\sqrt{\delta t}} * \sum_{j\neq i} \sigma_{i,j} \mathbf{1}_{\Sigma_j^k}. \tag{20}$$

2. Thresholding:

$$\Sigma_i^{k+1} = \left\{ x \,:\, \psi_i^k(x) \leq \min_{j\neq i} \psi_j^k(x) \right\}. \tag{21}$$

Various conditions are provided in [10] for ensuring that Algorithm 2 is unconditionally gradient stable (decreases energy (17) for any $\delta t > 0$). The question turns out to be interesting, with connections to isometric embeddability of finite metric spaces into Euclidean spaces. In particular, the triangle inequality (9) on $\sigma_{i,j}$ appears to be neither necessary nor sufficient.

Turning to anisotropy, i.e. normal dependent surface tensions $\sigma = \sigma(n)$ and the more general convolution kernels it requires, we recall the following facts from [9]:

Proposition 3.2. *(from [9]) Let Σ be a compact subset of $\mathbb{R}^d$ with smooth boundary. Let $K : \mathbb{R}^d \to \mathbb{R}$ be a kernel satisfying (12). Then:*

$$\lim_{\delta t \to 0^+} E_{\sqrt{\delta t}}(\Sigma, K_{\sqrt{\delta t}}) = \int_{\partial\Sigma} \sigma_K\big(n(x)\big)\, dH^{d-1}(x)$$

where the surface tension $\sigma_K : \mathbb{R}^d \to \mathbb{R}^+$ is defined as

$$\sigma_K(n) := \frac{1}{2} \int_{\mathbb{R}^d} |n \cdot x| K(x)\, dx. \tag{22}$$

We will in fact prove a stronger, Gamma convergence version of Proposition 3.2 below for a class of kernels that include sign changing ones. In polar coordinates, the expression for the surface tension σ_K that corresponds to a given convolution kernel K is:

$$\sigma_K(n) = \frac{1}{2} \int_0^\infty r^d \int_{\mathbb{S}^{d-1}} |n \cdot x| K(rx)\, dH^{d-1}(x)\, dr. \tag{23}$$

In [9], the following expression is obtained for the mobility μ_K associated with a given kernel K:

$$\mu_K(n) := \int_{n^\perp} K(x)\, dH^{d-1}(x). \tag{24}$$

Let us also note the following *Barrier Theorem* from [9] that places a restriction on the positivity of convolution kernels in terms of the Wulff shape W_σ of the given anisotropy σ.

Theorem 3.3. *(from [9]) Threshold dynamics algorithm (10) & (11) with a positive kernel K can approximate weighted mean curvature flow (5) associated with an anisotropic surface tension $\sigma : \mathbf{S}^{d-1} \to \mathbb{R}$ (for some choice of mobility $\mu : \mathbb{S}^{d-1} \to \mathbb{R}$) if and only if the corresponding Wulff shape W_σ is a* zonoid.

Zonoids are centrally symmetric convex bodies that are limits, in the Hausdorff topology, of *zonotopes*, which are defined as (finite) vector sums of line segments. In $\mathbb{R}^d$, a convex polytope with nonempty interior is a zonotope if and only if every $d-1$ dimensional face of it is a zonotope. Thus, for $d = 2$, any centrally symmetric, convex body is a zonoid. For $d = 3$, this is no longer the case. A simple example of a non-zonoid in $\mathbb{R}^3$ is the octahedron. Moreover, there exists a neighborhood of the octahedron that contains no zonoids. Theorem 3.3 implies that there is no monotone threshold dynamics scheme for an anisotropy σ the Wulff shape W_σ of which falls into such a neighborhood, even though W_σ may be smooth and strictly convex. See [14, 6] for these facts and much more information about zonoids. Thus, as promised in Section 1, this is an example of an application where positivity properties of the convolution kernel has to be compromised.

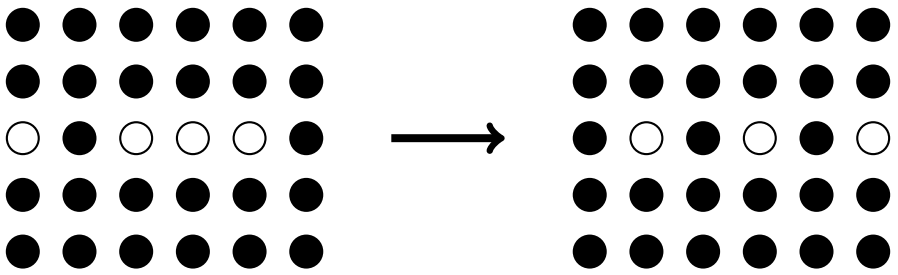

Figure 1: One step of the standard algorithm under the action of kernel K_1 on the periodic lattice $\mathbb{Z}/6\mathbb{Z} \bigoplus \mathbb{Z}/5\mathbb{Z}$. The updated configuration has a higher energy than the starting configuration.

4 New Variants of Threshold Dynamics

Here, we will consider extensions of the basic MBO algorithm (10) & (11) that allow us to dispense with various requirements on the convolution kernel K and, in the multiphase setting, on the surface tensions σ.

4.1 Non-monotone energy and oscillating solutions

We first establish a partial converse to Proposition 3.1, showing that assumption $\widehat{K} \geq 0$ on the Fourier transform of the kernel is not spurious.

Example 1:

Let the convolution kernel be given by

$$K_1\big((x_1, x_2)\big) = \begin{cases} 1/3 & (x_1, x_2) = (\pm 1, 0) \\ 1/9 & (x_1, x_2) = (0, \pm 1) \text{ or } (0, 0) \\ 0 & \text{otherwise} \end{cases}$$

Then the Fourier transform of K_1 must change signs since the origin is not the global maximum of K_1. Figure 1 shows an example where a step of the algorithm with K_1 increases the energy. The right hand side configuration has 6 broken horizontal bonds and 6 broken vertical bonds, while the left hand side has 4 broken horizontal bonds and 8 broken vertical bonds. K_1 assigns horizontal bonds a strength of $1/3$ and vertical bonds a strength of $1/9$ thus if we compare the two energies we see that $E_{RHS} - E_{LHS} = 6*1/3+6*1/9-(4*1/3+8*1/9) = 8/9$ thus the energy has increased under the algorithm.

Example 2:

In fact it is possible for the algorithm to get trapped in a periodic cycle where one of the configurations in the cycle has a higher energy than the others. Figure 2 shows an example of the algorithm with K_1 getting trapped in a 2-cycle where the right hand configuration has a higher energy. Both configurations have the same number of broken horizontal bonds. However the left hand side has 4 broken vertical bonds while the right hand side has 8 broken vertical bonds.

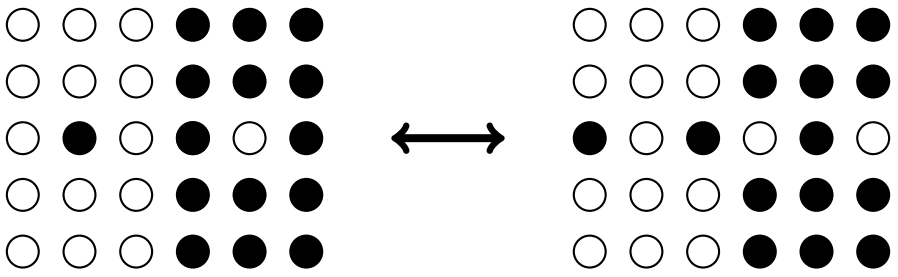

Figure 2: Behavior of the standard algorithm under the action of kernel K_1 on the periodic lattice $\mathbb{Z}/6\mathbb{Z} \bigoplus \mathbb{Z}/5\mathbb{Z}$. The algorithm gets trapped in a periodic loop between the two configurations above. The configuration on the right has a higher energy

4.2 A new variant: Single growth

Next, we show how to modify the original two-phase MBO algorithm, Algorithm 1, to a slightly more costly version (entailing two convolutions per time step as opposed to one) so that the assumption $\widehat{K} \geq 0$ can be dramatically relaxed while maintaining the energy dissipation property.

Algorithm 3: Alternate the following steps:

1. 1st Convolution:

$$\psi^{k+\frac{1}{2}} = \frac{1}{(\delta t)^{\frac{d}{2}}} K\left(\frac{x}{\sqrt{\delta t}}\right) * \mathbf{1}_{\Sigma^k} \tag{25}$$

2. 1st Thresholding:

$$\Sigma^{k+\frac{1}{2}} = \Sigma^k \cup \left\{x \,:\, \psi^{k+\frac{1}{2}}(x) \geq \frac{1}{2}\right\}. \tag{26}$$

3. 2nd Convolution:

$$\psi^{k+1} = \frac{1}{(\delta t)^{\frac{d}{2}}} K\left(\frac{x}{\sqrt{\delta t}}\right) * \mathbf{1}_{\Sigma^{k+\frac{1}{2}}} \tag{27}$$

4. 2nd Thresholding:

$$\Sigma^{k+1} = \Sigma^{k+\frac{1}{2}} \setminus \left\{x \,:\, \psi^{k+1}(x) \leq \frac{1}{2}\right\}. \tag{28}$$

Proposition 4.1. *If the convolution kernel K is of the form $K = K^1 + K^2$ with $K^1 \geq 0$ and $\widehat{K^2} \geq 0$, then Algorithm 3 dissipates energy (14) at every step. Furthermore if K^1 is positive in a neighborhood of the origin or $\hat{K}^2(0)$ is positive then if an iteration of Algorithm 1 changes a configuration Σ by a set of*

positive measure, an iteration of Algorithm 3 also changes Σ by a set of positive measure and strictly decreases the energy.

