

Upper Bounds on the Coarsening Rate of Discrete, Ill-Posed Nonlinear Diffusion Equations

Selim Esedoğlu and John B. Greer

September 20, 2006

Abstract

We prove a weak upper bound on the coarsening rate of the discrete-in-space version of an ill-posed, nonlinear diffusion equation. The continuum version of the equation violates parabolicity and lacks a complete well-posedness theory. In particular, numerical simulations indicate very sensitive dependence on initial data. Nevertheless, models based on its discrete-in-space version, which we study, are widely used in a number of applications, including population dynamics (chemotactic movement of bacteria), granular flow (formation of shear bands), and computer vision (image denoising and segmentation). Our bounds have implications for all three applications.

1 Introduction.

We prove a weak upper bound on the coarsening rate of a finite difference approximation of an ill-posed nonlinear parabolic partial differential equation (PDE) in one and two space dimensions. This finite difference approximation arises in the study of population dynamics, where it describes a class of reinforced random walks on a lattice, and in granular flow, where it is a simplified one dimensional model for the formation of shear bands by anti-plane shear in a granular medium. The scheme also has strong connections to a PDE and discrete algorithm introduced by Perona and Malik for image denoising. Our results apply to the Perona-Malik method in one space dimension.

Our proof relies on a method introduced by Kohn and Otto in [14] for proving weak upper bounds on the coarsening rates of energy driven systems. Given a system length scale L and energy E , their method requires only a dissipation inequality between $\frac{dL}{dt}$ and $\frac{dE}{dt}$ and a pointwise isoperimetric inequality relating L to E . These inequalities are combined with an ODE argument to prove a time-averaged lower bound on the energy that is conceptually equivalent to an upper bound on the coarsening rate. Kohn and Otto first applied the method to Cahn-Hilliard models for the coarsening of an equal-volume fraction binary mixture [14], and it has more recently been applied to both mean-field [8] and phase-field [9] models of phase separation, in addition to multicomponent phase separation [16], Mullins-Sekerka evolution of a binary mixture in the small volume fraction regime [7], epitaxial growth [15], and thin film droplets [18].

Although we follow the method of proof outlined in [14], the problem studied here requires new arguments to prove the isoperimetric inequality. In addition, the discrete setting distinguishes it from previous examples in a substantial way. The system's coarsening rate depends on the minimal length scale h (the grid cell width for the finite difference scheme) – numerical experiments show that the coarsening slows to a halt as $h \rightarrow 0$ [25]. Any good bound on the system's coarsening rate must reflect this dependence. We further remark that while the system studied here may be derived as a finite difference scheme for a PDE, our

primary interest is the discrete system itself, not the PDE. As is discussed in sections 1.2-1.3, the system is interesting in its own right, making it the subject of much recent research [10, 11, 19, 25].

We present results for one and two space dimensions, but for clarity, we limit discussion in the first few sections to the one dimensional result. The remainder of Section 1 describes the coarsening behavior observed in numerical experiments with the one dimensional scheme (Sections 1.1-1.3), and then explains the overall method of proof and states our main theorem (Section 1.4) in one dimension. Section 2 is devoted to establishing some fundamental properties of the scheme under study. One of the main ingredients of the proof of our main result, a decay relation that results from the gradient flow structure of the system, is established in Section 3. Section 4 is devoted to the second, more interesting, ingredient: an interpolation inequality. Section 5 then presents the proof of our main theorem in one space dimension by pulling together the results of the previous two sections. Numerical experiments corroborating our estimates follow in Section 6. Finally, Section 7 shows how to extend the interpolation inequality of Section 4 to two space dimensions. This in turn allows us to generalize our main theorem from Section 1.4 to two dimensions.

1.1 The Scheme.

We will start by considering the one dimensional system

$$\frac{dv_i}{dt} = D_h^+ D_h^- (R(v_i)) \quad i = 0, \dots, N-1 \quad (1)$$

where

$$D_h^+ v_i = \frac{v_{i+1} - v_i}{h}, \quad (2)$$

$$D_h^- v_i = \frac{v_i - v_{i-1}}{h}, \quad (3)$$

and $h = \frac{1}{N}$. To make sense of (1) when $i = 0$ or $i = N-1$, we set

$$v_i = v_j \quad j = i \bmod N \quad \text{when } i \neq 0, \dots, N-1. \quad (4)$$

This system has a non-increasing energy,

$$E(v) = \sum_{i=1}^{N-1} f(v_i)h, \quad (5)$$

where

$$f(s) = \int_0^s R(\xi) d\xi. \quad (6)$$

Requiring (4) ensures that the dynamics of (1) leave the total mass,

$$\mu := \frac{1}{N} \sum_i v_i, \quad (7)$$

unchanged.

For convex functions f , (1) is a convergent (as $h \rightarrow 0$) finite differences approximation of the parabolic PDE

$$v_t = (R(v))_{xx} \quad (8)$$

on the interval $[0, 1]$ with periodic boundary conditions. Such systems are well-understood; this paper is concerned instead with a special class of *non-convex* functions f that are convex on some interval $[\underline{v}, \hat{v})$ and concave on (\hat{v}, ∞) . We also assume that $R(s) = f'(s)$ is bounded on $[\underline{v}, \infty)$ with

$$R(\underline{v}) < R(s) \quad \text{for all } s \in [\underline{v}, \infty). \quad (9)$$

We allow \underline{v} to take on the value $-\infty$, in which case we require

$$\lim_{s \rightarrow -\infty} R(s) = \alpha \quad (10)$$

with

$$-\infty < \alpha < R(s) \text{ for all } s \in \mathbb{R}.$$

We will often refer to the example

$$R(s) = \frac{s}{1+s^2}, \quad (11)$$

with primitive

$$f(s) = \frac{1}{2} \log(1+s^2) \quad (12)$$

for $x \in [0, \infty)$. In this case, $\underline{v} = 0$ and $\hat{v} = 1$.

Writing (1) in the form

$$\frac{d}{dt} v_i = D_h^+ (R'(\xi) D_h^- v_i),$$

where ξ is between v_{i-1} and v_i , and similarly rewriting (8),

$$v_t = (R'(v) v_x)_x$$

shows that R' acts as diffusive flux. The switch of convexity in f corresponds to a maximum of R and a zero crossing of R' . If v is restricted so that $v < \hat{v}$, then (1) is parabolic, but for $v > \hat{v}$, $R'(v) < 0$ and the model behaves as a backwards diffusion equation. Figure 1 shows f , R , and R' for f given by (12).

1.2 Applications of (1).

Using f defined by (12), system (1) describes a reinforced random walk on a lattice and is used to model the movement of biological organisms affected by some external field that is in turn affected by the organisms' presence [11, 19]. Examples of such organisms are ants, which often follow ant trails, and the bacteria *Myxococcus Xanthus*, which glides along a substance produced by the bacteria. Denoting population density by v , (1) arises for external fields proportional to the number of entities present.

The dynamics of (8) share many qualitative features with an ill-posed nonlinear PDE that arises in image processing [21, 22] and granular flow [25, 24]:

$$u_t = (R(u_x))_x. \quad (13)$$

These similarities are expected, since (8) may be derived from (13) by letting $v = u_x$ and taking a spatial derivative of (13). For functions f such as (12), which have the additional properties of being nonnegative and even with

$$f(0) = 0,$$

(13) is a one-dimensional restriction of

$$u_t = \nabla \cdot (R(|\nabla u|) \nabla u), \quad (14)$$

which was introduced by Perona and Malik to denoise digital images represented by the function u [21, 22]. Intending to smooth noisy regions of the image without blurring edges (object boundaries), they required that f , the primitive of R , satisfy exactly the conditions discussed above.

Equation (13) has also been studied as a simplified one-dimensional model for the formation of shear bands in a granular medium [25]. In this case, f is defined by

$$f(s) = \sin \alpha \sin \phi \log \left| \cos \phi + s + \sqrt{1 + 2s \cos \phi + s^2} \right| + \cos \alpha \sqrt{1 + 2s \cos \phi + s^2}, \quad (15)$$

with $0 < \alpha < \frac{\pi}{2}$ and $0 \leq \phi \leq \pi$. This choice for f has $\underline{v} = -\infty$ and

$$\hat{v} = -\frac{\sin(\alpha - \phi)}{\sin \alpha}.$$

See [25] for an explanation of the parameters α and ϕ in relation to granular flow. Figure 2 includes a comparison of (12) with (15) for $\alpha = \frac{\pi}{4}$ and $\phi = \frac{\pi}{8}$.

