CONVERGENCE ANALYSIS OF THE GRAPH ALLEN-CAHN SCHEME

XIYANG LUO, ANDREA L. BERTOZZI

Abstract. Graph partitioning problems have a wide range of applications in machine learning. This work analyzes convergence conditions for a class of diffuse interface algorithm [A.L. Bertozzi and A. Flenner, *Multiscale Modeling & Simulation*, 10(3):10901118, 2012.] for binary and multi-class partitioning. Using techniques from numerical PDE and convex optimization, convergence and and monotonicity is shown for a class of schemes under a graph-independent timestep restriction. We also analyze the effects of spectral truncation, a technique used to save computation cost. Convergence of the scheme with spectral truncation is also proved under a timestep restriction inversely proportional to the size of the graph. Moreover, this restriction is shown to be sharp in the worst case. Various numerical experiments are done to compare theoretical results with practical performance.

Key words.

AMS subject classifications.

1. Introduction. Graph cut methods have been used widely used in data clustering and image segmentation. Recently, the reformulation of of graph cut problem in non-local total variation (TV) has lead to various fast approximations to the problem. In particular, a method inspired by the diffuse interface model in PDE was proposed [6]. This line of method has been applied to various applications in clustering, image segmentation, and image inpainting [20] [14] [16].

Diffuse interface models are often built around the Ginzburg-Landau functional in Euclidean space. Recall

$$GL(u) = \frac{\epsilon}{2}\int |\nabla u|^2 + \frac{1}{\epsilon}\int W(u(x))dx. \tag{1.1}$$

Where W is the double-well potential $W(u) = \frac{1}{4}(u^2-1)^2$, with minimizers 1 and -1. The first term is the H^1 semi-norm of the function u, which penalizes non-smoothness of u. The parameter ϵ controls the scale of the diffuse interface, namely, the sharpness of the transition between two phases.

The Ginzburg-Landau functional and TV is related through the notion of gamma-convergence:

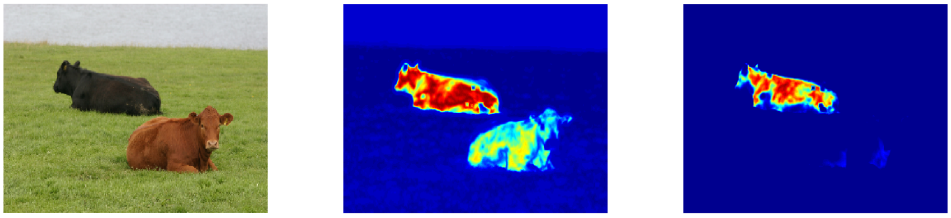
$$\lim_{\epsilon\to 0} GL_\epsilon(u) \to_\gamma CTV(u).$$

Evolution by the Ginzburg-Landau functional has been used to model dynamics of two phases in material science. The most common of which is the Allen-Cahn equation, the L^2 gradient flow on the Ginzburg-Landau functional. Another commonly used model is the Cahn-Hilliard equation, which is the H^{-1} gradient flow, both of which has been used in traditional image segmentation problems [3] [4]. Under suitable rescaling, the Allen-Cahn equation has been shown to converge to motion by mean curvature [26], [18], and thus fast numerical solvers for motion by mean curvature such as the MBO scheme [21] could be utilized to further approximate the Allen-Cahn equation.

A central notion in making the jump from PDEs to graphs is the Graph Laplacian. Graph Laplacians have been widely used in spectral clustering [28]. They share many similar characteristics with the continous Laplacian and can be shown to converge

to continuum limits under suitable assumptions [29]. Given a weight matrix W that is symmetric and has non-negative entries, one can construct three different types of graph Laplacians, namely,

1. Unnormalized : $L = D - W$
2. Random Walk : $L^{rw} = I - D^{-1}W$
3. Symmetric : $L^s = I - D^{-1/2}WD^{-1/2}$



(a) Original Image (b) Symmmetric Laplacian (c) Unnormalized Laplacian

Fig. 1.1: Comparison of Third Eigenvector of Graph Laplacians

where D is the diagonal matrix $d_{ii} = \sum_i w_{ij}$. The unnormalized graph Laplacian has the following nice formulae for its Dirichlet energy, which is analogous to that of the continuum Laplacian.

$$\frac{1}{2}\langle u, Lu\rangle = \frac{1}{2}\sum_{ij} w_{ij}(u_i - u_j)^2 \tag{1.2}$$

The random walk Laplacian L^{rw} has probabilistic interpretations, and can deal with outliers nicely [15]. The symmetric Laplacian L^s shares the same eigenvectors with L^{rw}, but is easier to deal with computationally being symmetric. In particular, the unnormalized Laplacian and the normalized Laplacian have very distinct eigenvectors, as can be seen from the visualization in Fig.1.1. The visualization is a plot of the third eigenvector for the nonlocal means graph formed from neighborhood patches of each pixel (see [6] for details). We shall discuss all three Laplacians whenever there is a distinction to be made.

In our proof, the only restrictions imposed on the weight matrix W are symmetry and non-negativity. In particular, we do not require triangle inequality. Under this assumption, a useful characterization of an unnormalized Laplacian is given by:

PROPOSITION 1.1 (Characterization of the Unnormalized Graph Laplacian). *A matrix L can be constructed as an unnormalized graph Laplacian from some valid weight matrix W if and only if L satisfies*

$$\begin{cases} L_{ij} = L_{ji} \\ L_{ij} \leq 0, i \neq j \\ L_{ii} = -\sum_{i \neq j} L_{ij} \end{cases} \tag{1.3}$$

We define the Ginzburg-Landau energy on graphs by replacing the spatial Laplacian with the graph Laplacian L

$$GL(u) = \frac{\epsilon}{2}\langle u, Lu\rangle + \frac{1}{\epsilon}\sum_i W(u_i). \tag{1.4}$$

The Allen-Cahn equation on graphs is defined as the gradient flow of the graph Ginzburg-Landau energy

$$u_t = -\nabla GL(u) = -\epsilon Lu - \frac{1}{\epsilon}W'(u). \tag{1.5}$$

In [6], a semi-implicit numerical scheme was used to counter the ill-conditioning of the graph Laplacian

$$\frac{u^{k+1}-u^k}{dt} = -\epsilon Lu^{k+1} - \frac{1}{\epsilon}W'(u^k). \tag{1.6}$$

Moreover, a convex penalty $\frac{c}{2}u^2$ is suggested to be added to form a "convex-concave" split, this gives us the actual scheme in [6]

$$\frac{u^{k+1}-u^k}{dt} = -\epsilon Lu^{k+1} - cu^{k+1} + cu^k - \frac{1}{\epsilon}W'(u^k). \tag{1.7}$$

Convex-concave splitting originated in an unpublished paper by Eyre [10], and has been used to resolve long-time solution of the Cahn-Hilliard equation in [27] [5]. However, for the particular scheme (1.7), we show in the next proposition that it's equivalent with (1.6) and thus we henceforth only analyze the scheme without convex splitting (1.6).

LEMMA 1.2 (Rescaled Timestep). *The scheme (1.7) could be rewritten as*

$$u^{k+1} - u^k = -\frac{dt}{1+cdt}\epsilon Lu^{k+1} - \frac{dt}{1+cdt}\frac{1}{\epsilon}W'(u^k). \tag{1.8}$$

Hence (1.7) is equivalent with (1.6) under a rescaling of stepsize.

As in the continuous case, the global minimizer of the Ginzburg-Landau functional is always the constant function 1 or -1, and hence the Allen-Cahn equation will eventually evolve to a constant. Thus in practice, a fidelity term $f(\phi(u), u)$ is often added for e.g. semi-supervised learning. Since the Ginzburg-Landau functional is a smoothed version of TV, this model is reminiscent of the classical ROF model [22] used in imaging when the fidelity term is quadratic. In the next sections, we analyze the scheme (1.6) without fidelity first, and later incorporate the fidelity term in the analysis.

2. Maximum Principle-L^∞ Estimates. The main result for this section is the following:

PROPOSITION 2.1 (A Priori Boundeness). *Define u^k by (1.6). Assume $\|u^0\|_\infty \leq 1$. If $dt \leq 0.5\epsilon$, then $\forall k$, $\|u^k\|_\infty \leq 1$.*

What is notable is that the timestep restriction is independent of the graph, i.e., this universal bound is guaranteed to work for *any* graph of any size. The constant $.5\epsilon$ is analogous to the ODE stiffness condition(see [24]). To prove the result, we need the maximum principle on graphs.

