



# Asymptotic Behaviour of Time Stepping Methods for Phase Field Models

Xinyu Cheng<sup>1</sup> · Dong Li<sup>2</sup> · Keith Promislow<sup>3</sup> · Brian Wetton<sup>1</sup> 

Received: 28 September 2020 / Revised: 9 December 2020 / Accepted: 12 December 2020 /  
Published online: 16 January 2021  
© The Author(s) 2021

## Abstract

Adaptive time stepping methods for metastable dynamics of the Allen–Cahn and Cahn–Hilliard equations are investigated in the spatially continuous, semi-discrete setting. We analyse the performance of a number of first and second order methods, formally predicting step sizes required to satisfy specified local truncation error  $\sigma$  in the limit of small length scale parameter  $\epsilon \rightarrow 0$  during meta-stable dynamics. The formal predictions are made under stability assumptions that include the preservation of the asymptotic structure of the diffuse interface, a concept we call profile fidelity. In this setting, definite statements about the relative behaviour of time stepping methods can be made. Some methods, including all so-called energy stable methods but also some fully implicit methods, require asymptotically more time steps than others. The formal analysis is confirmed in computational studies. We observe that some provably energy stable methods popular in the literature perform worse than some more standard schemes. We show further that when Backward Euler is applied to meta-stable Allen–Cahn dynamics, the energy decay and profile fidelity properties for these discretizations are preserved for much larger time steps than previous analysis would suggest. The results are established asymptotically for general interfaces, with a rigorous proof for radial interfaces. It is shown analytically and computationally that for most reaction terms, Eyre type time stepping performs asymptotically worse due to loss of profile fidelity.

**Keywords** Allen–Cahn equation · Cahn–Hilliard equation · Phase field model · Time stepping · Energy stability

---

✉ Brian Wetton  
wetton@math.ubc.ca

Xinyu Cheng  
xycheng@math.ubc.ca

Dong Li  
madli@ust.hk

Keith Promislow  
kpromisl@math.msu.edu

<sup>1</sup> Department of Mathematics, University of British Columbia, Vancouver V6T 1Z2, Canada

<sup>2</sup> Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong, Hong Kong

<sup>3</sup> Department of Mathematics, Michigan State University, East Lansing 48864, USA

## 1 Introduction

The mathematical literature for computational methods for Allen–Cahn (AC) dynamics [2], and its higher order relative Cahn–Hilliard (CH) dynamics [6], is dominated by the proposal, use, and analysis of so-called energy stable schemes [12,25,26,29]. AC and CH dynamics are gradient flows on an energy functional, and the solution should decrease that energy in time. Energy stable schemes guarantee that decrease no matter what time step is chosen. This is a desirable property not shared by standard fully implicit, semi-implicit (IMEX), or exponential integrator time stepping methods. We will show in this work that some (but not all) fully implicit methods can outperform energy stable schemes when subject to fixed accuracy requirements. The recent article [30] gives especially clear evidence that when time steps are chosen appropriately, fully implicit methods are conditionally energy stable, and further that the large time steps allowed by energy stable schemes can come at the cost of significant loss of accuracy. We show that in the metastable dynamic regime of AC and CH, some fully implicit methods can take optimally sized time steps. By optimal, we mean the asymptotically largest time steps as the order parameter  $\epsilon \rightarrow 0$  that satisfy a given local error tolerance. Here,  $\epsilon$  represents the width of interfacial layers in metastable dynamics and, like the authors of [30], we use the form of the equations scaled so that these dynamics transpire in an  $O(1)$  time scale. When the dynamics are in this metastable regime, which dominates the time of typical phenomena of interest, definite statements about the behaviour of different time stepping methods can be made. This criteria does not take into account solver efficiency. However, we can make definite statements on how efficient solvers for nonlinear implicit time stepping need to be to outperform other methods.

A combination of asymptotic analysis and careful computational work backs up our claims. The computational codes that generate the results shown in the paper are available by e-mail request from the last author. In addition, we present a rigorous result for implicit time stepping for meta-stable AC dynamics in radial geometry that shows that asymptotically larger time steps can be taken than previous analysis would suggest. These time steps preserve the diffuse interface structure (a property that we call profile fidelity) and also the energy decay property of the equations. This result is shown for a class of reaction terms. An interesting result in Sect. 6.2 shows that Eyre-type time stepping can perform asymptotically worse with most reaction terms, while implicit time stepping has uniform asymptotic behaviour over a class of reaction terms. This was predicted by the analysis and confirmed computationally.

Our study focuses on pure materials science applications rather than the use of Cahn–Hilliard equations to track interfaces in so-called *diffuse interface methods* [32] in which the CH dynamics are coupled to other physics. We consider the simplest form of AC and CH dynamics, whose Gamma limit (as  $\epsilon \rightarrow 0$ ) is well understood and use that well known structure to gain insight into the behaviour of the schemes. The authors believe that the insight gained from these studies will also apply to schemes used for other materials science models which are less well understood.

We consider a number of first and second order time stepping schemes: the energy stable Eyre’s method [14]; Backward (Implicit) Euler (BE) [17]; Trapezoidal Rule (TR) [17]; Second order Backward Differentiation Formula (BDF2) [17]; Secant [13]; standard semi-implicit (linear IMEX) methods of first and second order [3]; first and second order Scalar Auxiliary Methods (SAV) [25] for which a modified energy stability can be proved; and finally a second order Singular Diagonally Implicit Runge Kutta method with good stability properties (DIRK2) [17]. The resulting implicit systems are considered in the spatially continuous semi-discrete setting in a 2D periodic domain, with numerical validation done with

a suitably refined Fourier spectral approximation. Time step schemes that result in nonlinear systems are solved with Newton's iterations using the Preconditioned Conjugate Gradient Solver (PCG) developed in [9] at each iteration. Adaptive time stepping is done based on a user-specified local error tolerance  $\sigma$ . The variation of the number of time steps with  $\epsilon$  for fixed  $\sigma$  is predicted based on formal consideration of the local truncation error of the schemes in the metastable dynamics. The formal predictions are then validated in computational studies. With this criteria, first order BE performs better (asymptotically fewer time steps as  $\epsilon \rightarrow 0$ ) than Eyre and first order IMEX and SAV. Second order TR and BDF2 perform better than Secant, DIRK2, and second order IMEX and SAV. The difference in both cases is asymptotically larger for CH than AC. These comparisons are also valid for computational time, using PCG counts as the measure, to similar accuracy. It is seen that optimal numbers of time steps are obtained when the dominant local truncation error is a higher order time derivative. This observation may have application in other systems with metastable dynamics. We observe that standard IMEX methods perform almost identically to SAV methods of the same order in the scenario we consider, at reduced computational cost.

It is observed that the global accuracy of BE is better than a naïve prediction based on the size of the local truncation error would suggest. A formal analysis of the scheme for the AC case shows that the dominant error made in one time step is asymptotically smaller than expected. This is due to a special structure of the local truncation error for BE, in which the asymptotically largest term lies in a strongly damped space.

We introduce the equations and numerical schemes in Sect. 2 with some introductory analysis. The scaling for AC and CH is chosen so that the metastable interface dynamics (approximate curvature motion for AC and Mullins-Sekerka flow [22] for CH) occurs in  $O(1)$  time. In Sect. 3 we examine the metastable dynamics of the equations and make predictions for the behaviour of the time steps with  $\epsilon$  and local error tolerance  $\sigma$  under stability assumptions which are verified numerically in Sect. 4. We give an asymptotic analysis for the surprising accuracy and stability properties for BE with large time steps applied to AC in Sect. 5. In Sect. 6 we present the rigorous result for BE applied to AC with large time steps and also show the loss of profile fidelity for Eyre-type time stepping for most reaction terms. We end with a short discussion.

## 2 Equations and Schemes

We consider the simplest form of the AC dynamics for  $u(\mathbf{x}, t)$  given by

$$u_t = \Delta u - \frac{1}{\epsilon^2} f(u) \quad (1)$$

where  $f(u) = u^3 - u$  is the classical form of the reaction term. More general, smooth reaction terms are considered in Sect. 6. Non-smooth reaction terms and degenerate mobility are also of interest in some materials science applications and there are stability and convergence results for implicit time stepping applied to these problems in [4,5] for example. However, the asymptotic behaviour of time stepping schemes for these problems is not clear.

CH dynamics is described by a higher order partial differential equation

$$u_t = -\epsilon \Delta \Delta u + \frac{1}{\epsilon} f(u). \quad (2)$$

For computational simplicity, we consider the two-dimensional (2D) cases of these equations in a doubly periodic cell  $[0, 2\pi]^2$ . The time scaling in the equations above is chosen to give sharp interface (as  $\epsilon \rightarrow 0$ ) motion in  $O(1)$  time. The sharp interface limit yields curvature driven flow for AC and a nonlocal Mullins-Sekerka flow for CH [22]. Both types of dynamics have an associated energy functional

$$\mathcal{E} = \int (|\nabla u|^2/2 + W(u)/\epsilon^2) \quad (3)$$

where  $W(u) = \frac{1}{4}(u^2 - 1)^2$  and the reaction term  $f(u) = W'(u)$ . The energy  $\mathcal{E}(t)$  is monotonic decreasing due to the gradient flow nature of the dynamics. For AC the gradient is in  $L_2$  and for CH it is  $H^{-1}$ .

## 2.1 Time Stepping

### 2.1.1 Backward Euler

We consider the simplest implicit scheme, first order Backward Euler (BE), also known as Implicit Euler. Applied to (1) keeping space continuous, we have

$$\frac{u_{n+1} - u_n}{k_n} = \Delta u_{n+1} - \frac{1}{\epsilon^2} f(u_{n+1}) .$$

where  $u_n(\mathbf{x})$  approximates the exact solution  $u(\mathbf{x}, t_n)$  and  $k_n = t_{n+1} - t_n$  is the time step. We use the classical  $f(u) = u^3 - u$  as mentioned above. Dropping the subscript on the time step and the unknown solution at time level  $n + 1$  we have the nonlinear problem

$$u - k\Delta u + \frac{k}{\epsilon^2} f(u) = u_n \quad (4)$$

for  $u$  given  $u_n$ .

**Definition 1** A time stepping scheme is said to have the *energy decay property* if  $\mathcal{E}(u_{n+1}) \leq \mathcal{E}(u_n)$ .

This property could be conditional on the choice of time step size. Additionally, it could depend on  $u_n$ . If a scheme has the energy decay property for any  $u_n$  and  $k$ , the scheme is called *unconditionally energy stable*.

**Theorem 1** Consider (4), assume that  $u_n \in H^2(\Omega)$  and  $u_n$  takes values in  $[-1, 1]$ , then there exists  $u \in H^2(\Omega)$  that solves (4) with values in  $[-1, 1]$ . Define  $f_\infty := \max\{|f'(s)|, s \in [-1, 1]\}$ , then if  $k \leq 2\epsilon^2/f_\infty$  the solution  $u$  is unique and satisfies the energy decay property. Note that the energy stability result was established earlier in [30] with a different proof.

**Proof** The existence of  $u$  follows from the standard method of sub-/super-solutions applied to comparison functions  $-1$  and  $+1$ . To establish uniqueness, we assume  $u_1$  and  $u_2$  are solutions. Then their difference  $w = u_1 - u_2$  is a solution of

$$(1 - k\Delta)w = -k \cdot \frac{f(u_1) - f(u_2)}{\epsilon^2} = -\frac{k}{\epsilon^2} \cdot f'(s(x))w ,$$

where  $s$  takes values between  $u_1$  and  $u_2$ , and hence in  $[-1, 1]$ . Isolating  $w$  leads to the elliptic problem

$$\left[ 1 + \frac{kf'(s)}{\epsilon^2} - k\Delta \right] w = 0,$$

and if  $k < \epsilon^2/f_\infty$  then the corresponding elliptic operator is strictly positive and  $w$  is zero by the maximum principle. To establish energy decay, we take the inner product of (4) with the test function  $u - u_n$ :

$$\frac{1}{k} \int |u - u_n|^2 + \frac{1}{2} \int (|\nabla u|^2 - |\nabla u_n|^2 + |\nabla u - \nabla u_n|^2) = -\frac{1}{\epsilon^2} (f(u), (u - u_n)) .$$

From the Fundamental Theorem of Calculus we develop the expansion,

$$|F(u) - F(u_n) - f(u)(u - u_n)| = \left| \int_u^{u_n} f'(s)(s - u_n) ds \right| \leq \frac{f_\infty}{2} (u - u_n)^2.$$

Using this relation to eliminate  $f(u)$  yields the equality,

$$\left( \frac{1}{k} - \frac{f_\infty}{2\epsilon^2} \right) \int |u - u_n|^2 + E[u] - E[u_n] \leq 0,$$

which yields energy decay for  $k < 2\epsilon^2/f_\infty$ . The Theorem is also true when homogeneous Neumann boundary conditions are specified.  $\square$

Thus we have existence of solutions to (4) for any time step size, and uniqueness and energy stability under the restriction  $k \leq 2\epsilon^2/f_\infty$ . This is true for any  $u_n$  under the restrictions of the Theorem. We shall see in Sect. 6 that asymptotically larger time steps  $k = o(\epsilon)$  can be taken when the dynamics are slow (interface motion) with locally unique, energy stable solutions. This is verified in computational tests.

### 2.1.2 Eyre's Method

An alternative first order scheme to fully implicit BE was proposed by Eyre [14]:

$$u - k\Delta u + \frac{k}{\epsilon^2} u^3 = u_n + \frac{k}{\epsilon^2} u_n \quad (5)$$

The scheme is derived conceptually by keeping a convex part of the reaction term  $f(u) = u^3 - u$  implicit and a concave part explicit. In this sense, it is an IMEX method but an unusual one since a nonlinear term is kept implicit and a linear term is handled explicitly. The method has appealing properties:

**Theorem 2** (from [14]) *The time step (5) has a unique solution  $u$  for any  $u_n$  and  $k$  that is unconditionally energy stable.*

Additional first order schemes considered are the SAV scheme [25] and a linear IMEX method [3]:

$$u - k\Delta u + \frac{Sk}{\epsilon^2} u = u_n - \frac{k}{\epsilon^2} (u_n^3 - (S+1)u_n) \quad (6)$$

with  $S > 0$ , sometimes called a stabilization term. We take  $S = 2$  since that makes the left hand side a linearization about the far field values, but computational performance is relatively insensitive to  $S$ . The SAV scheme is energy stable with a modified energy. We use the same stabilization coefficient as above in the SAV scheme. There is a class of linearly implicit energy stable schemes [8,19,20] that require an asymptotically large stabilization term  $O(\epsilon^{-p})$  with  $p$  large and increasing from AC to CH and 2D to 3D for the analysis.