Much attention has been devoted to the ill-posed equations such as (8) and especially (13). In particular, there is a lot of recent effort that concerns developing an existence theory [13, 12, 26, 5]. Another topic of research has been appropriately regularized versions of the PDEs [6, 3, 2, 4, 1, 17]. But some work regarding the two PDEs has actually focused on discretizations like (1) and the similar discretization of (13),

$$\frac{d}{dt} u_i = D_h^- (R(D_h^+ u_i)). \quad (16)$$

We may derive (1) from (16) by letting

$$v_i = D_h^+ u_i \quad (17)$$

and applying D_h^+ to (16). Although the limiting behavior of (1) and (16) as $h \rightarrow 0$ is unclear for non-convex f (see [10] for a particular scaling limit), the schemes have nonetheless been studied in relation to each of the above mentioned applications. In image processing, the discretization is actually more important than the intended PDE model, as applications of the model on digital images involve only discretizations of (13). In population dynamics, (1) may be thought of as an example of a reinforced random walk on a lattice. The authors of [24, 25], pointed out that discrete models like (1) are of interest in of granular flow, which is inherently discrete with minimum length scales determined by the grain sizes.

1.3 Coarsening behavior.

System (1) may be derived as a gradient descent of (5) in the *discrete* H^{-1} norm (See the proof of Lemma 3 in this paper),

$$\|v_i\|_{H^{-1}} := \left\{ \sup_{\phi} \frac{1}{N} \sum_{i=0}^N (v_i - \mu) \phi_i, \frac{1}{N} \sum_{i=1}^N [(\phi_{i+1} - \phi_i) N]^2 \leq 1 \right\}. \quad (18)$$

As with all gradient descents, the energy landscape drives the system's dynamics as the v_i evolve towards a minimum of (5). The absolute minimum of (5) for a given set of initial data is determined entirely by μ , the initial data's mean value. If

$$\mu \leq \hat{v}, \quad (19)$$

where \hat{v} is the boundary between the convex and concave regions of f , then E is minimized when $v_i = \mu$ for all i . If

$$\mu > \hat{v}, \quad (20)$$

the concave region of f causes the least energy state to have all but one of the $v_i < \hat{v}$ leaving the remaining v_j to be as large as is necessary to conserve the total mass of the initial data.

Remembering that $\frac{dv}{dt} = 0$ at the energy minimum, one may use (1) and the structure of R to find the energy minimum when (20) holds. If $v_j = v_+ > \mu$ is the single remaining spike, then $v_i = v_- < \mu$ for all $i \neq j$, with

$$R(v_+) = R(v_-). \quad (21)$$

Conservation of mass implies

$$v_+ + (N-1)v_- = N\mu. \quad (22)$$

These two conditions suffice to solve for v_+ and v_- given μ . If f is given by (12), then (21) implies

$$v_+ = \frac{1}{v_-}.$$

Combining with (22) gives

$$v_{\pm} = \frac{N\mu \pm \sqrt{(N\mu)^2 - 4(N-1)}}{2}.$$

The gradient descent structure of (1) drives the v_i to this energy minimum and may be stopped only by the possible interference of a saddle point or local minimum corresponding to another stationary state of (1).

If the initial data of (1) satisfy (19) the energy minimization is expressed as a typical parabolic smoothing of the initial data. If (20) holds, then the energy minimization creates a more interesting dynamic. In this case, the conserved total mass of the data quickly aggregate to $K \ll N$ of the v_i , forming what may be aptly described as *spikes*. These spikes do not move, but their sizes do change. Gradually some spikes disappear while the remaining ones grow as the system converges to the energy minimum where a single v_j contains most of the mass (see Figure 3). The distance between spikes establishes a system length scale that coarsens, much like length scales in phase change models like the Allen-Cahn and the Cahn-Hilliard equations. The other steady state solutions of (1) may stop the coarsening process, but such interference is rare, since the single spike solution is the only stable steady state solution for initial data satisfying (20). A complete study of all steady state solutions of (1), including an investigation of stability, may be found in [19] and [25].

Though the initial aggregation of system mass to spikes occurs rapidly, the dynamics slow as the number of spikes decrease. The authors of [25] used numerical simulations of (1) to measure this coarsening rate and observed that the number of spikes scale like

$$K \sim \left(\frac{N}{t}\right)^{-\frac{1}{3}}. \quad (23)$$

Despite a remarkable correlation with the data (See [25] or Section 6 of this paper), ours is the first proof concerning this rate.

The physical meaning of the coarsening process of (1) depends on the particular application. For granular flow, where the energy function given by (15), the coarsening corresponds to a decrease in the number of shear bands in the granular medium (see Figure 5). When considering f given by (12), the coarsening represents an aggregation of biological organisms into population centers. For the nonlinearities used in image processing, our analysis applies only to entirely nonnegative data (and may be easily modified for entirely non-positive data). Since we study the signal's discrete derivative, we are examining the coarsening of the terraces produced by the Perona-Malik method along a single edge (see Figure 4). This coarsening corresponds to a simplification of the processed image. At early stages the image is close to the original, perhaps noisy image, while at later stages, fine structures (including noise) disappear and only the larger features remain. An accurate understanding of the coarsening speed of (16) might be used to estimate the computation time needed to process an image up to a desired state of complexity.

1.4 Our main result and method of proof.

We present bounds on the coarsening rate in one and two space dimensions. First, we prove the following time averaged lower bound on the energy of the one dimensional scheme (1):

Theorem 1. Suppose $\underline{v} \leq v_i(0)$ for $1 \leq i \leq N$,

$$\hat{v} < \mu = \frac{1}{N} \sum_i v_i, \quad (24)$$

and $v_i(0) > \hat{v}$ for at most $\frac{N}{2}$ values of i . Then there exists a universal constant $C_1 < \infty$ such that

$$\frac{1}{T} \int_0^T E^2 dt \geq 3^{\frac{1}{3}} C_1^{\frac{4}{3}} \left[(N^2 T)^{-\frac{1}{3}} \right]^2 \quad (25)$$

for $T \geq \frac{\sqrt{3}}{C_1} N L(0)^3$. The constant C_1 is provided by the interpolation estimate (32).

Remark. Theorem 1 may be generalized to bounds on more than the time average of E^2 . Combining Lemma 4 of this paper with Lemma 3 of [14] gives

$$\frac{1}{T} \int_0^T E^{\theta r} L^{-(1-\theta)r} dt \geq C \left[(N^2 T)^{-\frac{1}{3}} \right]^r$$

for any $0 \leq \theta \leq 1$ and r satisfying $r < 3$, $r\theta > 1$ and $(1-\theta)r < 2$. We restrict discussion to (25) only to simplify presentation.

Theorem 1 applies to large time behavior of (1), since $K \ll N$ (and is certainly less than $\frac{N}{2}$) after a transient initial period where the v_i rapidly separate into spikes and background. Our discussion of the system dynamics suggest that an appropriate measure of length scale is $\tilde{L} = \frac{1}{K}$, where K is the number of spikes in the system. Since f is an increasing function of v , on $[\underline{v}, \infty)$, the K spikes dominate the system energy, E (5):

$$E \sim K,$$

that is

$$E \sim \frac{1}{\tilde{L}}, \quad (26)$$

thus connecting (25), a bound on E , with the system length scale.

Given a system energy E and length scale L , Kohn and Otto introduced in [14] a method for proving time averaged bounds like (25). Their method requires two ingredients connecting E with L : a decay estimate that typically arises from the model's gradient descent structure and an interpolation estimate analogous to (26). Armed with these estimates, one applies an ODE argument like Lemma 3 in [14] or Lemma 4.2 in [18] to prove the time-averaged bound.

Although $\tilde{L} = \frac{1}{K}$ provides an intuitive measure of length scale, there is no clear decay estimate relating E with K . Instead we pick

$$L = \|v_i\|_{H^{-1}}, \quad (27)$$

which is similar to length scales used in [14, 15, 18]. Since (1) is a gradient descent of E with respect to the discrete H^{-1} norm, the system's decay estimate,

$$|\dot{L}| \leq (-\dot{E})^{\frac{1}{2}},$$

follows easily from arguments presented in [14]. To check that L is a valid measure of length scale, we provide numerical evidence in Section 6 that $L \sim \frac{1}{K}$.

The system's interpolation estimate,

$$EL \geq \frac{C}{N},$$

is more difficult to derive. The discrete setting of this problem requires arguments different from those used in other proofs of coarsening rate bounds [8, 9, 14, 15, 16, 18]; in particular, the dependence of the dynamics on the system size, N , distinguishes the system studied here. Although the decay estimate relating E with L does not depend N , the interpolation estimate does, and the coarsening rate reflects this dependence, precisely as observed in [25].

As discussed in [14], the methods used here will not provide a lower bound on the coarsening rate of (1). Proving a suitable lower bound would likely be more difficult, since the dynamics of (1) may be slowed by a variety of factors, including the interference of local minima and saddle points of the energy; hence there is in fact no lower bound in the naive sense. On the other hand, a system can not coarsen any faster than its energy landscape allows.