Proof. We will show that energy (14) is dissipated going from Σ^k to $\Sigma^{k+\frac{1}{2}}$. The argument going from $\Sigma^{k+\frac{1}{2}}$ to Σ^{k+1} is essentially the same. Let $\varphi = \mathbf{1}_{\Sigma^{k+\frac{1}{2}}} - \mathbf{1}_{\Sigma^k}$. Then $\varphi(x)$ is pointwise nonnegative since $\Sigma^k \subset \Sigma^{k+\frac{1}{2}}$. Comparing the energies we have:

$$\begin{aligned}
&E_{\sqrt{\delta t}}(\Sigma^{k+\frac{1}{2}}, K_{\sqrt{\delta t}}) - E_{\sqrt{\delta t}}(\Sigma^k, K_{\sqrt{\delta t}}) \\
&= \frac{1}{\sqrt{\delta t}}\Big(\int_D \varphi(x)(K_{\sqrt{\delta t}} * (\mathbf{1}_{(\Sigma^k)^c} - \mathbf{1}_{\Sigma^k}))(x)\, dx - \int_D \varphi(x)(K_{\sqrt{\delta t}} * \varphi)(x)\, dx \Big).
\end{aligned}$$

Note that $(K_{\sqrt{\delta t}} * (\mathbf{1}_{(\Sigma^k)^c} - \mathbf{1}_{\Sigma^k}))(x) > 0$ if and only if $\psi^{k+\frac{1}{2}}(x) < \frac{1}{2}$ and thus if and only if $\varphi(x) = 0$. Therefore

$$\frac{1}{\sqrt{\delta t}} \int_D \varphi(x)(K_{\sqrt{\delta t}} * (\mathbf{1}_{(\Sigma^k)^c} - \mathbf{1}_{\Sigma^k}))(x)\, dx \leq 0.$$

To establish the dissipation of energy, it remains to show that

$$-\frac{1}{\sqrt{\delta t}} \int_D \varphi(x)(K_{\sqrt{\delta t}} * \varphi)(x)\, dx \leq 0.$$

Let L be the periodic lattice associated to D. Then using the Fourier series expansion we have

$$-\int_D \varphi(x)(K_{\sqrt{\delta t}} * \varphi)(x)\, dx = -\sum_{\alpha \in L} \widehat{\varphi}(\alpha)^2 \widehat{K}(\alpha\sqrt{\delta t}). \tag{29}$$

If K is nonnegative then it is clear that the left hand side of the above equation is ≤ 0 and if $\widehat{K}$ is nonnegative then is clear that the right hand side is ≤ 0. Therefore if K can be split into a sum $K = K^1 + K^2$ where $K^1 \geq 0$ and $\widehat{K}^2 \geq 0$ we have

$$E_{\sqrt{\delta t}}(\Sigma^{k+\frac{1}{2}}, K_{\sqrt{\delta t}}) - E_{\sqrt{\delta t}}(\Sigma^k, K_{\sqrt{\delta t}}) \leq 0.$$

Now we prove the second statement. By the above it is enough to show that one of the steps of Algorithm 3 strictly decreases the energy. Let $\Sigma^0 = \Sigma$ and let Σ^1 and $\tilde{\Sigma}^1$ be the configurations obtained from Σ^0 after a single iteration of Algorithm 1 and Algorithm 3 respectively. Let $\varphi(x) = \mathbf{1}_{\Sigma^1} - \mathbf{1}_{\Sigma^0}$. By assumption the support of $\varphi(x)$ has positive measure. Let $\tilde{\Sigma}^{1/2}$ be the intermediate set obtained after applying the first two steps of Algorithm 3 to Σ^0. Then

$$\mathbf{1}_{\tilde{\Sigma}^{1/2}} - \mathbf{1}_{\Sigma^0} = \max(\varphi(x), 0) = \varphi_+(x).$$

First suppose that the support of $\varphi_+(x)$ also has positive measure. Consider the change in energy

$$E_{\sqrt{\delta t}}(\tilde{\Sigma}^{1/2}, K_{\sqrt{\delta t}}) - E_{\sqrt{\delta t}}(\Sigma^0, K_{\sqrt{\delta t}}) \leq -\frac{1}{\sqrt{\delta t}} \int_D \varphi_+(x)(K_{\sqrt{\delta t}} * \varphi_+)(x)\, dx =$$

$$-\frac{1}{\sqrt{\delta t}}\int_D \varphi_+(x)(K^1_{\sqrt{\delta t}} * \varphi_+)(x)\,dx - \frac{1}{\sqrt{\delta t}}\sum_{\alpha\in L}(\widehat{\varphi}_+(\alpha))^2\hat{K}^2(\alpha\sqrt{\delta t}).$$

It is enough to show that one of the two terms is strictly negative.

If $\hat{K}^2(0)$ is positive then we only need to show that $\hat{\varphi}_+(0) \neq 0$. This must be the case as φ_+ does not change signs and has support of positive measure.

If K^1 is positive in a neighborhood of the origin then there exists $\delta_0 > 0$ and $b_0 > 0$ such that $K^1(z) \geq b_0$ for all $z \in B(0,\delta_0)$. By the nonnegativity of K^1 we have

$$-\int_D \varphi_+(x)(K^1_{\sqrt{\delta t}} * \varphi_+)(x)\,dx \leq -b_0\int_D \varphi_+(x)\int_{B(0,\delta_0)} \varphi_+(x+z\sqrt{\delta t})\,dz\,dx.$$

By the Lebesgue differentiation theorem

$$\lim_{\delta_0\to 0}\frac{1}{m(B(0,\delta_0))}\int_{B(0,\delta_0)}\varphi_+(x+z\sqrt{\delta t})\,dz = 1$$

for almost every $x \in \mathrm{supp}(\varphi_+)$. Therefore

$$-b_0\int_D \varphi_+(x)\int_{B(0,\delta_0)}\varphi_+(x+z\sqrt{\delta t})\,dz < 0.$$

On the other hand if $\varphi_+(x)$ has support of measure zero then $\tilde{\Sigma}^{1/2}$ is equal to Σ^0 except on a set of measure zero. Therefore

$$\frac{1}{(\delta t)^{\frac{d}{2}}}K\left(\frac{x}{\sqrt{\delta t}}\right) * \mathbf{1}_{\tilde{\Sigma}^{1/2}} = \frac{1}{(\delta t)^{\frac{d}{2}}}K\left(\frac{x}{\sqrt{\delta t}}\right) * \mathbf{1}_{\Sigma^0}.$$

It then follows that

$$\mathbf{1}_{\tilde{\Sigma}^1} - \mathbf{1}_{\tilde{\Sigma}^{1/2}} = \min(\varphi(x), 0) = \varphi_-(x).$$

Since φ_+ had support of measure zero, φ_- must have positive support. An analogous argument to the above implies that the energy must decrease.

□

In addition to extending energy dissipation property to far more general kernels as described above in Proposition 4.1, Algorithm 3 maintains convergence to viscosity solution [12, 8] of the level set formulation of flow (5) in case the convolution kernel happens to be positive, with suitable decay and regularity, as we explain next. We will adapt to our new algorithm (3) the convergence argument that was given in [16] for the standard MBO scheme (1) for positive but otherwise fairly general convolution kernels. Hence, for the remainder of this subsection, we assume that kernel K satisfies the positivity, regularity, and decay properties (3.1) through (3.7) in [16], which are more stringent than assumptions needed elsewhere in this paper. In this framework, first threshold

dynamics is extended from sets (binary functions) to $L^1(\mathbb{R}^d)$ in a level set–by–level set fashion: For $\varphi \in L^1(\mathbb{R})$, let

$$S_h\varphi(x) = (K_{\sqrt{h}} * \varphi)(x) \tag{30}$$

$$G_h\varphi(x) = \sup\{\lambda \in \mathbb{R} : S_h \mathbf{1}_{\{y:\varphi(y)>\lambda\}}(x) \geq 1/2\} \tag{31}$$

in keeping with the notation of [16]. If φ happens to be a characteristic function, applying G_h gives one step of the standard MBO algorithm (1) with time step size h. The new, single growth version of threshold dynamics described in Algorithm 3 can be written in terms of G_h as well. To that end, define the following two new operators:

$$G_h^{\uparrow}\varphi(x) = \max(\varphi(x), G_h\varphi(x)) \tag{32}$$

$$G_h^{\downarrow}\varphi(x) = \min(\varphi(x), G_h\varphi(x)). \tag{33}$$

Then, one step of Algorithm 3 applied to a function $\varphi(x)$ is given by

$$G_h^{\downarrow}G_h^{\uparrow}\varphi(x). \tag{34}$$

Next, define a piecewise constant–in–time approximation to the propagator of the limiting continuum flow:

$$Q_t^h = \left(G_h^{\downarrow}G_h^{\uparrow}\right)^{j-1} \text{ if } (j-1)h \leq t < jh \text{ with } j \in \mathbb{N}. \tag{35}$$

We can now state our convergence result.

Theorem 4.2. *Let $g : \mathbb{R}^d \to \mathbb{R}$ be a bounded, uniformly continuous function. Let $u : \mathbb{R}^d \times [0,\infty) \to \mathbb{R}$ be the unique viscosity solution of the PDE*

$$\begin{cases} u_t = -F(D^2u, Du) \\ u(x,0) = g(x) \end{cases}$$

where F is given by

$$F(M,p) = -\left(\int_{p^\perp} K(x)\, dH^{d-1}(x)\right)^{-1} \left(\frac{1}{2}\int_{p^\perp} \langle Mx, x\rangle K(x)\, dH^{d-1}(x)\right) \tag{36}$$

for $d \times d$ symmetric matrices M and $p \in \mathbb{R}^d$. Then, for any $T \in [0,\infty)$,

$$Q_t^h g(x) \longrightarrow u(x,t) \text{ uniformly on } \mathbb{R}^d \times [0,T]$$

as $h \to 0^+$.

Operator $G_h^{\downarrow}G_h^{\uparrow}$ shares the following properties with G_h that are essential for the framework of [16]:

1. $G_h^{\downarrow}G_h^{\uparrow}(\rho \circ \varphi) = \rho \circ (G_h^{\downarrow}G_h^{\uparrow}\varphi)$ for all continuous, nondecreasing functions $\rho : \mathbb{R} \to \mathbb{R}$,

2. $G_h^{\downarrow} G_h^{\uparrow} \psi \geq G_h^{\downarrow} G_h^{\uparrow} \phi$ whenever $\psi \geq \phi$,

3. $G_h^{\downarrow} G_h^{\uparrow}(\phi + c) = G_h^{\downarrow} G_h^{\uparrow} \phi + c$, $G_h^{\downarrow} G_h^{\uparrow} c = c$, and $G_h^{\downarrow} G_h^{\uparrow} \phi(\cdot + y) = (G_h^{\downarrow} G_h^{\uparrow} \phi)(\cdot + y)$ for a constant $c \in \mathbb{R}$ and $y \in \mathbb{R}^d$.

