THE DYNAMICS OF THE THETA METHOD*

A. M. STUART[†] AND A. T. PEPLOW^{†‡}

Abstract. The dynamics of the theta method for arbitrary systems of nonlinear ordinary differential equations are analysed. Two scalar examples are presented to demonstrate the importance of spurious solutions in determining the dynamics of discretisations. A general system of differential equations is then considered. It is shown that the choice $\theta = \frac{1}{2}$ does not generate spurious solutions of period 2 in the timestep n. Using bifurcation theory, it is shown that for $\theta \neq \frac{1}{2}$ the theta method does generate spurious solutions of period 2. The existence and form of spurious solutions are examined in the limit $\Delta t \to 0$. The existence of spurious steady solutions in a predictor-corrector method is proved to be equivalent to the existence of spurious period 2 solutions. Numerical continuation is used to trace out the spurious solutions as Δt is varied. Timestepping experiments are presented to demonstrate the effect of the spurious solutions on the dynamics and some complementary theoretical results are proved. In particular, the linear stability restriction $\Delta t/\Delta x^2 \leq \frac{1}{2}$ for the Euler method applied to the heat equation is generalised to cope with a nonlinear problem. This naturally introduces a restriction on Δt in terms of the initial data; this restriction is necessary to avoid the effect of spurious solutions.

Key words. dynamics of numerical methods, asymptotic instabilities, dissipative problems, timestep as bifurcation parameter

AMS(MOS) subject classifications. 65M10, 35A40, 35K57

1. Introduction. In this paper we study the relationship between the asymptotic behaviour of a numerical simulation and that of the true solution itself, for fixed values of the mesh-spacing. Our viewpoint here is to consider the numerical method as a dynamical system and to analyse its possible asymptotic states. Our aim is to determine methods for which spurious asymptotic states are not introduced by the discretisation, either for all values of the timestep, or for all values of the timestep sufficiently small.

We study one-step methods for systems of nonlinear ordinary differential equations (ODEs). Such systems may be of interest in their own right or may arise from spatial discretisations of partial differential equations (PDEs). Many of the features we describe are qualitatively robust to spatial mesh refinement and this justifies studying the problem in isolation from the effect of spatial derivatives—essentially a method of lines approach. In fact, many of the results in this paper can be generalised from an ODE in \Re^n to an ODE in a Banach space.

We study the following ODE: find $u(t) \in \Re^m$ satisfying

(1.1)
$$u_t = G(u), \quad u(0) = u_0.$$

The regularity of G(u) will be specified when required. The theta method consists of finding a sequence $U_n \in \Re^m$ satisfying

(1.2)
$$U_{n+1} - U_n = \Delta t [(1 - \theta)G(U_n) + \theta G(U_{n+1})],$$

for $\theta \in [0, 1]$.

^{*} Received by the editors December 6, 1989; accepted for publication (in revised form) October 3, 1990.

[†] School of Mathematical Sciences, University of Bath, Bath BA2 7AY, United Kingdom.

[‡] Present address, Department of Civil Engineering, University of Bradford, Bradford BD7 1DP, United Kingdom. The work of this author was supported by the Science and Engineering Research Council, United Kingdom.

We also consider the following splitting of G:

(1.3)
$$G(u) = Au + H(u),$$

where both A and H map \Re^m into itself. A is assumed to be a matrix, whilst H is a nonlinear function. This splitting enables us to consider the semi-implicit method which is popular in the solution of many dissipative PDEs: find $U_n \in \Re^m$ satisfying

(1.4)
$$U_{n+1} - U_n = \Delta t [AU_{n+1} + H(U_n)].$$

Note that U_n approximates $u(n\Delta t)$. Under fairly mild assumptions, both the schemes (1.2) and (1.4) yield convergent solutions on compact time intervals. The main result of this paper is proved in Lemmas 3.1, 3.2, and 3.3. It is given in the following theorem.

MAIN THEOREM. If $\theta = \frac{1}{2}$, then the numerical method (1.2) cannot have period 2 solutions in n. The method (1.4) and the method (1.2) with $\theta \neq \frac{1}{2}$ can have period 2 solutions in n.

Such period 2 solutions are spurious and generated by discretisation. They are important because their presence can cause the dynamics of the discretisation to differ substantially from the dynamics of the underlying differential equation. This happens in the following two different ways:

(i) If the period 2 solutions are unstable (that is, repel data in their vicinity), then such solutions and their stable manifolds often delineate classes of initial data for which the correct asymptotic behaviour is reproduced by the numerical method and classes of initial data which lead to blow-up of the scheme or nonexistence of numerical solutions. (See Example 2.1 and Theorem 4.2.)

(ii) If the period 2 solutions are stable (that is, attract data in their vicinity), then U_n may converge to a period 2 solution as $n \to \infty$. All initial data lying within the domain of attraction of this period 2 solution will then yield spurious results. (See Example 2.2.)

In §2 we describe two simple examples which illustrate the importance of spurious periodic solutions in determining the dynamics of discretisations. These examples indicate why the nonexistence of period 2 solutions for the trapezoidal rule gives it an advantage over other methods. Section 3 contains the proof of the main theorem in the general case. In addition, we state and prove a theorem about the behaviour of spurious solutions as $\Delta t \to 0$. Specifically we prove that for C^1 nonlinearities G(u)spurious solutions approach infinity in norm as $\Delta t \to 0$ (if they exist for arbitrarily small Δt). This is essentially a consequence of the convergence of the schemes. We also show that the existence of period 2 solutions for the Euler method is equivalent to the existence of spurious steady solutions for a predictor-corrector method and hence the theory of spurious periodic solutions has far-reaching consequences. The results from $\S3$ are illustrated further in $\S4$ by means of several examples arising from discretisations of parabolic PDEs; numerical continuation is used to trace out the spurious solutions as Δt is varied, and timestepping experiments are presented which illustrate the effect of the spurious solutions. Complementary analysis of the specific problems is also presented. In particular, Theorem 4.2 provides a generalisation of the classical stability restriction for the numerical solution of the linear heat equation to a nonlinear problem. As would be expected, this nonlinear restriction involves dependence upon initial data, reflecting the existence of spurious solutions. Section 5 contains the conclusions.