2 Preliminaries.

We consider only initial data satisfying $v_i(0) \geq \underline{\nu}$. The following lemma shows that the v_i retain this lower bound at all later times.

Lemma 1. *Suppose v_i solves (1) for $t \in [0, T]$ and satisfies*

$$v_i(0) \geq \underline{\nu}$$

for $i = 0, \dots, N-1$. Then, v_i exists for all $t \geq 0$ and

$$v_i(t) \geq \underline{\nu}$$

for $i = 0, \dots, N-1$ all $t \geq 0$.

Proof. First note that standard ODE existence theory provides for the existence of v_i solving (1) on some time interval finite $[0, T]$. Pick $\varepsilon > 0$ and let v_i^ε solve

$$\frac{dv_i^\varepsilon}{dt} = \frac{R(v_{i+1}(t)) - 2R(v_i(t)) + R(v_{i-1}(t))}{h^2} + \varepsilon \quad (28a)$$

$$v_i^\varepsilon(0) = v_i(0) \quad (28b)$$

for $i = 0, \dots, N-1$. For small enough ε , continuous dependence of ODEs on parameters (See for example Theorem 2 in Section 2.3 of [20]) ensures the existence of such a v_i^ε with $v_i^\varepsilon(t) \rightarrow v_i(t)$ as $\varepsilon \rightarrow 0$ for each i and all $t \in [0, T]$. Suppose $v_i^\varepsilon(t) \geq \underline{\nu}$ for each i and all nonnegative $t \leq t_*$, and that

$$v_j(t_*) = \underline{\nu}$$

for some j . Then

$$\frac{dv_j^\varepsilon}{dt} = \frac{1}{h^2} (R(v_{j+1}) + R(v_{j-1}) - 2R(\underline{\nu})) + \varepsilon \geq \varepsilon > 0.$$

We thus have

$$v_j^\varepsilon(t) > \underline{\nu}$$

for each i and all $t \in [0, T]$, which is a contradiction: We must have that $v_j^\varepsilon(t_*) > \underline{\nu}$. Taking $\varepsilon \rightarrow 0$ proves

$$v_i(t) \geq \underline{\nu}$$

for all $t \geq 0$. Conservation of mass then implies that the v_i remain bounded, and we apply the maximal interval of existence theory for ODEs (see for example Theorem 3 in Section 2.4 of [20]) to complete the proof. \square

We next separate the v_i into *spikes* and background. We call v_i a spike if $v_i > \hat{\nu}$, where $\hat{\nu}$ is the parabolicity threshold discussed above. By selecting $\hat{\nu}$ as the dividing value, we ensure that the number of spikes does not increase:

Lemma 2. *If*

$$v_i(t_0) \leq \hat{v}$$

for some i and some $t_0 \geq 0$, then

$$v_i(t) \leq \hat{v}$$

for all $t \geq t_0$.

Proof. Apply the argument from Lemma 1, but now use the fact that R achieves its maximum at $v = \hat{v}$, so

$$\frac{d}{dt}v_i(t) \leq 0$$

whenever $v_i(t) = \hat{v}$.

3 Decay relation.

Lemma 3.

$$|\dot{L}| \leq (-\dot{E})^{\frac{1}{2}}. \quad (29)$$

Proof. Our selected length scale (18) allows a simple application of the proof of Lemma 2 in [14] to this discrete setting. We provide the details only for the reader's benefit. From (5), we have

$$-\dot{E} = -\sum_{i=0}^{N-1} R(v_i)\dot{v}_i h = -\sum_{i=0}^{N-1} R(v_i)D_h^+ D_h^- (R(v_i)) h.$$

Summing by parts and using the boundary condition (4) gives

$$-\dot{E} = \sum_{i=0}^{N-1} [D_h^- (R(v_i))]^2 h. \quad (30)$$

Pick t_1, t_2 with $0 \leq t_1 < t_2$ and let ξ be the optimal test function in the definition of $L(t_2)$ (see (18) and (27)) so that

$$L(t_2) = \|v(t_2)\|_{H^{-1}} = \sum_{i=0}^{N-1} (v_i(t_2) - \mu)\xi_i h$$

and

$$\sum_{i=0}^{N-1} |D_h^+ \xi_i|^2 h \leq 1.$$

Then we have

$$\begin{aligned} L(t_2) - L(t_1) &\leq \sum_{i=0}^{N-1} (v_i(t_2) - v_i(t_1))\xi_i h \\ &= \int_{t_1}^{t_2} \sum_{i=0}^{N-1} \dot{v}_i \xi_i h dt \\ &= \int_{t_1}^{t_2} \sum_{i=0}^{N-1} D_h^+ D_h^- (R(v_i)) \xi_i h dt \\ &= -\int_{t_1}^{t_2} \sum_{i=0}^{N-1} D_h^- (R(v_i)) D_h^- \xi_i h dt \\ &\leq \int_{t_1}^{t_2} \left(\sum_{i=0}^{N-1} |D_h^- (R(v_i))|^2 h \right)^{\frac{1}{2}} dt. \end{aligned}$$

Repeating the above with ξ'_i optimal in the definition of $L(t_1)$ gives

$$|L(t_2) - L(t_1)| \leq \int_{t_1}^{t_2} \left(\sum_{i=0}^{N-1} |D_h^-(R(v_i))|^2 h \right)^{\frac{1}{2}} dt. \quad (31)$$

Combining (31) with (30) shows that L is an absolutely continuous function of t satisfying (29).

4 An interpolation estimate.

In this section, we establish the second ingredient necessary for applying Kohn and Otto's technique to our problem: We need an interpolation inequality between our energy (5) and our length scale quantity (27). The following lemma establishes the required inequality in one space dimension; we present the two dimensional version in Section 7.1.

Lemma 4.

$$EL \geq \frac{C}{N} \quad (32)$$

for some $C > 0$ that depends only on μ .

Proof.

Suppose there are K spikes, that is K indices i where $v_i > \hat{v}$. Assume that $K \leq \frac{N}{2}$. Note that since f is a nonnegative function with $f(v) \geq f(\hat{v})$ for $v \geq \hat{v}$,

$$E \geq \frac{K}{N} f(\hat{v}). \quad (33)$$

Hence, to reach the desired conclusion, we only need to show that $L \geq C/K$ for some constant C . To that end, let l be the largest integer with

$$N = 2Kl + J \quad (34)$$

for any $J \in \mathbb{N}$ (using \mathbb{N} to denote the set of natural numbers). Notice

$$J \leq \frac{N}{2}, \quad (35)$$

otherwise

$$2Kl = N - J < \frac{N}{2}$$

so

$$2K(2l) < N,$$

contradicting the requirement that l be the largest integer satisfying (34).

We now associate the v_i with grid points on the unit interval by setting $x_i = \frac{i}{N}$ for $0 \leq i < N$. Defining $I_j = [x_{jl}, x_{(j+1)l}]$ for $0 \leq j \leq 2K$ gives $2K$ intervals within $[0, 1]$, each containing l grid points. Construct $g(x)$ on $[0, 1]$ by setting

$$g(x) = -\min((x - x_{jl}), (x_{(j+1)l} - x))$$

on each I_j that does *not* have a spike, letting $g(x) = 0$ everywhere else. Note that since there are $2K$ intervals I_j , at least K of them do not have a spike. Defining

$$g_i = N \int_{x_i}^{x_{i+1}} g(x) dx, \quad (36)$$

determines a grid function with

$$\|g_i\|_{H^1}^2 = N \sum_i (g_{i+1} - g_i)^2 \leq \frac{Kl}{N} = \frac{N-J}{2N}. \quad (37)$$

Pick any interval I_{j^*} on which $g_i \neq 0$ (i.e. one of the intervals that contain no spikes) and let $2\lambda = l$. Suppose $\lambda \in \mathbb{N}$, so that the minimum of $g(x)$ on I_{j^*} occurs at a grid point. Then

$$\frac{1}{N} \sum_{i \in I_{j^*}} |g_i| = \frac{2}{N} \left(\sum_{i=1}^{\lambda-1} \frac{i}{N} + \frac{\lambda}{2N} \right) = \frac{2}{N^2} \left(\frac{\lambda(\lambda-1)}{2} + \frac{\lambda}{2} \right) = \frac{\lambda^2}{N^2} = \left(\frac{N-J}{4KN} \right)^2.$$

If λ is not an integer (l is odd), then the minimum of $g(x)$ on I_{j^*} occurs halfway between two grid points, and

$$\frac{1}{N} \sum_{i \in I_{j^*}} |g_i| = \frac{2}{N^2} \left(\frac{(\lambda-1)((\lambda-1)-1)}{2} + \frac{\lambda-1}{2} \right) + \frac{1}{N} \left(\frac{2\lambda-1}{2N} + \frac{1}{4N} \right) = \frac{1}{N^2} \left(\lambda^2 - \lambda + \frac{3}{4} \right).$$

Noting that

$$f(y) = \frac{y^2}{3} - y + \frac{3}{4} \geq 0$$

for all y , we see

$$\lambda^2 - \lambda + \frac{3}{4} \geq \frac{2}{3}\lambda^2$$

and thus

$$\frac{1}{N} \sum_{i \in I_{j^*}} |g_i| \geq \frac{2}{3} \left(\frac{N-J}{4KN} \right)^2, \quad (38)$$

which holds for both odd and even l .