2.1. Maximum Principle. The classical maximal principle argument relies on the fact that $\Delta u(x_0) \geq 0$ for x_0 a local minimizer. This fact is also true for graphs.

PROPOSITION 2.2 (Second Order Condition on Graphs). *Let u be a function defined on a graph, and L be the unnormalized graph Laplacian . Any local minimizer i of u would satisfy $Lu|_i \leq 0$, where a local minimizer i is defined as $u_i \geq u_j, \forall w_{ij} > 0$*

Proof. Let i be a local minimizer. Then

$$\begin{aligned} Lu|_i &= d_i u_i - \sum_{j \neq i} w_{ij} u_j \\ &= \sum_{j \neq i} w_{ij}(u_i - u_j) \leq 0. \quad \square \end{aligned} \tag{2.1}$$

Next, we prove a discrete analogue of the continuous time maximum principle, which states that the implicitly discretized scheme for the heat equation on graphs is decreasing in the L_∞ norm. This line of thought is inspired by maximum principle for finite difference operators [9].

PROPOSITION 2.3 (Maximum Principle for Discrete Time). *For any $dt \geq 0$, let u be a solution to*

$$u = -dt * (Lu) + v \tag{2.2}$$

Then we have $\|u\|_\infty \leq \|v\|_\infty$.

Proof. Suppose i is a maximum of u. Then since $u(i) = dt * (-Lu)(i) + v(i)$ and $(-Lu)(i) \leq 0$, we have $max_u = u(i) \leq v(i) \leq max_v$. Arguing similarly with the minimum, we have that $\|u\|_\infty \leq \|v\|_\infty$. ∎

One thing to note is that the condition $L_{ij} \leq 0, i \neq j$ is *crucial* for the above analysis to hold. If L were replaced by a general positive-semidefinite matrix, we would still have an L^2 version of the proposition, namely, $\|u\|_2 \leq \|v\|_2$, but the L^∞ version would not be true.

2.2. A Priori Boundedness. We use the maximum principle to prove a priori boundedness. Consider splitting the one-line scheme (1.6) into two parts.

$$\begin{cases} v^k = u^k - dt * \dfrac{1}{\epsilon} W'(u^k) \\ u^{k+1} = -dt * (\epsilon L u^{k+1}) + v^k \end{cases} \tag{2.3}$$

Note that by maximum principle, $\|u^{k+1}\|_\infty \leq \|v^k\|_\infty$. We will need the lemma below to control the L^∞ norm of the first line in (2.3).

LEMMA 2.4. *Define the map $\mathcal{F}_{dt}(x) = x - dtW'(x)$. If $dt < 0.5$, $\mathcal{F}_{dt}$ maps $[-1, 1]$ to itself.*

Proof. Note $\mathcal{F}_{dt}(\pm 1) = \pm 1, \forall dt$. Moreover, $\mathcal{F}'_{dt}$ is quadratic and its two zeros $r_1(dt), r_2(dt)$ are a strictly increasing/decreasing function of dt respectively. Thus $\mathcal{F}_{dt}$ will map $[-1, 1]$ to itself if and only if $r_1(dt) \leq -1 \leq 1 \leq r_2(dt)$. Thus plugging in $\mathcal{F}'_{dt}(1) = 0$ gives $dt < 0.5$. ∎

Note that $\mathcal{F}_{dt}(x)$ here is the component-wise map of the first line in (2.3) and is the source of the stepsize restriction. Since estimates for $\mathcal{F}_{dt}(x)$ will follow the same idea as above and will involve only elementary calculations, we omit some of the details later.

Next, we prove our main conclusion below:

Proof. [Proof of Proposition 2.1] We prove by induction. Suppose $\|u^k\|_\infty \leq 1$. Then for each vertex i, we have $v^k(i) = u^k(i) - \frac{dt}{\epsilon} W'(u^k(i)) = \mathcal{F}_{dt/\epsilon}(u^k(i))$. By lemma 2.4, $|v^k(i)| < 1, \forall i$, given that $\frac{dt}{\epsilon} < 0.5$. Hence $\|v^k\|_\infty \leq 1$. Then by proposition 2.3, $\|u^{k+1}\|_\infty \leq \|v^k\|_\infty \leq 1$. ∎

If we are not so keen on keeping $u^k(i)$ in $[-1, 1]$ for each iteration, but merely care about whether the scheme is bounded or not, then we may relax the interval a bit to get a larger range of dt, as the next lemma shows.

LEMMA 2.5. *For $dt < 2.1$, $\mathcal{F}_{dt}$ maps $[-1.4, 1.4]$ to itself. Hence if $\|u^0\|_\infty < 1.4$, then $\{u^k\}$ is bounded for $dt < 2.1\epsilon$.*

Derivation of the constant: The constants are found using a computer program to find largest dt for which some $[-c, c]$ to itself. One could then prove the lemma by a brute force calculation.

For future reference, the 0.5ϵ bound will be called the "tight bound" where the 2.1ϵ bound will be called the "loose bound". Again, we must emphasis that the exact constants here do not matter so much as the fact that they're independent of the graph.

2.3. Variations of the Scheme. In this section, we prove boundedness results for both the random walk Laplacian and the symmetric graph Laplacian.

PROPOSITION 2.6 (Random Walk Graph Laplacian). *Let $\|u^0\|_\infty \leq 1$. If $dt < 0.5\epsilon$, the scheme (1.6) with $L = L^{rw}$ satisfies $\|u^k\|_\infty \leq 1, \forall k$.*

Proof. The proof could be copied verbatim by noting that the second order condition 2.2 holds also for random walk Laplacian L^{rw}. ∎

The case on symmetric graph Laplacian is a little different, and a uniformity condition on the graph must be added.

PROPOSITION 2.7 (Symmetric Graph Laplacian). *Let $\rho = \frac{\max_i d_i}{\min_i d_i}$. If $\rho \leq 2$, $\|u^0\|_\infty \leq \frac{1}{\sqrt{2}}$, $dt \leq 0.5\epsilon$, then the sequence $\{u^k\}$ is bounded.*

Proof. We note that while L^s_{ij} is no longer negative (thus losing the maximal principle), we still have the relation $L_s = D^{1/2} L_{rw} D^{-1/2}$. Thus line 2 of (2.3) becomes

$$D^{-1/2} u^{k+1} = -dt * L_{rw} D^{-1/2} u^{k+1} + D^{-1/2} v^k. \tag{2.4}$$

By maximum principle, $\|D^{-1/2} u^{k+1}\|_\infty \leq \|D^{-1/2} v^k\|_\infty$. Let $\tilde{u}^k = (\min_i d_i)^{1/2} D^{-1/2} u^{k+1}$, the first line of (2.3) becomes

$$\tilde{v}^k(i) = \frac{1}{c_i} \mathcal{F}_{dt/\epsilon}(c_i \tilde{u}^k(i)),$$

where $c_i = (\frac{d_i}{\min_j d_j})^{1/2}$. Note that by our assumptions on the graph, $c_i \in [1, \sqrt{2}], \forall i$.

Next, we prove $\|\tilde{u}^k\|_\infty \leq 1$ by induction. This is clearly true for $k = 0$ since $\|\tilde{u}^0\|_\infty \leq \sqrt{\rho} \|u^0\|_\infty \leq 1$. For general k, define $\Phi_i(x) = \frac{1}{c_i} \mathcal{F}_{dt/\epsilon}(c_i x)$ to be a rescaled version of the forward step. We claim Φ_i maps $[-1, 1]$ to itself for all i. This follows easily from the following lemma:

LEMMA 2.8 (Uniform Scaling). *For any $\gamma < 0.5$, $\mathcal{F}_\gamma$ maps $[-c, c]$ to itself for any $c \in [1, \sqrt{2}]$.*

Thus we get $\tilde{v}^k(i) \in [-1, 1]$. Applying maximum principle and equation (2.4), we finally get $\|\tilde{u}^{k+1}\|_\infty \leq 1$ and complete the induction. ∎

In previous work [6], [14], a fidelity term ϕ^0 was added to the scheme, where $\phi^0(i) \in \{1, -1\}$, for i belonging to a fidelity set Λ. We show boundedness of the graph Allen-Cahn scheme with fidelity. Restating from [6] the scheme with fidelity:

$$\begin{cases} v^k = u^k - dt * (\frac{1}{\epsilon} W'(u^k) + \eta\Lambda(u^k - \phi^0)) \\ u^{k+1} = -dt * (\epsilon L u^{k+1}) + v^k \end{cases} \tag{2.5}$$

PROPOSITION 2.9 (Graph Allen-Cahn with fidelity). *The graph Allen-Cahn scheme with fidelity (2.5) is bounded by 2 for $dt < \frac{1}{2+\eta}\epsilon$, for some c independent of the graph.*

Proof. Define the modified forward step $\Phi_{dt}(u) = u - dt[(u^2-1)u + \eta(u-1)]$. By the same argument as before, the limit stepsize is given by $\Phi'_{dt}(1) = 0$. Thus $dt = \frac{1}{2+\eta}\epsilon$. □

2.4. Graph Independent Convergence to ODE. As an interesting corollary of the Maximum Principle, we shall prove that the discrete scheme u^k converges to the continuous equation in a graph independent way.