The first reference to the relevance of period 2 solutions in generating spurious dynamics is by Newell [13]. This important work has not received the attention it deserves in the numerical analysis community, mainly because it is couched in the language of fluid mechanics. Many of the ideas in this and other papers are present in embryonic form in [13]; in particular, Newell recognized the importance of the linear stability limit in predicting the existence of spurious period 2 solutions. Newell's work on dissipative problems dealt mainly with a cubic viscous Burgers' equation and touched on the issue of implicit methods. Subsequently, further studies of period 2 solutions were undertaken by Mitchell and his co-workers ([6], [10], [11], [17], [19]), and the role of the linear stability limit in predicting their existence was firmly established. Mitchell's work has concentrated on explicit methods with quadratic nonlinearities (in particular, Fisher's equation) and the existence of stable period 2 solutions for Δt above the linear stability limit. Following the work of Newell and of Mitchell, a general result establishing the existence of spurious periodic solutions in the neighbourhood of the linear stability limit was proved for explicit discretisations of arbitrary reactiondiffusion-convection equations [22]. That work showed that spurious solutions can exist below the linear stability limit and in some circumstances for Δt arbitrarily small. This paper extends the work in [22] to more general numerical methods and to more general differential equations; we also include new results about the qualitative behaviour of the spurious solutions and several numerical examples which illustrate the application of the theory to PDEs. A unified treatment of spurious steady and periodic solutions in Runge-Kutta methods, linear multistep methods, and predictor-corrector methods is contained in [14]. The main result in this paper disproves a conjecture made by one of the present authors in [21], namely, that the Crank-Nicolson method for nonlinear parabolic equations can have period 2 solutions which bifurcate from $\Delta t = \infty$.

In general, dynamically varying timesteps should be used in time-dependent simulations of nonlinear problems, whenever feasible. Here we are concerned entirely with fixed timestepping strategies. There are two main reasons why we believe that the study of fixed timestepping strategies is important. First, time-like iterations are often used to solve steady problems (or to find other asymptotic states) and in such a context it is the asymptotic properties of the scheme that are of most importance; transient behaviour is irrelevant. Hence fixed timesteps are frequently used, the aim being to choose a timestepping algorithm which maximises the domain of attraction of the target solution; an important step in this direction is to avoid the existence of spurious solutions. Second, it is our belief that a thorough understanding of the problems associated with the fixed timestep strategy is necessary in order to develop simple variable-step strategies for large-scale PDEs; it seems likely that the emphasis in designing timestepping strategies will shift from schemes designed purely to control local error to schemes designed to reproduce the long-time dynamics of differential equations as closely as possible. Clearly these two approaches are intimately related, but further work is required on the subject. We claim that, for both fixed and variable timestepping strategies, bifurcation theory is an excellent tool for studying the dynamics of discretisations since it enables the study of changes in the topology of a dynamical system as a parameter (some measure of the timestep) is varied.

An alternative, and very powerful, algebraic approach to the study of spurious solutions is introduced in [8], where spurious steady solutions are considered. This approach can be generalised to cope with period 2 solutions; see [14]. The dynamics of variable timestepping algorithms are analysed in [5]

2. Motivation for the methods and results. In this section we present two examples which illustrate the effect of spurious period 2 solutions on the dynamics of discretisations. Another example illustrating the effect of unstable spurious solutions may be found in the paper of Brezzi, Ushiki, and Fujii [1]; and the effect of stable spurious solutions is discussed in the paper of Sleeman et al. [19].

Example 2.1. Unstable spurious solutions. This example illustrates the effect of unstable spurious solutions. Consider the ODE

$$(2.1) u_t = -u^3.$$

It is clear that $u(t) \to 0$ as $t \to 0$, for any initial data. Consider now the forward Euler discretisation of (2.1)

$$(2.2) U_{n+1} = U_n - \Delta t U_n^3.$$

It is straightforward to verify that (2.2) has a period 2 solution of the form $U_n = (-1)^n U_c$, where $U_c = \sqrt{2/\Delta t}$. The following result is proved in [23].

Result. The following three possibilities hold for (2.2):

- (i) If $|U_0| < |U_c|$, then $|U^n| \to 0$ as $n \to \infty$.
- (ii) If $|U_0| = U_c$, then $U_n = U_c$.
- (iii) If $|U_n| > U_c$, then $|U^n| \to \infty$ as $n \to \infty$.

Notice the important role of the unstable spurious period 2 solution: data lying below it has the correct asymptotic behaviour and data lying above it diverges to infinity. Thus the spurious solution divides the phase space into regions where correct and incorrect asymptotic behaviour are generated by the numerical method. Of course, this result is particularly simple since we are in one dimension, but similar results can be proved in higher dimensions (see Theorem 4.2.) Consequently we claim that unstable spurious solutions play a very important role in the dynamics of discretisations.

Example 2.2. Stable spurious solutions. This example illustrates the effect of stable spurious solutions. Consider the scalar ODE

$$(2.3) u_t = u^3.$$

The dynamics of this equation are very simple. The solution is

$$u(t) = rac{u(0)}{((1-2tu(0))^2)^{rac{1}{2}}}.$$

All initial data other than u(0) = 0 diverge to infinity monotonically in finite time $1/2u(0)^2$. The situation is summarised in Fig. 1; the variation with Δt is trivial for the differential equation and is included for the purpose of comparison with the discretisation.

Consider now the backward Euler discretisation of (2.3):

(2.4)
$$U_{n+1} = U_n + \Delta t U_{n+1}^3.$$

As for the differential equation, there is an unstable steady solution $U_n = 0$. It is a consequence of [24, Thm. 2.3] that there is no strictly positive or strictly negative sequence satisfying (2.4) for all $n \ge 0$. Thus the numerical solution cannot diverge to infinity monotonically as the true solution does. What happens is that initial data other than $U_0 = 0$ evolves towards a stable period 2 solution as n increases. It is straightforward to show that (2.4) has a period 2 solution of the form $U_n = (-1)^n U_c$,



FIG. 1. The dynamics of (2.3). Arrows indicate the evolution of |u(t)| with time.

with U_c as defined in the previous example. This solution is stable with a local contraction rate of $\frac{1}{25}$ over two timesteps. As in Example 2.1, the spurious solution exists for arbitrarily small Δt and moves off to infinity as $\Delta t \rightarrow 0$; see Theorem 3.4.