Finally, wherever g_i is nonzero, $v_i - \mu \leq \hat{v} - \mu < 0$ due to assumption (24). Using

$$\phi_i = \frac{g_i}{\|g_i\|_{H^{-1}}}$$

in the definition of L , (38) implies

$$\begin{aligned} \|g_i\|_{H^1} L &\geq \frac{1}{N} \sum_i g_i (v_i - \mu) \\ &= \frac{1}{N} \sum_{j=1}^{2K} \sum_{i \in I_j} |v_i - \mu| |g_i| \\ &\geq \frac{K(\mu - \hat{v})}{N} \sum_{i \in I_{j^*}} |g_i| \\ &\geq \frac{(\mu - \hat{v})}{6K} \left(\frac{N-J}{2N} \right)^2. \end{aligned}$$

Using (35) and (37) gives

$$L \geq \frac{(\mu - \hat{v})}{6K} \left(\frac{N-J}{2N} \right)^{\frac{3}{2}} \geq \frac{(\mu - \hat{v})}{48K}.$$

5 Proof of Theorem 1.

The remainder of the proof relies on Lemma 3 in [14]. We present the relevant details for the reader's benefit, paying particular attention to this problem's dependence on N . Suppose

$$L(T) > 2L(0). \quad (39)$$

From Lemma 3, we have

$$(\dot{L})^2 \leq -\dot{E}.$$

For this system, $\dot{E}(t) < 0$ until a steady state is reached so we may write t as a function of the energy. As in [14] we use the lower case 'e' when referring to the energy as an independent variable. Using this notation

$$\left(\frac{dL}{de}\right)^2 (-\dot{E}) \leq 1,$$

so

$$\int_0^T E^2 dt \geq \int_0^T e^2 \left(\frac{dL}{de}\right)^2 \left(-\frac{de}{dt}\right) dt = \int_{E(T)}^{E(0)} e^2 \left(\frac{dL}{de}\right)^2 de. \quad (40)$$

Since

$$\left| \int_{E(T)}^{E(0)} \frac{dL}{de} de \right|^2 \leq \int_{E(T)}^{E(0)} e^2 \left(\frac{dL}{de}\right)^2 de \int_{E(T)}^{E(0)} e^{-2} de,$$

(40) implies

$$\int_0^T E^2 dt \geq \left(\frac{1}{E(T)} - \frac{1}{E(0)}\right)^{-1} (L(T) - L(0))^2 \geq E(T) (L(T) - L(0))^2. \quad (41)$$

Suppose $L(T) \geq 2L(0)$. Lemma 4 and condition (39) then imply

$$\int_0^T E^2 dt \geq \left(\frac{C_I}{N}\right)^2 \frac{1}{E(T)}. \quad (42)$$

Defining

$$h(T) = \int_0^T E^2 dt,$$

(42) may be rewritten as

$$h(T) \geq \left(\frac{C_I}{N}\right)^2 (h'(T))^{-\frac{1}{2}}. \quad (43)$$

Grönwall's inequality [20] gives

$$h(T) \geq \left[3 \left(\frac{C_I}{N}\right)^4 T \right]^{\frac{1}{3}} \quad (44)$$

which is the same as (25).

Now consider T with $L(T) < 2L(0)$. In this case, (32) gives

$$h'(T) = E^2(T) \geq C(NL(T))^{-2} \geq C(NL(0))^{-2}.$$

Grönwall's inequality now gives

$$h(T) \geq \left(\frac{N}{C_I} L(0)\right)^{-2} T. \quad (45)$$

As long as $T > \frac{\sqrt{3}}{C_I} NL(0)$, then $(\frac{N}{C_I} L(0))^{-2} T \geq \left[3 \left(\frac{C_I}{N}\right)^4 T \right]^{\frac{1}{3}}$ and the theorem follows.

6 Numerical evidence.

We next demonstrate the actual coarsening rate of (1), by discretizing in time and solving numerically. In particular, we show a dependence of the coarsening rate on N , as our upper bound indicates. We use f defined by (12) in all examples and remark that similar results are discussed in [25] for f given by (15).

Let

$$F(v_i^n) = \frac{R(v_{i+1}^n) - 2R(v_i^n) + R(v_{i-1}^n)}{h^2}. \quad (46)$$

We first consider the Forward Euler Scheme,

$$\frac{v_i^{n+1} - v_i^n}{\tau} = F(v_i^n). \quad (47)$$

To demonstrate the coarsening rate of (47), we use a perturbation of $v_i = 3$ for initial data:

$$v_i(0) = 3 + 10^{-6} \sin\left(\frac{\pi i}{N}\right). \quad (48)$$

Notice that in a system with N grid points, $K = N$ for (48), since $\hat{v} = 1$ for f given by (12). Figure 6 shows a plot $\frac{K}{N}$ versus $N^2 t$ for the solution of (47) with $N = 200, 300, \dots, 1400, 1500$. After a transient initial period, the coarsening strongly corresponds with the predicted power law $(N^2 t)^{-\frac{1}{3}}$. Our analysis does not include the early time dynamics, since our results only hold for the later times when $K \leq \frac{N}{2}$. Figure 7 shows a plot of LN versus $N^2 t$. Its correspondence with the power law $(N^2 t)^{\frac{1}{3}}$ supports our claim that $L \sim \frac{1}{K}$. The quantity L as given by the H^{-1} norm (18) was computed using the following equivalent characterization:

$$\|v\|_{H^{-1}}^2 = \frac{1}{N} \sum_j (D_h^+ w_j)^2 \text{ where } D_h^+ D_h^- w = v \text{ with periodic b.c.} \quad (49)$$

The explicit scheme (47) has a strict time stepping restriction of $k \sim h^2$ which encourages using an implicit method like the Midpoint Scheme

$$\frac{v_i^{n+1} - v_i^n}{\tau} = F\left(\frac{v_i^n + v_i^{n+1}}{2}\right). \quad (50)$$

Figure 8 shows results for the implicit scheme, where time steps are not taken larger than $10h^2$. Significant slowdown in the coarsening may be seen, especially for large N . We use Newton's Method to solve (50) at each time step, using explicit time stepping to provide an initial guess for the Newton iterations. To ensure rapid convergence of the iterations, we use an adaptive time step – reducing the time step when many iterations are required, and increasing when only 0-5 iterations are required. We use the Sherman-Morrison Formula [23] and the tridiagonal matrix solver in LAPACK to solve the linear systems for each Newton iteration. Typically smaller time steps are required only at the beginning stages of evolution.

Unfortunately, though (50) has no stability restrictions on τ , large time steps cause errors that slow the coarsening rate. This slow down effect is easily demonstrated by fixing N and solving (50) to a fixed time t_* for a range of maximal time steps. We did so for $N = 200$ and $t_* = 10$. Figure 9 shows a plot of K at $t = 10$ as a function of the maximum time step used in computation. This slow-down of the dynamics should be of interest to the image processing community, for which the length scale is more important than the actual time, which is in fact artificial for most image processing applications. In particular, these results suggest that while implicit time stepping may remove time step restrictions, little real gain in processing speed might be made.

7 Two dimensions.

We easily generalize (1) to two dimensions by considering a system $v_{i,j}, i, j = 0, \dots, N-1$ satisfying

$$\dot{v}_{i,j} = D_{1,h}^+ D_{1,h}^- (R(v_{i,j})) + D_{2,h}^+ D_{2,h}^- (R(v_{i,j})) \quad (51)$$

with the rules

$$v_{i,j} = v_{i \bmod N, j \bmod N} \text{ for } i, j \neq 0, \dots, N-1 \quad (52)$$

and

$$v_{i,j} \geq \underline{v}. \quad (53)$$

Here, $D_{i,h}^+$ and $D_{i,h}^-$ are the forward and backward difference quotient operators, respectively, in the i -th coordinate direction. This system is a standard centered differencing discretization of

$$v_t = \Delta(R(v)) \quad (54)$$

with periodic boundary conditions. Although (51) has no relation to the Perona-Malik method for image denoising or to shearing in granular materials, it is a natural generalization of the reinforced random walk model to a two dimensional lattice, and has been studied in [11, 19]. The behavior of (51) is very similar to (1), with mass aggregating in spikes that coarsen with time. System (51) has

$$E(v_{i,j}) = \sum_{i,j=0}^{N-1} f(v_{i,j}) h^2 \quad (55)$$

for an energy. Condition (52) ensures that

$$\mu := \frac{1}{N^2} \sum_{i,j=1}^N v_{i,j} \quad (56)$$

does not change.