These methods are theoretically interesting but are extremely inaccurate and not useful for practical applications. We have further discussion of these schemes in Remark 3.

All time stepping schemes can be applied to CH (2), with BE and Eyre shown below:

$$u + k\Delta u - \frac{k}{\epsilon^2}\Delta f(u) = u_n \quad \text{BE}$$

$$u + k\Delta u - \frac{k}{\epsilon^2}\Delta u^3 = u_n - \frac{k}{\epsilon^2}\Delta u_n \quad \text{Eyre}$$

In this case, BE is known to have unique solutions with the energy decay property when  $k < \epsilon^3$  [30] and Eyre is unconditionally energy stable [14].

### 2.1.3 Second Order Schemes

We also consider the second order methods Trapezoidal Rule (TR), Secant (S) [13], Second Order Backward Differencing (BDF2), and Second Order Singular Diagonal Implicit Runge Kutta (DIRK2) [17] methods. These are described below for  $u_t = \mathcal{F}(u)$  with

$$\begin{aligned} \mathcal{F}(u) &= \Delta u - f(u)/\epsilon^2 && \text{for AC} \\ \text{and } \mathcal{F}(u) &= -\epsilon\Delta u + \Delta f(u)/\epsilon && \text{for CH} \end{aligned}$$

With this notation:

$$\begin{aligned} \text{(TR)} \quad u - \frac{k}{2}\mathcal{F}(u) &= u_n + \frac{k}{2}\mathcal{F}(u_n) \\ \text{(BDF2)} \quad \frac{3u}{2} - k\mathcal{F}(u) &= 2u_n - \frac{1}{2}u_{n-1}. \end{aligned}$$

Secant is a variant of TR with the term  $f(u) - f(u_n)$  replaced by

$$(W(u) - W(u_n))/(u - u_n)$$

where  $W$  is the energy term from (3). It is known to be conditionally energy stable [13]. For the simple form of  $W$  we have taken, the expression above can be factored explicitly. DIRK2 is a two stage method

$$\begin{aligned} u_* - \alpha k\mathcal{F}(u_*) &= u^n \\ u - \alpha k\mathcal{F}(u) &= u^n + (1 - \alpha)k\mathcal{F}(u_*) \end{aligned}$$

with  $\alpha = 1 - 1/\sqrt{2}$ . Both DIRK2 and BDF2 are A-stable, and so preferable to TR and Secant from the perspective of stiff ODE solver theory [17]. A second order linear IMEX method (SBDF2 [3]) and two variants of second order SAV methods based on BDF2 are also considered.

There are many other specialized schemes in the literature and we mention two second order unconditionally energy stable concave-convex splitting schemes here. They are non-linear two step schemes but the nonlinear problem at each time step is convex. One is based on TR, with the cubic term handled implicitly and the linear reaction term extrapolated from two previous time steps [11]. The second is a variant of SBDF2, again with the cubic term handled implicitly and an additional moderate stabilizing term [31]. Both these schemes have the same asymptotic error behaviour as the Secant, DIRK2, and SBDF2 methods shown in detail below.

## 2.2 Spatial Discretization and Solution Procedure

The current work concentrates on the time stepping errors, and it is convenient to consider the semi-discrete, spatially continuous approximation. This idealization is approximated well by the Fourier spectral spatial discretization. The computational results shown have sufficient spatial resolution that spatial errors do not affect the results in the digits shown.

We use the Preconditioned Conjugate Gradient (PCG) solvers developed in [9] for the schemes involving nonlinear implicit problems. We note that there has been recent promising work in the use of preconditioned steepest descent with approximate line search in solving these nonlinear problems [7,15]. Another approach has been to recast the implicit step as a minimization problem [30]. Both these techniques have the advantage that they look for local solutions which can be unique and have energy decay even for large time steps, as shown rigorously in Sect. 6.

The computations in this work are done in a full 2D setting, rather than in a reduced dimensional radial setting as could be done, in order to give PCG iteration counts for the nonlinear time stepping methods that have meaning for more general computations. Note that the PCG counts are independent of spatial resolution when the problem is resolved.

## 2.3 Error Estimation and Adaptive Time Stepping

We perform two time steps of the same size  $k$  in order to use a specialized predictor  $u_p$  for  $u_{n+2}$ .

$$u_p = u_n + \frac{k}{3} (\mathcal{F}(u_n) + 4\mathcal{F}(u_{n+1}) + \mathcal{F}(u_{n+2})) \quad (7)$$

where  $\mathcal{F}(u) = \Delta u - f(u)/\epsilon^2$  for AC and  $-\epsilon \Delta \Delta u + \Delta f(u)/\epsilon$  for CH as above. Time step sizes are adjusted so that

$$\|u_{n+2} - u_p\|_\infty \leq \sigma.$$

The predictor  $u_p$  is formally one order more accurate than the numerical approximation  $u_{n+2}$  from time stepping, up to fifth order. The predictor has an inherent dominant local error  $k^5 u_{ttttt}/90$  that is a pure time derivative of  $u$ , which is shown below in Sect. 3.1.3 to be a desirable property.

For the one step methods, the time step is adjusted adaptively to maintain a local error below  $\sigma$  as described in [9]. For BDF2 and its linear variants, time steps are only adjusted by a factor of two. When time steps are reduced (using Hermite cubic interpolation for the restart value) or increased, four time steps are taken before checking the local error to allow relaxation of the initial error layer.

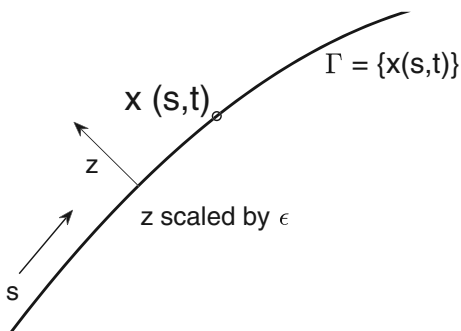
## 3 Local Truncation Errors in Metastable Dynamics

### 3.1 Metastable Dynamics

In our formulation, it is known that after a short time  $O(\epsilon^2)$  solutions to AC tend to interfaces between regions of solution near the equilibrium values,

$$u \approx \pm 1.$$

**Fig. 1** Sketch of the local coordinates of the metastable solution



These interfaces have width  $\epsilon$  and move approximately with curvature motion. We refer to this dynamics as metastable or slow, even though with the particular time scaling we have chosen it occurs in  $O(1)$  time. For the majority of the time, the solution will be in this regime, so we concentrate now on the expected and observed behaviour of time stepping in this setting. With the choice of  $f(u) = u^3 - u$ , we have

$$u(x, t) \approx g(z) \quad (8)$$

with  $g(z) := \tanh(z/\sqrt{2})$  and  $z = \text{dist}(x, \Gamma)/\epsilon$ , where  $\Gamma$  is the approximate interface with arc length parameter  $s$  moving with curvature motion (normal velocity equal to curvature). We fix its location at the  $u = 0$  level set. The local coordinates  $(s, z)$  are shown in Fig. 1. This structural result on the metastable solution can be obtained with formal asymptotics. In the outer asymptotic region for AC the solution takes the form  $u = \pm 1$  to all orders. Curvature motion as the limit  $\epsilon \rightarrow 0$  has been proven rigorously [1, 23].

CH has the same metastable solution structure (8) with normal interface velocity given by Mullins-Sekerka flow, in  $O(1)$  time in our scaling (2). We refer the reader to the review article [24] for details. It has been shown that numerical schemes can accurately approximate this limit with implicit time stepping with appropriate scaling of the time step with  $\epsilon$  [16]. We will show that this limit can be taken with asymptotically larger time steps than in that analysis.

From (8), we see that time and space derivatives are large near the interface. Starting with

$$u(x, t) \approx g(\text{dist}(x, \Gamma)/\epsilon)$$

we can take a time derivative to obtain:

$$u_t \approx g'(\text{dist}(x, \Gamma))V/\epsilon$$

where  $V$  is the normal velocity at the point on  $\Gamma$  closest to  $x$ . Formally taking higher derivatives in this pattern yields:

$$\frac{\partial^n u}{\partial t^n} = O(\epsilon^{-n}). \quad (9)$$

This is used to analyze the truncation error of the time stepping schemes.

### 3.1.1 Predicted Time Step Sizes for AC

A standard strategy for adaptive time stepping is to have a user specified local error tolerance of  $\sigma$ . The error for each time step is estimated and the time step adjusted so that there is



an estimated error in that single time step less than  $\sigma$ . It is known that the dominant local truncation error for BE is  $k^2 u_{tt}/2$  which in metastable dynamics is  $O(k^2/\epsilon^2)$  from (9). The local truncation error restriction then requires time steps of size

$$k = O(\sqrt{\sigma}\epsilon) \text{ (BE)}$$

We now proceed to determine the expected behaviour of time steps with  $\epsilon$  and  $\sigma$  from the other schemes. We can write the BE scheme (4) and Eyre's scheme (5) for AC in an instructive way

$$\begin{aligned} u - u^n - k\Delta u + k[u^3 - u]/\epsilon^2 &= 0 \text{ (BE)} \\ u - u^n - k\Delta u + k[u^3 - u]/\epsilon^2 + k(u - u_n)/\epsilon^2 &= 0 \text{ (Eyre)}. \end{aligned}$$

Knowing that the truncation error for BE is  $O(k^2/\epsilon^2)$  we see that the truncation error for the Eyre scheme is dominated by the last term in its expression above, which has leading order  $k^2 u_t/\epsilon^2 = O(k^2/\epsilon^3)$ . Our time step prediction in this case is

$$k = O(\sqrt{\sigma}\epsilon^{3/2}) \text{ (Eyre)}$$

Thus, the advantage of the Eyre scheme to be able to take large time steps and remain energy stable is never realized if accurate computational results are required. Reference [30] has an alternate way to view the loss of accuracy that does not highlight this asymptotic difference. The first order IMEX and SAV schemes have the same asymptotic behaviour as Eyre.

**Remark 1** It is well known that when large time steps are taken with Eyre's method, the dynamics occur in a slower time scale. This is an exact result for AC [30], qualitative for CH [10]. In this work, time steps are restricted by a specified local error tolerance. Thus, we do not see a change in time scale for the results of Eyre's method, rather we see decreased time step size.

**Remark 2** The formal local error analysis above relies on the stability of the schemes in metastable dynamics under the resulting time step restrictions. More than simple stability, the analysis requires that the time stepping preserves the asymptotic structure of the diffuse interface. This is the concept we have named *profile fidelity*. All predictions described in this section lead to time stepping that preserves profile fidelity for the classical choice of  $f(u) = u^3 - u$ . We observe the predicted time step behaviour in  $\epsilon$  and  $\sigma$  computationally. In Sect. 6.2 we show that for (most) other reaction terms, Eyre time stepping loses profile fidelity for time steps  $k = O(\epsilon^{3/2})$  and in these cases,  $k = O(\epsilon^2)$  is needed for accuracy.

**Remark 3** The first order, linearly implicit energy stable scheme for 2D AC is analyzed in [8]. The analysis requires a stabilization term of order  $\epsilon^{-2}|\ln \epsilon|$ . If such a scheme were implemented, the time steps required for a local error tolerance of  $\sigma$  would be  $k = O(\sqrt{\sigma}\epsilon^{5/2}/|\ln \epsilon|)$ , prohibitively small for practical computation.

We can determine the dominant term in the local truncation errors of the second order schemes applied to AC:

$$\begin{aligned} \text{(TR)} \quad k^3 u_{ttt}/12 &= O(k^3/\epsilon^3) \\ \text{(S)} \quad k^3 (u_{ttt}/12 + uu_t^2/(2\epsilon^2)) &= O(k^3/\epsilon^4) \\ \text{(DIRK2)} \quad k^3 ((\alpha^2(1-\alpha) + \alpha/2 - 1/6)u_{ttt} - 3\alpha^2(1-\alpha)uu_t^2/(2\epsilon^2)) &= O(k^3/\epsilon^4) \\ \text{(BDF2)} \quad -k^3 u_{ttt}/3 &= O(k^3/\epsilon^3) \end{aligned}$$

**Table 1** Order predictions for the behaviour of the numerical schemes with local error tolerance  $\sigma$  in the metastable regime of AC dynamics

Method (AC)	$L$	$k$	$M = O(1/k)$
BE	$k^2/\epsilon^2$	$\sqrt{\sigma}\epsilon$	$1/(\sqrt{\sigma}\epsilon)$
Eyre, IMEX1, SAV1	$k^2/\epsilon^3$	$\sqrt{\sigma}\epsilon^{3/2}$	$1/(\sqrt{\sigma}\epsilon^{3/2})$
TR, BDF2	$k^3/\epsilon^3$	$\sqrt[3]{\sigma}\epsilon$	$1/(\sqrt[3]{\sigma}\epsilon)$
S, DIRK2, SBDF2, SAV2-A	$k^3/\epsilon^4$	$\sqrt[3]{\sigma}\epsilon^{4/3}$	$1/(\sqrt[3]{\sigma}\epsilon^{4/3})$
SAV2-B	$k^3/\epsilon^5$	$\sqrt[3]{\sigma}\epsilon^{5/3}$	$1/(\sqrt[3]{\sigma}\epsilon^{5/3})$

Here,  $L$  is the local error,  $k$  is the time step size, and  $M$  is the number of time steps to reach a fixed end time

$$(\text{SBDF2}) \ k^3 (3u^2 + (M+1)) u_{tt}/\epsilon^2 = O(k^3/\epsilon^4)$$

We consider two second order SAV variants based on how an extrapolated approximation is computed. If the extrapolated value of  $u^{n+1}$  is taken as  $2u^n - u^{n-1}$  the scheme (referred to as SAV2-A) behaves similarly to SBDF2. If the extrapolated value is computed with a first order linear IMEX scheme as suggested in [25] (referred to as SAV2-B), the scheme has a local truncation error of order  $k^3/\epsilon^5$ . The results are summarized in Table 1. It is clear that BE takes asymptotically (as  $\epsilon \rightarrow 0$ ) fewer time steps than Eyre, although they are both first order in time step size. TR and BDF2 take asymptotically fewer time steps than Secant, DIRK2, SBDF2, SAV2-A and SAV2-B although they are all second order methods. The computations in Sect. 4 below show that these time step estimates correspond to real computational behaviour.