We solve (2.4) numerically using Newton iteration at each step, with an initial guess provided by the forward Euler method. Iteration is performed until convergence, for each value of n. All nonzero initial data evolves to the stable period 2 solution as n increases. Thus the dynamics of the numerical solution can be summarised as in Fig. 2, which should be compared with Fig. 1.

Figure 3 shows the solution of (2.4) with $U_0 = 1$ and $\Delta t = 0.0001$. The evolution towards the stable period 2 solution is clear.

Thus the spurious solution attracts all nonzero initial data. The correct asymptotics are not reproduced for any nonzero initial data. Consequently we claim that stable spurious solutions play a very important role in the dynamics of discretisations.

3. The Main Theorem and related results. In this section we prove the Main Theorem stated in §1. The proof is broken into three lemmas (Lemmas 3.1–3.3) concerned with the trapezoidal rule, the theta method for $\theta \neq \frac{1}{2}$, and the semi-implicit method. We also prove Theorem 3.4 concerning the behaviour of spurious solutions as $\Delta t \rightarrow 0$; we show that spurious solutions must tend to infinity in norm as $\Delta t \rightarrow 0$ (if they exist) and derive a condition under which spurious solutions do not exist for Δt arbitrarily small. In addition, we prove Theorem 3.5, which yields the same conclusion as Lemma 3.2, namely, the existence of spurious solutions, but with different assumptions. Finally, in Theorem 3.6, we prove that the existence of spurious period 2 solutions for $\theta = 0$ is equivalent to the existence of spurious steady solutions



FIG. 2. The dynamics of (2.4). Arrows indicate the evolution of $|U_n|$ with n.



FIG. 3. Solution of (2.4). $U_0 = 1, \Delta t = 0.01$.

for a predictor-corrector method. Extensions and alternative proofs of Theorems 3.4 and 3.6 can be found in [9].

Period 2 solutions of (1.2) are pairs $(U, V) \in \Re^m \times \Re^m$ with $U \neq V$, satisfying

(3.1)
$$V - U = \Delta t [(1 - \theta)G(U) + \theta G(V)],$$

(3.2)
$$U - V = \Delta t [(1 - \theta)G(V) + \theta G(U)].$$

Period 2 solutions of (1.4) are pairs $(U, V) \in \Re^m \times \Re^m$ with $U \neq V$, satisfying

(3.3)
$$V - U = \Delta t [AV + H(U)],$$

(3.4)
$$U - V = \Delta t [AU + H(V)].$$

LEMMA 3.1. For $\theta = \frac{1}{2}$, (1.2) cannot have period 2 solutions. Proof. When $\theta = \frac{1}{2}$, (3.1) and (3.2) give

$$2(V - U) = \Delta t[G(U) + G(V)],$$

$$2(U-V) = \Delta t[G(V) + G(U)].$$

Subtracting, we obtain U = V and so period 2 solutions cannot exist. \Box

Assumptions on G(u) for Lemmas 3.2 and 3.3.

- (i) G(0) = 0.
- (ii) $G(u) \in C^2(\Re^m, \Re^m)$ for u near 0.
- (iii) Let dG(0) denote the Jacobian of G at u = 0. Then we assume that dG(0) has a real, nonzero, simple eigenvalue η . The corresponding eigenvector is y. (This is required only for Lemma 3.2.)
- (iv) Let dH(0) denote the Jacobian of H(u) at u = 0. Then we assume that A dH(0) has a real, nonzero, simple eigenvalue γ . The corresponding eigenvector is w. (This is required only for Lemma 3.3.)
- (v) The matrix dG(0) is nonsingular.

We now discuss the necessity of the assumptions for the proof of the lemmas. Our method of proof is to consider the bifurcation of spurious period 2 solutions from steady solutions. Steady solutions u satisfy G(u) = 0 and without loss of generality we may assume that u = 0. Thus Assumption (i) is necessary for our method of proof. (However, it is possible for spurious period 2 solutions to exist which do not bifurcate from a steady solution at any finite value of Δt , as Examples 2.1 and 2.2 show.) Assumption (ii) is required so that we may employ standard bifurcation theory as in [3]. The existence of a real nonzero eigenvalue in Assumption (iii) (respectively, Assumption (iv)) for Lemma 3.2 (respectively, Lemma 3.3) is necessary. That the eigenvalue be simple is not strictly necessary, but simplifies the analysis. (Note that by a different method of proof we are able to relax the condition on the existence of a real eigenvalue, provided that G(u) has two zeros; see Theorem 3.5.) Assumption (v) precludes the bifurcation of steady solutions in the differential equation itself. This could be allowed for in the analysis, but would complicate it unnecessarily; the simultaneous bifurcation of steady and periodic solutions is analysed in [21].

Lemmas 3.2 and 3.3 are proved by application of standard "bifurcation from a simple eigenvalue" arguments; see [3, Chap. 5] or [7]. Here Δt plays the role of

1357

bifurcation parameter. Note that the lemmas provide information about the structure of the period 2 solutions near to the bifurcation point. This is required to initiate numerical continuation procedures which trace out the spurious solutions; see §5.

LEMMA 3.2. Period 2 solutions of (1.2) bifurcate from the trivial solution (U, V) = (0,0) at $\Delta t = 2/(2\theta - 1)\eta$. Furthermore, the branches of nontrivial solutions satisfy, for $\mu \ll 1$,

$$U = -V = \mu y + O(|\mu|^2),$$

$$\Delta t = \frac{2}{(2\theta - 1)\eta} + O(|\mu|).$$

Proof. Consider (3.1) and (3.2). Clearly these are satisfied by (U, V) = (0, 0) for all Δt . Straightforward application of [3, Thm. 5.3, Chap. 5] shows that bifurcation occurs at simple eigenvalues Δt of the following problem:

$$eta - lpha = \Delta t[(1 - heta)dGlpha + heta dGeta],$$

 $lpha - eta = \Delta t[(1 - heta)dGeta + heta dGlpha].$

Throughout this proof dG is evaluated at u = 0. Adding the two equations shows that $dG(\alpha + \beta) = 0$. Since dG is invertible, we obtain $\alpha = -\beta$. Thus the equations reduce to

$$[2 - \Delta t(2\theta - 1)dG]\beta = 0.$$

By assumption, dG has a real, nonzero, simple eigenvalue η and so we deduce that the linearised system is singular at

$$\Delta t = \frac{2}{(2\theta - 1)\eta}$$

A little calculation shows that this eigenvalue is simple (in the sense defined in [3, Chap. 5]) and the conclusions of the lemma follow in a straightforward way. \Box