We generalize the definition of L to this two dimensional lattice:

$$L(v_{i,j}) = \|v_{i,j}\|_{H^{-1}} := \sup \left\{ \frac{1}{N^2} \sum_{i,j=0}^{N-1} (v_{i,j} - \mu) \xi_{i,j}, \sum_{i,j=0}^{N-1} [(\xi_{i+1,j} - \xi_{i,j})^2 + (\xi_{i,j+1} - \xi_{i,j})^2] \leq 1 \right\}. \quad (57)$$

In this case, we expect $L \sim \frac{1}{\sqrt{K}}$ where K is again the number of $v_{i,j} > \hat{v}$. Lemmas 1-3 all hold for (51), however the interpolation inequality is different.

7.1 An interpolation estimate.

Lemma 5. *Let $v_{i,j}$ solve (51) with (52) and (53) holding. For L defined by (57) and E defined by (55),*

$$EL^2 \geq \frac{C}{N^2}$$

for some $C > 0$ that depends only on μ .

Proof.

Let Q denote the periodic unit square in \mathbb{R}^2 . We associate the system $v_{i,j}$ with a grid on Q by letting $x_i = \frac{i}{N}$ and $y_j = \frac{j}{N}$. Assume we have K spikes where $v_{i,j} > \hat{v}$ and that $2K \leq N^2$. Divide Q into $2K$ squares Q_v with side length $\frac{1}{\sqrt{2K}}$ and let $l = \frac{N}{\sqrt{2K}}$. Let p be the largest integer with

$$l = p + r$$

for some nonnegative $r < 1$. Note that

$$r < \frac{l}{2}, \quad (58)$$

otherwise

$$l \geq p + \frac{l}{2},$$

thus implying $l \geq 2p$ which contradicts the assumption on p .

One may think of l^2 as an estimate on the number of grid points in each of the Q_v , bearing in mind that unless $\sqrt{2K}$ is an integer that divides N exactly, some of the Q_v will contain more grid points than others. To establish uniformity of sub-squares, we define \hat{Q}_v to be the square contained strictly within Q_v containing exactly $p \times p$ grid points with one of those grid points exactly on the bottom left-hand corner of \hat{Q}_v . Note that each Q_v contains a \hat{Q}_v , so there are exactly $2K$ squares \hat{Q}_v .

Construct $g(x)$ on Q by setting

$$g(x, y) = -\text{dist}(\partial\hat{Q}_v)$$

on each \hat{Q}_v that does *not* have a spike, letting $g(x) = 0$ everywhere else. Note that at least K of the \hat{Q}_v do not have a spike. Defining

$$g_{i,j} = N^2 \int_{x_i}^{x_{i+1}} \int_{y_j}^{y_{j+1}} g(x, y) dx dy, \quad (59)$$

determines a grid function with

$$\|g_{i,j}\|_{H^1}^2 \leq 1. \quad (60)$$

Pick any \hat{Q}_v on which $g_{i,j} \neq 0$ and assume for simplicity that p is even (the odd case is similar). Using C as a ‘‘universal’’ constant, we compute

$$\frac{1}{N^2} \sum_{i,j \in \hat{Q}_v} |g_{i,j}| \geq \frac{C}{N^3} \sum_{j=1}^{\frac{p}{2}} j^2 = \frac{Cp}{N^3} \left(\frac{p^2}{2} + \frac{3p}{2} + 1 \right) \geq C \left(\frac{p}{N} \right)^3.$$

Using (58) gives

$$p = l - r \geq \frac{l}{2},$$

so

$$\frac{1}{N^2} \sum_{i,j \in \hat{Q}_v} |g_{i,j}| \geq C \left(\frac{l}{N} \right)^3 = \frac{C}{2K\sqrt{2K}}. \quad (61)$$

Finally, wherever g_i is nonzero, $v_i - \mu \leq \hat{v} - \mu < 0$ by assumption on μ . The estimates (60) and (61) show

$$\begin{aligned} L &\geq \frac{1}{N^2} \sum_{i,j} g_{i,j} (v_{i,j} - \mu) \\ &= \frac{1}{N^2} \sum_{\hat{Q}_v} \sum_{i,j \in \hat{Q}_v} |v_{i,j} - \mu| |g_{i,j}| \\ &\geq \frac{K(\mu - \hat{v})}{N^2} \sum_{i,j \in \hat{Q}_v} |g_{i,j}| \\ &\geq \frac{C(\mu - \hat{v})}{\sqrt{K}}. \end{aligned}$$

We complete the proof by combining this lower bound on L with the observation

$$E \geq K \frac{f(\hat{v})}{N^2}.$$

7.2 Final result in two dimensions.

Theorem 2. Suppose $\underline{v} \leq v_{i,j}(0)$ for $1 \leq i, j \leq N$,

$$\hat{v} < \mu = \frac{1}{N^2} \sum_{i,j=1}^N v_{i,j}, \quad (62)$$

and $v_{i,j}(0) > \hat{v}$ for at most $\frac{N^2}{2}$ ordered pairs (i, j) . Then for each $\sigma \in (1, 2]$, there exist universal constants $C_1(\sigma), C_2(\sigma) < \infty$ such that for all $0 < T_0 \leq T$,

$$\frac{1}{T - T_0} \int_{T_0}^T E^\sigma dt \geq C_1(\sigma) \left[(N^2(T - T_0))^{-\frac{1}{2}} \right]^\sigma \quad (63)$$

for $(T - T_0) \geq C_2(\sigma) N^2 L(T_0)^4$. If $1 < \sigma < 2$, (63) also holds for $T_0 = 0$.

Proof. The proof can be obtained by following the proof of Lemma 4.2 in [18] for the case $\alpha = 2$. We do not provide the details here because of their similarities with the proof of Theorem 1,

7.3 Numerical evidence for the 2D case.

We use Forward Euler time stepping (47) to solve (51). Although we observed a stability restriction on the explicit scheme of about $\tau \leq \frac{h^2}{4}$, standard implicit schemes provide little speed-up, due to the computation required by repeated Newton iterations and the slow down of coarsening caused by large time steps (see Section 6).

We evolve an $N \times N$ system with $N = 2^n \times 10$ for $1 \leq n \leq 6$. Figure 10 shows $\frac{K}{N^2}$ versus Nt compared with the power law $(N^2 t)^{-\frac{1}{2}}$. Each simulation has the same initial condition,

$$v_{i,j}(0) = 3 + 10^{-6} \sin\left(\frac{i\pi}{N}\right) \sin\left(\frac{j\pi}{N}\right). \quad (64)$$

The results match our proved bound (63).

8 Conclusion.

We proved a rigorous, weak upper bound on the coarsening rate for the discrete in space version of an ill-posed nonlinear diffusion equation that appears in applied contexts as varied as population dynamics, granular flow, and image processing. Despite the lack of a complete continuum theory, the discrete version we analyzed remains in widespread use in these areas. From an applied perspective, coarsening of the solutions is one of the primary features of interest. Moreover, our numerical experiments indicate that not only are the upper bounds we establish attained, but they are in fact generically observed. We therefore hope that our analysis will be directly relevant to applications of these equations.

The lack of a complete continuum theory for the PDEs we deal with forced us to base our analysis on the discrete-in-space version of the PDEs. It would be interesting to see whether our results can be extended to the notions of solution developed for these ill-posed PDEs in recent literature (e.g. [5, 26]), or for regularized versions of these PDEs, such as those proposed and analyzed in [1, 2, 3, 4, 6, 17].

Acknowledgments. We thank Bob Kohn for recommending this problem, and for his advice and encouragement throughout the paper's development. We also thank Felix Otto and Dejan Slepčev for helpful comments.