PROPOSITION 2.10 (Convergence to ODE). *Let U be the continuous solution of (1.5) defined on a fixed interval $[0,T]$. Assume $\|U(.,0)\|_\infty \leq 1$. Set $U^k = U(.,kdt)$ Define $e^k = u^k - U^k$ to be the global error. Then $\|e_k\| \leq Mdt = O(dt)$, where M is independent of dt and the graph Laplacian L.*

Proof. Note that if we assume $\|u^0\|_\infty \leq 1$, we have that u^ks are bounded. Since U is a gradient flow, $GL(U(.,t))$ is decreasing in time and thus U is also bounded. Therefore, we may assume W'' to be bounded (by a graph-free constant). In the following notation, M and the constant for big O is graph-independent. We first compute the local truncation error:

LEMMA 2.11. *Define the local truncation error τ^k as below:*

$$\frac{U^{k+1} - U^k}{dt} = -LU^{k+1} - W'(U^k) + \tau^k. \tag{2.6}$$

Then $\|\tau^k\|_\infty \leq Mdt$ Subtract (2.6) from the discrete scheme, and we get an evolution equation for the error term

$$\frac{e^{k+1} - e^k}{dt} = Le^{k+1} - (W'(u^k) - W'(U^k)) + \tau^k \tag{2.7}$$

By using Taylor expansion on W', we have

$$e^{k+1} = -dtLe^{k+1} - dtW''(\xi)e^k + e^k + \tau^k \tag{2.8}$$

Here W'' is the diagonal matrix $W''_{jj} = W''(\xi_j)$. By applying the maximum principle, taking L^∞ norm gives us

$$\|e^{k+1}\|_\infty \leq \| - dtW''(\xi)e^k + e^k + \tau^k\|_\infty \leq (1 + Mdt)\|e^k\|_\infty \tag{2.9}$$

for some graph independent constant M.

Hence by iterating the equation above, we get

$$\|e^k\|_\infty \leq (Me^{MT})dt = O(dt). \quad \square$$

3. Energy Method-L^2 Estimates. In this section, we take a different view and do estimates in terms of the L^2 norm. We will use this to prove convergence to stationary point for the full scheme and also convergence of scheme with spectral truncation.

Energy estimates for the semi-implicit scheme has been done in [30] [10] in terms of a *convex-concave* splitting. In an unpublished script that is frequently cited [10] , Eyre proved:

PROPOSITION 3.1 (Eyre). *Let $E = E_1 + E_2$ be a splitting with E_1 convex and E_2 concave. Then for any dt, the semi-implicit scheme $u^{k+1} = u^k - dt\nabla E_1(u^{k+1}) - dt\nabla E_2(u^k)$ is monotone in E, namely,*

$$E(u^{k+1}) \leq E(u^k), \quad \forall k \geq 0.$$

The semi-implicit scheme in fact coincides with the notion of proximal gradient method for minimizing the splitting $E = E_1 + E_2$. Recall that proximal gradient iteration is given as

$$u^{k+1} = Prox_{dtE_1}(u^k - dt\nabla E_2(u^k)), \tag{3.1}$$

where the $Prox$ operator is defined as $Prox_{\gamma f}(x) = argmin_u\{f(u)+\frac{1}{2\gamma}\|u-x\|^2\}$. The $Prox$ operator is defined for all proper convex functions taking extended real values, and its connection to the semi-implicit scheme is clear from its implicit gradient interpretation. Namely, if $y = Prox_{\gamma f}(x)$,

$$y = x - \gamma\partial f(y). \tag{3.2}$$

∂f is the *subgradient* of f, which coincides with the gradient if f is differentiable. Classical results for convergence of proximal gradient method can be found in [7],which requires both E_1, E_2 to be convex (instead of E_2 concave as in Eyre), and ∇E_2 be Lipschitz continuous.

In our case, $E = GL(u)$, $E_1 = \frac{\epsilon}{2}\langle u, Lu\rangle$ and $E_2 = \frac{1}{4\epsilon}\sum_i W(u_i)$. However, our E_2 is neither concave nor convex. Instead of using quadratic penalty to force convex-concavity (which is shown to be a rescaling of time in (1.2)) we will use a result that deals with non-convex proximal gradient method. The estimate below is a simplified version of the proof in [23], and is analogous to that in [10]. General discussions of non-convex proximal gradient is found in [1].

PROPOSITION 3.2 (Energy Estimate). *Let $E = E_1 + E_2$. Define x^{k+1} by $x^{k+1} \in x^k - dt\partial E_1(x^{k+1}) - dt\nabla E_2(x^k)$. Suppose E_1 is convex, E_2 smooth and ∇E_2 Lipschitz continuous with Lipschitz constant M, we have*

$$E(x^k) - E(x^{k+1}) \geq (\frac{1}{dt} - \frac{M}{2})\|x^{k+1} - x^k\|^2. \tag{3.3}$$

Proof.

$$\begin{aligned}
E(x^k) - E(x^{k+1}) &= E_1(x^k) - E_1(x^{k+1}) + E_2(x^k) - E_2(x^{k+1}) \\
&\geq \langle\partial E_1(x^{k+1}), x^k - x^{k+1}\rangle + E_2(x^k) - E_2(x^{k+1}) \\
&= \frac{1}{dt}\|x^{k+1} - x^k\|^2 + \langle\nabla E_2(x^k), x^k - x^{k+1}\rangle + E_2(x^k) - E_2(x^{k+1}) \\
&\geq \frac{1}{dt}\|x^{k+1} - x^k\|^2 - \frac{M}{2}\|x^{k+1} - x^k\|^2
\end{aligned}$$

The second line is by convexity of E_1, the third by plugging in the the definition of x^{k+1}, the fourth by simple Taylor expansion of the function E_2. □

Even though the proof is simple, we have the freedom of choosing E_1 to be a general non-smooth convex function. In particular, we shall later set $E_1(u) = I_V(u) + \frac{\epsilon}{2}\langle u, Lu\rangle$, where I is the indicator function and V an eigenspace.

3.1. Convergence of Scheme. In this section, we use the estimate 3.2 to extend our result of boundedness to convergence under the same stepsize restriction.

PROPOSITION 3.3 (Convergence of Graph Allen-Cahn). *Let $\|u^0\|_\infty \leq 1$. Under the strict bound $dt < 0.5\epsilon$, the scheme (1.6) converges to a stationary point and is monotone in the Ginzburg-Landau energy $GL(u) = \frac{\epsilon}{2}\langle u, Lu\rangle + \sum_i W(u_i)$.*

Proof. From proposition 2.1, $dt < 0.5\epsilon$, implies $\|u^k\|_\infty \leq 1, \forall k$. Since the L^∞ ball in $\mathbb{R}^n$ is convex, all points ξ that lie in the line segment from u^k to u^{k+1} satisfy $\|\xi\|_\infty \leq 1$. Thus we can WLOG assume the Lipschitz constant M of ∇E_2 to satisfy $M \leq \frac{1}{\epsilon}\max_{|x|_\infty \leq 1}|W''(x)| = \frac{2}{\epsilon}$. Since $dt < 0.5\epsilon < \frac{2}{M}$, by summing up the inequality in (3.2) and cancelling the telescopic series on the left, we have

$$GL(u^0) - GL(u^n) \geq c\sum_{i\leq n}\|u^i - u^{i-1}\|^2, \tag{3.4}$$

where $c = \frac{1}{dt} - \frac{M}{2} > 0$. Since $GL(u^n) \geq 0$, we have that the sequence $\{u^i - u^{i-1}\}$ is square summable, thus $\lim_{i\to\infty}\|u^i - u^{i-1}\| = 0$. Since $\{u^i\}$ is bounded, there exists a limit point u^* for the sequence, and u^* in unique by the condition $\lim_{i\to\infty}\|u^i - u^{i-1}\| = 0$. Hence $\{u^i\}$ converges to u^*. By continuity, u^* is a stationary point, i.e., $\nabla GL(u^*) = 0$. □

Remark: The argument does not work for the "loose bound" $dt < 2.1\epsilon$, since the Lipschitz restriction $\frac{2}{M}$ will no longer be greater than 2.1ϵ. There indeed exists cases where u^ks are bounded but not convergent, namely, when they cycle around. However, in practice the scheme tends to converge under larger timestep than the tight bound 0.5ϵ.