**Remark 4** We predict the number  $M$  of time steps in Tables 1 and 2 and how it varies with  $\epsilon$  and  $\sigma$ . As shown in Fig. 2 we are also predicting how a profile of time steps  $k(t)$  behaves with  $\epsilon$  and  $\sigma$ .

### 3.1.2 Predicted Time Step Sizes for CH

The same local truncation analysis can be done for the CH in the metastable regime where the solution has the same interface structure (8) with the interface  $\Gamma$  moving approximately with Mullins-Sekerka flow in  $O(1)$  time. BE, TR, BDF2, and SBDF2 have the same error expressions as above, but Eyre, Secant and DIRK2 have local truncation errors when applied to CH listed below:

$$(\text{Eyre}) \ k^2 (u_{tt}/2 - \Delta u_t/\epsilon) = O(k^2/\epsilon^4)$$

$$(\text{S}) \ k^3 (u_{ttt}/12 - \Delta(uu_t^2)/(2\epsilon)) = O(k^3/\epsilon^5)$$

$$(\text{DIRK2}) \ k^3 ((\alpha^2(1-\alpha) + \alpha/2 - 1/6)u_{ttt} + 3\alpha^2(1-\alpha)\Delta(uu_t^2)/(2\epsilon)) = O(k^3/\epsilon^5)$$

$$(\text{SBDF2}) \ k^3 (3u^2 + (M+1)) \Delta u_{tt}/\epsilon = O(k^3/\epsilon^5)$$

where we have used the fact that the Laplacian  $\Delta$  increases the size of terms by  $1/\epsilon^2$  near the interface. The first order IMEX and SAV schemes have the same asymptotic behaviour as Eyre. SAV2-A behaves similarly to SBDF2 as before, with SAV2-B worse by a power of  $\epsilon$  as for the AC case above. The results are summarized in Table 2. The predictions in this table are validated in the numerical experiments in the next section. Although the methods all have the formal order of accuracy in terms of time step size, the behaviour as  $\epsilon \rightarrow 0$  varies

**Table 2** Order predictions for the behaviour of the numerical schemes with local error tolerance  $\sigma$  in the metastable regime of CH dynamics

Method (CH)	$L$	$k$	$M = O(1/k)$
BE	$k^2/\epsilon^2$	$\sqrt{\sigma}\epsilon$	$1/(\sqrt{\sigma}\epsilon)$
Eyre, IMEX1, SAV1	$k^2/\epsilon^4$	$\sqrt{\sigma}\epsilon^2$	$1/(\sqrt{\sigma}\epsilon^2)$
TR, BDF2	$k^3/\epsilon^3$	$\sqrt[3]{\sigma}\epsilon$	$1/(\sqrt[3]{\sigma}\epsilon)$
S, DIRK2, SBDF2, SAV2-A	$k^3/\epsilon^5$	$\sqrt[3]{\sigma}\epsilon^{5/3}$	$1/(\sqrt[3]{\sigma}\epsilon^{5/3})$

Here,  $L$  is the local error,  $k$  is the time step size, and  $M$  is the number of time steps to reach a fixed end time

significantly. Note that the gap between BE and the other first order schemes, and between TR/BDF2 and Secant/DIRK2/SBDF2/SAV2-A is wider for CH dynamics than it was for AC.

### 3.1.3 Discussion: The Source of Increased Local Error

In the metastable regime, the two terms in AC and CH (diffusion and nonlinear reaction) are both large but approximately cancel to give the slow dynamics. The methods with asymptotically (as  $\epsilon \rightarrow 0$ ) small local errors (BE, TR, BDF2) have dominant truncation errors that are pure time derivatives of the solution, which inherit this high order cancellation. The other methods which have large local errors have truncation errors that involve the reaction term individually. This imbalance amplifies the size of the error. As an example, DIRK2 applied to  $u_t = \mathcal{F}(u)$  has an error proportional to  $\mathcal{F}''u_t^2$ . From this discussion, we believe the ranking of the schemes in this work will also apply to other nonlinear problems with metastable dynamics.

## 4 Computational Results

### 4.1 Allen–Cahn

We take initial conditions in the form of a radial front

$$\tanh \frac{\sqrt{(x - \pi)^2 + (y - \pi)^2} - 2}{\epsilon\sqrt{2}}$$

and compute with  $\epsilon = 0.2, 0.1, 0.05$  and  $0.025$ . The benchmark for accuracy is the time  $T$  at which the value at the domain centre  $(\pi, \pi)$  changes from negative to positive. Except for the exponentially small (in  $\epsilon$ ) derivative discontinuities at the periodic boundaries, the dynamics approximate the sharp interface limit of curvature motion of a circle shrinking to a point at the domain centre. The expectation from asymptotic analysis of the sharp interface limit is that

$$T = 2 + O(\epsilon^2).$$

This is confirmed by the numerical solutions below. A video of the dynamics is available [27].

**Table 3** Computational results for the AC benchmark problem with fixed  $\epsilon = 0.2$  and local error tolerance  $\sigma$  varied. BE results are on the left, Eyre on the right

$\sigma$	BE			Eyre		
	$M$	CG	$E$	$M$	CG	$E$
$1e-4$	717	5348 [7.46]	0.003	2,350	14,856 [6.32]	0.047
$1e-5$	2225 (3.10)	9448 [4.24]	0.001	7351 (3.12)	28,263 [3.85]	0.014
$1e-6$	7010 (3.15)	23,017 [3.28]	0.001	23,172 (3.15)	68,148 [2.94]	0.004

Here,  $M$  is the total number of time steps taken (with the ratio to the value above in brackets), CG is the number of conjugate iterations (with the ratio to the number of time steps in brackets),  $E$  is the error in the benchmark time

**Table 4** Computational results for the AC benchmark problem with fixed local error tolerance  $\sigma = 10^{-4}$  and  $\epsilon$  varied

$\epsilon$	BE			Eyre		
	$M$	CG	$E$	$M$	CG	$E$
0.2	717	5,348 [7.46]	0.003	2350	14,856 [6.32]	0.047
0.1	1291 (1.80)	12,354 [9.57]	0.001	6463 (2.75)	44,717 [6.92]	0.069
0.05	2412 (1.87)	27,782 [11.52]	0.001	18,218 (2.83)	143,416 [7.87]	0.099
0.025	4630 (1.92)	64,884 [14.01]	*	52,595 (2.89)	497,846 [9.47]	0.141

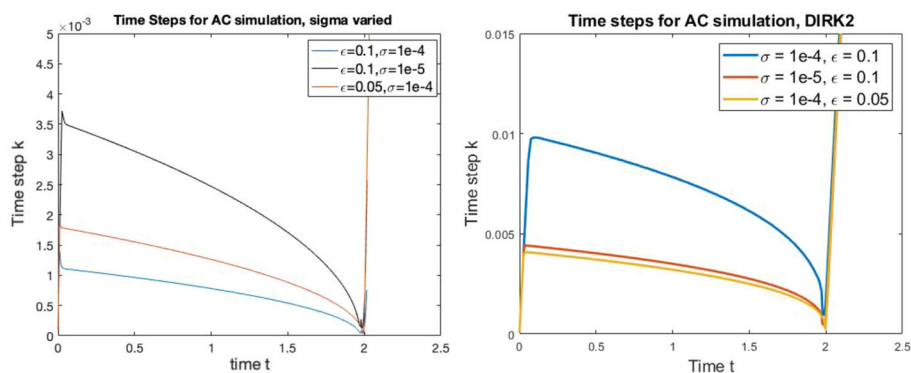
Here,  $M$  is the total number of time steps taken (with the ratio to the value above in brackets) and CG is the number of preconditioned conjugate gradient iterations (with the ratio to the number of time steps in brackets),  $E$  is the error in the benchmark time with \* denoting a result correct to three decimal places

#### 4.1.1 First Order Methods

The PCG approach is known to have bounded condition number under the scaling  $k = C\epsilon^2$  for BE with  $C < 1$  [30] and we observe good behaviour in the example below even with  $C > 1$  in the metastable regime. It is observed computationally in this work that the PCG for Eyre's method is independent of  $k$  and  $\epsilon$  although the authors are not aware of a proof in the literature. PCG counts can be used as a proxy for computational time when comparing methods.

Results of the numerical experiments in which  $\sigma$  and  $\epsilon$  were varied for BE and Eyre are shown in Tables 3 and 4. Spatial errors do not affect the digits shown in any of the computational results in this paper.

Table 3 validates the second order  $O(k^2)$  local truncation error since the number of time steps was predicted to be  $M = O(1/\sqrt{\sigma})$  for both methods with  $\epsilon$  constant, noting that  $\sqrt{10} \approx 3.16$ . Such results for other schemes and for the CH benchmark problem below are not shown, but verify the formal accuracy of the schemes. Table 4 validates the prediction of  $M = O(1/\epsilon)$  for BE and  $M = O(1/\epsilon^{3/2})$  for Eyre with  $\sigma$  constant, noting that  $2^{3/2} \approx 2.83$ . Both tables validate the prediction that for the same local tolerance  $\sigma$ , Eyre involves more computational work than BE and gives less accurate answers. CG counts for both methods are small as expected. You see (unexpectedly) that the final accuracy of BE does not seem to degrade as  $\epsilon \rightarrow 0$  for fixed  $\sigma$ . This is discussed in Sect. 5 below. Although BE does not guarantee energy stability, no step accepted by the local error tolerance exhibited an energy increase.



**Fig. 2** Time steps  $k$  for Allen–Cahn dynamics with  $\epsilon$  and  $\sigma$  varied using BE (left) and DIRK2 (right). The time steps decrease in size as the simulation approaches the topological singularity at  $t \approx 2$ . Note that for each method, the profiles  $k(t)$  have the same *shape* as  $\sigma$  and  $\epsilon$  are varied and scale with these quantities according to our theoretical predictions. In particular, note that time steps decrease more quickly for BE as  $\sigma$  is decreased but more quickly for DIRK2 as  $\epsilon$  is decreased, as we predict

**Table 5** Computational results for the AC benchmark problem with fixed local error tolerance  $\sigma = 10^{-4}$  and  $\epsilon$  varied

$\epsilon$	IMEX1		SAV1	
	$M$	$E$	$M$	$E$
0.2	3932	0.067	3936	0.067
0.1	11,110 (2.83)	0.096	11,112 (2.82)	0.096
0.05	31,676 (2.85)	0.138	31,682 (2.85)	0.138
0.025	90,748 (2.86)	0.198	90,760 (2.86)	0.198

Here,  $M$  is the total number of time steps taken (with the ratio to the value above in brackets) and  $E$  is the error in the benchmark time

For completeness, we show the time step sizes as a function of time for BE in Fig. 2 with  $\epsilon$  and  $\sigma$  varied. As mentioned in Remark 4 our predictions for the behaviour of the time steps sizes  $k$  as  $\epsilon$  and  $\sigma$  are varied describe a profile  $k(t)$ .

We repeat the  $\epsilon \rightarrow 0$  study for IMEX1 and SAV1 in Table 5. These methods require a fixed number of FFT calculations per time step to invert the constant coefficient linear implicit aspect of the schemes, with SAV1 requiring four times as many solves as IMEX1. It is seen that IMEX1 behaves almost identically to SAV1 and both are superior to Eyre's method when computational cost is considered. In the context of this study, there is no benefit from the theoretical guarantees of energy stable schemes and BE is the optimal (with our asymptotic definition) first order scheme with IMEX1 the runner up. This will remain true for other nonlinear solver strategies for BE as long as they require fewer than  $O(1/\sqrt{\epsilon})$  iterations when adaptive time steps are taken.

**Remark 5** Note that for the BE computation for  $\epsilon = 0.025$  we can still get reasonable accuracy taking  $\sigma = 10^{-2}$ . In this case, the maximum value of  $k/\epsilon^2$  is 14.6. Clearly, the theory which guarantees existence of solutions and energy decay for  $k < \epsilon^2$  [30] can be improved for metastable dynamics. This is explored in the analysis in Sects. 5 and 6 below.

### 4.1.2 Second Order Methods

The CG counts of all the nonlinear second order methods are relatively insensitive to  $\epsilon$ , similar to the first order methods shown above. We show the number of time steps used for the seven methods in Table 6, for  $\sigma = 10^{-4}$  fixed and  $\epsilon$  varied. All second order methods give at least three digits of accuracy to the benchmark time with this tolerance  $\sigma$ . The superiority of TR and BDF2 is clearly seen with  $M = O(1/\epsilon)$ , compared to  $M = O(1/\epsilon^{4/3})$  (noting that  $2^{4/3} \approx 2.52$ ) for Secant, DIRK2, SBDF2, SAV2-A and  $M = O(1/\epsilon^{5/3})$  (noting that  $2^{5/3} \approx 3.18$ ) for SAV-B as predicted above. The pattern in the number of time steps for the multi-step methods is a bit rougher due to the strict criteria we have used for adaptive time step change. As above, we see no benefit from the theoretical guarantees of energy stable schemes. Fully implicit methods TR and BDF2 are asymptotically optimal in terms of the number of time steps and are computationally optimal if the solvers require fewer than  $O(1/\sqrt[3]{\epsilon})$  iterations when adaptive time steps are taken (which appears to be the case with the Newton PCG solver we used). SBDF2 is the runner up and notably it is comparable to the fully implicit DIRK2 method but does not have the overhead of a nonlinear solve.

It is interesting to note that the slight change in the extrapolation procedure in the SAV2 schemes makes such a difference to their asymptotic performance. It is confirmation that merely considering the order of time stepping scheme and its theoretical energy stability properties is not the whole story.