Property 2, in particular, says that $G_h^{\downarrow} G_h^{\uparrow}$ is, just like G_h, monotone. Thanks to these properties, it follows from [5, 16] that to prove convergence of Algorithm 3, it is sufficient to establish the following consistency lemma:

Lemma 4.3. *For $\varphi \in C^2(D)$ for every $z \in D$ such that $D\varphi(z) \neq 0$ and for $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x \in B(z, \delta)$ and $h \leq \delta$ we have the following inequalities:*

$$G_h^{\downarrow} G_h^{\uparrow} \varphi(x) \leq \varphi(x) + (\varepsilon - F(D^2\varphi(z), D\varphi(z)))h \tag{37}$$

$$G_h^{\downarrow} G_h^{\uparrow} \varphi(x) \geq \varphi(x) + (-\varepsilon - F(D^2\varphi(z), D\varphi(z)))h \tag{38}$$

Furthermore if $\varphi(x) = \sqrt{x^2 + 1}$ then there exists $\delta > 0$ and $C > 0$ such that for every x and $h \leq \delta$

$$G_h^{\downarrow} G_h^{\uparrow}(\varphi)(x) \leq \varphi(x) + Ch \tag{39}$$

$$G_h^{\downarrow} G_h^{\uparrow}(-\varphi)(x) \geq -\varphi(x) - Ch \tag{40}$$

Lemma 4.3 will follow from the following analogous statement for the operator G_h that can be found in [16], where it plays a pivotal role:

Lemma 4.4 ([16])**.** *If $\varphi \in C^2(D)$ then for every $z \in D$ such that $D\varphi(z) \neq 0$ and $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x \in B(z, \delta)$ and $h \leq \delta$ we have the following inequalities:*

$$G_h \varphi(x) \leq \varphi(x) + (\varepsilon - F(D^2\varphi(z), D\varphi(z)))h \tag{41}$$

$$G_h \varphi(x) \geq \varphi(x) + (-\varepsilon - F(D^2\varphi(z), D\varphi(z)))h \tag{42}$$

Furthermore if $\varphi(x) = \sqrt{x^2 + 1}$ then there exists $\delta > 0$ and $C > 0$ such that for every x and for $h \leq \delta$

$$G_h(\varphi)(x) \leq \varphi(x) + Ch \tag{43}$$

$$G_h(-\varphi)(x) \geq -\varphi(x) - Ch \tag{44}$$

We now show how to obtain Lemma 4.3 from Lemma 4.4:

Proof of Lemma 4.3. First, observe that

$$G_h^{\downarrow} G_h^{\uparrow} \varphi \geq G_h \varphi \text{ for any } \varphi. \tag{45}$$

Indeed,

$$\begin{aligned} G_h^{\downarrow} G_h^{\uparrow} \varphi &= \min\Big(\max(G_h\varphi, \varphi)\,,\, G_h \max(G_h\varphi, \varphi)\Big) \\ &\geq \min\Big(\max(G_h\varphi, \varphi)\,,\, G_h\varphi\Big) \\ &\geq \min\Big(G_h\varphi\,,\, G_h\varphi\Big) = G_h\varphi \end{aligned}$$

where we used the monotonicity of G_h to get the first inequality. Inequality (38) now follows from (45) and inequality (42) of Lemma 4.4.

Next, observe that if all lower (or upper) level set of φ are strictly convex, then $G_h^{\downarrow} G_h^{\uparrow} \varphi = G_h \varphi$. Thus, inequalities (39) & (40) follow immediately from inequalities (43) & (44) of Lemma 4.4.

What remains is inequality (37). Observe that if $\varepsilon - F(D^2\varphi(z), D\phi(z)) \geq 0$, then

$$\begin{aligned} G_h^{\downarrow} G_h^{\uparrow} \varphi &\leq \max(\varphi, G_h\varphi) && (46) \\ &\leq \max\big(\varphi(x), \varphi(x) + (\varepsilon - F(D^2\varphi(z), D\varphi(z)))h\big) && (47) \\ &= \varphi(x) + (\varepsilon - F(D^2\varphi(z), D\varphi(z)))h. && (48) \end{aligned}$$

Hence, all that remains is to establish inequality (37) in the case $\varepsilon - F(D^2\varphi(z), D\varphi(z)) < 0$. For the remainder of the argument we will write $F = F(D^2\varphi(z), D\varphi(z)))$ to simplify notation. By Lemma 4.4 there exists δ_0 such that for all $x \in \overline{B(z, \delta_0)}$ we have $G_h\varphi(x) \leq \varphi(x) + (\varepsilon/2 - F)h$. Then let

$$E_x = \{y : \varphi(y) \geq \varphi(x) + (\varepsilon - F)h\}$$

and $\theta(x) = S_h \mathbf{1}_{E_x}(x)$. It follows that $\theta(x) < 1/2$ for every $x \in \overline{B(z, \delta_0)}$. Thus $\theta_c = \sup_{x \in B(z,\delta_0)} \theta(x) < 1/2$. Therefore we may choose δ so small that for every $x \in B(z, \delta)$ and $h \leq \delta$

$$\int_{B(z,\delta_0)^c} K_{\sqrt{h}}((x - y))\, dy < 1/2 - \theta_c.$$

Since $(\varepsilon - F)h < 0$ we know that for every $y \in B(z, \delta_0)$ we must have $G_h^{\uparrow}\varphi(y) = \varphi(y)$. Therefore if we consider the set

$$E_x^{\uparrow} = \{y : G_h^{\uparrow}\varphi(y) \geq \varphi(x) + (\varepsilon - F)h\}$$

it can only differ from E_x on $B(z, \delta_0)^c$. Thus $E_x^{\uparrow} \subset E_x \cup B(z, \delta_0)^c$. Taking $x \in B(z, \delta)$ we get the chain of inequalities

$$S_h \mathbf{1}_{E_x^{\uparrow}}(x) \leq S_h \mathbf{1}_{E_x}(x) + S_h \mathbf{1}_{B(z,\delta_0)^c}(x) \leq \theta_c + S_h \mathbf{1}_{B(z,\delta_0)^c}(x) < 1/2.$$

Therefore $G_h^{\downarrow} G_h^{\uparrow} \varphi(x) \leq \varphi(x) + (\varepsilon - F)h$ for $x \in B(z, \delta)$ and $h \leq \delta$ as desired. □

4.3 Multiphase Single Growth Algorithm

In this section we describe versions of the single growth algorithm that dissipate the multiphase MBO energy, with quite general interfacial energies, at every step. Assume that D is partitioned into $N > 2$ sets, and recall that we denote the partition by $\mathbf{\Sigma} = (\Sigma_1, \dots \Sigma_N)$.

As noted in [9], the natural candidate for approximating the most general form of multiphase interfacial energy (4) in the style of Lyapunov functionals (14) and (17) is

$$E_\varepsilon(\mathbf{\Sigma}, \mathbf{K}_\varepsilon) = \frac{1}{\varepsilon} \sum_{(i,j)\in\mathcal{I}_N} \int_{\Sigma_j} (K_{i,j})_\varepsilon * \mathbf{1}_{\Sigma_i}\, dx \tag{49}$$

which requires choosing a possibly different convolution kernel for the anisotropy $\sigma_{i,j} : \mathbb{R}^N \to \mathbb{R}^+$ associated with each interface $(\partial\Sigma_i) \cap (\partial\Sigma_j)$ in the network. Here, we only require that each $K_{i,j}$ satisfy

$$K_{i,j}(x) = K_{j,i}(x) = K_{i,j}(-x) \tag{50}$$

for all $i \neq j$ and all x.

Following the general strategy described in [10] for deriving threshold dynamics-type algorithms from non-local approximate energies such as (49), we first extend energy (49) to functions $\mathbf{u} \in \mathcal{K}$, with time step δt satisfying $\varepsilon = \sqrt{\delta t}$:

$$E_{\sqrt{\delta t}}(\mathbf{u}, \mathbf{K}_{\sqrt{\delta t}}) = \frac{1}{\sqrt{\delta t}} \sum_{(i,j)\in\mathcal{I}_N} \int_D u_j (K_{i,j})_{\sqrt{\delta t}} * u_i\, dx \tag{51}$$

Then, a threshold dynamics algorithm can be systematically derived by linearizing (51) at a given configuration and minimizing it over the entire box constraint set $\mathcal{K}$. Fix a partition $\mathbf{\Sigma}$, and let $u_i = \mathbf{1}_{\Sigma_i}$. The linearization of relaxed energy (51) at $\mathbf{u} = (u_1, \dots, u_N)$, evaluated at some function $\varphi = (\varphi_1, \dots, \varphi_N)$ turns out to be:

$$L_{\mathbf{u},\sqrt{\delta t}}(\varphi) = E_{\sqrt{\delta t}}(\mathbf{u}, \mathbf{K}_{\sqrt{\delta \mathbf{t}}}) + \frac{2}{\sqrt{\delta t}} \sum_{i=1}^{N} \int_D \varphi_i \sum_{j\neq i} (K_{i,j})_{\sqrt{\delta t}} * u_j\, dx \tag{52}$$

Minimizing (52) over (19) yields the following algorithm from [9], which is the obvious extension of Algorithm 2 to normal dependent surface tensions:

Algorithm 4: ([9]): Given a time step size $\delta t > 0$, alternate the following steps:

1. Convolution:

$$\psi_i^k = \sum_{j \neq i} (K_{i,j})_{\sqrt{\delta t}} * \mathbf{1}_{\Sigma_j^k}. \tag{53}$$

2. Thresholding:

$$\Sigma_i^{k+1} = \left\{ x \,:\, \psi_i^k(x) \leq \min_{j \neq i} \psi_j^k(x) \right\}. \tag{54}$$

Algorithm 4 is natural, and appears to work well in practice; see [9] for some examples. However, the question of whether it in fact decreases the corresponding energy (49) for any choice of $\delta t > 0$ is now an even more complicated problem than in the case of Algorithm 2 for energy (17), not least because there are multiple ways to construct a convolution kernel corresponding to a given anisotropy: the stability of the algorithm is likely to depend not only on the properties of the surface tensions $\sigma_{i,j}$, but also the particular convolution kernels $K_{i,j}$ used to approximate them.

To make some headway, here we will instead consider new and slightly more expensive versions of Algorithm 4 that are motivated by the Gauss-Seidel version given in [10] of Algorithm 2, as well as Algorithm 3 of the previous section. To that end, given a partition $\mathbf{\Sigma}$, define

$$i_{\min}(x) = \operatorname*{arg\,min}_{1 \leq i \leq N} \sum_{j \neq i} \int (K_{i,j})_{\sqrt{\delta t}} * \mathbf{1}_{\Sigma_j} \, dx$$

so that $x \in \Sigma_{i_{\min}(x)}$ after one step of Algorithm 4. Also, let $i_\Sigma(x)$ denote the unique i such that $x \in \Sigma_i$, and let $\mathbf{e}_n \in \mathbb{R}^N$ denote the n^{th} standard basis vector. Then the direction of perturbation affected by Algorithm 4 on the current configuration is given by

$$\varphi(x) = \mathbf{e}_{i_{\min}}(x) - \mathbf{e}_{i_\Sigma}(x) \tag{55}$$

for each $x \in D$.