Note that by section 2, boundedness can be shown for L^{rw}, L^s that satisfy a condition, and for the scheme with a fidelity term. Hence modifying the constants if necessary, we immediately have:

THEOREM 3.4 (Main Convergence Result). *There exists c independent of the graph such that $\forall dt < c\epsilon$, the scheme (1.6), with or without fidelity, with the graph Laplacian equal to L, L^{rw} or L^s satisfying uniformity condition, converges.*

3.2. Convergence Results Directly from the Energy Estimate. In this section, we digress from our analysis on the graph Allen-Cahn scheme and instead study the scheme (1.6) where L is replaced by an *arbitrary* symmetric semi-positive definite matrix. Note that we would still have the energy estimate (3.2), but not the maximal principle. We will show that by using *only* (1.6), we are still able to achieve convergence.

There are two reasons for generalizing L to an arbitrary semi-positive definite matrix. First, this serves as a baseline convergence result if we're to study other schemes (e.g., the Cahn-Hilliard equation or if we are to analyze approximations such as the Nystrom extension). The second reason is that the proof here can be carried over almost verbatim to show convergence of (1.6) under spectral truncation. Our main result is below:

THEOREM 3.5. *Let L be an arbitrary symmetric and semi-positive definite matrix such that $\rho_L \leq C$ for some C independent of N, where $\rho_L = \max_i|\lambda_i|$. Define u^k by*

the scheme (1.6). Suppose $\|u^0\|_\infty \leq 1$, then the scheme is monotone in the Ginzburg-Landau energy for timestep $dt = O(N^{-1})$, where N is the size of the system.

Since the result is an analysis on dt vs N, we allow the constants to depend on ϵ, and WLOG set $\epsilon = 1$ in the proof.

Our strategy here is to choose dt so small and apply proposition 3.2 to force *monotonicity* in the Ginzburg-Landau functional GL. Since $GL(u) = O(\|u\|^4)$, boundedness in function value will imply boundedness in the variable u. For our purpose, we will need a more refined version of this argument.

LEMMA 3.6 (Inverse Bound). *Let M be any positive constant. If $GL(u) \leq M$, then $\|u\|_2^2 \leq N + \sqrt{NM}$, where N is the dimension of u.*

Proof. By assumption, $\sum_i (u_i^2 - 1)^2 \leq M$. Then from Cauchy-Schwarz, $\sum_i (u_i^2 - 1) \leq \sqrt{NM}$, hence our lemma. □

To prove the theorem, we will need a direct estimate on the norm of u^{k+1} from u^k. The proof of the lemma is an application of norm conversions in 7.1 in the Appendix.

LEMMA 3.7. *Let u^k and u^{k+1} be successive iterates of scheme (1.6). Then their function value satisfies the inequality below:*

$$\|u^{k+1}\|_2 \leq (1 + dt)\|u^k\|_2 + dt\|u^k\|_2^3. \tag{3.5}$$

Proof. Since L is symmetric semi-positive definite, we have $\|u^{k+1}\|_2 \leq \|v^k\|_2$. Then since $v^k(i) = u^k(i) - dt * [u^k(i)(u^k(i)^2 - 1)]$, we have $\|v^k\|_2 \leq (1 + dt)\|u^k\|_2 + dt\|u^k\|_6^3 \leq (1 + dt)\|u^k\|_2 + dt\|u^k\|_2^3$ □

Proof. [Proof of Proposition 3.5.] Since $\|u^0\|_\infty \leq 1$, we have $\|u^0\|_2 \leq \sqrt{N}$. Moreover, $GL(u^0) \leq \rho_L \|u^0\|_2^2 + \frac{1}{4\epsilon} \sum_i 1 \leq C_1 N$. By lemma 3.6, for any v s.t. $GL(v) \leq GL(u^0)$, we have $\|v\| \leq C_2\sqrt{N}$, $C_2 \geq 1$.

We claim that there exists stepsize $dt \leq \delta N^{-1}$, such that the following holds for all k

$$\begin{aligned} GL(u^k) \leq & GL(u^0) \leq C_1 N \\ \|u^k\|_2 \leq & C_2\sqrt{N} \end{aligned} \tag{3.6}$$

We argue by induction. The case $k = 0$ has been proved above. Suppose this is valid for k, then we have $\|u^k\| \leq C_2\sqrt{N}$. Allowing an abuse of notation in C, we have $\|u^{k+1}\| \leq C(1 + dt)N^{1/2} + CdtN^{3/2}$ by lemma 3.7. If we choose $dt \leq \delta N^{-1}$, δ small ,we have $\|u^{k+1}\| \leq CN^{1/2}$.

The crux of the proof is applying the energy estimate (3.2) to u^{k+1}. Note the estimate (3.2) is valid with constant $M \leq \max_{\|\xi\|_\infty \leq R} \|\nabla^2 W(\xi)\|, R \leq CN^{1/2}$. Simplifying terms, $M \leq CN$. Thus by further adjusting δ if necessary, we have for $dt < \delta N^{-1}$, $GL(u^{k+1}) \leq GL(u^k) \leq GL(u^0)$. However, this would mean that $GL(u^{k+1}) \leq C_1 N$, and thus by the inverse bound lemma 3.6, $\|u^k\|_2 \leq C_2\sqrt{N}$. This completes the induction step. □

Remark: The convergence stepsize *is* graph-size dependent. However, we will show in the next section that the dependence in unavoidable if we're dealing with arbitrary L.

4. Analysis on Spectral Truncation. In many applications, the number of nodes N on a graph will be huge, and it will be infeasible to invert L every iteration in (1.6). In [6] [20], a strategy proposed was to project u onto the first m eigenvectors. This approach is called *spectral truncation*. In practice, spectral truncation gives accurate segmentation results but is computationally much cheaper.

4.1. Convergence Results for Spectral Truncation. Let us formally define the spectral truncated version of scheme (1.6). Define $V_m = span\{\phi^1, \phi^2, \ldots, \phi^m\}$ to be space spanned by one choice of m eigenvectors with the smallest eigenvalues. Define P_m to be the projection operator onto the space V_m. Then the spectral truncated scheme is defined as

$$\begin{cases} v^k = u^k - dt * \frac{1}{\epsilon} W'(u^k) \\ u^{k+1} = P_m[-dt * (\epsilon L u^{k+1}) + v^k] \end{cases} \tag{4.1}$$

Just as in the previous analysis, we first show boundedness.

PROPOSITION 4.1 (Boundedness of Spectral Truncation). *Let the graph Laplacian satisfy $\rho_L \leq C$ for some C independent of N. Define u^k by the scheme (4.1). Suppose $\|u^0\|_\infty \leq 1$, then the scheme is monotone in the Ginzburg-Landau energy for timestep $dt = O(N^{-1})$, where N is the size of the system.*

The key int the proof is the following observation, which links spectral truncation to the proximal gradient method.

LEMMA 4.2 (Reformulation of Spectral Truncation). *If $u^0 \in V_m$, the spectral truncated scheme (4.1) is equivalent to the proximal gradient scheme (3.1) with $E_1 = \frac{\epsilon}{2}\langle u, Lu\rangle + I_{V_m}$, $E_2 = \frac{1}{4\epsilon}\sum_i W(u_i)$, where I_{V_m} is the indicator function of the m-th eigenspace which is 0 inside V_m and $+\infty$ outside.*

Proof. Since the explicit step (line 1) is the same between the two schemes, we only have to show the following: Define u, u' as

$$\begin{aligned} u =& \operatorname*{argmin}_y \frac{\epsilon}{2}\langle y, Ly\rangle + \frac{1}{dt}\|y - v\|^2 \\ u' =& \operatorname*{argmin}_y \frac{\epsilon}{2}\langle y, Ly\rangle + I_{V_k}(y) + \frac{1}{dt}\|y - v\|^2 \end{aligned}$$

We want to show that u, u' satisfy the relation $P_m(u) = u'$. Project onto eigenvectors and we have $u = \sum_{i=1}^N c_i\phi^i$, $u' = \sum_{i=1}^N c_i'\phi^i$. Since $u' \in V_m$, we have $c_i' = 0, \forall i > m$. Moreover, letting $d_i = \langle v, \phi^i\rangle$,

$$\frac{\epsilon}{2}\langle u, Lu\rangle + \frac{1}{dt}\|u - v\|^2 = \sum_{i=1}^N (\frac{\epsilon}{2}\lambda_i c_i^2 + (c_i - d_i)^2)$$

Thus the minimization is done component-wise in c_i, and it's easy to see that $c_i = c_i', \forall i \leq m$. Thus $P_m(u) = u$. ∎

We may now prove our main result:

Proof. [Proof of Proposition 4.1] The proof is almost identical to that in proposition 3.5. We again argue inductively that (3.6) holds for $dt \leq \delta N^{-1}$.