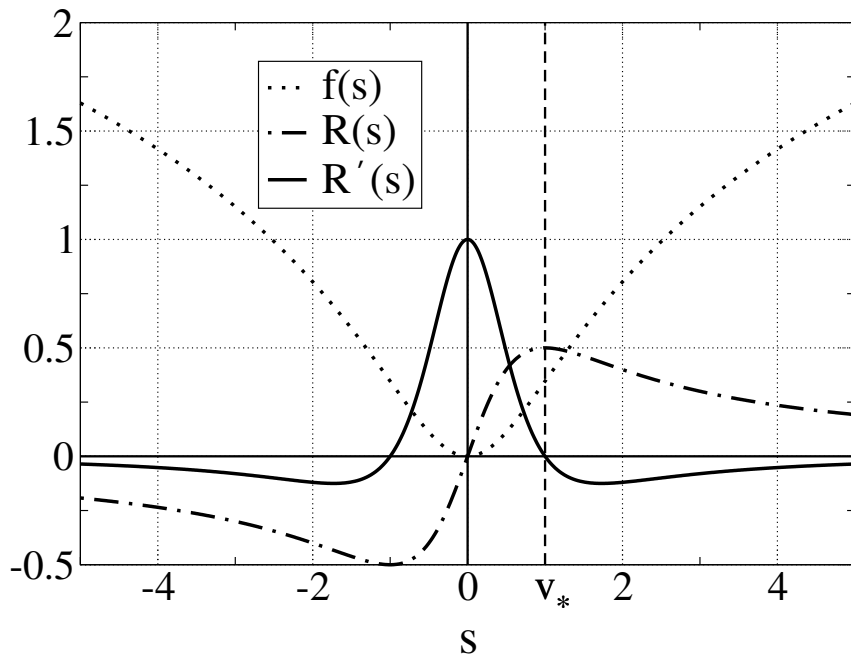


Figure 1: The functions f , R , and R' for $f(s) = \frac{1}{2} \log(1 + s^2)$.

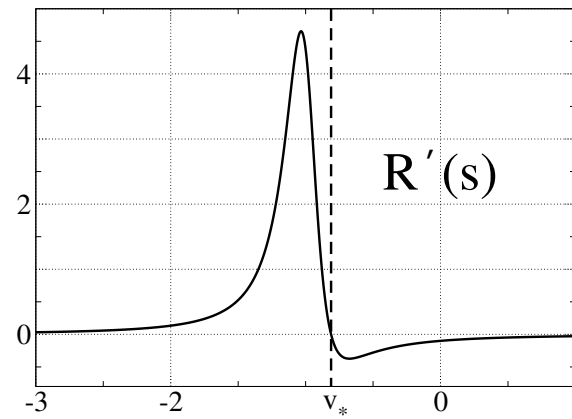
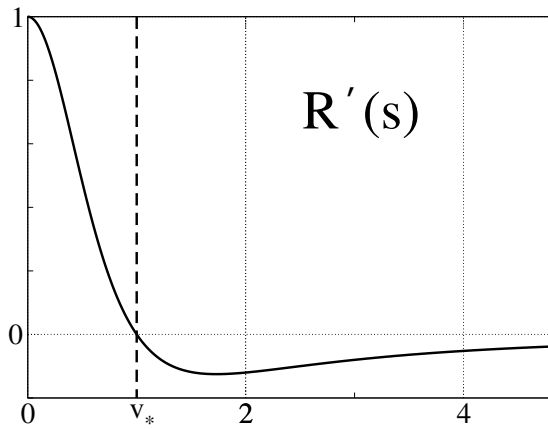
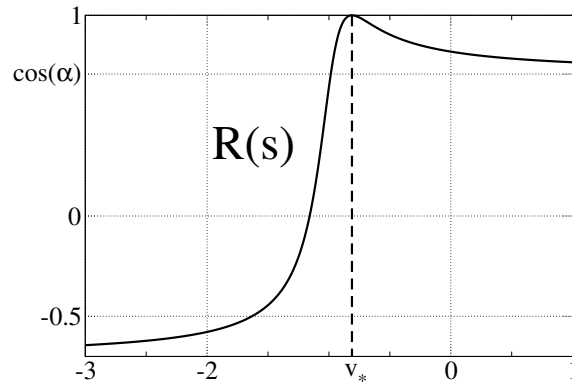
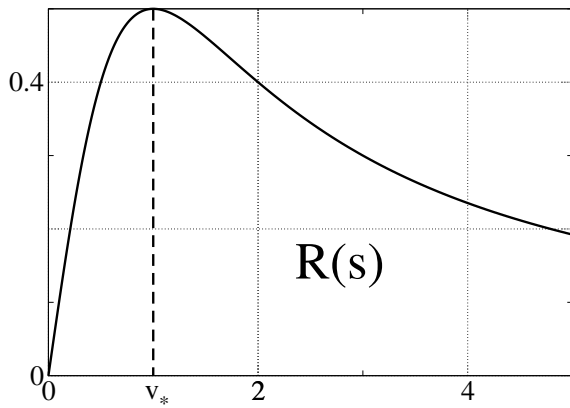


Figure 2: Comparison of R and R' when f is defined by (12) (left) and (15) (right).

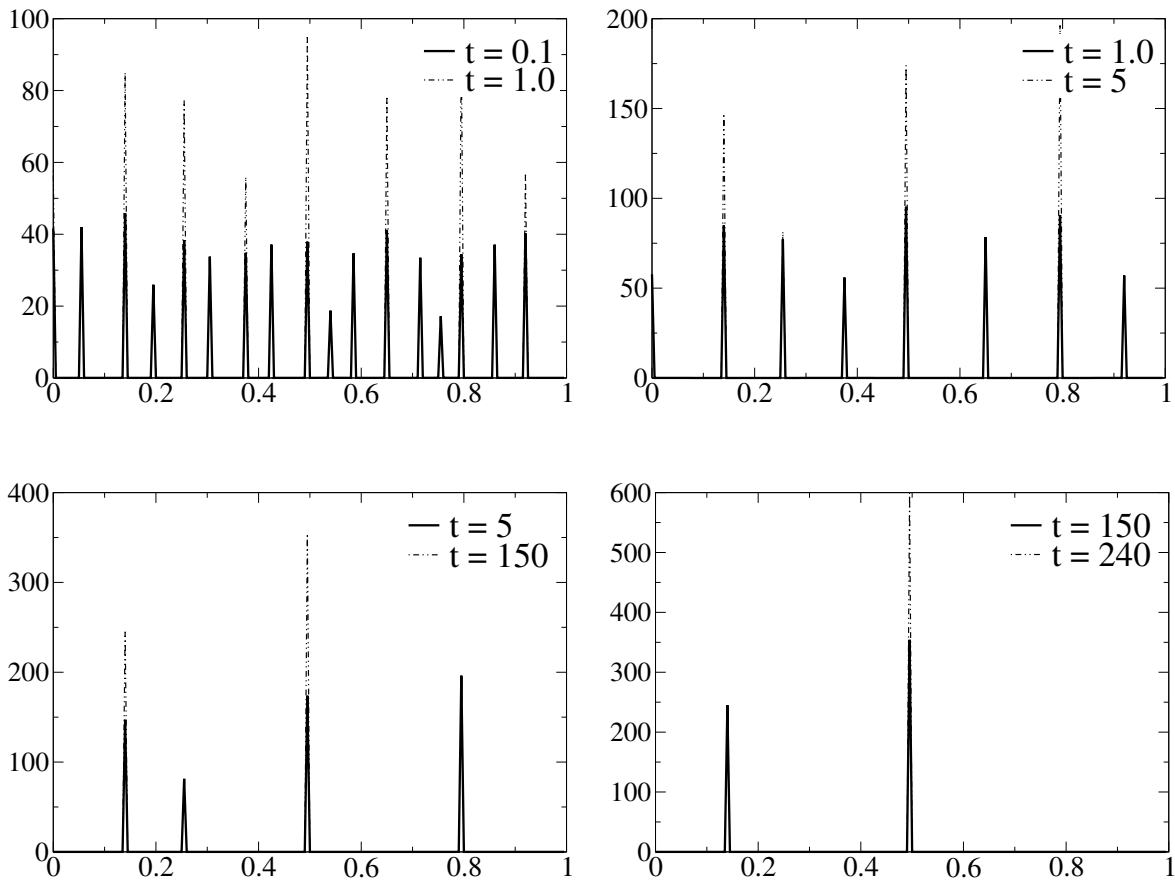


Figure 3: The dynamics of (1) for $N = 200$ and $v_i(0) = 3 + \eta_i$ with a random perturbation $|\eta_i| \leq 10^{-5}$. The v_i quickly aggregate to spikes that decrease in number until a single spike contains most of the mass.