LEMMA 3.3. Period 2 solutions of (1.4) bifurcate from the trivial solution (U, V) = (0,0) at $\Delta t = \frac{2}{\gamma}$. Furthermore, the branches of nontrivial solutions satisfy, for $\mu \ll 1$,

$$egin{aligned} U &= -V = \gamma w + O(|\mu|^2), \ \Delta t &= rac{2}{\gamma} + O(|\mu|). \end{aligned}$$

Proof. The proof is very similar to Lemma 3.2 and is omitted.

An important question for the numerical analyst is to determine to where the branches of solutions $(U(\mu), V(\mu), \Delta t(\mu))$ go to in function and parameter space. In particular, it is important to know whether the branches extend to arbitrarily small values of Δt . This depends on the global structure of the nonlinear terms and no general answer is possible. It is, however, an important area for future research.

We now prove a result about the behaviour of spurious periodic solutions as $\Delta t \rightarrow 0$. Roughly, Theorem 3.4 states that if spurious solutions exist for arbitrarily small Δt then they must move off to infinity in norm—essentially a consequence of the convergence of the scheme.

THEOREM 3.4 Consider genuine period 2 solutions (U, V) of (1.2) and of (1.4) with $U \neq V$. Assume that $G(U) \in C^1(\Re^m, \Re^m)$. The following results hold:

- (i) If a pair (U, V) can be found for arbitrarily small Δt, then ||U||_∞ and ||V||_∞ → ∞ as Δt → 0.
- (ii) Furthermore, if G(U) = BU + J(U), where B is an invertible matrix and J(U) is uniformly bounded for all U, then period 2 solutions cannot exist for arbitrarily small Δt .

Note on Theorem 3.4. The assumption that G(u) is a C^1 function is strictly necessary. See [9, Ex. 4.8] for an illustration of this fact. It is shown that, for $G(u) = -\sqrt{(u)}, u > 0$ and $G(u) = \sqrt{(|u|)}, u < 0$ the Euler method has period 2 solutions of the form $U_n = (-1)^n \Delta t^2/4$.

Proof. We restrict our proofs to the theta method (3.1), (3.2) since the proofs for the semi-implicit method (3.3), (3.4) are slight modifications. Consider (1.2). Period 2 solutions satisfy (3.1), (3.2). Adding these equations shows that for $\Delta t \neq 0$,

$$G(U) + G(V) = 0.$$

Using this result in (3.1), (3.2), we obtain

(3.5)
$$V - U = \Delta t (1 - 2\theta) G(U)$$

and

(3.6)
$$U - V = \Delta t (1 - 2\theta) G(V).$$

Subtracting (3.6) from (3.5) and using the Mean Value Theorem gives

(3.7)
$$\left[2I + \Delta t(1-2\theta) \int_0^1 dG(tU + (1-t)V)dt\right](V-U) = 0.$$

We first prove part (i) of Theorem 3.4. We assume that U remains bounded in norm as $\Delta t \to 0$ and we obtain a contadiction. From (3.5) we deduce that V remains bounded. Thus the linear operator in (3.7) is invertible for all Δt sufficiently small and U = V. Hence period 2 solutions do not exist for all Δt sufficiently small. This is the required contradiction which establishes that U becomes unbounded. Similarly, it may be shown that V becomes unbounded.

We now prove Theorem 3.4(ii). We assume that a pair (U, V) with $U \neq V$ can be found for Δt arbitrarily small and we obtain a contradiction. Period 2 solutions satisfy (3.5). Using (3.5) to eliminate V, and using G(U) = BU + J(U), we obtain

$$BU + J(U) + B[U + \Delta t BU + \Delta t J(U)] + J(V) = 0.$$

Here $\hat{\Delta t} = (1 - 2\theta) \Delta t$. Rearranging gives

$$(2B + \hat{\Delta t}B^2)U = -J(U) - J(V) - \hat{\Delta t}BJ(U).$$

From the boundedness of J and the invertibility of B we deduce that U remains bounded as $\Delta t \to 0$. By (3.5) so too does V. Hence, by (3.7), U = V for all Δt sufficiently small so that period 2 solutions cannot exist. \Box

We now prove another theorem on the existence of spurious periodic solutions in the θ method. Lemma 3.2 proves the existence of spurious periodic solutions which bifurcate from a steady solution by means of a simple symmetry-breaking bifurcation. However, to prove Lemma 3.2 it is necessary to assume that the Jacobian dG(0) is invertible and that dG(0) has a real, nonzero eigenvalue. We now prove the existence of spurious period 2 solutions assuming the existence of two steady solutions but under the weaker assumption that the Jacobian is invertible at both steady solutions; we do not need to assume the existence of a real eigenvalue. The result holds for $\Delta t \gg 1$ but may be useful as a starting point for numerical continuation to small values of Δt .

THEOREM 3.5. Let $\theta \neq \frac{1}{2}$. Assume that there exist U_0, V_0 with $U_0 \neq V_0$ satisfying $G(U_0) = G(V_0) = 0$ and that $dG(U_0)$ and $dG(V_0)$ are invertible. Then for $|\Delta t|$ sufficiently large, there exist spurious period 2 solutions of the discretisation (1.2).