Suppose (3.6) is true for k. Since the the projection operator P_k does not increase L^2 norm, we would still have $\|u^{k+1}\| \leq (1 + dt)\|u^k\| + dt\|u^k\|^3$, and thus $\|u^{k+1}\| \leq C\sqrt{N}$. By the equivalence of spectral truncation with proximal gradient (4.2), we may use the energy estimate (3.2), and argue that $GL(u^{k+1}) \leq GL(u^k)$ under such stepsize. This in turn will force $\|u^{k+1}\| \leq C_2\sqrt{N}$, ending the induction. ∎

Since we now know a priori that $\|u^k\|$ is bounded by $O(\sqrt{N})$, we can WLOG assume the Lipschitz constant $M = O(N)$. Thus by copying verbatim the proof of proposition 3.3, we have

PROPOSITION 4.3 (Convergence Result). *The truncated scheme is convergent under the stepsize restriction $dt \leq \delta N^{-1}$.*

4.2. A Counter Example for Graph-Independent Timestep Restriction. In the previous subsection, we proved that the spectral truncated scheme is bounded under stepsize restriction $dt = O(N^{-1})$. One would hope to achieve a graph-free stepsize rule as in the case of the full scheme (1.6). However, as we will show in our example below, uniform convergence stepsize over all graph Laplacian of all sizes is not possible.

PROPOSITION 4.4 (Sub-Optimality of Estimate 4.1). *For any $dt = \delta N^{-\alpha}, \alpha < 1$, we can always find a graph Laplacian $L_{N\times N}$ with $\rho_L \leq 1$, and an initial condition $\|u^0\|_\infty = 1$ such that the scheme in (4.1) with truncation level $m = 2$ has $\lim_{k\to\infty} \|u^k\|_\infty = \infty$. Hence our estimate for the stepsize restriction in (4.1) is sub-optimal.*

We will explicitly construct a graph and a graph Laplacian to attain the worst case bound. Graph construction is as follows, which is illustrated in Fig.4.1.

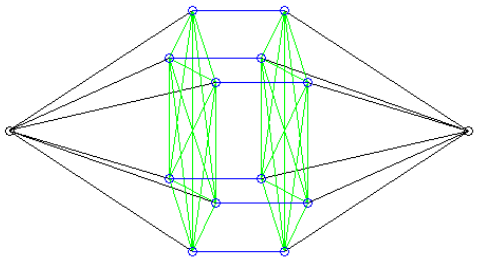


Fig. 4.1: Illustration of Limit Case Graph with $N = 7$

Construction of Graph

1. *Nodes*: The nodes consists of two "clusters" (blue nodes) and two outlier points (black nodes). Each cluster contains $N-1$ nodes and thus the graph contains a total of $2N$ nodes.

2. *Edge Weights*: Connect all nodes to each other within clusters and set edge weights to 1 (green edges). Connect the inter cluster nodes in a pairwise fashion and set weights to 0.1 (black edges). Finally, connect the outlier node with the clusters and set edge weights to $0.1/N$

3. *Indexing*: Nodes on the left will be indexed by odd numbers and nodes on the right even. The first and the second node will be the two outlier nodes respectively.

4. *Graph Laplacian*: The graph Laplacian is taken to be $\frac{1}{\rho_L}L$, where L is the unnormalized graph Laplacian $D - W$.

There is lots of freedom in choosing the edge weights, what's important is the ratio between the blue and the black edges must be close to N. We shall choose our initialization by thresholding the second eigenvector, namely, $u^0 = Sign(\phi^2)$. The key property of the graph is its second eigenvector, the computation of which could be found in 7.2 in the Appendix :

PROPOSITION 4.5. *Under the setup above, the second eigenvector of the graph*

Laplacian is

$$\phi^2 = \left(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2\sqrt{N}}, -\frac{1}{2\sqrt{N}}, \ldots, \frac{1}{2\sqrt{N}}, -\frac{1}{2\sqrt{N}}\right)^t.$$

Moreover, projection of u^0 *onto the first two eigenvectors satisfies* $P_2(u^0) = C\sqrt{N}\phi^2$, *where* C *is approximately* 0.5.

Next, we shall give a proof of our main conclusion. The idea is that after the first iteration, the values of u^1 on the outlier nodes will be high enough such that the scheme will diverge. For clarity, we omit simple calculations.

Proof. [Proof of proposition 4.4] We run through the scheme with $u^0 = Sgn(\phi^2)$, $dt = \delta N^{-\alpha}$.

(Step1) We compute u^1. Since u^0 is ± 1 valued, $v^0 = u^0$. By boundedness of eigenvalues, $u^1 = C_0\sqrt{N}\phi^2 = O(\sqrt{N})\phi^2$, since C_0 is bounded from below with respect to N.

(Step 2) We compute u^2. Note since $u^1(1) = -u^1(2) = O(\sqrt{N})$, and $dt \leq N^{-\alpha}$, we have $|v_1^1| = |(1-dt)u_1^1 + dt(u_1^1)^3| = O(\sqrt{N}) + O(N^{3/2}/N^{-\alpha}) = O(N^{\theta/2})$, where $\theta > 1$. Similarly, $v^1(j) = O(1)$ for $j \neq 2$. Moreover, by symmetry, $u^1(2k) = -u^1(2k+1)$, and hence we have $v^1 = O(N^{\theta/2})\phi^2 + O(N^{\frac{\theta-1}{2}})$. We have $u^2 = O(N^{\theta/2})\phi^2 + O(N^{\frac{\theta-1}{2}})$

(Step 3) Inductively, we can show that $u^k = O(N^{\frac{\theta^{k-1}}{2}})\phi^2 + O(N^{\frac{\theta^{k-1}-1}{2}})$. And thus letting $k \to \infty$, $u^k \to \infty$. ∎

Whether the theoretical worst case bound is attained if we truncate more than 2 eigenvectors is not proved here and could be done in future work. However, due to the freedom of constructing such graphs, the thought is that there are more complicated examples such that the bound is attained for truncation level $m > 2$.

4.3. Heuristic Explanation for Good Average Behavior. Despite of the theoretical worst case example given above, the timestep for spectral truncation does not depend badly on the size N of the problem. In this section, we attempt to give a heuristic explanation of this from two viewpoints.

First view is to analyze the projection operator P_m in the L^∞ norm. The reason why maximum principle fails is because P_m is not non-expansive in the L^∞ norm. Namely, for some vector $\|v\|_\infty \leq 1$, we have $\|P_m(v)\|_\infty = O(\sqrt{N})$ in the worst case. However, an easy analysis shows the probability of attaining such $O(\sqrt{N})$ bound decays exponentially as N grows large. And thus in practice, it is very rare that adding P_m would violate the maximal principle"too much".

The second view is to restrict our attention to data that come from a random sample. Namely, we assume our data points x^i are sampled i.i.d. from a probability distribution p, and that the graph Laplacian is computed from the Euclidean distance $\|x^i - x^j\|$.

In [29], it was proven that under very general assumptions, the discrete eigenfunctions, eigenvalues will converge to continuous limits almost surely. Moreover, the projection operators P_k converges compactly almost surely to their continuous limits. Moreover, results for continuous limits of graph-cut problems can be found in [25]. Under this set up, we can define the Allen-Cahn scheme on the continuous domain and discuss its properties on suitable function spaces. The spectral truncated scheme *still* wouldn't satisfy maximal principle, but at least it evolves in a sample-size independent fashion. Of course a rigorous proof would require heavy functional analysis.

5. Results for Multiclass Classification. The previous analysis can be carried over in a straight forward fashion to the multiclass Ginzburg-Landau.