Below, we present Algorithms 5 and 6 which differ from Algorithm 4 by placing a single growth constraint on the perturbation direction φ. For each $x \in D$, if $i_{\min}(x)$ and $i_\Sigma(x)$ fall into certain classes (that depend on the iteration number) then $\varphi(x)$ is chosen as in equation (55), otherwise $\varphi(x)$ is set to $\mathbf{0}$. In other words, only a subset of the points $x \in D$ are redistributed among the phases as indicated by (55). The essential advantage this brings is to compute convolutions with the phases more frequently, thus yielding a more reliable descent direction. It turns out that this simple modification guarantees energy dissipation for a wide class of kernels as we describe below.

Algorithm 5: Given an initial partition of D into N sets $\mathbf{\Sigma}^0 = \{\Sigma_i^0\}_{i=1}^N$ and a time step δt the $(k+1)^{th}$ iteration $\mathbf{\Sigma}^{k+1}$ is obtained from $\mathbf{\Sigma}^k$ by a series of substeps indexed by $(m,n) \in \mathcal{I}_N$ For $(m,n) \neq (1,2)$ let $p(m,n)$ denote the predecessor of (m,n) in the dictionary ordering of $\mathcal{I}_N$ and define $\mathbf{\Sigma}^{k,p(1,2)} = \mathbf{\Sigma}^k$ and $\mathbf{\Sigma}^{k,(N,N-1)} = \mathbf{\Sigma}^{k+1}$. Then $\mathbf{\Sigma}^{k,(m,n)}$ is obtained from $\mathbf{\Sigma}^{k,p(m,n)}$ as follows:

1. For each $(i,j) \in \mathcal{I}_N$ form the convolutions:

$$\psi_{i,j}^{k,(m,n)}(x) = (K_{i,j})_{\sqrt{\delta t}} * \mathbf{1}_{\Sigma_j^{k,p(m,n)}} \tag{56}$$

2. For each i form the sums:

$$\Psi_i^{k,(m,n)}(x) = \sum_{j \neq i} \psi_{(i,j)}^{k,(m,n)}(x) \tag{57}$$

3. Threshold the m^{th} function:

$$G^{k,(m,n)} = \{x \in D : \min_i \Psi_i^{k,(m,n)}(x) = \Psi_m^{k,(m,n)}(x)\} \tag{58}$$

4. Grow set m into set n only:

$$\Sigma_m^{k,(m,n)} = \Sigma_m^{k,p(m,n)} \cup (G^{k,(m,n)} \cap \Sigma_n^{k,p(m,n)}) \tag{59}$$

5. Update set n:

$$\Sigma_n^{k,(m,n)} = \Sigma_n^{k,(m,n)} \setminus (G^{k,(m,n)} \cap \Sigma_n^{k,p(m,n)}) \tag{60}$$

Proposition 4.5. *Suppose that each kernel $K_{i,j}$ may be split into a sum $K_{i,j} = K_{i,j}^1 + K_{i,j}^2$ such that $K_{i,j}^1 \geq 0$ almost everywhere and $\widehat{K}_{i,j}^2 \geq 0$ almost everywhere then Algorithm 5 dissipates the energy (49) at each step. Furthermore if for every $(i,j) \in \mathcal{I}_N$ either $K_{i,j}^1$ is positive in a neighborhood of the origin or $\hat{K}_{i,j}^2(0)$ is positive then if an iteration of Algorithm 4 changes a configuration $\mathbf{\Sigma}$ by a set of positive measure, an iteration of Algorithm 5 also changes $\mathbf{\Sigma}$ by a set of positive measure and strictly decreases the energy.*

Proof. We show that at each substep the energy is dissipated moving from $\mathbf{\Sigma}^{k,p(m,n)}$ to $\mathbf{\Sigma}^{k,(m,n)}$. Set $\varphi^{k,(mn)} = \mathbf{1}_{\mathbf{\Sigma}^{k,(m,n)}} - \mathbf{1}_{\mathbf{\Sigma}^{k,p(m,n)}}$. Observe that we may write

$$E_{\sqrt{\delta t}}(\mathbf{\Sigma}^{k,(m,n)}, \mathbf{K}_{\sqrt{\delta t}}) - E_{\sqrt{\delta t}}(\mathbf{\Sigma}^{k,p(m,n)}, \mathbf{K}_{\sqrt{\delta t}}) =$$

$$L_{\mathbf{\Sigma}^{k,p(m,n)},\sqrt{\delta t}}(\varphi^{k,(m,n)}) - E_{\sqrt{\delta t}}(\mathbf{\Sigma}^{k,p(m,n)}, K_{\sqrt{\delta t}}) + Q_{\sqrt{\delta t}}(\varphi^{k,(m,n)}). \tag{61}$$

where $L_{\mathbf{\Sigma}^{k,m},\sqrt{\delta t}}$ is given in equation (52) and $Q_{\sqrt{\delta t}}$ is a quadratic function that does not depend on $\mathbf{\Sigma}^{k,m}$. By the discussion preceding Algorithm 5 we know

that $L_{\mathbf{\Sigma}^{k,m},\sqrt{\delta t}}(\varphi^{k,m+1}) - E_{\sqrt{\delta t}}(\mathbf{\Sigma}^{k,m+1}, \mathbf{K}_{\sqrt{\delta t}}) \leq 0$. Therefore it suffices to show that $Q_{\sqrt{\delta t}}(\varphi^{k,m+1}) \leq 0$. This term is given by

$$Q_{\sqrt{\delta t}}(\varphi^{k,(m,n)}) = \frac{1}{\sqrt{\delta t}} \sum_{(i,j)\in\mathcal{I}_N} \int_D \varphi_i^{k,(m,n)}(x) \Big((K_{i,j})_{\sqrt{\delta t}} * \varphi_j^{k,(m,n)} \Big)(x)\, dx. \quad (62)$$

This formula actually simplifies dramatically as $\varphi_i^{k,(m,n)} \equiv 0$ unless $i = m$ or $i = n$. Furthermore $\varphi_m^{k,(m,n)}$ is nonnegative pointwise and $\varphi_n^{k,(m,n)} = -\varphi_m^{k,(m,n)}$. Thus nearly every term of $Q_{\sqrt{\delta t}}(\varphi^{k,(m,n)})$ is zero and we get

$$Q_{\sqrt{\delta t}}(\varphi^{k,(m,n)}) = -2 \int_D \varphi_m^{k,(m,n)}(x)((K_{m,n})_{\delta t} * \varphi_m^{k,(m,n)})(x)\, dx.$$

Recalling equation (29) and the subsequent argument in Proposition 4.1 energy dissipation is proven.

Now we turn to the second statement. Let $\mathbf{\Sigma}^0 = \mathbf{\Sigma}$. Let $\mathbf{\Sigma}^1$ be the configuration obtained from a single iteration of Algorithm 4 and let $\mathbf{\Sigma}^{0,(m,n)}$ be the configurations obtained from the substeps of Algorithm 5. As before set

$$\varphi(x) = \mathbf{1}_{\mathbf{\Sigma}^1}(x) - \mathbf{1}_{\mathbf{\Sigma}^0}(x)$$

and

$$\varphi^{0,(m,n)}(x) = \mathbf{1}_{\mathbf{\Sigma}^{0,(m,n)}}(x) - \mathbf{1}_{\mathbf{\Sigma}^{0,p(m,n)}}(x).$$

Using the above inequality for the quadratic term and following the proof of the second statement of Proposition 4.1, we only need to show that for some $(m,n) \in \mathcal{I}_N$ the function $\varphi_m^{0,(m,n)}(x)$ has support of positive measure.

Suppose that for every $(m,n) \in \mathcal{I}_N$ the function $\varphi_m^{0,(m,n)}$ has support of zero measure. In this case it follows that no set has grown or shrank by more than a set of measure zero. Thus for any function $f \in L^1(D)$ and any label $1 \leq i \leq N$ we have

$$\int_{\Sigma_i^0} f(x)\, dx = \int_{\Sigma_i^{0,(1,2)}} f(x)\, dx = \cdots = \int_{\Sigma_i^{0,(N-1,N)}} f(x)\, dx$$

As a result we may compute every convolution in the substeps of Algorithm 5 against $\mathbf{\Sigma}^0$ without changing the result. This allows us to write $\varphi_m^{0,(m,n)}$ in terms of φ:

$$\varphi_m^{0,(m,n)} = \max(\varphi_m(x), 0) |\min(\varphi_n(x), 0)|.$$

If we then sum over $n \neq m$ we get

$$\sum_{n\neq m} \varphi_m^{0,(m,n)} = \max(\varphi_m(x), 0).$$

The support of φ may be decomposed as

$$\operatorname{supp}(\varphi) = \bigcup_{1\leq m\leq N} \operatorname{supp}\big(\max(\varphi_m(x), 0)\big).$$

Thus φ has support of measure zero a contradiction.

□

Below we describe a variant of Algorithm 5 that requires fewer convolutions but imposes a more restrictive condition on the kernels.

Algorithm 6: Given an initial partition of D into N sets $\mathbf{\Sigma}^0 = \{\Sigma_i^0\}_{i=1}^N$ and a time step δt the $(k+1)^{th}$ iteration $\mathbf{\Sigma}^{k+1}$ is obtained from $\mathbf{\Sigma}^k$ by computing the following N substeps $\mathbf{\Sigma}^{k,0}, \dots, \mathbf{\Sigma}^{k,N}$ where $\mathbf{\Sigma}^{k,0} = \mathbf{\Sigma}^k$ and $\mathbf{\Sigma}^{k,N} = \mathbf{\Sigma}^{k+1}$. For each $0 \leq m \leq \ell - 1$ the partitions $\mathbf{\Sigma}^{k,m+1}$ are obtained from $\mathbf{\Sigma}^{k,m}$ as follows:

1. For each $(i,j) \in \mathcal{I}_N$ form the convolutions:

$$\psi_{i,j}^{k,m+1}(x) = (K_{i,j})_{\sqrt{\delta t}} * \mathbf{1}_{\Sigma_j^{k,m}} \tag{63}$$

2. For each i form the sums:

$$\Psi_i^{k,m+1}(x) = \sum_{j \neq i} \psi_{(i,j)}^{k,m+1}(x) \tag{64}$$

3. Threshold the $(m+1)^{th}$ function:

$$G^{k,m+1} = \{x \in D : \min_i \Psi_i^{k,m+1}(x) = \Psi_{m+1}^{k,m+1}(x)\} \tag{65}$$

4. Grow the $(m+1)^{th}$ set:

$$\Sigma_{m+1}^{k,m+1} = \Sigma_{m+1}^{k,m} \cup G^{k,m+1} \tag{66}$$

5. Update the other sets:

$$\Sigma_i^{k,m+1} = \Sigma_i^{k,m} \setminus G^{k,m+1} \quad \forall i \neq m+1 \tag{67}$$

Proposition 4.6. *Suppose that each kernel $K_{i,j}$ may be split into a sum $K_{i,j} = K_{i,j}^1 + K_{i,j}^2$ such that $K_{i,j}^1 \geq 0$ almost everywhere, $\widehat{K}_{i,j}^2 \geq 0$ and for almost every $x \in \mathbb{R}^d$ and i,j,k distinct we have the pointwise triangle inequality*

$$K_{i,k}(x) \leq K_{i,j}(x) + K_{j,k}(x) \tag{68}$$

then Algorithm 6 dissipates the energy (49) at each step. Furthermore if for every $(i,j) \in \mathcal{I}_N$ either $K_{i,j}^1$ is positive in a neighborhood of the origin or $\hat{K}_{i,j}^2(0)$ is positive then if an iteration of Algorithm 4 changes a configuration $\mathbf{\Sigma}$ by a set of positive measure, an iteration of Algorithm 6 also changes $\mathbf{\Sigma}$ by a set of positive measure and strictly decreases the energy.