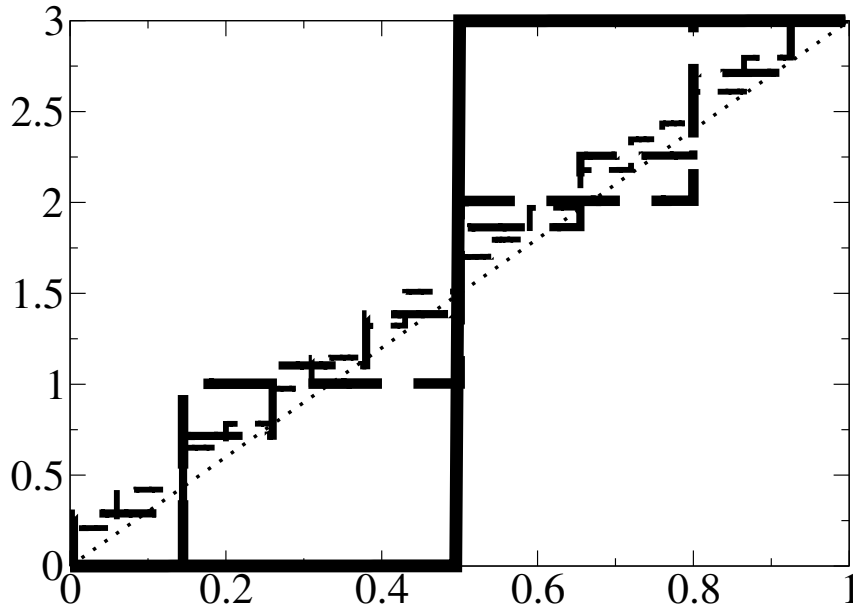


Figure 4: **Perona-Malik equation.** The dynamics of (16) for $N = 200$ and f given by (12). Thicker lines denote later stages of the evolution.

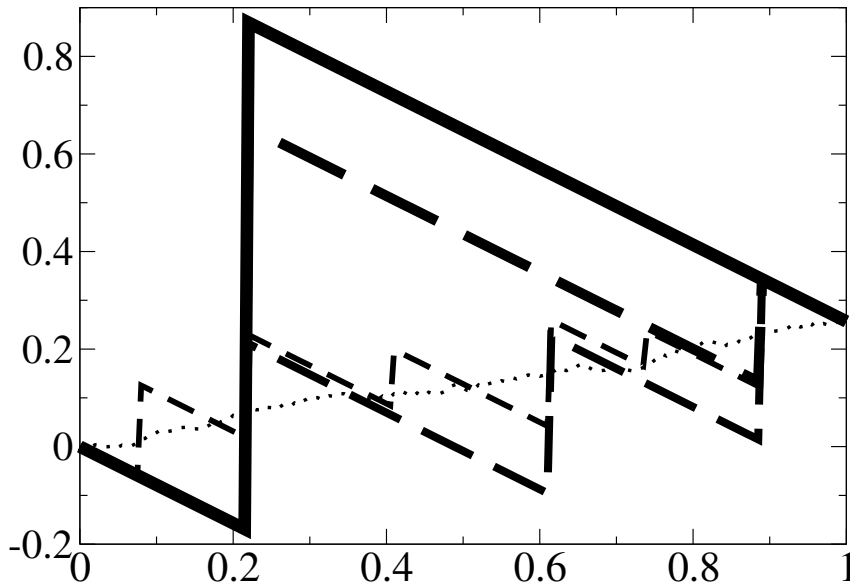


Figure 5: **Shear bands in granular flow.** The dynamics of (16) for $N = 200$ and f given by (15). Thicker lines denote later stages of the evolution.

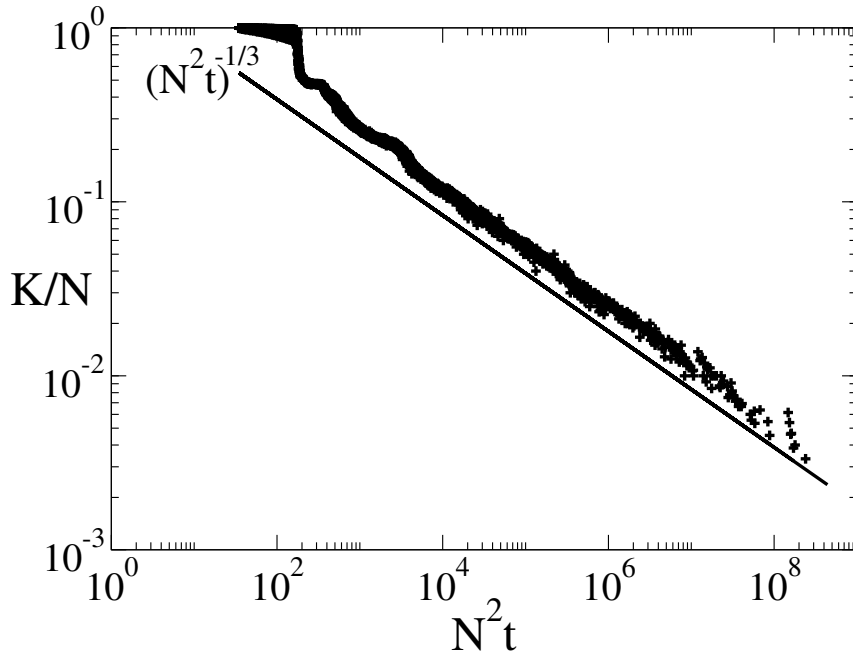


Figure 6: After an initial transient period, the jump density, $\frac{K}{N}$ corresponds closely with the proved coarsening bound (25).

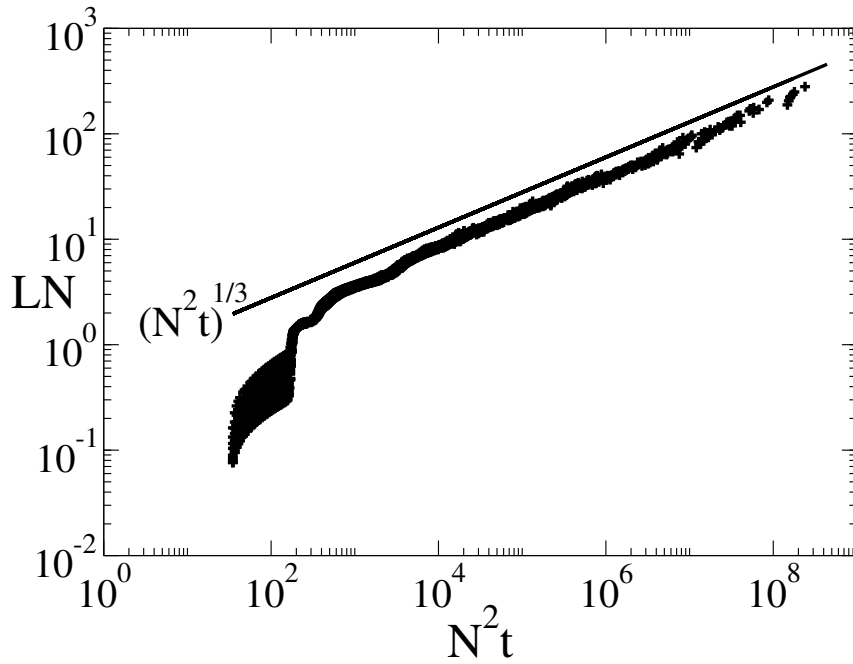


Figure 7: The length scale behaves like $L \sim \frac{1}{K}$, supporting our claim that (27) defines a valid measure of length scale.

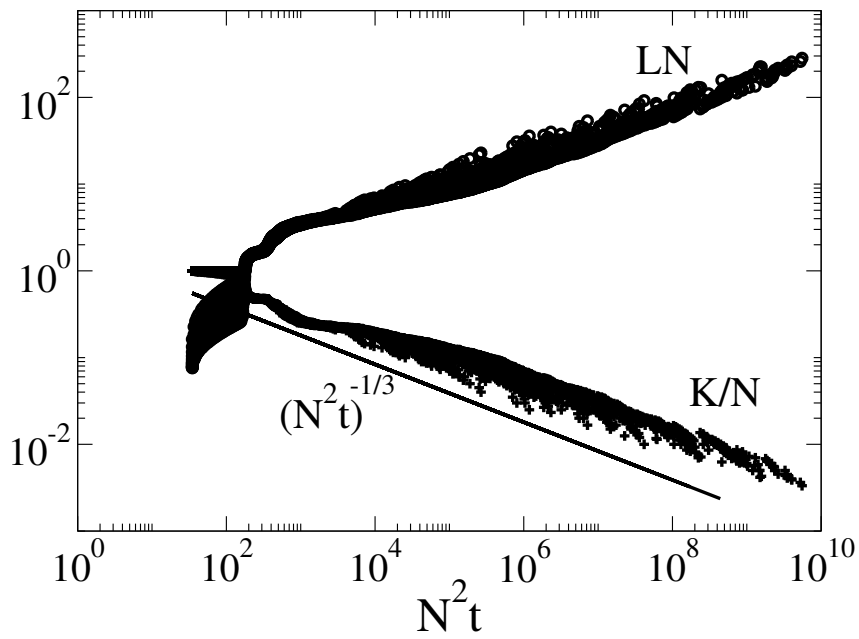


Figure 8: Evolving length scale (top) and jump density (bottom) for the implicit scheme (50). Although implicit schemes have no time step restriction for stability, the coarsening rate badly reflects the error caused by larger time steps.