Multiclass diffuse interface algorithm on graphs can be found in [13] [17] [20]. In most of the algorithms, the labels are vectorized into separate coordinates. To be more precise, given K the number of classes, and N the number of nodes on the graph, we define an $N \times K$ matrix $\boldsymbol{u}$, where each entry $\boldsymbol{u_{ij}}$ stands for the "score" of the ith node belonging to the jth class. Often one would project $\boldsymbol{u}$ onto the Gibbs simplex $\mathcal{G} = \{\sum x_i = 1 | x_i \geq 0\}$ [14] to make $\boldsymbol{u_{ij}}$ into a probability distribution.

The Ginzburg Landau functional for multiclass is defined as

$$GL(\boldsymbol{u}) = \frac{\epsilon}{2} tr(\boldsymbol{u} L \boldsymbol{u}) + \frac{1}{2\epsilon} \sum_{i=1}^{N} W(\boldsymbol{u_i}) \tag{5.1}$$

Here, $\boldsymbol{u_i}$ stands for the ith row of $\boldsymbol{u}$, and W should be a "multi-well" function that is analogous to the double-well in the binary case. The function should have local minima near the unit vectors $e_k = (0, 0, \ldots, 1, \ldots, 0)^t$, and grows fast when u is far from the origin. We will use the L^2 double well, namely,

$$W(\boldsymbol{u_i}) = (\Pi_{k=1}^{K} \|\boldsymbol{u_i} - e_k\|_2^2). \tag{5.2}$$

In [14], a different double well is used where L^1 norms are taken instead of L^2. The paper claimed that L^2 double well suffers from the problem that the function value of W in the center of the Gibbs simplex is small. This problem could be alleviated if we rescale distance by a suitable function $\rho(x)$. Namely, replacing $\|\boldsymbol{u_i} - e_k\|_2^2$ by $\rho(\|\boldsymbol{u_i} - e_k\|_2^2)$. Moreover, when k is reasonably small, even such adjustments are unnecessary. However, choosing the L^2 well comes with the bonus advantage that the problem is smooth. This gives better convergence guarantees as well as makes the problem easier to compute numerically. For example, the L^2 double well doesn't require projection onto the Gibbs simplex $\mathcal{G}$ in every iteration as in [14].

We minimize GL using the forward-backward method as in (2.3).

$$\begin{cases} \boldsymbol{v^k} = \boldsymbol{u^k} - dt * \frac{1}{2\epsilon} \sum_i \nabla W(\boldsymbol{u_i^k}) \\ \boldsymbol{u^{k+1}} = -dt * (\epsilon L \boldsymbol{u^{k+1}}) + \boldsymbol{v^k} \end{cases} \tag{5.3}$$

Since the diffusion step is done columns-wise, the maximum principle carries over naturally. Namely, we have

PROPOSITION 5.1 (Maximum Principle Multiclass). *Let $\boldsymbol{u}, \boldsymbol{v}$ be $K \times N$ matrices.*

$$\boldsymbol{u} = -dt * (\epsilon L \boldsymbol{u}) + \boldsymbol{v} \tag{5.4}$$

Then $\|u_j\|_\infty \leq \|v_j\|_\infty$, where u_j, v_j are the jth column of u, v respectively.

With the exact same reasoning as in the binary case, we need a range of stepsize dt for which the forward gradient step $\mathcal{F}_t$ of the "multi-well" maps $[-R, R]^{N \times K}$ onto itself, as the next lemma shows.

LEMMA 5.2. *Define $\mathcal{F}_t : \boldsymbol{u_i} \mapsto \boldsymbol{u_i} - dt * \frac{1}{2\epsilon} \nabla W(\boldsymbol{u_i})$. Then $\exists R(K)$ and $\exists c(R, K)$ independent of N such that for $dt < c(R, K)$, $\mathcal{F}_t([-R, R]^K) \subset [-R, R]^K$*

Proof. Since the double well W doesn't depend on N, the constants R and dt will then naturally be independent of N if we prove its existence.

We define a new map ϕ_{dt} to be $\phi_{dt}(R) = \sup\{\|F_{dt}(u)\|_\infty, u \in \partial[-R, R]^K\}$. It can be shown that ϕ_{dt} is continuous. Since the double well W is nearly quadratic

when R is large, we have that $\exists R_1$ such that $\nabla W|_{\partial[-R_2,R_2]^K}$ points inward of the box $[-R_2, R_2]^K$, for all $R_2 \geq R_1$. Thus we can find c dependent on R_2 such that $\phi_{dt}(R_2) \leq R_2, \forall R_2 \geq R_1, dt < c$. Take $R = \max \phi_c([0, R_1])$, by shrinking c if necessary, we have $\max \phi_c([0, R]) \leq R$, and thus $\mathcal{F}_t([-R, R]^K) \subset [-R, R]^K, \forall dt < c$. □

The proof works for any function that acts independently on each component $\boldsymbol{u_i}$ and has fast growth towards infinity. The estimates here are not as precise as the 0.5ϵ bound in the binary case, since an explicit calculation will be a rather complicated formulae that involves K. However, in practice, the stepsize restriction is also comparable to 0.5ϵ, at least when the number of clusters K is moderate.

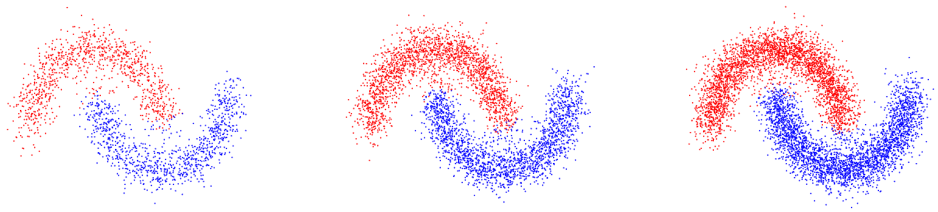
Additional fidelity terms and alternative graph Laplacian could be handled the same way as in the binary case. Hence we have,

THEOREM 5.3 (Convergence). *The multiclass graph Allen-Cahn scheme (with or without fidelity) is convergent for stepsize $dt < c\epsilon$.*

6. Numerical Results. In this section, we will construct various numerical experiments of increasingly larger scales. This helps demonstrate our theory, and also have some implication on the real world performance of the schemes.

6.1. Two Moons. The two moons data was used by Buhler et al [8] in exploring spectral clustering with p-Laplacians. It is constructed from sampling from two half circles of radius one on $\mathbb{R}^2$, centered at (0,0) and (1,0.5). Gaussian noise of standard deviation 0.02 in $\mathbb{R}^{100}$ is then added to each of the points. The weight matrix was constructed using Zelnik-Manor and Perona's procedure [31]. Namely, set $w_{ij} = e^{-d(i,j)/\sqrt{\tau_i \tau_j}}$, where τ_i is the Mth closest distance to i. W is further symmetrized by taking the max between two symmetric entries.

Fig.6.1 is an illustration of the data set of three different sizes being segmented perfectly under a uniform stepsize. A 0 mass constraint was used instead of fidelity points, and random initialization is chosen. The parameters for the experiment is $dt = 0.5, \epsilon = 1$, which is exactly the tight bound.

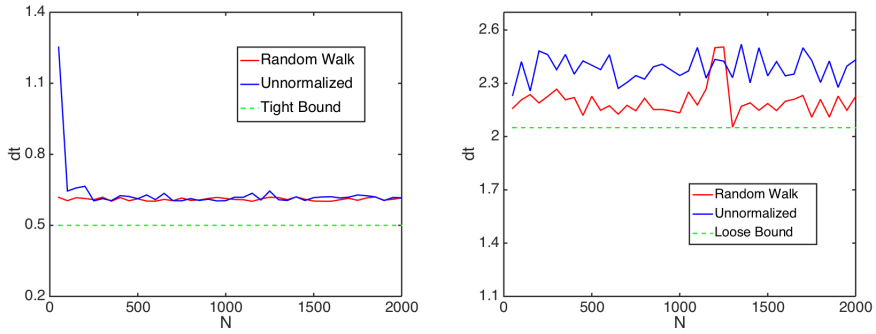


(a) N=1000 (b) N=2000 (c) N=3000

Fig. 6.1: Different data sizes segmented under same stepsize Allen-Cahn

To test the theory in a more rigours manner, we will compute several "maximum stepsizes" that ensures some criterion (e.g. bounded after 500 iterations, etc.), and compare this with the stepsize predicted by the theory. Bisection with 1e-5 accuracy is used to determine the maximum stepsize that satisfies the criterion given.