## 4.2 Cahn–Hilliard

For the initial conditions we take

$$\tanh\left(\frac{r - 5/2}{\epsilon\sqrt{2}}\right) + \tanh\left(\frac{3/2 - r}{\epsilon\sqrt{2}}\right) + 1$$

with  $r = \sqrt{(x - \pi)^2 + (y - \pi)^2}$  and compute with  $\epsilon = 0.2, 0.1, 0.05$  and  $0.025$ . The dynamics approximate the sharp interface limit of two concentric circles, with the inner circle shrinking. As before, the benchmark is the time  $T$  at which the value at the domain centre  $(\pi, \pi)$  changes from negative to positive. A video of the dynamics is available [28].

### 4.2.1 First Order Methods

Results of the numerical experiments in which  $\epsilon$  is varied for the first order methods are shown in Table 7. These validate the prediction of  $M = O(1/\epsilon)$  for BE and  $M = O(1/\epsilon^2)$  for Eyre and IMEX1 with  $\sigma$  constant. As for the AC case, SAV1 behaves similarly to IMEX1 at increased computational cost. For CH, the implicit problem for BE is more difficult to solve as  $\epsilon \rightarrow 0$  with fixed  $\sigma$ , but it is still more accurate than Eyre stepping for equivalent computational cost. It will be asymptotically more efficient as long as the solution strategy for the nonlinear problem requires fewer than  $O(1/\epsilon)$  iterations with adaptive time stepping. As with AC, we see that BE does not suffer from global accuracy decrease as  $\epsilon \rightarrow 0$ .

### 4.2.2 Second Order Methods

The CG counts for the second order methods behave like those of BE with  $\epsilon$  as shown above. We show the number of time steps used for the four methods in Table 8, for  $\sigma = 10^{-4}$  fixed and  $\epsilon$  varied. The superiority of TR and BDF2 is clearly seen, consistent with  $M = O(1/\epsilon)$

**Table 6** Computational results for the second order methods applied to the AC benchmark problem with fixed local error tolerance  $\sigma = 10^{-4}$  and  $\epsilon$  varied

$\epsilon$	TR	S	BDF2	DIRK2	SBDF2	SAV2-A	SAV2-B
0.2	170	236	280	180	588	768	1,572
0.1	278 (1.64)	512 (2.16)	472 (1.69)	364 (2.02)	1384 (2.35)	1572 (2.04)	5436 (3.46)
0.05	492 (1.77)	1208 (2.36)	860 (1.82)	814 (2.24)	3260 (2.35)	3392 (2.16)	15,088 (2.78)
0.025	916 (1.86)	2,960 (2.45)	1632 (1.90)	1894 (2.33)	7600 (2.33)	7980 (2.35)	48,048 (3.18)

Shown are the total number of time steps taken (with the ratio to the value above in brackets)

**Table 7** Computational results for the first order methods applied to the CH benchmark problem with fixed local error tolerance  $\sigma = 10^{-4}$  and  $\epsilon$  varied

$\epsilon$	BE		Eyre		CG		IMEX1	
	$M$	$E$	$M$	$E$	$M$	$E$	$M$	$E$
0.2	730	5348 [7.33]	*	3055	36,684 [12.0]	0.019	9982	0.014
0.1	1184 (1.62)	24,778 [20.9]	0.001	12,751 (4.17)	190,204 [14.0]	0.021	43,332 (0.015)	0.015
0.05	2068 (1.75)	66,307 [32.1]	*	52,753 (4.13)	937,774 [17.8]	0.022	181,234 (4.18)	0.015
0.025	3768 (1.82)	198,771 [52.8]	*	215,443 (4.08)	4,504,278 [20.9]	0.022	740,366 (4.09)	0.015

Here,  $M$  is the total number of time steps taken (with the ratio to the value above in brackets), CG is the number of conjugate iterations (with the ratio to the number of time steps in brackets), and  $E$  is the error in the benchmark time with \* denoting a result correct to three decimal places



**Table 8** Computational results for the second order methods applied to the CH benchmark problem with fixed local error tolerance  $\sigma = 10^{-4}$  and  $\epsilon$  varied

$\epsilon$	TR	S	BDF2	DIRK2	SBDF2
0.2	230	534	320	378	1388
0.1	314 (1.36)	1530 (2.87)	468 (1.46)	788 (2.08)	4108 (2.96)
0.05	474 (1.51)	4722 (3.08)	748 (1.60)	1906 (2.42)	12,352 (3.01)
0.025	792 (1.67)	14,924 (3.16)	1312 (1.75)	6048 (3.17)	44,060 (3.57)

Shown are the total number of time steps taken (with the ratio to the value above in brackets)

, compared to  $M = O(1/\epsilon^{5/3})$  (noting that  $2^{5/3} \approx 3.17$ ) for Secant, DIRK2, and SBDF2 as predicted above. Results for SAV2-A are comparable to those for SBDF2. Again, the implications for the asymptotic computational superiority of fully implicit TR and BDF2 under the assumption of sufficient solver efficiency are clear.

## 5 Asymptotic Analysis of Properties of BE AC Solutions

The results in Table 4 present the accuracy for BE applied to AC with fixed local error tolerance  $\sigma = 10^{-4}$  under various values of  $\epsilon$ . It is remarkable the accuracy in the benchmark time does not degrade as  $\epsilon \rightarrow 0$ . This is unexpected, as a naïve prediction would be that the final accuracy scaled like  $M\sigma = O(\sqrt{\sigma}/\epsilon)$  where  $M$  is the number of time steps. It is clear that the resulting solution accuracy for the schemes under specified local error tolerance is a nontrivial question.

We present below the asymptotic analysis of a fully implicit BE time step (4) in two dimensions assuming the solution is in the meta-stable regime. That is,  $u_n$  is approximately described as a curve  $\mathbf{x}_n(s)$  parametrized by arc length with normal  $\hat{n}$ , dressed with the heteroclinic profile (8). We take the scaling  $k = c\epsilon$  with  $c$  independent of  $\epsilon$ , both sufficiently small depending only on the curve  $\mathbf{x}_n$ . We consider the formal asymptotics for the implicit time step  $u$  of (4) in this setting, anticipating that  $u$  will have the same local dependence  $u(s, z)$ . Using

$$\Delta \approx \frac{1}{\epsilon^2} \frac{\partial^2}{\partial z^2} + \frac{\kappa}{\epsilon} \frac{\partial}{\partial z}$$

where  $\kappa$  is the curvature of the interface, we find at leading order  $O(\epsilon^{-1})$  that  $u$  has the same homoclinic structure around a new curve  $\mathbf{x}(s)$ . That is,

$$u_{n+1} \approx g(z) + \epsilon v(z, s) \quad (10)$$

with  $g(z) = \tanh(z/\sqrt{2})$  and where we have changed coordinates to  $(s, z)$  with

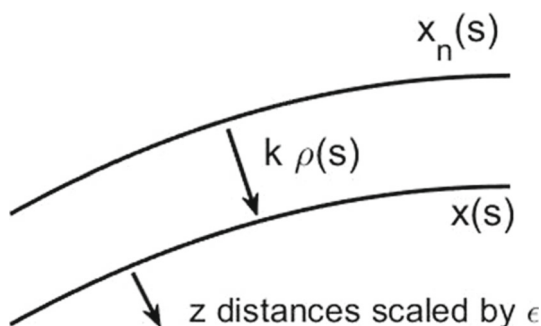
$$(x, y) = \mathbf{x}(s) + \epsilon z \hat{n}$$

based on the curve  $\mathbf{x}(s)$  after the implicit time step. In the language of Remark 2 we predict that the scheme preserves profile fidelity and show below that this is asymptotically consistent. In (10),  $v(z, s)$  is the correction to the leading order solution. We will identify the size and structure of this term below.

We take

$$\mathbf{x}_n = \mathbf{x} - k\rho(s)\hat{n} \quad (11)$$

**Fig. 3** Sketch of the asymptotic analysis of the fully implicit problem. Here,  $\rho(s)$  the average normal speed of the interface between time steps



where  $\rho$  is the average normal speed through the time step. Recalling that  $k = c\epsilon$  and the spatial scaling of  $z$ , we have

$$u_n \approx g(z - c\rho(s)). \quad (12)$$

A diagram is shown in Fig. 3. Note that the variation in normal direction appears in higher order asymptotic terms, so it is consistent in what follows to use the same  $\hat{n}$  as normal direction for both curves, i.e. the same “ $z$ ”.

Considering now the next order term  $O(1)$  in (4) with the forms (10) and (12):

$$g'c\rho + \frac{1}{2}g''c^2\rho^2 + \frac{1}{6}g'''c^3\rho^3 \approx c\kappa g' - c\mathcal{L}v \quad (13)$$

where  $\mathcal{L} := \partial^2/\partial z^2 + f'(g) \cdot$  and we have used the smallness of  $c$  for the cubic Taylor approximation of  $g(z) - g_n(z)$  on the right hand side. We consider (13) at each  $s$  in the  $L^2(\mathbb{R})$  orthogonal decomposition of  $G := \text{span}\{g'(z)\}$  and  $G^\perp$ . Note that  $g'' \in G^\perp$  (this does not depend on the specific reaction term  $f = u^3 - u$  chosen here) and  $\mathcal{L}$  has  $G$  as its kernel and has bounded inverse on  $G^\perp$  from standard Fredholm theory [18]. Thus we have  $\rho = \kappa + O(c^2)$  and  $v = O(c)$  in  $G^\perp$ . Careful examination of these results shows that the errors in  $G^\perp$  do not accumulate and are globally of size  $O(c\epsilon) = O(k)$  and so decrease as  $\epsilon \rightarrow 0$ . Global errors in interface position after  $O(1/k)$  time steps are of size  $c^2$ , independent of  $\epsilon$ . Global solution value errors due to the position error have size  $O(c^2/\epsilon) = O(k^2/\epsilon^3)$  and so it is seen that BE behaves like a second order method in this scaling. This explains the unexpected accuracy in AC BE computations as  $\epsilon \rightarrow 0$ .

**Remark 6** Note that the error estimator (7) uses  $\mathcal{F}(u)$  which sees the undamped dominant truncation error term, which is why the number of time steps behaves with  $\epsilon$  in the manner predicted in Sect. 3.1.1. Thus for BE applied to AC in the metastable regime, the estimator asymptotically over-estimates the local errors actually made.

The formal asymptotic results can also be used to show that the implicit time steps in this scaling lead to energy decrease. Neglecting the  $O(c^2)$  terms in the interface motion, we have from (11)

$$\mathbf{x}_n = \mathbf{x} - k\kappa\hat{n}.$$

Using the identities for arc length parametrized curves  $|\mathbf{x}_s| = 1$ ,  $\kappa\hat{n} = \mathbf{x}_{ss}$  and  $\mathbf{x}_s \cdot \mathbf{x}_{sss} = -\kappa^2$  it follows by taking the  $s$  derivative of the equation above and the dot product with  $\mathbf{x}_s$  at each  $s$  that

$$|\mathbf{x}_{n,s}| \geq 1 + k\kappa^2 \geq 1 = |\mathbf{x}_s|.$$

This shows that the metastable curve at time  $n$  is longer than at the next step  $n + 1$ . Since the energy  $\mathcal{E}$  is proportional to curve length to highest order in the metastable regime [21], we have shown formally that implicit time stepping for AC has the energy decay property under this time step scaling. Large, accurate, fully implicit time steps can be taken in computations validated in Sect. 4.

In the next section we show a closely related rigorous result in a radial geometry. The main result is in Proposition 1. A key ingredient is an identification of a dominant term in the space  $G$  that represents the interface motion, separate from heavily damped terms in the perpendicular space, as shown here. Care must be taken to control the size of terms which are formally neglected in this asymptotic analysis.

## 6 Rigorous Radial Analysis of AC With BE and Eyre Time Stepping

We derive rigorous asymptotic evolution of a radially symmetric profile for BE and first order Eyre-type methods for the Allen–Cahn equation in  $\mathbb{R}^2$ . Extensions to radial profiles in  $\mathbb{R}^d$  is immediate. More precisely we consider a splitting  $f = f_+ - f_-$  and study the iterative scheme

$$\begin{aligned} \frac{u - u_n}{k} &= u_{rr} + \frac{1}{r}u_r - \frac{1}{\varepsilon^2}(f_+(u) - f_-(u_n)), \quad r \in [0, \infty) \\ u_r(0) &= 0, u(\infty) = 1. \end{aligned}$$

For simplicity, we assume that  $f$  is smooth, odd about  $u = 0$ , has precisely three simple zeros at  $u = \pm 1$  and at  $u = 0$ , and tends to  $\pm\infty$  as  $u \rightarrow \pm\infty$ . This includes the classical choice of  $f(u) = u^3 - u$  but we consider other reaction terms in this class since Eyre's method can have quite different behaviour as shown in Sect. 6.2. The BE scheme corresponds to the choice  $f_- \equiv 0$  while Eyre-type schemes take  $f'_+, f'_- \geq 0$ . We pose the problem on the affine space

$$Y := \{u + 1 \in H_R^1(0, \infty) \mid \partial_r u(0) = 0.\},$$

with  $u_n \in Y$  as a given. The assumptions on  $f$  imply that the continuous 1D Allen–Cahn equation has a steady state solution

$$g_{zz} = f(u), \quad (14)$$

which is heteroclinic to  $\pm 1$ ; that is  $g \rightarrow \pm 1$  as  $z \rightarrow \pm\infty$ . Considering  $R > 1$ , we modify this  $g$  at the order  $O(e^{-1/\epsilon})$  so that  $g' = 0$  on  $(-\infty, -1/\epsilon)$  for  $\epsilon \ll 1$ . This introduces exponentially small residuals in the sequel that have no impact upon the salient results of our analysis.