Proof. Period 2 solutions of (1.2) are pairs (U, V) satisfying (3.1) and (3.2). For $\Delta t = \infty$ these equations have the solution $U = U_0$ and $V = V_0$. We now prove that these solutions can be uniquely extended to solutions for $|\Delta t|$ sufficiently large. By the Implicit Function Theorem, this can be done provided that the following problem for (α, β) has the unique solution (0, 0):

$$(1-\theta)dG(U_0)\alpha + \theta dG(V_0)\beta = 0,$$

$$(1-\theta)dG(V_0)\beta + \theta dG(U_0)\alpha = 0.$$

If $\theta = 0$ or 1 we deduce that $\alpha = \beta = 0$ automatically. If $\theta \neq 0, 1$ then simple manipulations show that $\alpha = \beta = 0$ provided that $\theta \neq \frac{1}{2}$. \Box

The final theorem in this section relates the existence of spurious periodic solutions in the Euler method to the existence of spurious steady solutions in the following predictor-corrector method for (1.1):

(3.8)
$$U_{n+1}^p = U_n + \Delta t G(U_n),$$

(3.9)
$$U_{n+1} = U_n + \frac{\Delta t}{2} [G(U_{n+1}^p) + G(U_n)].$$

THEOREM 3.6. Consider the predictor-corrector method (3.8), (3.9). This method has a spurious steady solution satisfying $U_{n+1} = U_n$ whenever the theta method with $\theta = 0$ has a spurious period 2 solution.

Proof. Steady solutions of (3.8), (3.9) satisfy

$$(3.10) V = U + \Delta t G(U),$$

(3.11)
$$0 = G(U) + G(V).$$

Using (3.11) in (3.10) we obtain

$$(3.12) U = V + \Delta t G(V).$$

Equations (3.10) and (3.12) are identical to (3.1) and (3.2) with $\theta = 0$. Hence the result follows.

4. Application to nonlinear parabolic equations. Here we apply the results in §4 to nonlinear parabolic equations of the form

$$(4.1) u_t = u_{xx} + \lambda u + h(u, u_x)$$

with boundary conditions

$$(4.2) u(0,t) = u(1,t) = 0$$

and some initial conditions. For simplicity we shall assume that h(0,0) = 0 and that h(a,b) is superlinear in its arguments a and b. These assumptions are not necessary (see [22]) but simplify the analysis considerably; we seek period 2 solutions bifurcating from zero and our assumptions ensure that the linear eigenvalue problem governing bifurcation is symmetric and explicitly solvable.

We discretise (4.1), (4.2) in space using a centred approximation to the second derivative and a consistent approximation to the nonlinear term (which may or may not involve upwinding). We obtain the system of J - 1 nonlinear ODEs

(4.3)
$$\frac{du_j}{dt} = \frac{\delta_x^2 u_j}{\Delta x^2} + \lambda u_j + g(u_{j-1}, u_j, u_{j+1}),$$

for $j = 1, \dots, J - 1$. Here $J\Delta x = 1, \delta_x^2 u_j = u_{j+1} - 2u_j + u_{j-1}$, and

$$(4.4) u_0 = u_J = 0$$

From our assumptions on h(a, b) we deduce that g(0, 0, 0) = 0 and that g is superlinear in its three arguments. (Note that g may depend upon Δx but that we supress this for notational convenience.)

We can apply Lemma 3.2 directly to (4.3), (4.4). We let $U = [U_1, U_2, \dots, U_{J-1}]^T$ and $V = [V_1, V_2, \dots, V_{J-1}]^T$. At zero, the Jacobian of the right-hand side is a symmetric, tridiagonal matrix with J - 1 distinct eigenvalues $\eta^{(k)}$ and corresponding eigenvectors $y^{(k)}$ [4]. These are

(4.5)
$$\eta^{(k)} = \lambda - \frac{4\sin^2(k\pi/2J)}{\Delta x^2},$$

with corresponding eigenvectors given componentwise by

(4.6)
$$y_j^{(k)} = \sin(k\pi j/J).$$

The index $k = 1, \dots, J-1$. For $y_j^{(k)}$ the subscript j denotes the jth component of the vector $y^{(k)}$. By Lemma 3.2 we can find J-1 branches of period 2 solutions when applying the θ method to (4.2) and locally these satisfy

(4.7)
$$U_j = -V_j = \mu y_j^{(k)} + O(|\mu^2|).$$

(4.8)
$$\Delta t_k = \frac{2}{(2\theta - 1)\eta^{(k)}} + O(|\mu|).$$

Note that the smallest of the positive values of Δt is the asymptotic stability limit for the numerical method linearised about zero.

The local results (4.7) and (4.8) provide the starting points for a numerical continuation procedure which traces out spurious period 2 solutions in the theta method as Δt is varied. We now describe several examples corresponding to different choices of nonlinearity h(a, b). Throughout the following, we use the nonlinear solver PITCON [16] to follow the solution branches of the nonlinear equations governing existence of period 2 solutions. Note that the bifurcation of period 2 solutions is necessarily of pitchfork type [7], [22] and so there are two branches emanating from each bifurcation point. It is possible to move from one branch to the other by interchanging the roles of U and V. For this reason the two branches are indistinguishable in the figures.

Example 4.1. Consider the equation

$$(4.9) u_t = u_{xx} - u^3$$



FIG. 4. l_2 norm of spurious periodic solutions versus $1/\Delta t$ for equations (4.10) and (4.11).

together with boundary conditions (4.2). A Liapunov functional approach shows that all solutions converge to zero as $t \to \infty$. We discretise the equation using explicit Euler timestepping (θ =0) and we take $g(a, b, c) = -b^3$. We obtain, for $j = 1, \dots, J-1$,

(4.10)
$$u_{j}^{n+1} = u_{j}^{n} + r\delta_{x}^{2}u_{j}^{n} - \Delta t(u_{j}^{n})^{3},$$

boundary conditions

(4.11)
$$u_0^n = u_J^n = 0,$$

and some initial condition on u_j^0 . Here $r = \Delta t / \Delta x^2$.

We seek period 2 solutions satisfying the Z_2 symmetry U = -V and find that they satisfy

(4.12)
$$\frac{\delta_x^2 U_j}{\Delta x^2} - U_j^3 + \frac{2U_j}{\Delta t} = 0, \qquad j = 1, \cdots, J - 1$$

together with

$$(4.13) U_0 = U_J = 0.$$

Using the estimate (4.7), (4.8) we can follow the branches of spurious periodic solutions satisfying (4.12), (4.13). Figure 4 shows a plot of these branches in the case $\Delta t =$ 0.1; thus J = 10 and there are nine branches. Note that the branches exist for Δt arbitrarily small. This is proved in [22], where it is shown that the solution branches approach infinity at a rate proportional to $\sqrt{1/\Delta t}$. This scaling follows from a balance between U_j^3 and $2U_j/\Delta t$. The fact that they approach infinity is a consequence of Theorem 3.4(i). Recall that the solution of the differential equation tends to zero as t tends to infinity. The following theorem gives conditions for the Euler approximation to behave similarly. The theorem reflects the effect of the spurious periodic solutions in that Δt must be restricted in terms of the reciprocal of the square of the magnitude of the initial data; this is precisely the scaling of the spurious solutions with Δt . The theorem shows that the phase space can be divided into regions where correct and incorrect asymptotic behaviour is observed.