Fig.6.2 (a),(b) plots the maximum stepsize for the scheme (1.6) to be bounded by 1.0005, 10 respectively. Random $-1, 1$ initial conditions are chosen. No fidelity terms are added and the diffuse parameter $\epsilon = 1$. We also compute results for the random



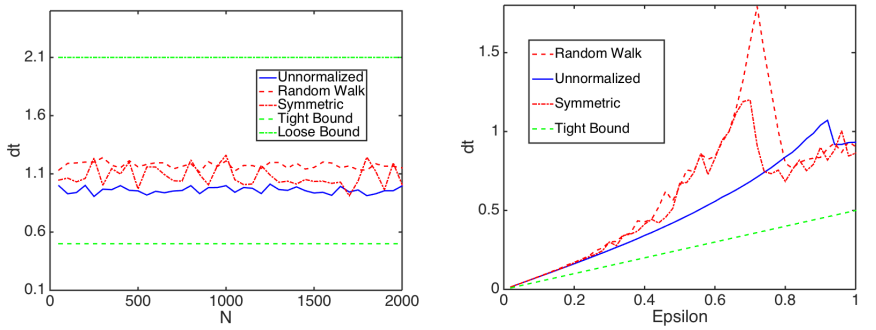
(a) Small Tolerance (b) Big Tolerance

Fig. 6.2: Testing Tight and Loose Bound

walk Laplacian and the unnormalized Laplacian as comparison. The actual results are independent of graph size, and also match the tight and loose bound nicely.

Since we are interested in convergence stepsize, we switch our criterion from boundedness to convergence, namely, we compute the stepsizes for which the scheme has iterative difference less than 1e-4 in 1000 iterations. ϵ is still chosen to be 1.

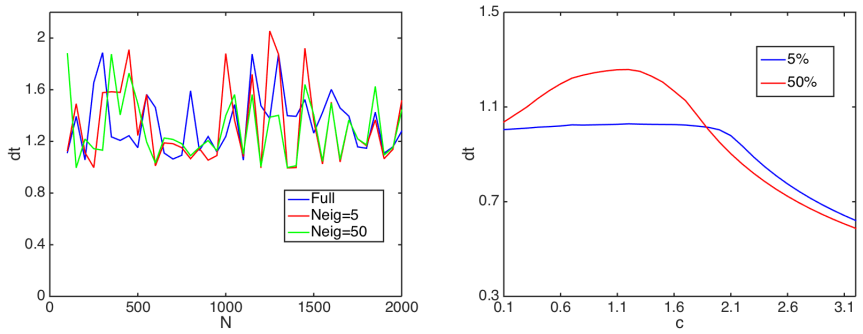
Fig.6.3 (a) plots the limit convergence stepsize for the scheme with the three different Laplacians. As we can see, the typical limit stepsize is between the tight and loose bound. (b) fixes $N = 2000$ and varies ϵ to plot the relation between dt and ϵ. They're almost linear as predicted by the 0.5ϵ bound.



(a) Fixed ϵ Varying N (b) Fixed N Varying ϵ

Fig. 6.3: Profile of maximum convergent stepsize

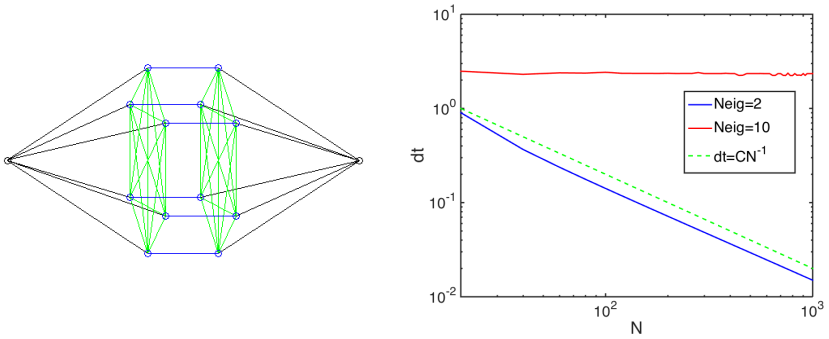
Fig.6.4 (a) plots the scheme with spectral truncation. The results are compared with the full scheme, and are roughly in the same range. Fig.6.4(b) plots the effects of adding a quadratic fidelity term with power c while keeping $\epsilon = 1$ fixed. As we can see from the result, the fidelity term does constitute an additional restriction when c is large. However, stepsizes remain roughly the same for small c. It is hard to analyze the exact effect when c and ϵ are comparable (dt could even increase for small c).



(a) Eigenvector Truncation (b) c vs dt

Fig. 6.4

6.2. Worst Case Graph. Despite the good practical behaviour of spectral truncation, this experiment shown in Fig.6.5 is a realization of the worst case stepsize restriction for spectral truncation. The plot in log-log axis shows the convergent dt vs the size of the problem N. The scheme is initialized by thresholding the 2nd eigenvector. The slope of the descent matches the theoretical $k = -1$ line almost exactly, proving the sub-optimality of the theoretical result.



(a) Illustration of Graph (b) Maximum Stepsize

Fig. 6.5: Worst case stepsize for Spectral Truncation

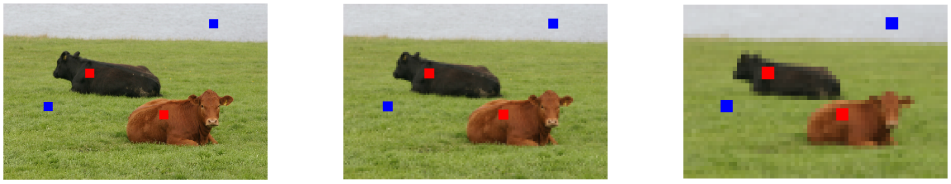
6.3. Two Cows. The point of this experiment is to test the effects of Nystrom extension on the stepsize. The images of the two cows are from the Microsoft Database. Nystrom extension is a technique used to approximate eigenvectors of the graph Laplacian [11] [2] [12].

From the original 312×280 image, we generate 10 images with successily lower resolution of $(312/k) \times (280/k)$, $k = 1, \ldots 10$. A non-local graph constructed from feature windows of size 7×7 is used, and weights are constructed by the standard Gaussian Kernel $w_{ij} = e^{-d_{ij}/\sigma^2}$. The eigenvectors are constructed by using Nystrom extension, the details of which could be found in [6].

Nystrom extension produces an orthogonal set of vectors that approximates the true eigenvectors by subsampling from the original graph. The following examples show that this imprecision does not cause numerical instability.

Fig.6.6 illustrates three images with 1,1/2,1/5 times original resolution being segmented under the uniform condition $dt = 2$, $\epsilon = 4$. The blue and red box corresponds to fidelity points of the two classes, the constant in front of the fidelity are $c_1 = 1$ and $c_2 = 0.4$ for the cows and the background respectively

Fig.6.7 is a profile of N vs dt. To ensure segmentation quality, smaller epsilon had to be chosen for images of lower resolution, and the final result is displayed in terms of the dt/ϵ ratio.



(a) Original Image (b) 1/2 Res Image (c) 1/5 Res Image

(d) Segmentation Result 1 (e) Segmentation Result 2 (f) Segmentation Result 3

Fig. 6.6: Images of different resolution segmented under same stepsize

6.4. MNIST. This experiment is used to demonstrate the case of multiclass clustering by the L^2 multiclass Ginzburg-Landau functional.

The MNIST database [19], found at http://yann.lecun.com/exdb/mnist/, is a data set of 70000 28×28 images of handwritten digits from 0-9 (Fig.6.8). The graph is constructed by first doing a PCA dimension reduction and use the standard Gaussian kernel with $\sigma = 1$ on the principle components.

For our purpose, we will focus on clusters of size three. One reason is that for larger number of clusters $K = 6$, the original L^2 well will be stuck at the center of the simplex, and would need to be rescaled accordingly, adding unnecessary complications to the demo. Another reason is that it is easy to pick three digits that look similar as seen in the collumns of Fig.6.8. We shall pick the tuples that are hardest to cluster and use it to check the validity of the algorithm.