We introduce  $z = \frac{r-R}{\epsilon}$ , the weighted inner product

$$\langle u, v \rangle_R := \int_{-R/\epsilon}^{\infty} u(z)v(z) (R + \epsilon z) dz,$$

and the associated spaces  $L_R^2$  and  $H_R^1$  with  $B_{H_R^1}(\delta)$  the ball that is centered at the origin with radius  $\delta$  in the space  $H_R^1$ . We rewrite the iterative equation as

$$\frac{u - u_n}{k} = \epsilon^{-2} (u_{zz} - (f_+(u) - f_-(u_n))) + \frac{\epsilon^{-1}u_z}{R + \epsilon z}, \quad (15)$$

on the domain  $z \in (-R/\epsilon, \infty)$ . We decompose  $u_n$  and  $u$  as

$$\begin{aligned} u_n &= g \left( z + \frac{R - R_n}{\epsilon} \right) + v_n, \\ u &= g(z) + v, \end{aligned}$$

where  $R_n$  and  $v_n$  are taken as given and  $R$  and  $v$  are to be determined. The profile associated to  $u_n$  is denoted  $g_n$  and observe that it admits the expansion

$$g_n = g \left( z + \frac{R - R_n}{\epsilon} \right) = g + g' \frac{R - R_n}{\epsilon} + O \left( \left( \frac{R - R_n}{\epsilon} \right)^2 \right).$$

In the sequel we will enforce the orthogonality conditions

$$\langle v, g' \rangle_R = 0, \quad \langle v_n, g'_n \rangle_R = 0, \quad (16)$$

and denote the corresponding subspaces of  $L_R^2$  by  $X^\perp$  and  $X_n^\perp$  respectively with the associated orthogonal projections  $\Pi$  and  $\Pi_n$ .

At this point the analysis of the implicit and Eyre-type schemes diverges sufficiently that we approach them distinctly.

## 6.1 Backward Euler Estimates

For BE we take  $f'_- \equiv 0$ ,  $f = f_+$ , and write the iterative map as

$$v + \frac{k}{\epsilon^2} Lv = v_n - (g - g_n) + \frac{kg'}{\epsilon(R + \epsilon z)} - \frac{k}{\epsilon^2} \mathcal{N}, \quad (17)$$

where we have introduced the linear operator

$$L := - \left( \partial_z^2 + \frac{\epsilon}{R + \epsilon z} \partial_z \right) + f'(g) = - \frac{1}{R + \epsilon z} \partial_z ((R + \epsilon z) \partial_z) + f'(g), \quad (18)$$

and the nonlinearity

$$\mathcal{N}(v) := f(g + v) - (f(g) + f'(g)v).$$

The operator  $L$  is self-adjoint in the weighted inner product for which the eigenvalue problem takes the form

$$L\psi = \frac{\lambda}{R + \epsilon z} \psi,$$

subject to  $\partial_z \psi(-R/\epsilon) = 0$  and  $\psi \rightarrow 0$  as  $z \rightarrow \infty$ . Since the profile  $g$  solves (14), it will be useful to compare  $L$  to the simpler operator

$$L_0 := -\partial_z^2 + f'(g), \quad (19)$$

arising as the linearization of (14) about  $g$  in  $L^2(\mathbb{R})$ . The operator  $L_0$  is self-adjoint on  $L^2(\mathbb{R})$ , and since  $g$  is heteroclinic with  $g' > 0$ , the Sturm-Liouville theory on  $L^2(\mathbb{R})$  implies that  $L_0$  has a simple, ground-state eigenvalue at  $\lambda = 0$  with eigenfunction  $g'$  and the remainder of the spectrum of  $L_0$  is strictly positive, in particular  $L_0$  is uniformly coercive on the space  $\{g'\}^\perp_{L^2(\mathbb{R})}$ . While  $L$  does not generically have a kernel, it does have an eigenspace with a small associated eigenvalue. However, for  $\epsilon$  sufficiently small, it inherits the coercivity of  $L_0$ .

**Lemma 1** Fix  $\epsilon_0 > 0$  sufficiently small, then there exists  $\alpha > 0$ , independent of  $R \geq 1$  and of  $\epsilon \in (0, \epsilon_0)$ , such that

$$\langle Lv, v \rangle_R \geq \alpha \|v\|_{H_R^1}^2. \quad (20)$$

for all  $v \in H_R^1$  satisfying  $\langle v, g' \rangle_R = 0$ .

**Proof** We defer the proof of  $L_R^2$  coercivity to the appendix. To extend coercivity to  $H_R^1$  we observe that

$$\langle Lv, v \rangle_R = \int_{-R/\epsilon}^{\infty} (R + \epsilon z) (|v'|^2 + f'(g)|v|^2) dz,$$

so that for any  $t \in (0, 1)$  we may write

$$\begin{aligned} \langle Lv, v \rangle_R &= t \langle Lv, v \rangle_R + (1-t) \langle Lv, v \rangle_R, \\ &\geq \int_{-R/\epsilon}^{\infty} (R + \epsilon z) (t|v'|^2 + ((1-t)\alpha - t\|f'(g)\|_{\infty})|v|^2) dz, \\ &\geq \tilde{\alpha} \|v\|_{H_R^1}^2, \end{aligned}$$

where we have introduced  $\tilde{\alpha} := \alpha/(1 + \alpha + \|f'(g)\|_{\infty}) > 0$ . Dropping the tilde, we have (20) with  $\alpha$  independent of  $\epsilon > 0$  and  $R > 1$ .  $\square$

We assume throughout our analysis that  $\|v\|_{H_R^1}$  and  $\|v_n\|_{H_R^1}$  are uniformly bounded by  $\delta \ll 1$ . Returning to (17), we denote its right-hand side as  $\mathcal{F}_{\text{BE}}$ . To have the inversion of the operator on the left-hand side be contractive the term  $\mathcal{F}_{\text{BE}}$  must be approximately orthogonal to the small eigenspace of  $L$ . As Lemma 1 shows it is sufficient to be  $L_R$ -orthogonal to  $g'$ , the kernel of  $L_0$ . To this end we determine  $R = \hat{R}_{\text{BE}}(v, v_n, R_n)$  such that  $\mathcal{F}_{\text{BE}} \in X^{\perp}$ , or equivalently

$$\langle \mathcal{F}_{\text{BE}}, g' \rangle_R = 0. \quad (21)$$

Assuming this condition has been enforced we introduce

$$M := I + \frac{k}{\epsilon^2} L,$$

and may rewrite the BE iteration in the equivalent formulation

$$v = \mathcal{G}_{\text{BE}}(v, v_n, R - R_n) := M^{-1} \Pi \mathcal{F}_{\text{BE}}(v, v_n, R - R_n). \quad (22)$$

The key step is the introduction of the operator  $\Pi$ , the orthogonal projection onto  $X^{\perp}$ . This is redundant when  $\mathcal{F}_{\text{BE}} \in X^{\perp}$ , but preserves contractivity for choices of  $(v, v_n, R - R_n)$  when it is not. Our goal is to show the function  $\mathcal{G}_{\text{BE}}$  is a contraction mapping and to develop asymptotic formula for  $R$  and  $v$ .

**Lemma 2** The function  $R = \hat{R}_{\text{BE}}$  satisfies the implicit relation

$$\frac{R - R_n}{k} = -\frac{1}{R} + \frac{k}{4R^3} - \frac{b_1 k^2}{\epsilon^2 R^3} + O\left(\delta, \frac{k^3}{\epsilon^2}, \frac{\delta^2}{\epsilon}\right). \quad (23)$$

where

$$b_1 := \frac{\|g''\|_R^2}{6\|g'\|_R^2} > 0. \quad (24)$$

Moreover we have the Lipschitz estimate

$$|\hat{R}_{\text{BE}}(v; v_n, R_n) - \hat{R}_{\text{BE}}(\tilde{v}; v_n, R_n)| \leq c \frac{k\delta}{\epsilon} \|v - \tilde{v}\|_R, \quad (25)$$

so long as  $k\delta^2 \ll \epsilon^2$ .

**Proof** Due to parity considerations, we remark that  $\|g'\|_R^2 = R\|g'\|_{L^2(\mathbb{R})}^2$ , up to exponentially small terms. For brevity, and as an element of foreshadowing, we approximate  $(R - R_n)$  by  $k$  in the  $O$ -error terms. We address the terms in  $\mathcal{F}_{\text{BE}}$  and derive the following elementary estimates,

$$\langle v_n, g' \rangle_R = \langle v_n, (g' - g'_n) \rangle_R = -\langle v_n, g'' \rangle_R \frac{R - R_n}{\epsilon} + O\left(\delta \frac{k^2}{\epsilon^2}\right), \quad (26)$$

$$\begin{aligned} \langle g - g_n, g' \rangle_R &= -\|g'\|_R^2 \left( \frac{(R - R_n)}{\epsilon} - \frac{(R - R_n)^2}{4R\epsilon} \right) + \\ &\quad + \frac{\|g''\|_R^2 (R - R_n)^3}{6\epsilon^3} + O\left(\frac{k^4}{\epsilon^3}\right), \end{aligned} \quad (27)$$

$$\left\langle \frac{g'}{R + \epsilon z}, g' \right\rangle_R = \|g'\|_{L^2(\mathbb{R})}^2 = \frac{\|g'\|_R^2}{R}, \quad (28)$$

$$|\langle \mathcal{N}, g' \rangle_R| \leq c\delta^2. \quad (29)$$

For this scheme,  $\mathcal{F}_{\text{BE}}$  depends upon  $v$  only through  $\mathcal{N}$ . Collecting terms in the orthogonality condition that are linear in  $R - R_n$  and identifying relevant higher order terms yields the relation

$$\frac{R - R_n}{k} = -\frac{1}{R} + \frac{(R - R_n)^2}{4Rk} + \frac{b_1(R - R_n)^3}{k\epsilon^2} + O\left(\delta, \frac{k^3}{\epsilon^2}, \frac{\delta^2}{\epsilon}\right) \quad (30)$$

where  $b_1$  is given in (24). Under the assumptions on  $k$  and  $\delta$  we have the leading order result  $R - R_n = -k/R$ . Substituting this relation into (30) yields the result (23).

To obtain the Lipschitz estimate we observe from the estimates above that

$$|\hat{R}_{\text{BE}}(v) - \hat{R}_{\text{BE}}(\tilde{v})| \leq c \frac{k}{\epsilon \|g'\|_R^2} |\langle \mathcal{N}(v), g' \rangle_R - \langle \mathcal{N}(\tilde{v}), \tilde{g}' \rangle_R|.$$

The nonlinearity satisfies the Lipschitz properties

$$\|\mathcal{N}(v) - \mathcal{N}(\tilde{v})\|_R \leq c\delta \|v - \tilde{v}\|_R,$$

while

$$\|g' - \tilde{g}'\|_R \leq c \frac{|\hat{R}_{\text{BE}}(v) - \hat{R}_{\text{BE}}(\tilde{v})|}{\epsilon}.$$

Adding and subtracting  $\langle \mathcal{N}(\tilde{v}), g' \rangle_R$  and using (29), we arrive at the estimates

$$|\hat{R}_{\text{BE}}(v) - \hat{R}_{\text{BE}}(\tilde{v})| \leq c \left( \frac{k\delta}{\epsilon} \|v - \tilde{v}\|_R + \frac{k\delta^2}{\epsilon^2} |\hat{R}_{\text{BE}}(v) - \hat{R}_{\text{BE}}(\tilde{v})| \right).$$

Imposing the condition  $k\delta^2 \ll \epsilon^2$  yields (25).  $\square$

To establish bounds on the map  $\mathcal{G}_{\text{BE}}$  defined in (22) we apply  $M$  to both sides of the relation and take the  $L_R^2$  inner product with respect to  $\mathcal{G}_{\text{BE}}$ . Using the coercivity estimate (20) we find

$$\|\mathcal{G}_{\text{BE}}\|_R^2 + \alpha \frac{k}{\epsilon^2} \|\mathcal{G}_{\text{BE}}\|_{H_R^1}^2 \leq \|\Pi \mathcal{F}_{\text{BE}}\|_R \|\mathcal{G}_{\text{BE}}\|_R.$$

Taking  $v, v_n \in B_{H_R^1}(\delta)$  for  $\delta \ll 1$  and recalling that the projection  $\Pi$  crucially cancels the leading order term in  $g - g_n$ , we estimate

$$\|\mathcal{G}_{\text{BE}}\|_R + \alpha \frac{k}{\epsilon^2} \|\mathcal{G}_{\text{BE}}\|_{H_R^1} \leq c \left( \delta + \frac{k^2}{\epsilon^2} + k + \frac{k}{\epsilon^2} \delta^2 \right). \quad (31)$$

For the BE system we examine distinguished limits  $k = \epsilon^s$ , for  $s \in (1, 2)$ , which we call the large time-step regime, for which the  $H_R^1$  term is dominant on the left-hand side of (31). We drop the  $L_R^2$  term to find,

$$\|\mathcal{G}_{\text{BE}}\|_{H_R^1} \leq c \left( \delta \epsilon^{2-s} + \epsilon^s + \delta^2 \right).$$

Taking  $\delta = \epsilon^{s'}$  for any  $s' > \max\{s/2, 2(s-1)\}$  then we determine that

$$\|\mathcal{G}_{\text{BE}}\|_{H_R^1} \leq c(\epsilon^{2-s+s'} + \epsilon^s + \epsilon^{2s'}) \leq \delta,$$

for  $\epsilon$  sufficiently small. In particular, since  $s > \max\{s/2, 2(s-1)\}$  in the large time-stepping regime, we may take  $\delta = k = \epsilon^s$ , so that, viewing  $\mathcal{G}_{\text{BE}}$  as a map on  $(v, v_n)$ , we have  $\mathcal{G}_{\text{BE}} : B_{H_R^1}(k) \times B_{H_R^1}(k) \mapsto B_{H_R^1}(k)$ , for all  $s$  in the large time-step regime.

**Proposition 1** Fix  $1 < s < 2$ , then in the distinguished limit  $k = \epsilon^s$  the function  $\mathcal{G}_{\text{BE}}$  defined in (22) with  $R := R_{n+1} = \hat{R}_{\text{BE}}(v; v_n, R_n)$  maps  $B_{H_R^1}(k) \times B_{H_R^1}(k)$  into  $B_{H_R^1}(k)$  and is a strict contraction. In particular it has a unique solution  $v \in B_{H_R^1}(k)$ , denoted by  $v_{n+1}$  which satisfies

$$\left\| v_{n+1} - \frac{k}{R^2} L \Pi g'' \right\|_{H_R^1} \leq c \frac{\epsilon^2}{k} \|v_n\|_R + O(k^2).$$

In particular there exists  $c > 0$  such that for all  $v_0 \in B_{H_R^1}(ck)$  and  $R_0 > 1$  the sequence  $\{(v_n, R_n)\}_{n=1}^N$  satisfies  $v_n \in B_{H_R^1}(ck)$  while  $\{R_n\}_{n=0}^N$  satisfies the backwards Euler iteration

$$\frac{R_{n+1} - R_n}{k} = -\frac{1}{R} - \frac{b_1 k^2}{\epsilon^2 R^3} + O(k), \quad (32)$$

where  $b_1 > 0$  is given by (24). Here  $N$  is the iteration number such that  $R_N > 1$  and  $R_{N+1} < 1$ .