THEOREM 4.2. Consider the solution of (4.10), (4.11). Let

$$u_{\max} = \max_{0 \le j \le J} |u_j^0|.$$

Then we find the following: (i) If

$$\Delta t \le \frac{\Delta x^2}{2 + 3(\Delta x u_{\max})^2}$$

Then $u_j^n \to 0$ as $n \to \infty$ for all initial data (correct asymptotics). (ii) If $J = 0 \pmod{3}$, then there exists initial data such that if

$$\Delta t > \frac{2\Delta x^2}{3 + (\Delta x u_{\max})^2}$$

then $|u_j^n| \to \infty$ as $n \to \infty$ (spurious asymptotics). Proof. (i) Define

$$u_{\max}^n = \max_{0 \le j \le J} u_j^n, \qquad u_{\min}^n = \min_{0 \le j \le J} u_j^n.$$

Note that $u_{\max}^n \ge 0$ and $u_{\min}^n \le 0$ by virtue of (4.11). The proof proceeds by induction. Assume that

(4.14)
$$\Delta t \le \frac{\Delta x^2}{2 + 3(\Delta x u_{\max}^n)^2},$$

(4.15)
$$\Delta t \le \frac{\Delta x^2}{2 + 3(\Delta x u_{\min}^n)^2}$$

This holds for n = 0 by assumption. Rearranging (4.10), we obtain, for $j = 1, \dots, J-1$

$$u_j^{n+1} = ru_{j-1}^n + [1 - 2r - \Delta t(u_j^n)^2]u_j^n + ru_{j+1}^n.$$

Using the monotonicity of $(1-2r)u - \Delta tu^3$ implied by (4.14), (4.15), we deduce that

(4.16)
$$u_j^{n+1} \le [1 - \Delta t (u_{\max}^n)^2] u_{\max}^n$$

Similarly,

(4.17)
$$u_j^{n+1} \ge [1 - \Delta t (u_{\min}^n)^2] u_{\min}^n.$$

This shows that, if u_{\max}^{n+1} (respectively, u_{\min}^{n+1}) is nonzero and attained in the interior $(j \neq 0, J)$, then it strictly decreases (respectively, increases) over one step. If u_{\max}^{n+1} (respectively, u_{\min}^{n+1}) is attained at the boundary (j = 0 or j = J), then $u_{\max}^{n+1} = 0$

(respectively, $u_{\min}^{n+1} = 0$.) Thus equations (4.14) and (4.15) hold for n+1. By induction we deduce that $u_i^n \to 0$ as $n \to \infty$.

(ii) For this part we assume that $J = 0 \pmod{3}$. With this assumption it may be shown by substitution that

(4.18)
$$u_j^n = A^n \sin(2\pi j/3)$$

is an exact solution of (4.10), (4.11) provided that A^n satisfies the recursion

(4.19)
$$A^{n+1} = [1 - 3r - \frac{3}{4}\Delta t (A^n)^2]A^n$$

This exact nonlinear solution satisfies the recursion because of the property of aliasing: to establish (4.19) we use $\sin^3 x = \frac{3}{4} \sin x - \frac{1}{4} \sin 3x$. For $x = 2\pi j/3$, $\sin 3x = 0$ and hence (4.19) follows. If

(4.20)
$$1 - 3r - \frac{3}{4}\Delta t (A^n)^2 < -1,$$

it is clear that $|A^{n+1}| > |A^n|$ and hence, by induction, that $|A^n| \to \infty$ as $n \to \infty$. Hence $u_j^n \to \infty$ as $n \to \infty$ provided that

$$(4.21) \qquad \qquad \Delta t > \frac{8\Delta x^2}{12 + 3(\Delta x A^0)^2}.$$

However, $u_{\text{max}} = A^0 \sin(2\pi/3)$. Thus (4.21) gives

$$\Delta t > rac{2\Delta x^2}{3+(\Delta x u_{ ext{max}})^2}$$

as required. \Box

We now describe some timestepping experiments which illustrate Theorem 4.2. We solve the recursion (4.10), (4.11) with $1/\Delta t = 300.557$ and with J = 10. We use initial data of the form

$$u_j^0 = aU_j,$$

where U_j is the spurious periodic solution lying on the uppermost branch in Figure 4 at $\frac{1}{\Delta t} = 300.557$. Figures 5 and 6 show plots of the l_2 norm of u_j^n ,

$$\sqrt{\sum_{j=0}^{J-1} (u_j^n)^2},$$

versus n for two different values of a. In Fig. 5 we take a = 1. The true solution is pure period 2 in n. However, rounding errors excite an instability and the solution diverges to infinity (spurious asymptotics.) This occurs in a very small number of steps; the l_2 norm is approximately constant for eight steps (since we start on the period 2 solution which has the symmetry $U_j = -V_j$) and in the last two steps the instability rapidly takes over and the solution diverges to infinity. In Fig. 6 we take a = 0.99. Here the solution lies just beneath the spurious periodic solution and it is attracted to zero (correct asymptotics).

Example 4.3. In this example we study the Ginzburg-Landau equation

$$u_t = u_{xx} + \lambda(u - u^3),$$



FIG. 5. a = 1.0. Plots of the l_2 norm of u_j^n versus $t = n\Delta t$.



Fig. 6. a = 0.99. Plots of the l_2 norm of u_j^n versus $t = n\Delta t$.

together with boundary conditions (4.2). The dynamics of this problem are very well understood: the solution converges to zero for λ sufficiently small and to the unique positive or negative solution for larger values of λ for almost all initial data. See [2].