Table 6.1 shows the limit stepsizes of various tuples, and the error rate when segmented under a uniform stepsize. 5% fidelity points are used, and $\epsilon = 1$. The scheme is projected onto the first 100 eigenvectors. An interesting remark is unlike

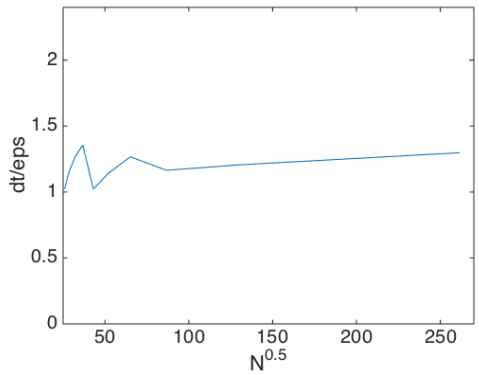


Fig. 6.7: Maximum Convergence Stepsize for Two Cows with Varying Resolution

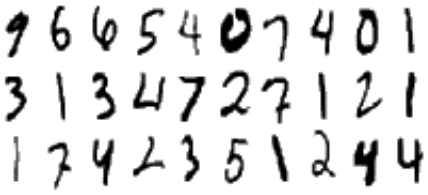


Fig. 6.8: Illustration of MNIST digits

previous data sets, the different graphs here are *not* sampled uniformly from a fixed distribution. However, it is shown here that they are still segmented around the same stepsize.

Tuples	{4,6,7}	{3,5,8}	{1,0,9}	{0,6,1}	{2,7,1}
Max dt	0.5823	0.5914	0.5716	0.5701	0.5755
Err (dt=0.5)	2.08%	2.42%	4.01%	3.64%	1.78%

Table 6.1: Clustering Stepsize of MNIST

7. Appendix. LEMMA 7.1 (Norm Conversions). *Let* $1 \leq p < q \leq +\infty$. *Then the formulae below explicitly states the equivalence between norms:*

$$\|u\|_q \leq \|u\|_p \leq \|u\|_q N^{1/p-1/q}$$

.

Proof. The right hand side is by a generalization of Holder's inequality. The left hand side is by simple polynomial expansion. □

LEMMA 7.2 (Computation of Second Eigenvector of Graph 4.1). *The second eigenvector of the graph in Fig. 4.1 is*

$$\phi^2 = \left(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2\sqrt{N}}, -\frac{1}{2\sqrt{N}}, \ldots, \frac{1}{2\sqrt{N}}, -\frac{1}{2\sqrt{N}}\right)^t$$

Proof. We set the blue edges have weights α, and the yellow edges β/n. Recall the variational formulation of the second eigenvector

$$\operatorname*{argmin}_u Dir(u) \quad s.t. \quad \langle u, e^1 \rangle = 0, \|u\|_2 = 1$$

Note that by symmetry, we can assume $\phi^2 = (a, -a, b, -b, \ldots, b, -b)^t$. Under this parameterization, we have that the Dirichlet energy is

$$Dir(\alpha, \beta) = \frac{\alpha}{n}(b-a)^2 \times n + \beta(2b)^2 n \tag{7.1}$$

Hence by computing the Lagrange multipliers, we have

$$\begin{cases} nkb = 2\gamma n b - (b-a) \\ ka = a - b \end{cases} \tag{7.2}$$

where $\gamma = \frac{\beta}{\alpha}$, and k is the lagrange multiplier. The equation for k is

$$k^2 - (\frac{1}{n} + 2\gamma + 1)k + 2\gamma = 0 \tag{7.3}$$

Setting $2\gamma = 1+\theta$, and computing the roots k, we have $k = 1-(\sqrt{\frac{1}{n} - \theta + (\frac{1}{n} + \theta)^2/4 - \theta^2} - \frac{\theta}{2} - \frac{1}{2n})$. Idealy, $k = 1 - \frac{1}{\sqrt{n}}$ would yield our desired vector. However, we see this can be done by setting the correction term $\theta = O(\frac{1}{n^2})$. Thus $2\gamma = 1$. ∎

REFERENCES

[1] Hedy Attouch, Jérôme Bolte, and Benar Fux Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Mathematical Programming*, 137(1-2):91–129, 2013.

[2] Serge Belongie, Charless Fowlkes, Fan Chung, and Jitendra Malik. Spectral partitioning with indefinite kernels using the nyström extension. In *Computer VisionECCV 2002*, pages 531–542. Springer, 2002.

[3] Michal Beneš, Vladimır Chalupeckỳ, and Karol Mikula. Geometrical image segmentation by the allen–cahn equation. *Applied Numerical Mathematics*, 51(2):187–205, 2004.

[4] Andrea Bertozzi, Selim Esedo g$^-$ lu, and Alan Gillette. Analysis of a two-scale cahn-hilliard model for binary image inpainting. *Multiscale Modeling & Simulation*, 6(3):913–936, 2007.

[5] Andrea L Bertozzi, Selim Esedoglu, and Alan Gillette. Inpainting of binary images using the Cahn-Hilliard equation. *IEEE Transactions on image processing*, 16(1):285–291, 2007.

[6] Andrea L Bertozzi and Arjuna Flenner. Diffuse interface models on graphs for classification of high dimensional data. *Multiscale Modeling & Simulation*, 10(3):1090–1118, 2012.

[7] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2009.

[8] Thomas Bühler and Matthias Hein. Spectral clustering based on the graph p-laplacian. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 81–88. ACM, 2009.

[9] Philippe G Ciarlet. Discrete maximum principle for finite-difference operators. *Aequationes mathematicae*, 4(3):338–352, 1970.

[10] David J Eyre. An unconditionally stable one-step scheme for gradient systems. *Unpublished article*, 1998.

[11] Charless Fowlkes, Serge Belongie, Fan Chung, and Jitendra Malik. Spectral grouping using the Nystrom method. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(2):214–225, 2004.

[12] Charless Fowlkes, Serge Belongie, and Jitendra Malik. Efficient spatiotemporal grouping using the nystrom method. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–231. IEEE, 2001.

[13] Cristina Garcia-Cardona, Arjuna Flenner, and Allon G Percus. Multiclass semi-supervised learning on graphs using ginzburg-landau functional minimization. In *Pattern Recognition Applications and Methods*, pages 119–135. Springer, 2015.

[14] Cristina Garcia-Cardona, Ekaterina Merkurjev, Andrea L Bertozzi, Arjuna Flenner, and Allon G Percus. Multiclass data segmentation using diffuse interface methods on graphs. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(8):1600–1613, 2014.

[15] Stephen Guattery and Gary L Miller. On the quality of spectral separators. *SIAM Journal on Matrix Analysis and Applications*, 19(3):701–719, 1998.

[16] Huiyi Hu, Thomas Laurent, Mason A Porter, and Andrea L Bertozzi. A method based on total variation for network modularity optimization using the mbo scheme. *SIAM Journal on Applied Mathematics*, 73(6):2224–2246, 2013.

[17] Huiyi Hu, Justin Sunu, and Andrea L Bertozzi. Multi-class graph mumford-shah model for plume detection using the mbo scheme. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 209–222. Springer, 2015.

[18] Tom Ilmanen. Convergence of the Allen-Cahn equation to Brakke's motion by mean curvature. *J. Differential Geom.*, 38(2):417–461, 1993.

[19] Yann LeCun and Corinna Cortes. The MNIST database of handwritten digits, 1998.

[20] Ekaterina Merkurjev, Tijana Kostic, and Andrea L Bertozzi. An MBO scheme on graphs for classification and image processing. *SIAM Journal on Imaging Sciences*, 6(4):1903–1930, 2013.

[21] Barry Merriman, James K Bence, and Stanley J Osher. Motion of multiple junctions: A level set approach. *Journal of Computational Physics*, 112(2):334–363, 1994.

[22] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[23] Suvrit Sra. Scalable nonconvex inexact proximal splitting. In *Advances in Neural Information Processing Systems*, pages 530–538, 2012.

[24] John C Strikwerda. *Finite difference schemes and partial differential equations*. Siam, 2004.

[25] Nicolas Garcia Trillos, Dejan Slepcev, James von Brecht, Thomas Laurent, and Xavier Bresson. Consistency of Cheeger and Ratio graph cuts. *arXiv preprint arXiv:1411.6590*, 2014.

[26] Yves van Gennip, Nestor Guillen, Braxton Osting, and Andrea L Bertozzi. Mean curvature, threshold dynamics, and phase field theory on finite graphs. *Milan Journal of Mathematics*, 82(1):3–65, 2014.

[27] Benjamin P Vollmayr-Lee and Andrew D Rutenberg. Fast and accurate coarsening simulation with an unconditionally stable time step. *Physical Review E*, 68(6):066703, 2003.

[28] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[29] Ulrike Von Luxburg, Mikhail Belkin, and Olivier Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, pages 555–586, 2008.

[30] Alan L Yuille, Anand Rangarajan, and AL Yuille. The concave-convex procedure (CCCP). *Advances in neural information processing systems*, 2:1033–1040, 2002.

[31] Lihi Zelnik-Manor and Pietro Perona. Self-tuning spectral clustering. In *Advances in neural information processing systems*, pages 1601–1608, 2004.