**Proof** We have established the mapping property. To establish the contractivity we must control the impact of  $f$  upon the projection  $\Pi$  through the motion of the front  $R$ . We assume that  $v, \tilde{v}, v_n \in B_{H_R^1}(k)$  and denote  $R = R(v)$  and  $\tilde{R} = R(\tilde{v})$ , with the associated front profiles denoted by  $g$  and  $\tilde{g}$ . The estimate (25) establishes that  $\hat{R}_{\text{BE}}$  is Lipschitz with constant  $ck\delta/\epsilon$ , which in the large time-step regime reduces to  $ck^2/\epsilon$ . Following the proof of (25) we find that

$$\|\mathcal{F}_{\text{BE}}(v) - \mathcal{F}_{\text{BE}}(\tilde{v})\|_{H_R^1} \leq c \frac{k^2}{\epsilon^2} \|v - \tilde{v}\|_{H_R^1}. \quad (33)$$

In the large time-step regime, using (20) we deduce the bound

$$\|M^{-1}\Pi f\|_{H_R^1} \leq \alpha^{-1} \frac{\epsilon^2}{k} \|f\|_R \quad (34)$$

We wish to obtain a bound on the difference of  $\mathcal{G}_{\text{BE}}$  at two values of  $v$ :

$$\mathcal{G}_{\text{BE}}(v, v_n) - \mathcal{G}_{\text{BE}}(\tilde{v}, v_n) = M^{-1}\Pi\mathcal{F}_{\text{BE}} - \tilde{M}^{-1}\tilde{\Pi}\tilde{\mathcal{F}}_{\text{BE}}.$$

We first bound the difference

$$\mathbf{g}_{\text{BE}} := (M^{-1}\Pi - \tilde{M}^{-1}\tilde{\Pi})\mathcal{F}_{\text{BE}}. \quad (35)$$

The analysis is complicated by the fact that  $M$  is only uniformly invertible on the range of  $\Pi$ . To factor these projected inverses we act with  $M$ , observing

$$M\mathbf{g}_{\text{BE}} = (\Pi - M\tilde{M}^{-1}\tilde{\Pi})\mathcal{F}_{\text{BE}} = (\Pi\tilde{M} - M)\tilde{M}^{-1}\tilde{\Pi}\mathcal{F}_{\text{BE}} + \Pi(I - \tilde{\Pi})\mathcal{F}_{\text{BE}}, \quad (36)$$

where we used that fact that  $\tilde{M}\tilde{M}^{-1}\tilde{\Pi} = \tilde{\Pi}$  and hence  $\tilde{M}\tilde{M}^{-1}\tilde{\Pi} + (I - \tilde{\Pi}) = I$ . Since the right-hand side of (36) lies in the range of  $\Pi$  we may invert boundedly,

$$\Pi\mathbf{g}_{\text{BE}} = M^{-1}\Pi(\tilde{M} - M)\tilde{M}^{-1}\tilde{\Pi}\mathcal{F}_{\text{BE}} + M^{-1}\Pi(I - \tilde{\Pi})\mathcal{F}_{\text{BE}}. \quad (37)$$

To recover the whole  $\mathbf{g}_{\text{BE}}$  we act with  $(I - \Pi)$  on (35) obtaining

$$(I - \Pi)\mathbf{g}_{\text{BE}} = -(I - \Pi)\tilde{M}^{-1}\tilde{\Pi}\mathcal{F}_{\text{BE}} = -(\tilde{\Pi} - \Pi)\tilde{M}^{-1}\tilde{\Pi}\mathcal{F}_{\text{BE}}. \quad (38)$$

Adding (38) to (37) yields a regularized expression that accounts for the shifts in the projections

$$\mathbf{g}_{\text{BE}} = M^{-1}\Pi(\tilde{M} - M)\tilde{M}^{-1}\tilde{\Pi}\mathcal{F}_{\text{BE}} + M^{-1}\Pi(\Pi - \tilde{\Pi})\mathcal{F}_{\text{BE}} + (\Pi - \tilde{\Pi})\tilde{M}^{-1}\tilde{\Pi}\mathcal{F}_{\text{BE}}. \quad (39)$$

The operators  $M^{-1}\Pi$  and  $\tilde{M}^{-1}\tilde{\Pi}$  are bounded using (34), while

$$\begin{aligned} \|\tilde{M} - M\|_{R*} &= \frac{k}{\epsilon} \|f'(g) - f'(\tilde{g})\|_{R*} \leq c\|g - \tilde{g}\|_{R*}, \\ &\leq c \frac{|R - \tilde{R}|}{\epsilon} \leq c \frac{k^2}{\epsilon^2} \|v - \tilde{v}\|_{H_R^1}, \end{aligned}$$

where  $\|\cdot\|_{R*}$  denotes the operator norm from  $L_R^2$  into itself. The projections satisfy

$$\begin{aligned} \|(\Pi - \tilde{\Pi})f\|_R &= \|g'\langle g', f \rangle_R - \tilde{g}'\langle \tilde{g}', f \rangle_R\|_R, \\ &\leq c \frac{k^2}{\epsilon^2} \|v - \tilde{v}\|_{H_R^1} \|f\|_R + c \frac{\epsilon^2 k^2}{k \epsilon^2} \|v - \tilde{v}\|_{H_R^1} \end{aligned}$$

Applying these estimates to (39) and using (31) to estimate  $\Pi\mathcal{F}_{\text{BE}}$  we obtain

$$\begin{aligned} \|\mathbf{g}_{\text{BE}}\|_{H_R^1} &\leq c \left( \frac{\epsilon^4 k^2}{k^2 \epsilon^2} + \frac{\epsilon^2 k^2}{k \epsilon^2} \right) \|v - \tilde{v}\|_{H_R^1} \|\Pi\mathcal{F}_{\text{BE}}\|_{L^2}, \\ &\leq c (\epsilon^2 + k) \frac{k^2}{\epsilon^2} \|v - \tilde{v}\|_{H_R^1}. \end{aligned} \quad (40)$$

Finally we write

$$\mathcal{G}_{\text{BE}}(v, v_n) - \mathcal{G}_{\text{BE}}(\tilde{v}, v_n) = \mathbf{g}_{\text{BE}} + \tilde{M}^{-1}\tilde{\Pi}(\mathcal{F}_{\text{BE}} - \tilde{\mathcal{F}}_{\text{BE}}),$$



and using (33), (40) estimate

$$\|\mathcal{G}_{\text{BE}}(v, v_n) - \mathcal{G}_{\text{BE}}(\tilde{v}, v_n)\|_{H_R^1} \leq c \left( \frac{k^3}{\epsilon^2} + k \right) \|v - \tilde{v}\|_{H_R^1},$$

which is contractive so long as  $k \ll \epsilon^{\frac{2}{3}}$  which holds with the large time-step regime.

Within the large time-step regime the leading order iteration (30) simplifies as  $k \ll k^2/\epsilon^2$  and the dominant correction is given by the  $b_1$  term. To compare to standard notation we rewrite the regime as  $\epsilon^2 \ll k = \delta \ll 1$  and replace the internal parameter  $\delta$  with  $k$ , the result is the large time-step iteration (32).  $\square$

## 6.2 Eyre-Type Iterations

For an Eyre iteration the map (15) takes the form

$$v + \frac{k}{\epsilon^2} L_+ v = v_n - (g - g_n) + \frac{k g'}{\epsilon(R + \epsilon z)} + \frac{k}{\epsilon^2} (\mathcal{R} - \mathcal{N}), \quad (41)$$

where we have introduced the Eyre linear operator

$$L_+ := - \left( \partial_z^2 + \frac{\epsilon}{R + \epsilon z} \partial_z \right) + f'_+(g) = - \frac{1}{R + \epsilon z} \partial_z ((R + \epsilon z) \partial_z) + f'_+(g), \quad (42)$$

the explicit-term residual

$$\mathcal{R}(v, v_n) := f_-(g_n) - f_-(g) + f'_-(g_n) v_n,$$

and the nonlinearity

$$\mathcal{N}(v, v_n) := \mathcal{N}_+(v) - \mathcal{N}_-(v_n),$$

which we further decompose into implicit and explicit parts

$$\begin{aligned} \mathcal{N}_+(v) &:= f_+(g + v) - (f_+(g) + f'_+(g)v), \\ \mathcal{N}_-(v_n) &:= f_-(g_n + v_n) - (f_-(g_n) + f'_-(g_n)v_n). \end{aligned}$$

The operator  $L_+$  is self-adjoint in the weighted inner product for which the eigenvalue problem takes the form

$$L_+ \psi = \frac{\lambda}{R + \epsilon z} \psi,$$

subject to  $\partial_z \psi(-R/\epsilon) = 0$  and  $\psi \rightarrow 0$  as  $z \rightarrow \infty$ . The coercivity estimate is substantially simpler than for BE as the operator  $L_+$  is strictly positive without constraint.

**Lemma 3** *There exists  $\alpha_+ > 0$ , independent of  $R \geq 1$ , such that*

$$\langle L_+ v, v \rangle_R \geq \alpha_+ \|v\|_{H_R^1}^2. \quad (43)$$

for all  $v \in H_R^1$ .

**Proof** Since  $f'_+ \geq 0$  the normalized ground-state eigenfunction  $\psi_0$  of  $L_+$ , satisfies

$$\lambda_0^+ = \langle L_+ \psi_0, \psi_0 \rangle_R = \int_{-R/\epsilon}^{\infty} ((\partial_z \psi_0)^2 + f'_+(g) \psi_0^2) (R + \epsilon z) dz > 0.$$

Since the ground-state eigenvalue is strictly positive, this establishes the  $L_R^2$  coercivity of  $L_+$  with  $\alpha_+ = \lambda_0^+$ . The  $H_R^1$  coercivity follows as in Lemma 1.  $\square$

We assume throughout our analysis that  $\|v\|_{H_R^1}$  and  $\|v_n\|_{H_R^1}$  are uniformly bounded by  $\delta \ll 1$ . We denote the right-hand side of (41) by  $\mathcal{F}_E$  and introduce

$$M_+ := I + \frac{k}{\epsilon^2} L_+,$$

which is strictly contractive on the full space  $L_R^2$ , and re-write the Eyre iteration as

$$v = \mathcal{G}_E(v, v_n, R - R_n) := M_+^{-1} \Pi \mathcal{F}_E(v, v_n, R - R_n). \quad (44)$$

For the Eyre iteration the role of the projection  $\Pi$  is diminished as  $M_+$  is contractive without it. Our goal is to show the existence of a map  $R = \hat{R}_E(v, v_n, R_n)$ , for which

$$\langle \mathcal{F}_E, g' \rangle_R = 0, \quad (45)$$

to establish the contractive mapping properties of  $\mathcal{G}_E$ , and to develop asymptotic formula for  $R$  and  $v$ . We do this in the long time-stepping regime,  $k \gg \epsilon^2$ , which has no lower bound for the Eyre scheme.

**Lemma 4** Assume  $k \gg \epsilon^2$ . There exists a smooth function  $\hat{R}_E : B_{H_R^1}(\delta) \times B_{H_R^1}(\delta) \times \mathbb{R} \mapsto \mathbb{R}$  such that the profile  $g = g(z; R)$  satisfies (45). The function  $R = \hat{R}_E$  satisfies the implicit relation

$$R - R_n = -\frac{\epsilon^2}{c_- R} + O\left(\epsilon^3, \delta\epsilon, \frac{\epsilon^4}{k}\right). \quad (46)$$

where we have introduced the leading order Eyre time constant

$$c_- := \frac{\langle f'_-(g)g', g' \rangle_R}{\|g'\|_R^2} > 0 \quad (47)$$

when  $f'_- \neq 0$ . Moreover we have the Lipschitz estimate

$$|\hat{R}_E(v; v_n, R_n) - \hat{R}_E(\tilde{v}; v_n, R_n)| \leq c\epsilon\delta\|v - \tilde{v}\|_R, \quad (48)$$

so long as  $\delta \ll 1$ .

**Proof** Due to parity considerations, we remark that  $\|g'\|_R^2 = R\|g'\|_{L^2(\mathbb{R})}^2$ , up to exponentially small terms. For brevity, and as an element of foreshadowing, we approximate  $(R - R_n)$  by  $\epsilon^2$  in the  $O$ -error terms. Addressing the terms in  $\mathcal{F}_{BE}$  one by one, we record

$$\langle v_n, g' \rangle_R = \langle v_n, (g' - g'_n) \rangle_R = O(\delta\epsilon) \quad (49)$$

$$\langle g - g_n, g' \rangle_R = -\|g'\|_R^2 \frac{(R - R_n)}{\epsilon} + O(\epsilon^3), \quad (50)$$

$$\left\langle \frac{g'}{R + \epsilon z}, g' \right\rangle_R = \|g'\|_{L^2(\mathbb{R})}^2 = \frac{\|g'\|_R^2}{R}, \quad (51)$$

$$\langle \mathcal{R}, g' \rangle_R = \frac{\langle f'_-(g)g', g' \rangle_R (R - R_n)}{\epsilon} + \langle f'_-(g)v_n, g' \rangle_R + O(\epsilon^2, \epsilon\delta), \quad (52)$$

$$|\langle \mathcal{N}, g' \rangle_R| \leq c\delta^2. \quad (53)$$

With these reductions we can simplify the orthogonality condition, identifying terms that are linear in  $R - R_n$  and most relevant higher order terms. The result is the balance

$$\frac{R - R_n}{k} \left(1 + \frac{c_- k}{\epsilon^2}\right) = -\frac{1}{R} - \frac{\langle f'_-(g)g', v_n \rangle_R}{\epsilon\|g'\|_R^2} + O\left(\epsilon, \delta, \frac{\delta^2}{\epsilon}\right), \quad (54)$$

where  $c_-$ , introduced in (47) is positive since  $f'_- \geq 0$  by assumption. The largest terms and error terms come from the residual, and we kept the lower order constant on the left-hand side to emphasize that in the long time-stepping regime, the residual dominates the natural time-step term. Indeed, the iteration is independent of step size,  $k$ , given at leading order by (54).