We discretise the equation using explicit Euler timestepping ($\theta = 0$) and seek period 2 solutions satisfying the Z_2 symmetry U = -V. We obtain the defining problem

(4.22)
$$\frac{\delta_x^2 U_j}{\Delta x^2} + \lambda (U_j - U_j^3) + \frac{2U_j}{\Delta t} = 0, \qquad j = 0, \cdots, J - 1,$$

together with boundary conditions (4.13). This problem is related to (4.12), (4.13) by a simple group of continuous transformations: we let $\gamma = 2/\Delta t$; then (4.22), (4.13) can be transformed into (4.12), (4.13) by

$$(4.23) U_j \to \lambda^{-\frac{1}{2}} U_j$$

and

$$(4.24) \qquad \qquad \gamma \to \gamma - \lambda.$$

Thus we expect the spurious periodic solutions of the Ginzburg-Landau equation to be of the same form as the solutions of Example 4.1 but scaled and shifted. The bifurcation points are shifted to the left (in $1/\Delta t$) and the branches are scaled down. This can be seen in Fig. 7, which should be compared with Fig. 4. The spurious periodic solutions affect the dynamics of the Euler method in a similar way to Example 4.1; it is necessary to restrict Δt in terms of the magnitude of the initial data to obtain satisfactory results.

Example 4.4. Consider the equation

(4.25)
$$u_t = u_{xx} - \frac{u^3}{1 + \delta u^4},$$

together with boundary conditions (4.2). When discretised in space alone, the system of nonlinear ODEs satisfies the criteria of Theorem 3.4(ii). Consequently we expect that branches of spurious periodic solutions cannot extend back to arbitrarily small Δt .

We discretise the equation using explicit Euler timestepping ($\theta = 0$) and seek period 2 solutions satisfying the Z_2 symmetry U = -V. We obtain the problem

(4.26)
$$\frac{\delta_x^2 U_j}{\Delta x^2} - \frac{U_j^3}{1 + \delta U_j^4} + \frac{2U_j}{\Delta t} = 0, \qquad j = 1, \cdots, J - 1,$$

together with boundary conditions (4.13). The numerical solution of this problem is shown in Fig. 8. Notice that the branches of spurious periodic solutions are bounded uniformly from $\Delta t = 0$. This is in accordance with Theorem 3.4(ii).

Notice also that the branches of solutions in Fig. 8 emanating from ||U|| = 0, $\Delta t = \Delta t_k$ are asymptotic to the same value Δt_k at $||U|| = \infty$. This results from the fact that the nonlinear source term is asymptotically negligible for $||U|| \ll 1$ and for $||U|| \gg 1$. This form of bifurcation from infinity can be established rigorously by applying the Asymptotic Bifurcation Theorem of Krasnosel'skii quoted in [25].

Example 4.5. Consider the viscous Burgers' equation

$$(4.27) u_t = u_{xx} + u u_x,$$



FIG. 7. l_2 norm of spurious periodic solutions versus Δt for the Ginzburg–Landau equation. $\lambda = 90.$



FIG. 8. l_2 norm of spurious periodic solutions versus Δt for Euler discretisation of (4.25).



FIG. 9. l_2 norm of spurious periodic solutions versus $1/\Delta t$ for the upwind scheme (4.27).

together with boundary conditions (4.2). A Liapunov functional approach shows that all solutions converge to zero as $t \to \infty$. We consider two different schemes for the solution of (4.27), (4.2). Both involve Euler timestepping but in one we use an upwind scheme and in the other a centred scheme.

Consider first the *upwind scheme* for which we take

$$g(a, b, c) = \frac{c^2 - b^2}{2\Delta x}$$
 for $j = 1, \dots, J - 1$.

Numerical continuation was implemented to follow the nontrivial branches away from the bifurcation points given by (4.7), (4.8). Figure 9 shows a plot of these branches in the case $\Delta x = 0.1$. Most of the branches are irrelevant since they exist only above the linear stability limit $(1/\Delta t \approx 200)$.

The branch emanating from the eighth bifurcation point (from left to right in $1/\Delta t$) enters into the region of linearised stability for the scheme. This branch is shown in Fig. 10. Note that the branch seems to terminate at $1/\Delta t \approx 360$, contradicting the Rabinowitz global bifuration theory [15]. What actually happens is that the solutions U and V satisfy U = V at this point and beyond this the roles of U and V interchange and the solution follows a symmetric transformation of the original branch back to the bifurcation point; in $(U, V, \Delta t)$ space it is actually a loop and there is no contradiction. This branch is of particular interest. At the point where U = V, we have obtained a steady solution of the discretisation of Burgers' equation. This steady solution is spurious and caused by *spatial discretisation* (in contrast to the spurious periodic solutions which are caused by temporal discretisation). The existence of spurious steady solutions for Burgers' equation (with nonhomogeneous boundary conditions)



FIG. 10. l_2 of a spurious periodic solution versus $1/\Delta t$ for the upwind scheme (4.27).

is discussed in [20], for the driven cavity problem in [18], and for nonlinear elliptic problems in [12].

We now discretise Burgers' equation by centred differencing. Thus we take

(4.28)
$$g(a,b,c) = \frac{c^2 - a^2}{4\Delta x}$$
 for $j = 1, \dots, J - 1$.

The spurious periodic solutions are shown in Fig. 11. The linear theory determining the bifurcation points is identical to that for the upwind scheme. Note, however, that the behaviour of the solution branches is entirely different far from the bifurcation points. This illustrates an important point: the interaction between spatial and temporal discretisation is crucial in the design of schemes which minimise the effect of spurious solutions.

Example 4.6. Consider the equation

(4.29)
$$u_t = u_{xx} + \lambda(u+u^3),$$

together with boundary conditions (4.2). We discretise the equation using backward Euler timestepping ($\theta = 1$) and seek period 2 solutions. The points at which spurious period 2 solutions bifurcate from the trivial solution are given by (4.8), which reduces to

(4.30)
$$\Delta t_k = -\frac{2\Delta x^2}{4\sin^2(k\pi/2J) - \lambda\Delta x^2} \quad \text{for} \quad i = 1, \cdots, J - 1.$$

We choose λ sufficiently large that $\Delta t_1 > 0$. Figure 12 shows the spurious periodic solution emanating from Δt_1 . Note that it can be found for all Δt sufficiently small and moves off to infinity in norm as predicted by Theorem 3.4(i).