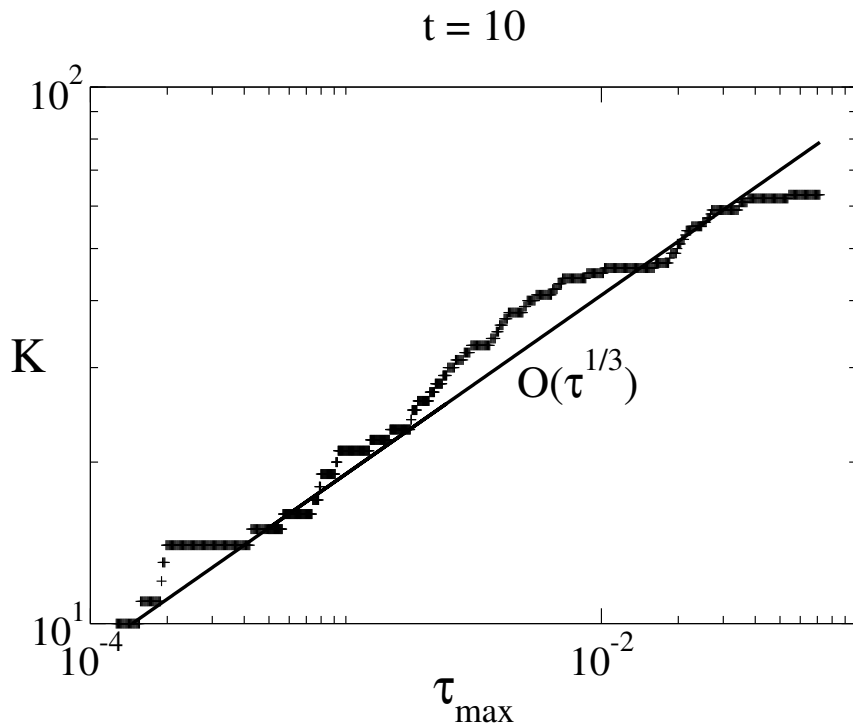


Figure 9: **Coarsening slow-down.** Although implicit schemes have no time step requirements for stability, errors due to large time steps have a significant effect on the coarsening rate. Using the initial data given by (48), we set $N = 200$ and plot K at $t = 10$ as a function of the maximal time step used in (50) (this is the time step used except at the earliest stages of evolution, where τ may be smaller to ensure convergence of the Newton iterations).

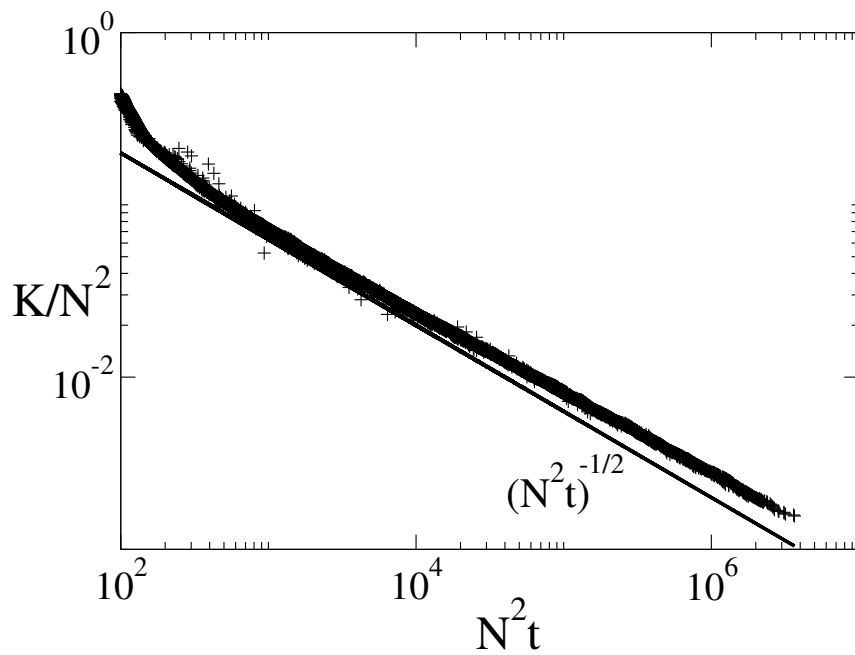


Figure 10: A comparison of the evolving jump density $\frac{K}{N^2}$ with the expected coarsening rate (63) for the two-dimensional system (51).

References

- [1] H. Amann. Time-delayed perona-malik type problems. *Preprint*, 2006.
- [2] A. Belahmidi and A. Chambolle. Time-delay regularization of anisotropic diffusion and image processing. *M2AN*, 39(2):231–251, 2005.
- [3] G. Bellettini and G. Fusco. A regularized perona-malik functional: some aspects of the gradient dynamics. In H. Broer, J. Mahwin, A. Vanderbauwhede, and S. V. Lunel, editors, *EQUADIFF 2003 Proceedings of the International Conference on Differential Equations*, Hasselt, Belgium, July 2003.
- [4] G. Bellettini and G. Fusco. The gamma limit and the related gradient flow for singular perturbation functionals of perona-malik type. *Preprint*, 2005.
- [5] G. Bellettini, M. Novaga, and E. Paolini. Global solutions to the gradient flow equation of a nonconvex functional. *To appear in SIAM J. Math. Anal.*, 2005.
- [6] F. Catte, P.-L. Lions, J.-M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal.*, 29(1):182–193, 1992.
- [7] Sergio Conti, Barbara Niethammer, and Felix Otto. Coarsening rates in off-critical mixtures. *SIAM J. Math. Anal.*, 37(6):1732–1741 (electronic), 2006.
- [8] S. Dai and R. L. Pego. Universal bounds on coarsening rates for mean-field models of phase transitions. *SIAM J. Math. Anal.*, 37(2):347–371 (electronic), 2005.
- [9] S. Dai and R. L. Pego. An upper bound on the coarsening rate for mushy zones in a phase-field model. *Interfaces Free Bound.*, 7(2):187–197, 2005.
- [10] S. Esedoğlu. An analysis of the Perona-Malik scheme. *Comm. Pure Appl. Math.*, 54(12):1442–1487, 2001.
- [11] D. Horstmann, K. J. Painter, and H. G. Othmer. Aggregation under local reinforcement: from lattice to continuum. *European J. Appl. Math.*, 15(5):546–576, 2004.
- [12] B. Kawohl and N. Kutev. Maximum and comparison principle for one-dimensional anisotropic diffusion. *Math. Ann.*, 311:107–123, 1998.
- [13] S. Kichenassamy. The perona-malik paradox. *SIAM J. Appl. Math.*, 57(5):1328–1342, 1997.
- [14] R. V. Kohn and F. Otto. Upper bounds on coarsening rates. *Comm. Math. Phys.*, 229(3):375–395, 2002.
- [15] R. V. Kohn and X. Yan. Upper bound on the coarsening rate for an epitaxial growth model. *Comm. Pure Appl. Math.*, 56(11):1549–1564, 2003.
- [16] R. V. Kohn and X. Yan. Coarsening rates for models of multicomponent phase separation. *Interfaces Free Bound.*, 6(1):135–149, 2004.
- [17] M. Nitzberg and T. Shiota. Nonlinear image filtering with edge and corner enhancement. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14:826–833, 1992.
- [18] F. Otto, T. Rump, and D. Slepčev. Coarsening rates for a droplet model: Rigorous upper bounds. *SIAM J. Math. Anal.*, 2006. To appear.
- [19] K. J. Painter, D. Horstmann, and H. G. Othmer. Localization in lattice and continuum models of reinforced random walks. *Appl. Math. Lett.*, 16(3):375–381, 2003.
- [20] L. Perko. *Differential equations and dynamical systems*, volume 7 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edition, 1996.
- [21] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. Technical report, Dept. of EECS Technical Report, U.C. Berkeley, 1987.
- [22] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern. Anal. Machine Intell.*, 12:629–639, 1990.

- [23] W. H. Press, S.A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C*. Cambridge University Press, Cambridge, second edition, 1992. The art of scientific computing.
- [24] M. Shearer, D. G. Schaeffer, and T. P. Witelski. Stability of shear bands in an elastoplastic model for granular flow: the role of discreteness. *Math. Models Methods Appl. Sci.*, 13(11):1629–1671, 2003.
- [25] T. P. Witelski, D. G. Schaeffer, and M. Shearer. A discrete model for an ill-posed nonlinear parabolic PDE. *Phys. D*, 160(3-4):189–221, 2001.
- [26] K. Zhang. Existence of infinitely many solutions for the one-dimensional perona-malik model. *Calc. Var. Partial Differential Equations*, 26(2):171–199, 2006.