To obtain the Lipschitz estimate we observe from the bounds above that dependence of  $\hat{R}_E$  on  $v$  arises from the balance of the linear  $R - R_n$  term in the residual against the nonlinearity. Since both these terms are multiplied by  $k/\epsilon^2$  this factor cancels and we have the balance

$$|\hat{R}_E(v) - \hat{R}_E(\tilde{v})| \leq \frac{c\epsilon}{\langle f'_-(g)g', g' \rangle_R} |\langle \mathcal{N}(v), g' \rangle_R - \langle \mathcal{N}(\tilde{v}), \tilde{g}' \rangle_R|.$$

The nonlinearity satisfies the Lipschitz properties

$$\|\mathcal{N}(v, v_n) - \mathcal{N}(\tilde{v}, v_n)\|_R \leq c\delta \|v - \tilde{v}\|_R,$$

while

$$\|g' - \tilde{g}'\|_{L^2} \leq c \frac{|\hat{R}_E(v) - \hat{R}_E(\tilde{v})|}{\epsilon}.$$

Adding and subtracting  $\langle \mathcal{N}(\tilde{v}, v_n), g' \rangle_R$  and using (53), we arrive at the estimates

$$|\hat{R}_E(v) - \hat{R}_E(\tilde{v})| \leq c \left( \epsilon\delta \|v - \tilde{v}\|_R + \delta^2 |\hat{R}_{BE}(v) - \hat{R}_{BE}(\tilde{v})| \right).$$

Imposing the condition  $\delta \ll 1$  yields (48).  $\square$

We may now establish the main result on the Eyre sequence.

**Proposition 2** *There exists  $c > 0$  such that for any  $k \gg \epsilon^2$  the function  $\mathcal{G}_E$  defined in (44) with  $R := \hat{R}_E(v; v_n, R_n)$  maps  $B_{H_R^1}(c\epsilon) \times B_{H_R^1}(c\epsilon)$  into  $B_{H_R^1}(\mathcal{G}_E(0, v_n), \epsilon^2)$  and is a strict contraction, satisfying*

$$\|\mathcal{G}_E(v, v_n) - \mathcal{G}_E(\tilde{v}, v_n)\|_{H_R^1} \leq c\epsilon^2 \|v - \tilde{v}\|_{H_R^1}. \quad (55)$$

*In particular  $\mathcal{G}_E$  has a unique fixed point in that set, which we denote  $v_{n+1}$ . Moreover, if the Eyre balance parameter*

$$\gamma := \|L_+^{-1} \Pi \circ f'_-(g)g'\|_{H_R^{1*}} < 1 \quad (56)$$

*then for any  $\rho \in (0, 1)$  there exists  $c > 0$  such that for all  $v_0 \in B_{H_R^1}(c\epsilon)$  and  $R_0 > 1$ , the sequence  $\{(v_n, R_n)\}_{n=1}^N$  satisfies  $v_n \in B_{H_R^1}(c\epsilon)$  and*

$$\frac{R_{n+1} - R_n}{\epsilon^2} = -\frac{c_E}{R_{n+1}} + O(\epsilon^{1-\rho}).$$

*where the Eyre number,  $c_E$ , is defined by*

$$c_E := \frac{\|g'\|_R^2}{\langle f'_-(g)g', g' \rangle_R + \langle K_+ L_+^{-1} \Pi f'_-(g)g', f'_-(g)g' \rangle_R} > 0, \quad (57)$$

*where  $K_+ > 0$  is defined in (65) and  $K_+ L_+^{-1} \Pi > 0$  is self-adjoint.*

**Proof** To establish the contractivity of  $\mathcal{G}_E$  we follow the arguments for backward Euler, sketching only the differences. We introduce

$$g_E := (M_+^{-1}\Pi - \tilde{M}_+^{-1}\tilde{\Pi})\mathcal{F}_E; \quad (58)$$

and derive the expression

$$\begin{aligned} g_E &= M_+^{-1}\Pi(\tilde{M}_+ - M_+)\tilde{M}_+^{-1}\tilde{\Pi}\mathcal{F}_E \\ &\quad + M_+^{-1}\Pi(\Pi - \tilde{\Pi})\mathcal{F}_E + (\Pi - \tilde{\Pi})\tilde{M}_+^{-1}\tilde{\Pi}\mathcal{F}_E. \end{aligned} \quad (59)$$

The operators  $M_+^{-1}\Pi$  and  $\tilde{M}_+^{-1}\tilde{\Pi}$  are bounded as  $L_+$  has no small eigenvalues. Using (48) we estimate

$$\begin{aligned} \|\tilde{M}_+ - M_+\|_{R*} &= \frac{k}{\epsilon} \|f'_+(g) - f'_+(\tilde{g})\|_{R*} \leq c\|g - \tilde{g}\|_{R*}, \\ &\leq c \frac{|R - \tilde{R}|}{\epsilon} \leq c\delta\|v - \tilde{v}\|_{H_R^1}. \end{aligned}$$

Similarly the projections satisfy

$$\begin{aligned} \|(\Pi - \tilde{\Pi})\|_{R*} &= \|g' \langle g', \cdot \rangle_R - \tilde{g}' \langle \tilde{g}', \cdot \rangle_R\|_{R*} \\ &\leq c\delta\|v - \tilde{v}\|_{H_R^1} \end{aligned}$$

Applying these estimates to (58,59) and following the proof of (48) to estimate  $\Pi\mathcal{F}_E$  we obtain

$$\begin{aligned} \|g_E\|_{H_R^1} &\leq c \frac{\epsilon^2}{k} \delta \|v - \tilde{v}\|_{H_R^1} \|\Pi\mathcal{F}_{BE}\|_{L^2}, \\ &\leq c \left( \frac{\epsilon^2 \delta^2}{k} + \frac{\epsilon^4 \delta}{k} + \epsilon \delta + \delta^2 \right) \|v - \tilde{v}\|_{H_R^1}. \end{aligned} \quad (60)$$

Finally we write

$$\mathcal{G}_E(v, v_n) - \mathcal{G}_E(\tilde{v}, v_n) = g_E + \tilde{M}^{-1}\tilde{\Pi}(\mathcal{F}_E - \tilde{\mathcal{F}}_E), \quad (61)$$

and estimate the  $\mathcal{F}_E$  term from which the dominant contribution comes from the residual

$$\begin{aligned} \|\mathcal{F}_E - \tilde{\mathcal{F}}_E\|_R &\leq c \frac{k}{\epsilon^2} \|f_-(g) - f_-(\tilde{g})\|_R \leq c \frac{k}{\epsilon^2} \frac{|R - \tilde{R}|}{\epsilon}, \\ &\leq c \frac{k\delta}{\epsilon} \|v - \tilde{v}\|_R, \end{aligned}$$

where we used (48) in the last inequality. In particular we deduce that

$$\|\tilde{M}_+^{-1}\tilde{\Pi}(\mathcal{F}_E - \tilde{\mathcal{F}}_E)\|_{H_R^1} \leq c\epsilon\delta\|v - \tilde{v}\|_R. \quad (62)$$

Combining (60), (62) and (61), imposing  $\delta = \epsilon$ , and using  $k \gg \epsilon^2$  we arrive at strict contractivity on  $B_{H_R^1}(c\delta)$  for any fixed  $c > 0$ .

To establish bounds on the the fixed point  $v_{n+1}$  of  $\mathcal{G}_E(\cdot; v_n)$  we observe from (43) that in the large time-stepping regime

$$\|M_+^{-1}\Pi\|_{H_R^{1*}} \leq c \frac{\epsilon^2}{k}.$$

Using this result we expand

$$\Pi \mathcal{F}_E = \frac{k}{\epsilon^2} \Pi \left( f'_-(g) g' \frac{R - R_n}{\epsilon} + f'_-(g) v_n \right) + O(\delta, \epsilon^2, k).$$

Inverting  $M_+$  we find, at leading order

$$v_{n+1} = \frac{R - R_n}{\epsilon} L_+^{-1} \Pi f'_-(g) g' + L_+^{-1} \Pi f'_-(g) v_n + O(\epsilon^2, \delta^2)$$

In particular we deduce that

$$\left\| v_{n+1} - \frac{R - R_n}{\epsilon} L_+^{-1} \Pi f'_-(g) g' \right\|_{H_R^1} \leq \gamma \|v_n\|_{H_R^1} + O(\epsilon^2, \delta^2).$$

Arguing inductively, since the Eyre balance parameter  $\gamma < 1$  and the functions  $\|L_+^{-1} \Pi f'_-(g) g'\|_{H_R^1}$  are uniformly bounded for all  $R \geq 1$ , we deduce that if  $\delta := \|v_0\|_{H_R^1} = O(\epsilon)$  then the sequences  $\{(R - R_n)\epsilon^{-2}\}_0^N$  and  $\{\epsilon^{-1}\|v_n\|_{H_R^1}\}_0^N$  are uniformly bounded, independent of  $\epsilon \ll 1$  and  $k \gg \epsilon^2$  for all  $n \leq N$  so long as  $R_n > 1$  for all  $n = 0, \dots, N$ .

To improve this bound we require Lipschitz estimates on the  $v_n$  component of  $\mathcal{G}_E$ . To this end we find

$$\begin{aligned} \|\mathcal{G}_E(v; v_n) - \mathcal{G}_E(v; \tilde{v}_n)\|_{H_R^1} &\leq \|M_+^{-1} \Pi(I + \frac{k}{\epsilon^2} f'_-(g))\|_{H_R^{1*}} \|v_n - \tilde{v}_n\|_{H_R^1}, \\ &\leq \left( \gamma + O\left(\frac{\epsilon^2}{k}\right) \right) \|v_n - \tilde{v}_n\|_{H_R^1}. \end{aligned} \quad (63)$$

Here we introduce the quasi-steady parameter  $\rho \in (0, 1)$ . Since  $|R_n - R_m| = O(\epsilon^{2-\rho})$  for  $|n - m| \leq N_\rho := \epsilon^{-\rho}$  we infer that

$$\left\| L_{+,n}^{-1} \Pi f'_-(g_n) f'_n - L_{+,m}^{-1} \Pi f'_-(g_m) g'_m \right\|_{H_R^1} \leq c\sqrt{\epsilon},$$

for all such  $n$  and  $m$ . For  $n > N_\rho$  we define the quasi-equilibrium

$$v_{n*} := \frac{R_n - R_{n-1}}{\epsilon} E_n(z)$$

where  $E_n$  is the  $R = R_n$  translate of

$$E := K_+ L_+^{-1} \Pi f'_-(g) g'. \quad (64)$$

Here the self-adjoint operator

$$K_+ := (I - L_+ \Pi \circ f'_-(g))^{-1} > 0, \quad (65)$$

is well defined since  $\|L_+ \Pi \circ f'_-(g)\|_{H_R^{1*}} = \gamma < 1$  by assumption. Using the Lipschitz property (63) of  $\mathcal{G}_E$  and the quasi equilibrium relation

$$\|v_{n*} - \mathcal{G}_E(v_{n*}; v_{n*})\|_{H_R^1} = O(\epsilon^2),$$

we deduce that

$$\|v_{k+1} - v_{n*}\|_{H_R^1} \leq \gamma \|v_k - v_{n*}\|_{H_R^1} + O(\epsilon^{2-\rho}).$$

for  $k = n - m_\epsilon, \dots, n$ . Since  $\gamma^{N_\rho} \ll \epsilon$  we deduce from an inductive argument that

$$\left\| v_n - \frac{R_n - R_{n-1}}{\epsilon} E_n \right\|_{H_R^1} = O(\epsilon^{2-\rho}),$$

for all  $n > N_s$ . Inserting this result in (46) we arrive at the leading order Eyre iteration (57).  $\square$

**Remark 7** There are two examples of particular relevance

$$f(u) = u^3 - u,$$

with the decomposition  $f_+ = (1 + \beta)u^3$  and  $f_- = u + \beta u^3$  for  $\beta > 0$ . The choice  $\beta = 0$  is classical and very degenerate, as in this case  $f'_-(u) = 1$  and the corresponding Eyre balance parameter  $\gamma$ , defined in (56) is zero, and the Eyre number, (57) is 1. In this case it is possible to rewrite Eyre's method as backward Euler with a rescaled time. In particular the slow convergence to equilibrium will not be in evidence. For larger values of  $\beta$  the balance parameter increases from zero and the Eyre number decreases from 1. As the balance parameter increases through 1 we anticipate enhanced slowing of the front profile as the Eyre number tends to zero. The choice of non-zero  $\beta$  can be viewed as spurious, a deliberate attempt to foul the method. A more robust example of non-zero balance arises naturally through the model

$$f(u) = u^5 - \beta u^3,$$

with  $\beta \geq 1$ . This suggests the optimal decomposition  $f_+ = u^5$  and  $f_- = \beta u^3$ . Here, unambiguously, increasing  $\beta$  increases the balance parameter and will lead to non-trivial enhanced slow-down with potential instability as  $\gamma$  increases through 1. These analytic predictions are validated in a computational study below.

**Remark 8** To leading order, in the large time-stepping regime  $k \gg \epsilon^2$ , the Eyre iteration recovers backward Euler with the substitution  $k \mapsto c_E \epsilon^2$ . This reduces to the exact result for the case  $f(u) = u^3 - u$  and  $f_-(u) = u$ , for which  $f'_- = 1$ , as the Eyre constant reduces to 1 since  $\Pi f_-(g)g' = \Pi g' = 0$ .

The strong contractivity of  $\mathcal{G}_E$  with respect to  $v$ , given in (55), arises from the strong convexity with respect to  $v$ , but the slow evolution and marginal convergence to the quasi-equilibrium, given in (63) arises from the balance between the implicit and explicit terms. The parameter  $\gamma$  measures this balance, with the quasi-equilibrium structure lost as  $\gamma$  increases towards 1. Indeed, since  $\|K_+\|_{H_R^{1*}} \sim (1 - \gamma)^{-1}$ , the Eyre constant will generically tend to zero as  $\gamma \rightarrow 1$ .

### 6.2.1 Computational Validation of Remark 7

We perform computations for AC with the non-classical  $f(u) = u^5 - u^3$  (which also leads to meta-stable dynamics of curvature motion) using the same initial conditions and accuracy criteria as described in Sect. 4.1. BE performs almost identically to the results shown in Tables 3 and 4 for the classical  $f(u) = u^3 - u$  in terms of accuracy and variation of time steps with  $\epsilon$  and  $\sigma$ . This matches the theory in Sect. 6.1 which can be summarized as BE has profile fidelity when  $k = o(\epsilon)$ .