FIG. 11. l_2 norm of spurious periodic solutions versus $1/\Delta t$ for the centred scheme.



FIG. 12. l_2 norm of spurious periodic solutions versus $\frac{1}{\Delta t}$ for backward Euler discretisation of (4.29).

5. Conclusions. In this paper we have investigated the dynamics of numerical methods for nonlinear differential equations. We have shown, by means of illustrative examples, that spurious periodic solutions can seriously degrade the performance of the numerical methods (§2). Consequently, it is important to design schemes which minimise the effect of spurious solutions. We have shown that the trapezoidal rule never possesses spurious solutions of period 2 in n; on the other hand, the theta method with $\theta \neq \frac{1}{2}$ does possess spurious solutions of period 2 in n. (This and other related results are in §3.) We have applied our theory to a number of specific examples arising from discretisation of semilinear parabolic equations (§4). In a specific case, we have generalised the asymptotic stability criterion $\Delta t/\Delta x^2 \leq \frac{1}{2}$ for the Euler discretisation of the heat equation to a nonlinear problem (Theorem 4.2). This result reflects the dependence upon initial conditions required in the nonlinear case; the precise form of the restriction for the nonlinear problem is intimately related to the existence of spurious periodic solutions in the discretisation.

Acknowledgements. The authors thank both referees, whose comments have improved the presentation. A.M. Stuart is grateful to A.R. Mitchell and C.M. Elliott for several helpful discussions.

REFERENCES

- F. BREZZI, S. USHIKI, AND H. FUJII, Real and ghost bifurcation dynamics in difference schemes for ordinary differential equations, in Numerical Methods for Bifurcation Problems, T. Kupper, H. D. Mittlemann, and H. Weber, eds., Birkhäuser-Verlag, Boston, 1984.
- N. CHAFEE AND E. INFANTE, A bifurcation problem for a nonlinear parabolic problem, J. Applic. Anal., 4 (1974), pp. 17–37.
- [3] S.-N. CHOW AND J. HALE, Methods of Bifurcation Theory, Springer-Verlag, Berlin, New York, 1982.
- [4] S. D. CONTE AND C. DE BOOR, *Elementary Numerical Analysis*, Third Edition, McGraw-Hill, New York, 1980.
- [5] D. F. GRIFFITHS, The dynamics of some linear multistep methods with step-size control, in Numerical Analysis, D. F. Griffiths and G. A. Watson, eds., Pitman, Boston, 1987.
- [6] D. F. GRIFFITHS AND A. R. MITCHELL, Stable periodic bifurcations of an explicit discretisation of a nonlinear partial differential equation in reaction diffusion, IMA J. Numer. Anal., 8 (1988), pp. 435–454.
- [7] J. GUCKENHEIMER AND P. HOLMES, Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields, in Applied Mathematical Sciences, 42, Springer-Verlag, Berlin, New York, 1983.
- [8] E. HAIRER, A. ISERLES, AND J. M. SANZ-SERNA, Equilibria of Runge-Kutta methods, Numer. Math., to appear.
- [9] A. R. HUMPHRIES, Numerical methods for ordinary differential equations and dynamical systems, Mathematical Tripos, Part III, University of Cambridge, Cambridge, 1990.
- [10] A. R. MITCHELL AND D. F. GRIFFITHS, Beyond the linearised stability limit in nonlinear problems, in Numerical Analysis, D. F. Griffiths and G. A. Watson, eds., Research Notes in Mathematics, 140, Pitman, Boston, 1986.
- [11] A. R. MITCHELL AND S. W. SCHOOMBIE, Nonlinear diffusion and stable period 2 solutions of a discrete reaction-diffusion model, J. Comput. Appl. Math., 25 (1989), pp. 363–372.
- [12] T. MURDOCH AND C. J. BUDD, Convergent and spurious solutions of nonlinear elliptic equations, in preparation.
- [13] A. C. NEWELL Finite amplitude instabilities of partial difference equations, SIAM J. Appl. Math., 33 (1977), pp. 133–160.
- [14] A. ISERLES, A. T. PEPLOW, AND A. M. STUART, A unified approach to spurious solutions introduced by time-discretisation, Part I, SIAM J. Numer. Anal., 28 (1991), to appear.
- [15] P. H. RABINOWITZ, Some global results for nonlinear eigenvalue problems, J. Funct. Anal., 7 (1971), pp. 487–513.
- [16] W. C. RHEINBOLDT, Numerical Analysis of Parameterised Nonlinear Equations, Wiley-Interscience, New York, 1986.

- [17] S. W. SCHOOMBIE AND A. R. MITCHELL, Numerical studies of a discretised version of Fisher's equation beyond the linearised stability limit, SIAM J. Sci. Statist. Comput., submitted.
- [18] R. SHREIBER AND H. B. KELLER, Spurious solutions of the driven cavity calculations, J. Comput. Phys., 49 (1983), pp. 165–172.
- [19] B. D. SLEEMAN, D. F. GRIFFITHS, A. R. MITCHELL, AND P. D. SMITH, Stable periodic solutions in nonlinear difference equations, SIAM J. Sci. Statist. Comput., 9 (1988), pp. 543–557.
- [20] A. B. STEPHENS AND G. R. SHUBIN, Multiple solutions and bifurcations of finite difference approximations to some steady problems of fluid mechanics, SIAM J. Sci. Statist. Comput., 2 (1981), pp. 404-415.
- [21] A. M. STUART Nonlinear instability in dissipative finite difference schemes, SIAM Rev., 31 (1989), pp. 191–220.
- [22] —, Linear instability implies spurious periodic solutions, IMA J. Numer. Anal., 9 (1989), pp. 465–486.
- [23] ——, The global attractor under discretisation, in Continuation and Bifurcations: Numerical Techniques and Applications, D. Roose, B. De Dier, and A. Spence, eds., NATO ASI Series, Kluwer, the Netherlands, 1990.
- [24] A. M. STUART AND M. S. FLOATER, On the computation of blow-up, European J. Appl. Math., 1 (1990), pp. 47-71.
- [25] J. F. TOLAND, Asymptotic linearity and nonlinear eigenvalue problems, Quart. J. Math., 24 (1973), pp. 241–250.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.