Proof. We proceed by showing that the energy is dissipated moving from substep $\mathbf{\Sigma}^{k,m}$ to $\mathbf{\Sigma}^{k,m+1}$. Let $\varphi^{k,m+1}(x) = \mathbf{1}_{\mathbf{\Sigma}^{k,m+1}}(x) - \mathbf{1}_{\mathbf{\Sigma}^{k,m}}(x)$.

As in the argument of Proposition 4.5 the change in energy

$$E_{\sqrt{\delta t}}(\mathbf{\Sigma}^{k,m+1}, \mathbf{K}_{\sqrt{\delta t}}) - E_{\sqrt{\delta t}}(\mathbf{\Sigma}^{k,m}, \mathbf{K}_{\sqrt{\delta t}})$$

will be nonnegative as long as the quadratic term in the difference $Q_{\sqrt{\delta t}}(\varphi^{k,m})$ is nonnegative. Equations (66) and (67) show that $\varphi_{m+1}^{k,m+1}(x)$ is pointwise nonnegative and $\varphi_i^{k,m+1}(x)$ is pointwise nonpositive for $i \neq m+1$. In fact $\varphi_{m+1}^{k,m+1}(x) = -\sum_{i \neq m+1} \varphi_i^{k,m+1}(x)$. Plugging this into the formula for Q from equation (62) and defining

$$\mathcal{I}_N(m+1) = \{(i,j) \in \mathcal{I}_N : i, j \neq m+1\}$$

we get

$$-\sum_{i \neq m+1} \sum_{j \neq m+1} \int_D |\varphi_i^{k,m+1}(x)| \left(\left((K_{i,m+1})_{\sqrt{\delta t}} + (K_{j,m+1})_{\sqrt{\delta t}} \right) * |\varphi_j^{k,m+1}| \right)(x)\, dx +$$

$$\sum_{(i,j) \in \mathcal{I}_N(m+1)} \int_D |\varphi_i^{k,m+1}(x)| \left((K_{i,j})_{\sqrt{\delta t}} * |\varphi_j^{k,m+1}| \right)(x)\, dx.$$

Applying (68) the above is

$$\leq -2 \sum_{i \neq m+1} \int_D |\varphi_i^{k,m+1}(x)| \left((K_{i,m+1})_{\sqrt{\delta t}} * |\varphi_i^{k,m+1}| \right)(x)\, dx$$

The remainder of the argument proceeds exactly as in the proof of Proposition 4.5

□

4.4 Extension to General Graphs

The MBO Algorithm extends naturally to partitioning problems formulated on graphs, see e.g. [13]. In this section, we briefly discuss extending to the setting of graphs the new variants of threshold dynamics (and their associated stability statements) presented in previous sections.

We define an undirected graph $G = (\mathcal{V}, \mathcal{E})$ as a collection of vertices $\mathcal{V} = \{1, 2, \ldots, n\}$ and an edge set consisting of unordered pairs of vertices $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. An undirected weighted graph $G = (\mathcal{V}, \mathcal{E}, W)$ has the extra structure of a symmetric weight matrix $W : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$. W describes the connection strength between two vertices in $\mathcal{V}$. Note that even if some pair $(v_1, v_2) \notin \mathcal{E}$ it is still possible that $W(v_1, v_2) \neq 0$. Indeed in many applications it is useful to assign a nonzero weight to a pair of vertices that share a neighbor but not necessarily an edge, or more generally to a pair that are separated by a sufficiently short path.

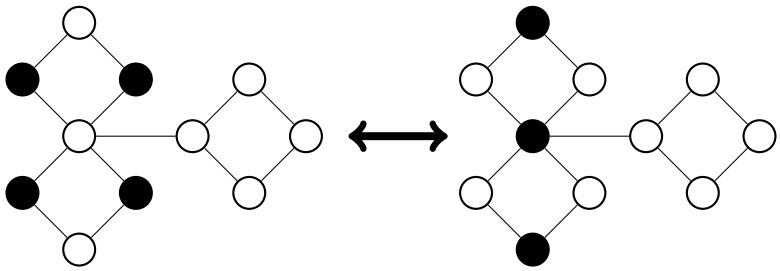

Figure 3: Behavior of the standard algorithm under the action of the weights W_1 on a graph. The algorithm gets trapped in a periodic loop between the two configurations above. The configuration on the right has a higher energy.

Now we are ready to discuss the extension of MBO to graphs. For a given undirected weighted graph $G = (\mathcal{V}, \mathcal{E}, W)$ we define the binary MBO energy of a partition $\Sigma \cup \Sigma^c = \mathcal{V}$ as

$$E(\Sigma, G, W) = \frac{1}{|\mathcal{V}|^2} \sum_{v_1 \in \Sigma} \sum_{v_2 \in \Sigma^c} W(v_1, v_2). \tag{69}$$

Algorithms 1 and 3 are easily adapted to the graph case by replacing the convolutions with the discrete sums

$$\psi(v_1) = \sum_{v_2 \in \Sigma} W(v_1, v_2). \tag{70}$$

Recall that by Proposition 3.1 and Example 1, the continuous version of Algorithm 1 decreases the energy (14) at every step if the convolution kernel K has positive Fourier transform. Unsurprisingly an analogous result holds for the graph version of the algorithm as well. Algorithm 1 decreases the energy (69) at every step if W is a positive semidefinite matrix. As in the continuous version, it is easy to describe an example where the failure of this condition leads to the algorithm increasing the energy.

Example 3:

Let A be the adjacency matrix of the graph in Figure 3, I the identity matrix and take $W_1 = A + \lambda I$ for any $\lambda < 2$. Applying Algorithm 1 to the graph causes it to get trapped in a periodic loop between the two configurations above. The left hand side has 8 edges between vertices in different classes while the right hand side has 9 edges between vertices in different classes. Thus the algorithm increases the energy every time it moves from left to right.

The failure of Algorithm 1 to dissipate energy for weights as simple as a multiple of the identity plus the adjacency matrix of the graph is a serious detriment. Indeed for many graphs working with a PSD weight matrix may be either computationally time consuming or unnatural for the application. On the other hand a simple modification of Proposition 4.1 shows that Algorithm 3 will decrease the energy for a wide choice of weights.

Proposition 4.7. *Suppose that the weight matrix W may be decomposed into a sum $W = W^1 + W^2$ where $W^1 \geq 0$ entrywise and W^2 is PSD. Then the graph analogue of Algorithm 3 decreases the energy (69) at every step. Furthermore if either $W^1(v, v) > 0$ for each $v \in \mathcal{V}$ or W^2 is PD then if an iteration of Algorithm 1 changes the class of one or more vertices then an iteration of Algorithm 3 also changes the class of one or more vertices and strictly decreases the energy.*

We may also define a multiclass MBO energy for graphs. Given an N class partition $\mathbf{\Sigma} = (\Sigma_1, \ldots, \Sigma_N)$ of $\mathcal{V}$ and for each $(i, j) \in \mathcal{I}_N$ a symmetric weight matrix $W_{i,j}$ with $W_{i,j} = W_{j,i}$ the multiclass MBO energy is defined as

$$E(\Sigma, G, W) = \frac{1}{|\mathcal{V}|^2} \sum_{(i,j)\in\mathcal{I}_N} \sum_{v_1\in\Sigma_i} \sum_{v_2\in\Sigma_j} W_{i,j}(v_1, v_2). \tag{71}$$

As with the binary algorithms the multiclass Algorithms 5 and 6 may be extended to the graph case by replacing the convolutions with the discrete sums

$$\psi_{i,j}(v_1) = \sum_{v_2\in\Sigma_j} W_{i,j}(v_1, v_2). \tag{72}$$

Minor modifications to Propositions 4.5 and 4.6 give the following statements.

Proposition 4.8. *Suppose that for each $(i, j) \in \mathcal{I}_N$ the weight matrix $W_{i,j}$ can be decomposed into a sum $W^1_{i,j} + W^2_{i,j}$ such that $W^1_{i,j}$ has nonnegative entries and $W^2_{i,j}$ is PSD. Then the graph analogue of Algorithm 5 dissipates energy (71) at every step. Furthermore if for each $(i, j) \in \mathcal{I}_N$ either $W^1_{i,j}(v, v) > 0$ for every $v \in \mathcal{V}$ or $W^2_{i,j}$ is PD then if an iteration of Algorithm 4 changes the class of one or more vertices then an iteration of Algorithm 5 changes the class of one or more vertices and strictly decreases the energy.*

Proposition 4.9. *Suppose that for each $(i, j) \in \mathcal{I}_N$ the weight matrix $W_{i,j}$ can be decomposed into a sum $W^1_{i,j} + W^2_{i,j}$ such that $W^1_{i,j}$ has nonnegative entries and $W^2_{i,j}$ is PSD and for all $v_1, v_2 \in \mathcal{V}$ and i, j, k distinct we have the triangle inequality*

$$W_{i,j}(v_1, v_2) \leq W_{i,k}(v_1, v_2) + W_{j,k}(v_1, v_2).$$

Then the graph analogue of Algorithm 6 dissipates energy (71) at every step. Furthermore if for each $(i, j) \in \mathcal{I}_N$ either $W^1_{i,j}(v, v) > 0$ for every $v \in \mathcal{V}$ or $W^2_{i,j}$ is PD then if an iteration of Algorithm 4 changes the class of one or more vertices then an iteration of Algorithm 6 changes the class of one or more vertices and strictly decreases the energy.

5 Convergence of non-local energies

In [10], Gamma convergence of the Lyapunov functional (17) to the interfacial energy (16) is established for radially monotonic and symmetric, nonnegative kernels. However, Algorithms 3, 5, and 6 guarantee energy dissipation for a

much larger class of kernels. Thus it is desirable to have a more general Gamma convergence result.