When Eyre's method is applied to the dynamics with  $f(u) = u^5 - u^3$ , with the natural splitting suggested in Remark 7, profile fidelity is lost as predicted. The formal prediction of

**Table 9** Computational results for the AC benchmark problem with fixed local error tolerance  $\sigma = 10^{-4}$  and  $\epsilon$  varied, using Eyre's method with reaction term  $f(u) = u^5 - u^3$

$\epsilon$	Eyre with $f(u) = u^5 - u^3$	
	$M$	$E$
0.2	5726	0.001
0.1	21,947 (3.83)	0.005
0.05	86,499 (3.94)	0.007
0.025	343,525 (3.97)	0.007

Here,  $M$  is the total number of time steps taken (with the ratio to the value above in brackets) and  $E$  is the error in the benchmark time

$k = O(\epsilon^{3/2})$  which was seen computationally for  $f(u) = u^3 - u$  in Table 4 is not observed for  $f(u) = u^5 - u^3$ . Rather, we see  $k = O(\epsilon^2)$  as predicted by the theory in the previous section. The numerical results are shown in Table 9.

## 7 Summary and Future Work

We have identified the time step scaling for several first and second order schemes for AC and CH under the restriction of fixed local truncation error,  $\sigma$ . In particular, we derive the asymptotic behaviour of time-step number with  $\sigma$  and asymptotic parameter  $\epsilon$  during meta-stable dynamics. These predictions are made under the assumption that the time steps preserve the asymptotic structure of the diffuse interface, a concept we refer to as *profile fidelity*. The predictions are verified in numerical experiments. We see that methods whose dominant local truncation error can be expressed as a pure time derivative have optimal asymptotic performance in this particular limit. BE, TR, and BDF2 all have this desirable property. We believe these methods will also have superior performance for other problems with metastable dynamics. Our numerical results show that BE performs better than expected and we have shown an explanation of this behaviour with formal asymptotics.

The optimal fully implicit methods asymptotically computationally outperform all linearly implicit methods in the limit we consider. We present precise criteria on the computational cost of nonlinear solvers for this comparison. The provably energy stable first and second order SAV schemes had higher computational cost than standard IMEX methods for similar results. As a final result, we present a rigorous proof that large time steps with fully implicit BE can be taken with locally unique solutions that are energy stable. This is done for the 2D radial AC equation in meta-stable dynamics. Eyre-type iteration is also considered in this analytic framework, and it is shown that in general this approach loses profile stability unless very small time steps are taken.

Our work gives strong evidence that some fully implicit schemes for phase field models should be given more consideration and that provably energy stable schemes are inaccurate and give no benefits during meta-stable dynamics. Perhaps a hybrid scheme that switches between the two approaches depending on the dynamic regime should be considered.

Extending the analysis to the non-radial case and to CH is an interesting question. We observed that the question of global accuracy is not trivial in Sect. 5 and should be considered for other schemes. Accurate local error estimation for these problems is another interesting question to pursue.

**Acknowledgements** KP recognizes support from the NSF DMS under award 1813203. BW acknowledges support from an NSERC Canada grant. XC is supported by a UBC International Doctoral Fellowship.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix: $L_R^2$ Coercivity of $L$

Here we show the technical argument for Lemma 1. For  $u, v \in H^1(\mathbb{R})$  we define the inner product

$$\langle u, v \rangle_\ell := \int_{-\ell}^{\ell} u(s)v(s) ds,$$

with the standard norms  $L_\ell^2$  and  $H_\ell^1$  while  $L_{\ell^c}^2$  is defined in  $\mathbb{R} \setminus [-\ell, \ell]$ . Let  $L_0$  be as defined in (19).

**Lemma 5** Fix  $\ell_0 > 0$  sufficiently large there exists  $\alpha > 0$  such that for all  $\ell > \ell_0$

$$\langle L_0 u, u \rangle_\ell \geq \alpha, \quad (66)$$

for all  $u \in H_\ell^1 \cap L^2(\mathbb{R})$  satisfying  $\langle u, g' \rangle_{L^2(\mathbb{R})} = 0$  and  $1 = \|u\|_{L_\ell^2} \geq \|u\|_{L_{\ell^c}^2}$ .

**Proof** Let  $\phi$  be the minimizer of  $\langle L_0 u, u \rangle_\ell$  over  $H^1(\mathbb{R})$  subject to  $\|u\|_{L_\ell^2} = 1$  and the full-line orthogonality  $\langle u, \psi_0 \rangle_{L^2(\mathbb{R})} = 0$ . By scaling, the minima is attained with  $\|\phi\|_{L_\ell^2} = \frac{1}{2}$  and satisfies

$$L_0 \phi = \lambda \phi, \quad \text{on } [-\ell, \ell],$$

subject to Neumann boundary conditions  $\phi_x(\pm\ell) = 0$ , in addition to the full line orthogonality condition. The operator  $L_0$  on the truncated domain has eigenvalues  $\lambda_0^\ell < \lambda_1^\ell < \dots$  which are  $O(e^{-d\ell})$  far away from the eigenvalues of  $L_0$  on the full line. In particular  $\lambda_0^\ell$  may be negative, but the rest are uniformly positive. In  $L_\ell^2$  we partition  $\phi = \beta\psi_0^\ell + \phi^\perp$ , where  $\psi_0^\ell$  is the  $L_\ell^2$  ground state of  $L_0$  and  $\phi^\perp \in L_\ell^2$  is  $L_\ell^2$  orthogonal to  $\psi_0^\ell$ . Then we have

$$\langle L_0 \phi, \phi \rangle_\ell \geq \lambda_0 \beta^2 + \lambda_1^\ell \|\phi^\perp\|_{L_\ell^2}^2. \quad (67)$$

On the other hand the orthogonality condition implies that

$$\langle \phi, g' \rangle_{\mathbb{R}} = 0 = \beta + \langle \phi, g' \rangle_{\ell^c}.$$

where the subscript  $\ell^c$  denotes integration over  $\mathbb{R} \setminus [-\ell, \ell]$  with the corresponding norms. In particular we deduce that

$$|\beta| \leq \|\phi\|_{L_{\ell^c}^2} \|g'\|_{L_{\ell^c}^2} \leq \|g'\|_{L_{\ell^c}^2} \|\phi\|_{L_\ell^2}.$$

Since  $g'$  decays exponentially at  $\pm\infty$ , is complementary norm is exponentially small in  $\ell$ . From orthogonality of  $\psi_0^\ell$  and  $\phi^\perp$  we have

$$\|\phi\|_{L_\ell^2}^2 = \beta^2 + \|\phi^\perp\|_{L_\ell^2}^2 \leq \|\phi^\perp\|_{L_\ell^2}^2 + \|g'\|_{L_{\ell^c}^2}^2 \|\phi\|_{L_\ell^2}^2.$$



or equivalently

$$1 = \|\phi\|_{L^2_\ell}^2 \leq \frac{1}{1 - \|g'\|_{L^2_\ell}^2} \|\phi^\perp\|_{L^2_\ell}^2,$$

and taking  $\ell$  large enough we use these bound in (67) to show that  $\alpha$  is exponentially close to  $\lambda_1^\ell > 0$ .  $\square$

To complete the proof of Lemma 1 we take  $\ell$  sufficiently large to apply Lemma 5 and then take  $\epsilon$  sufficiently small that  $\epsilon|z| \leq \epsilon\ell \ll 1$ . Under these conditions  $L_\ell^2$  and  $L_R^2(-\ell, \ell)$  are equivalent norms, uniformly in  $\epsilon$ , and we have uniform  $L_R^2(-\ell, \ell)$  coercivity of  $L$ . Conversely,  $L$  is clearly  $L_R^2$  coercive on  $[-R/\epsilon, \infty) \setminus [-\ell, \ell]$  since  $f'(g)$  is strictly positive there. Clearly  $L$  is uniformly coercive on function with more than half their  $L_R^2$  mass in  $[-R/\epsilon, \infty) \setminus [-\ell, \ell]$ . The  $g'$  orthogonality condition implies approximate orthogonality to  $\psi_0$  for  $\ell$  large. From these we deduce the full  $L_R^2$ -coercivity of  $L$  over  $X$ .

## References

1. Alikakos, N., Bates, P., Chen, X.: Convergence of the cahn-hilliard equation to the hele-shaw model. *Arch. Rational Mech. Anal.* **128**(2), 165–205 (1994)
2. Allen, S.M., Cahn, J.W.: A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. *Acta Metall.* **27**(6), 1085–1095 (1979)
3. Ascher, U.M., Ruuth, S.J., Wetton, B.T.R.: Implicit-explicit methods for time-dependent partial differential equations. *SIAM J. Numer. Anal.* **32**(3), 797–823 (1995)
4. Barrett, J.W., Blowey, J.F., Garcke, H.: Finite element approximation of the cahn-hilliard equation with degenerate mobility. *SIAM J. Numer. Anal.* **37**(1), 286–318 (2000)
5. Blowey, J.F., Copetti, M.I.M., Elliott, C.M.: Numerical analysis of a model for phase separation of a multicomponent alloy. *IMA J. Numer. Anal.* **16**(1), 111–139 (1996)
6. Cahn, J.W., Hilliard, J.E.: Free energy of a nonuniform system. i. interfacial free energy. *J. Chem. Phys.* **28**(2), 258–267 (1958)
7. Chen, L., Hu, X., Wise, S.: Convergence analysis of the fast subspace descent method for convex optimization problems. *Math. Comput.* **89**, 1 (2020)
8. Cheng, X.: On the stability of a semi-implicit scheme of Cahn-Hilliard type equations. PhD thesis, University of British Columbia (2017)
9. Christlieb, A., Jones, J., Promislow, K., Wetton, B., Willoughby, M.: High accuracy solutions to energy gradient flows from material science models. *J. Comput. Phys.* **257**, 193–215 (2014)
10. Christlieb, A., Promislow, K., Xu, Z.: On the unconditionally gradient stable scheme for the cahn-hilliard equation and its implementation with fourier method. *Commun. Math. Sci.* **11**, 06 (2013)
11. Diegel, A.E., Wang, C., Wise, S.M.: Stability and convergence of a second-order mixed finite element method for the Cahn-Hilliard equation. *IMA J. Numer. Anal.* **36**(4), 1867–1897 (2015)
12. Du, Q., Ju, L., Li, X., Qiao, Z.: Stabilized linear semi-implicit schemes for the nonlocal Cahn-Hilliard equation. *J. Comput. Phys.* **363**, 39–54 (2018)
13. Du, Q., Nicolaides, R.A.: Numerical analysis of a continuum model of phase transition. *SIAM J. Numer. Anal.* **28**(5), 1310–1322 (1991)
14. Eyre, D.J.: Unconditionally gradient stable time marching the cahn-hilliard equation. *MRS Proc.* **529**, 39 (1998)
15. Feng, W., Salgado, A.J., Wang, C., Wise, S.M.: Preconditioned steepest descent methods for some nonlinear elliptic equations involving p-laplacian terms. *J. Comput. Phys.* **334**, 45–67 (2017)
16. Feng, X., Prohl, A.: Numerical analysis of the cahn-hilliard equation and approximation for the hele-shaw problem. *Interfaces Free Bound.* **7**, 1–28 (2005)
17. Hairer, E., Wanner, G.: Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems. Springer, Berlin (1996)
18. Kapitula, T., Promislow, K.: Spectral and Dynamical Stability of Nonlinear Waves. Applied Mathematical Sciences, vol. 185. Springer, New York (2013)
19. Li, D., Qiao, Z.: On second order semi-implicit fourier spectral methods for 2d cahn-hilliard equations. *J. Sci. Comput.* **70**(1), 301–341 (2017)

20. Li, D., Qiao, Z.: On the stabilization size of semi-implicit fourier-spectral methods for 3d cahn hilliard equations. *Commun. Math. Sci.* **15**, 1489–1506 (2017)
21. Modica, L., Mortola, S.: Un esempio di  $\gamma$ -convergenza. *Boll. Un. Mat. Ital.* **14**(5), 285–299 (1977)
22. Mullins, W.W., Sekerka, R.F.: Morphological stability of a particle growing by diffusion or heat flow. *J. Appl. Phys.* **34**(2), 323–329 (1963)
23. Pego, R.L.: Front migration in the nonlinear cahn-hilliard equation. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **422**(1863), 261–278 (1989)
24. Savin, O.: Phase transitions, minimal surfaces and a conjecture of de giorgi. *Curr. Dev. Math.* **101**(3), 59–113 (2010)
25. Shen, J., Xu, J., Yang, J.: The scalar auxiliary variable (sav) approach for gradient flows. *J. Comput. Phys.* **353**, 407–416 (2018)
26. Vollmayr-Lee, B.P., Rutenberg, A.D.: Fast and accurate coarsening simulation with an unconditionally stable time step. *Phys. Rev. E* **68**, 066703 (2003)
27. Wetton, B.: 2D Allen Cahn Simulation (YouTube Video). [https://youtu.be/W7oNaJQ4\\_kc](https://youtu.be/W7oNaJQ4_kc), (December 2018)
28. Wetton, B.: 2D periodic Cahn Hilliard Simulation (YouTube Video). <https://youtu.be/gI-S7MfWN5I> (March 2018)
29. Wise, S.: Unconditionally stable finite difference, nonlinear multigrid simulation of the Cahn-Hilliard-Hele-Shaw system of equations. *J. Sci. Comput.* **44**, 38–68 (2010)
30. Xu, J., Li, Y., Wu, S., Bousquet, A.: On the stability and accuracy of partially and fully implicit schemes for phase field modeling. *Comput. Methods Appl. Mech. Eng.* **345**, 826–853 (2019)
31. Yan, Y., Chen, W., Wang, C., Wise, S.: A second-order energy stable bdf numerical scheme for the cahn-hilliard equation. *Commun. Comput. Phys.* **23**, 01 (2018)
32. Yue, P., Feng, J.J., Liu, C., Shen, J.: A diffuse-interface method for simulating two-phase flows of complex fluids. *J. Fluid Mech.* **515**, 293–317 (2004)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.