In this section, we establish the Gamma limit of (17) for a much wider class of kernels, including *sign changing* kernels. The key property that we require of the kernel is a strong *positive core* near the origin; otherwise, the kernel is free to oscillate above and below zero at the outskirts. The positive core ensures that any negative mass further out will be sufficiently counterbalanced. To that end, let

$$BV_{\mathcal{K}} = \Big\{ \mathbf{u} \in \mathcal{K} \text{ s.t. } u_i(x) \in BV(D) \text{ for } i \in \{1, 2, \dots, N\} \Big\}$$

and for any $\mathbf{u} \in \mathcal{K}$ define the energy

$$E(\mathbf{u}, \sigma) = \begin{cases} \displaystyle\sum_{(i,j) \in \mathcal{I}_N} \int_D \sigma_{i,j}(\nabla u_i) + \sigma_{i,j}(\nabla u_j) - \sigma_{i,j}\big(\nabla(u_i + u_j)\big) & \text{if } \mathbf{u} \in BV_{\mathcal{K}}, \\ +\infty & \text{otherwise.} \end{cases} \tag{73}$$

Theorem 5.1. *Let the kernel K satisfy (12). In addition, assume that there exist positive constants a, α, β such that the following conditions hold:*

1. $\alpha(\frac{2}{a})^{d+2} \leq \beta$,
2. $K(z) \geq \beta$ *for* $|z| \leq a$,
3. $|\min(K(z), 0)| \leq \alpha |z|^{-(d+2)}$ *for all* z.

Given constant surface tensions $\sigma_{i,j} > 0$ satisfying the triangle inequality (9), define

$$\sigma_{i,j}(n) = \sigma_{i,j} \int_{\mathbb{R}^d} K(z) |z \cdot n| \, dz$$

Then as $\varepsilon \to 0$ the Lyapunov functionals $E_\varepsilon(\cdot, K_\varepsilon)$ given in (51) Gamma converge in the L^1 topology over $\mathcal{K}$ to the energy $E(\cdot, \sigma)$ given in (73). Furthermore if for some sequence $\mathbf{u}_\varepsilon$ we have $\sup_{\varepsilon > 0} E_\varepsilon(\mathbf{u}_\varepsilon, K_\varepsilon) < \infty$ then $\mathbf{u}_\varepsilon$ is pre-compact in $L^1(D)$ and the set of accumulation points is contained in $BV_{\mathcal{K}}(D)$.

The proof of Theorem 5.1 will be built over the following lemmas and propositions. First we will prove the theorem for nonnegative kernels that satisfy (12) and are positive in a neighborhood of the origin.

Lemma 5.2. *Suppose that K is a nonnegative kernel that satisfies (12) and is positive in a neighborhood of the origin. If for some sequence $\mathbf{u}_\varepsilon$ we have $\sup_{\varepsilon > 0} E_\varepsilon(\mathbf{u}_\varepsilon, K_\varepsilon) < \infty$ then $\mathbf{u}_\varepsilon$ is pre-compact in $L^1(D)$ and the set of accumulation points is contained in $BV_{\mathcal{K}}(D)$.*

Proof. K is strictly positive in a neighborhood of the origin thus there exists some $s, t \in (0, 1)$ such that $K(z) \geq s$ for all $|z| \leq t$. Let $J(z) = cs(1 - |z/t|)$

for $|z| < t$ and 0 otherwise, where c is chosen so that J has unit mass. Then $\frac{1}{c}J(z) \leq K(z)$ for all $z \in \mathbb{R}^d$. Therefore

$$\sup_{\varepsilon>0} E_\varepsilon(\mathbf{u}_\varepsilon, \frac{1}{c}J_\varepsilon) \leq \sup_{\varepsilon>0} E_\varepsilon(\mathbf{u}_\varepsilon, K_\varepsilon).$$

Since the energy is linear in the kernel it follows that $\sup_{\varepsilon>0} E_\varepsilon(\mathbf{u}_\varepsilon, J_\varepsilon)$ is bounded. In addition $|\nabla J(z)| = \frac{cs}{t}$ for $|z| < t$ and 0 for $|z| > t$. Thus $|\nabla J(z)| \leq \frac{2}{t}J(z/2)$ for all z. Now J fits into the framework of Lemma 5 in [10], which gives the desired result.

□

Next we show that the functional (18) converges pointwise to $E(\mathbf{u}, \sigma)$. This ensures that lim sup inequality needed for the Gamma convergence argument will be satisfied. If $\mathbf{u} \notin BV_{\mathcal{K}}(D)$ then the pointwise convergence follows immediately from Lemma 5.2. Indeed we must have $\lim_{\varepsilon\to 0} E_\varepsilon(\mathbf{u}, K_\varepsilon) = \infty$ for otherwise the constant sequence $\mathbf{u}$ would have an accumulation point in $BV_{\mathcal{K}}(D)$ implying $\mathbf{u} \in BV_{\mathcal{K}}(D)$.

For $\mathbf{u} \in BV_{\mathcal{K}}(D)$ we recall Lemma 4 from [10] which gives pointwise convergence under very mild conditions on the kernel. Although the argument in [10] is given for radially symmetric kernels, the modification to general kernels is straight forward.

Lemma 5.3. *(from [10]) Let K be a kernel satisfying (12) and $\mathbf{u} \in BV_{\mathcal{K}}$ then* $\lim_{\varepsilon\to 0} E_\varepsilon(\mathbf{u}, K_\varepsilon) = E(\mathbf{u}, \sigma)$.

To complete the Gamma convergence argument for nonnegative kernels we only have left to prove the lim inf inequality. A key tool that we will need is the following Lemma from [10] that says that integer scalings of the parameter ε are guaranteed to decrease the energy.

Lemma 5.4. *(from [10]) If the kernel K is nonnegative and the $\sigma_{i,j}$ satisfy the triangle inequality then for every $N \in \mathbb{Z}_+$ we have $E_{N\varepsilon}(\mathbf{u}, K_{N\varepsilon}) \leq E_\varepsilon(\mathbf{u}, K_\varepsilon)$.*

Now we are ready to present the lim inf argument.

Proposition 5.5. *If K satisfies (12) and in addition K is nonnegative then for any sequence $\mathbf{u}_\varepsilon$ converging to $\mathbf{u}$ in L^1 the inequality* $\liminf_{\varepsilon\to 0} E_\varepsilon(\mathbf{u}_\varepsilon, K_\varepsilon) \geq E(\mathbf{u}, \sigma)$ *holds.*

Proof. K is pointwise nonnegative therefore if we fix some $L > 0$ and let

$$K^L(z) = K(z)\mathbf{1}_{B(0,L)}(z)$$

then we decrease the energy by replacing $E_\varepsilon(\mathbf{u}_\varepsilon, K_\varepsilon)$ with $E_\varepsilon(\mathbf{u}_\varepsilon, K^L_\varepsilon)$. Now fix $\delta > 0$ and for each ε let $\delta_\varepsilon = n_\varepsilon\varepsilon$ where $n_\varepsilon \in \mathbb{Z}_+$ is chosen such that $|\delta - \delta_\varepsilon|$ is minimized. It follows immediately that $|\delta - \delta_\varepsilon| \leq \varepsilon/2$. Since δ_ε is obtained from ε by an integer scaling we may use Lemma 5.4 to get the inequality

$$E_\varepsilon(\mathbf{u}_\varepsilon, K^L_\varepsilon) \geq E_{\delta_\varepsilon}(\mathbf{u}_\varepsilon, K^L_{\delta_\varepsilon}).$$

Now we wish to replace δ_ε with δ. Thus we must estimate the resulting error

$$R_{\varepsilon,\delta} = |E_{\delta_\varepsilon}(\mathbf{u}_\varepsilon, K^L_{\delta_\varepsilon}) - E_\delta(\mathbf{u}_\varepsilon, K^L_\delta)|.$$

We have:

$$\begin{aligned} R_{\varepsilon,\delta} &\leq \sum_{(i,j)\in\mathcal{I}_N} \sigma_{i,j} \int_D \int_{B(0,L)} u_{\varepsilon,i}(x) u_{\varepsilon,j}(x+z) \left| \delta_\varepsilon^{-(d+1)} K\left(\frac{z}{\delta_\varepsilon}\right) - \delta^{-(d+1)} K\left(\frac{z}{\delta}\right) \right| dz\, dx \\ &\leq \max_{(i,j)\in\mathcal{I}_N} N^2 \sigma_{i,j} \int_D \int_{B(0,L/\delta)} \delta^{-1} \left| \left(\frac{\delta}{\delta_\varepsilon}\right)^{(d+1)} K\left(\frac{\delta}{\delta_\varepsilon} z\right) - K(z) \right| dz\, dx. \end{aligned}$$

Since smooth functions are dense in $L^1(\mathbb{R}^d)$ we may approximate the above integral to arbitrary precision by replacing K with a smooth function f. The spaces D and $B(0, L/\delta)$ have finite measure thus uniform continuity shows that

$$\lim_{\varepsilon\to 0} \int_D \int_{B(0,L/\delta)} \delta^{-1} \left| \left(\frac{\delta}{\delta_\varepsilon}\right)^{(d+1)} f\left(\frac{\delta}{\delta_\varepsilon} z\right) - f(z) \right| dz\, dx = 0.$$

It then follows that

$$\lim_{\varepsilon\to 0} R_{\varepsilon,\delta} = 0.$$

Combining the above with the L^1 convergence of $\mathbf{u}_\varepsilon$ to $\mathbf{u}$ we obtain:

$$\liminf_{\varepsilon\to 0} E_\varepsilon(\mathbf{u}_\varepsilon, K^L_\varepsilon) \geq \liminf_{\varepsilon\to 0} E_\delta(\mathbf{u}_\varepsilon, K^L_\delta) = E_\delta(\mathbf{u}, K^L_\delta)$$

Now if we allow δ to go to zero the question of the liminf inequality has been reduced to the question of pointwise convergence of the functional. However this is already covered above thus

$$\lim_{\delta\to 0} E_\delta(\mathbf{u}, K^L_\delta) = E(\mathbf{u}, \sigma^L)$$

where

$$\sigma^L_{i,j}(n) = \sigma_{i,j} \int_{\mathbb{R}^d} K^L(z) |z \cdot n|\, dz.$$

By monotone convergence $\lim_{L\to\infty} E(\mathbf{u}, \sigma^L) = E(\mathbf{u}, \sigma)$.

□

This completes the proof of Theorem 5.1 for nonnegative kernels that are positive in a neighborhood of the origin. To extend the result to kernels that change sign we show that it is possible to rearrange the mass of the kernel so that it becomes nonnegative, while also decreasing the energy functional (18). It is essential that this process does not change the limiting energy $E(\mathbf{u}, \sigma)$ and that the rearranged kernel still is positive in a neighborhood of the origin and satisfies (12). The following lemma shows that these goals can be accomplished simultaneously.

Lemma 5.6. *Suppose that K satisfies the conditions in Theorem 5.1. Then there exists a nonnegative kernel $\tilde{K}$ such that $\tilde{K}$ is positive in a neighborhood of the origin, $\tilde{K}$ satisfies (12), the inequality $E_\varepsilon(\mathbf{u}, K_\varepsilon) \geq E_\varepsilon(\mathbf{u}, \tilde{K}_\varepsilon)$ holds and for every $n \in S^{d-1}$*

$$\int_{\mathbb{R}^d} |z \cdot n| K(z) dz = \int_{\mathbb{R}^d} |z \cdot n| \tilde{K}(z)\, dz.$$

Proof. Split K into $K^+ = \max(K, 0)$ and $K^- = \min(K, 0)$. Recall the constants a, α, β from Theorem 5.1. For $j \in \mathbb{Z}_+$ let

$$A_j = \{z \in \mathbb{R}^d : |z| \in (a2^{j-1}, a2^j)\}$$

Define

$$\tilde{K}(z) = K^+(z) + \sum_{j=1}^{\infty} 2^{j(d+1)} K^-(2^j z) \mathbf{1}_{A_j}(2^j z).$$

From this construction we see that $\tilde{K}$ is possibly negative only for z satisfying $|z| \in (a/2, a)$. Recall that $|K^-(z)| \leq \alpha |z|^{-(d+2)}$ and $K(z) \geq \beta$ for $|z| < a$. Choose some z_0 such that $|z_0| \in (a/2, a)$ then using these inequalities we see

$$\tilde{K}(z_0) = K^+(z_0) + \sum_{j=1}^{\infty} 2^{j(d+1)} K^-(2^j z_0) \mathbf{1}_{A_j}(2^j z_0) \geq \beta - \alpha (\frac{2}{a})^{d+2} \sum_{j=1}^{\infty} 2^{-j} = \beta - \alpha (\frac{2}{a})^{d+2} \geq 0.$$

It is clear that $\tilde{K}$ satisfies the symmetry condition $\tilde{K}(z) = \tilde{K}(-z)$ since K^+ and K^- satisfy this condition and the A_j are radially symmetric sets. Near the origin $\tilde{K} = K^+$ so there must be a neighborhood where $\tilde{K}$ is strictly positive.

Next we recall that $E_\varepsilon(\mathbf{u}, K_\varepsilon)$ is linear in the kernel. Therefore using the linearity and Fubini's theorem we may write

$$E_\varepsilon\big(\mathbf{u}, \tilde{K}_\varepsilon(z)\big) = E_\varepsilon\big(\mathbf{u}, K^+_\varepsilon(z)\big) + \sum_{j=1}^{\infty} 2^{j(d+1)} E_\varepsilon\big(\mathbf{u}, (K^- \mathbf{1}_{A_j})_\varepsilon(2^j z)\big).$$

All of the terms in the infinite sum are negative, thus if we decrease their magnitude the overall energy will increase. By Lemma 5.4 we know that

$$|E_\varepsilon\big(\mathbf{u}, (K^- \mathbf{1}_{A_j})_\varepsilon(2^j z)\big)| \geq |E_{\varepsilon 2^j}\big(\mathbf{u}, (K^- \mathbf{1}_{A_j})_{\varepsilon 2^j}(2^j z)\big)|.$$

Writing out the formula for the energy functional

$$\begin{aligned} &E_{\varepsilon 2^j}\big(\mathbf{u}, (K^- \mathbf{1}_{A_j})_{\varepsilon 2^j}(2^j z)\big) \\ &\quad = \sum_{(i,j) \in \mathcal{I}_N} \sigma_{i,j} \frac{1}{\varepsilon 2^j} \int_D u_i(x) \int_{\mathbb{R}^d} \frac{1}{(\varepsilon 2^j)^d} K(\frac{z}{\varepsilon}) \mathbf{1}_{A_j}(z/\varepsilon) u_j(x+z)\, dz\, dx \\ &\quad = \frac{1}{2^{j(d+1)}} E_\varepsilon\big(\mathbf{u}, (K^- \mathbf{1}_{A_j})_\varepsilon(z)\big) \end{aligned}$$

Therefore,

$$E_\varepsilon\big(\mathbf{u}, K_\varepsilon^+(z)\big) + \sum_{j=1}^{\infty} 2^{j(d+1)} E_\varepsilon\big(\mathbf{u}, (K^-\mathbf{1}_{A_j})_\varepsilon(2^j z)\big) \\ \leq E_\varepsilon\big(\mathbf{u}, K_\varepsilon^+(z)\big) + \sum_{j=1}^{\infty} E_\varepsilon\big(\mathbf{u}, (K^-\mathbf{1}_{A_j})_\varepsilon(z)\big).$$

Now note that $K_\varepsilon^+(z) + \sum_{j=1}^{\infty}(K^-\mathbf{1}_{A_j})_\varepsilon(z) = K_\varepsilon(z)$ almost everywhere. Therefore

$$E_\varepsilon\big(\mathbf{u}, K_\varepsilon^+(z)\big) + \sum_{j=1}^{\infty} E_\varepsilon\big(\mathbf{u}, (K^-\mathbf{1}_{A_j})_\varepsilon(z)\big) = E_\varepsilon(\mathbf{u}, K_\varepsilon)$$

and thus we have the desired result

$$E_\varepsilon(\mathbf{u}, K_\varepsilon) \geq E_\varepsilon(\mathbf{u}, \tilde{K}_\varepsilon).$$

It remains to show $\int_{\mathbb{R}^d} |z \cdot n| K(z) dz = \int_{\mathbb{R}^d} |z \cdot n| \tilde{K}(z)\, dz$. This reduces to showing that

$$\int_{\mathbb{R}^d} |z \cdot n| K^-(z)\, dz = \sum_{j=1}^{\infty} 2^{j(d+1)} \int_{\mathbb{R}^d} |z \cdot n| K^-(z2^j)\mathbf{1}_{A_j}(z2^j)\, dz.$$

Changing variables $z' = z2^j$ for each integral on the right hand side and then summing the results gives the equality. The equality implies that $z\tilde{K} \in L^1(\mathbb{R}^d)$ and since $\tilde{K} = K^+$ near the origin it also follows that $\tilde{K} \in L^1(\mathbb{R}^d)$.

□

This completes the proof of Theorem 5.1.

Note that the kernel inequalities given in Theorem 5.1 were only used to show that a certain rearrangement of the negative mass of the kernel K could produce a nonnegative kernel $\tilde{K}$. Indeed the actual necessary conditions on K needed to find a properly rearranged nonnegative $\tilde{K}$ are much weaker than the given inequalities. However, a necessary and sufficient condition is extremely difficult to describe in terms of the physical properties of the kernel. Thus in the next lemma we instead describe all rearrangements that will decrease (18) and preserve $E(\mathbf{u}, \sigma)$ along with (12) and positivity in a neighborhood of the origin. As a result if for some kernel K one of the following rearrangements produces a nonnegative $\tilde{K}$ then $E_\varepsilon(\mathbf{u}, K_\varepsilon)$ Gamma converges to $E(\mathbf{u}, \sigma)$.

Lemma 5.7. *Given a kernel K that satisfies (12) and has a neighborhood of the origin U such that $U = -U$ and $K(U) \subset (0, \infty)$, let $\{\Omega_m^+\}$ and $\{\Omega_j^-\}$ be measurable decomposition of* $\operatorname{supp}(K^+) \setminus U$ *and* $\operatorname{supp}(K^-)$ *respectively and let $\{\psi_m^+\}$ and $\{\psi_j^-\}$ be sequences of nonnegative functions supported on* $\operatorname{supp}(K^+)$ *and* $\operatorname{supp}(K^-)$ *respectively such that*

$$\Omega_m^+ = -\Omega_m^+ \text{ and } \Omega_j^- = -\Omega_j^- \text{ for all } m, j$$

$$\frac{1}{j}\cdot\Omega_j^-\cap U=\varnothing \ \textit{ for all } \ j$$

$$\sum_{n=1}^{\infty}\psi_m^+(z)=1 \ \textit{ for all } \ z\in\operatorname{supp}(K^+)\setminus U$$

$$\sum_{m=1}^{\infty}\psi_j^-(z)=1 \ \textit{ for all } \ z\in\operatorname{supp}(K^-).$$

$$\psi_m^+(z)=\psi_m^+(-z) \ \textit{ and } \ \psi_j^-(z)=\psi_j^-(-z) \ \textit{ for all } \ m,j,z$$

Define

$$\begin{aligned}\tilde{K}(z)=K^+(z)\mathbf{1}_U(z)+\sum_{m=1}^{\infty}\frac{1}{m^{d+1}}K^+(z/m)\mathbf{1}_{\Omega_m^+}(z/m)\psi_m^+(z/m)\\ +\sum_{j=1}^{\infty}j^{d+1}K^-(jz)\mathbf{1}_{\Omega_j^-}(jz)\psi_j^-(jz).\end{aligned}$$

Then $\tilde{K}$ satisfies (12), $\tilde{K}$ is positive in a neighborhood of the origin, the energies satisfy the inequality $E_\varepsilon(\mathbf{u},K_\varepsilon)\geq E_\varepsilon(\mathbf{u},\tilde{K}_\varepsilon)$ and for every $n\in S^{d-1}$

$$\int_{\mathbb{R}^d}|z\cdot n|K(z)dz=\int_{\mathbb{R}^d}|z\cdot n|\tilde{K}(z)\,dz.$$

Proof. By design $\tilde{K}$ is strictly positive in a neighborhood of the origin and $\tilde{K}(z)=\tilde{K}(-z)$. As in the proof of Lemma 5.6 the equality $\int_{\mathbb{R}^d}|z\cdot n|K(z)dz=\int_{\mathbb{R}^d}|z\cdot n|\tilde{K}(z)\,dz$ will follow by changing variables and recollecting the various terms in the sums. Therefore we know $z\tilde{K}(z)\in L^1(\mathbb{R}^d)$. Since $\tilde{K}$ and K are equal on U it also follows that $\tilde{K}\in L^1(\mathbb{R}^d)$.

Again we use the fact that the energy is linear in the kernel to write

$$\begin{aligned}E_\varepsilon(\mathbf{u},\tilde{K}_\varepsilon)=E\big(\mathbf{u},(K^+\mathbf{1}_U)_\varepsilon\big)+\sum_{m=1}^{\infty}\frac{1}{m^{d+1}}E_\varepsilon\big(\mathbf{u},(K^+\mathbf{1}_{\Omega_m^+}\psi_m^+)_\varepsilon)(z/m)\big)\\ +\sum_{j=1}^{\infty}j^{d+1}E_\varepsilon\big(\mathbf{u},(K^-\mathbf{1}_{\Omega_j^-}\psi_j^-)_\varepsilon(jz)\big)\end{aligned}$$

as well as

$$\begin{aligned}E_\varepsilon(\mathbf{u},K)=E\big(\mathbf{u},(K^+\mathbf{1}_U)_\varepsilon\big)+\sum_{m=1}^{\infty}E_\varepsilon\big(\mathbf{u},(K^+\mathbf{1}_{\Omega_m^+}\psi_m^+)_\varepsilon(z)\big)\\ +\sum_{j=1}^{\infty}E_\varepsilon\big(\mathbf{u},(K^-\mathbf{1}_{\Omega_j^-}\psi_j^-)_\varepsilon(z)\big).\end{aligned}$$

Thus we only need to show

$$\frac{1}{m^{d+1}} E_\varepsilon\big(\mathbf{u}, (K^+ \mathbf{1}_{\Omega_m^+} \psi_m^+)_\varepsilon)(z/m)\big) \leq E_\varepsilon\big(\mathbf{u}, (K^+ \mathbf{1}_{\Omega_m^+} \psi_m^+)_\varepsilon(z)\big)$$

and

$$j^{d+1} E_\varepsilon\big(\mathbf{u}, (K^- \mathbf{1}_{\Omega_j^-} \psi_j^-)_\varepsilon(jz)\big) \leq E_\varepsilon\big(\mathbf{u}(K^- \mathbf{1}_{\Omega_j^-} \psi_j^-)_\varepsilon(z)\big).$$

The second inequality follows from an argument identical to the one in Lemma 5.6. For the first inequality we can use Lemma 5.4 in the opposite direction to get

$$\frac{1}{m^{d+1}} E_\varepsilon\big(\mathbf{u}, (K^+ \mathbf{1}_{\Omega_m^+} \psi_m^+)_\varepsilon)(z/m)\big) \leq \frac{1}{m^{d+1}} E_{\varepsilon/m}\big(\mathbf{u}, (K^+ \mathbf{1}_{\Omega_m^+} \psi_m^+)_{\varepsilon/m})(z/m)\big).$$

Simplifying the right hand side of the above equation we see that it is indeed $E_\varepsilon\big(\mathbf{u}, (K^+ \mathbf{1}_{\Omega_m^+} \psi_m^+)_\varepsilon(z)\big)$. Thus the argument is complete.

□

The discussion preceding Lemma 5.7 proves the following more general version of Theorem 5.1

Theorem 5.8. *Let $K(z)$ be a kernel such that K is positive in a neighborhood of the origin, K satisfies (12) and for some rearrangement described in Lemma 5.7 the resulting $\tilde{K}$ is nonnegative. Given constant surface tensions $\sigma_{i,j} > 0$ satisfying the triangle inequality (9), define*

$$\sigma_{i,j}(n) = \sigma_{i,j} \int_{\mathbb{R}^d} K(z)|z \cdot n| \, dz$$

Then as $\varepsilon \to 0$ the Lyapunov functionals $E_\varepsilon(\cdot, K_\varepsilon)$ given in (51) Gamma converge in the L^1 topology over $\mathcal{K}$ to the energy $E(\cdot, \sigma)$ given in (73). Furthermore if for some sequence $\mathbf{u}_\varepsilon$ we have $\sup_{\varepsilon>0} E_\varepsilon(\mathbf{u}_\varepsilon, K_\varepsilon) < \infty$ then $\mathbf{u}_\varepsilon$ is pre-compact in $L^1(D)$ and the set of accumulation points is contained in $BV_{\mathcal{K}}(D)$.

In practice it is difficult to check whether a given kernel K has a nonnegative rearrangement $\tilde{K}$. However the following proposition gives a very simple necessary condition.

Proposition 5.9. *Suppose that some rearrangement of K produces a nonnegative $\tilde{K}$. Then for every $s \in (-\infty, 1]$ and every $X \subset \mathbb{R}^d$ that is star shaped with respect to the origin the integral $\int_X |x|^s K(x) \, dx$ is nonnegative.*

Proof. For $t \in \mathbb{R}^+$ let $t \cdot X = \{x \in \mathbb{R}^d : x/t \in X\}$. Then since X is star shaped

with respect to the origin $t \cdot X \subset X$ for $t < 1$ and $X \subset t \cdot X$ for $t > 1$.

$$\begin{aligned}
\int_X |x|^s K(x)\,dx =& \sum_{j=1}^{\infty} \int_X |x|^s K(x) \mathbf{1}_{\Omega_j^+}(x) \psi_j^+(x)\,dx \\
&- \sum_{m=1}^{\infty} \int_X |x|^s |K(x)| \mathbf{1}_{\Omega_m^-}(x) \psi_m^-(x)\,dx \\
=& \sum_{j=1}^{\infty} \frac{1}{j^{s+d}} \int_{j\cdot X} |x|^s K(x/j) \mathbf{1}_{\Omega_j^+}(x/j) \psi_j^+(x/j)\,dx \\
&- \sum_{m=1}^{\infty} m^{s+d} \int_{\frac{1}{m}\cdot X} |x|^s |K(mx)| \mathbf{1}_{\Omega_m^-}(mx) \psi_m^-(mx)\,dx \\
\geq& \int_X |x|^s \tilde{K}(x)\,dx \geq 0
\end{aligned}$$

□

6 Conclusions

Recent developments in our understanding of threshold dynamics entail various assumptions on the convolution kernel used. However, much of the theory can be extended to a greater variety of kernels; the analysis presented in this paper provides several examples. The added flexibility in the choice of the kernel is significant in applications ranging from machine learning to materials science. Moreover, restrictions can often be sidestepped by alternate forms of the basic algorithm that maintain its most beneficial qualities. Several such variants were introduced and rigorously studied in this paper.

7 Acknowledgments

The authors were supported by NSF DMS-1317730. Part of the work of Selim Esedoḡlu was carried out during a visit to Centre de Recherches Mathematiques and McGill University in Montreal. The hospitality of these institutions is gratefully acknowledged.

References

[1] G. Alberti and G. Bellettini. A non-local anisotropic model for phase transitions: asymptotic behavior of rescaled energies. *European J. Appl. Math.*, 9:261–284, 1998.

[2] F. Almgren, J. E. Taylor, and L.-H. Wang. Curvature-driven flows: a variational approach. *SIAM Journal on Control and Optimization*, 31(2):387–438, 1993.

[3] L. Ambrosio and A. Braides. Functionals defined on partitions in sets of finite perimeter II: semicontinuity, relaxation, and homogenization. *J. Math. Pures et Appl.*, 69:307–333., 1990.

[4] G. Barles and C. Georgelin. A simple proof of convergence for an approximation scheme for computing motions by mean curvature. *SIAM J. Numer. Anal.*, 32:484–500, 1995.

[5] G. Barles and P. Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic Analysis*, 4:271–283, 1991.

[6] Ethan D. Bolker. A class of convex bodies. *Transactions of the American Mathematical Society*, 145:323–345, 1969.

[7] E. Bonnetier, E. Bretin, and A. Chambolle. Consistency result for a non-monotone scheme for anisotropic mean curvature flow. *Interfaces and Free Boundaries*, 14(1):1–35, 2012.

[8] Y.-G. Chen, Y. Giga, and S. Goto. Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations. *Journal of Differential Geometry*, 33:749–786, 1991.

[9] M. Elsey and S. Esedoglu. Threshold dynamics for anisotropic surface energies. Technical report, UM, 2016. Under review.

[10] S. Esedoḡlu and F. Otto. Threshold dynamics for networks with arbitrary surface tensions. *Communications on Pure and Applied Mathematics*, 68(5):808–864, 2015.

[11] L. C. Evans. Convergence of an algorithm for mean curvature motion. *Indiana University Mathematics Journal*, 42:553–557, 1993.

[12] L. C. Evans and J. Spruck. Motion of level sets by mean curvature. I. *Journal of Differential Geometry*, 33:635–681, 1991.

[13] C. Garcia-Cardona, E. Merkurjev, A. L. Bertozzi, A. Flenner, and A. G. Percus. Multiclass data segmentation using diffuse interface methods on graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8):1600–1613, 2014.

[14] Paul Goodey and Wolfgang Weil. Centrally symmetric convex bodies and the spherical Radon transform. *Journal of Differential Geometry*, 35:675–688, 1992.

[15] C. Herring. *The Physics of Powder Metallurgy*, chapter Surface tension as a motivation for sintering, pages 143–179. McGraw Hill, 1951.

[16] H. Ishii, G. E. Pires, and P. E. Souganidis. Threshold dynamics type approximation schemes for propagating fronts. *Journal of the Mathematical Society of Japan*, 51:267–308, 1999.

[17] T. Laux and F. Otto. Convergence of the thresholding scheme for multi-phase mean-curvature flow. Technical Report 33/2015, Max Planck Institute for Mathematics in the Sciences, 2015.

[18] S. Luckhaus and T. Sturzenhecker. Implicit time discretization for the mean curvature flow equation. *Calculus of Variations and Partial Differential Equations*, 3(2):253–271, 1995.

[19] P. Mascarenhas. Diffusion generated motion by mean curvature. CAM Report 92-33, UCLA, July 1992. (URL = http://www.math.ucla.edu/applied/cam/index.html).

[20] B. Merriman, J. K. Bence, and S. Osher. Motion of multiple junctions: a level set approach. *Journal of Computational Physics*, 112(2):334–363, 1994.

[21] B. Merriman, J. K. Bence, and S. J. Osher. Diffusion generated motion by mean curvature. In J. Taylor, editor, *Proceedings of the Computational Crystal Growers Workshop*, pages 73–83. AMS, 1992.

[22] M. Miranda, D. Pallara, F. Paronetto, and M. Preunkert. Short-time heat flow and functions of bounded variation in $\mathbb{R}^N$. *Ann. Fac. Sci. Toulouse, Mathematiques*, 16(1):125–145, 2007.

[23] W. W. Mullins. Two dimensional motion of idealized grain boundaries. *J. Appl. Phys.*, 27:900–904, 1956.

[24] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42:577–685, 1989.

[25] S. J. Ruuth. A diffusion generated approach to multiphase motion. *Journal of Computational Physics*, 145:166–192, 1998.

[26] S. J. Ruuth. Efficient algorithms for diffusion-generated motion by mean curvature. *Journal of Computational Physics*, 144:603–625, 1998.