B-Methods for the Numerical Solution of Evolution Problems with Blow-up Solutions
Part II: Splitting Methods

Mélanie Beck[a], Martin J. Gander[b,*], Felix Kwok[c]

[a] *Dawson College, Department of Mathematics, H3Z 1A4, Montreal, Canada*
[b] *University of Geneva, Section of Mathematics, 1205, Geneva, Switzerland*
[c] *Université Laval, Département de mathématiques et de statistique, G1V 0A6, Québec, Canada*

Abstract

B-methods are numerical methods which are especially tailored to solve nonlinear partial differential equation that have blow up solutions. We have presented in Part I a systematic construction of B-methods based on the variation of constants formula. Here, we use splitting methods as a second way to construct B-methods, and we prove several special properties of such methods. We illustrate our analysis with numerical experiments.

Keywords: Geometric Integration, Blow-up Solutions, Non-Linear Partial Differential Equations, Nonlinear Systems of Equations, Splitting Methods

1. Introduction

Nonlinear partial differential equations (PDEs) arise in many important models in science and engineering, and very few of those models have closed form solutions. One therefore has to resort to numerical methods to compute approximations. If the partial differential equation has further geometric properties, it is often an advantage for the numerical approximation to also have the same geometric property, which led to the research field of geometric numerical integration. Much progress has been made over the last two decades in this area, see for example [28, 40, 27, 15, 7] and references therein.

In specific applications, the nonlinear PDE models can have solutions that blow up in finite time. This is in particular the case for combustion models [23, 17, 31, 33], turbulent flow [38], nonlinear optics [36, 41, 42] and population dynamics [46, 26]. This blow-up indicates in general that the model is losing its validity, and it is therefore important to understand the precise behavior

*Corresponding Author

Email addresses: mbeck@dawsoncollege.qc.ca (Mélanie Beck), martin.gander@unige.ch (Martin J. Gander), felix.kwok@mat.ulaval.ca (Felix Kwok)

of the model when the blow-up time is approached. Studying such blow-up phenomena is necessarily done on a case-by-case basis, see for example [35, 49, 24, 50, 25, 20, 2, 19, 16], and the reviews [3, 21]. Blow-up can even happen when first integrals are conserved, see [9] and references therein. The analysis of blow-up phenomena is an active field of research, and many results have been obtained over the past two decades, see [8, 43, 37, 44, 29, 13, 14, 47] and references therein.

The construction of numerical methods to approximate the blow-up time and rate of such models focuses in general on adaptive techniques. Very successful are moving mesh methods, see [10]. In [11] self-similar solution techniques are employed for obtaining a scale invariant adaptive numerical method. A more direct numerical time stepping approach can be found in [30], where a numerical method using arclength ingredients is constructed and analyzed, and in [48], compactification of base spaces is combined with the validation of Lyapunov functions. Adaptive time stepping is also a very successful technique, where the time step is proportional to the inverse of the norm of the solution, see for example [45].

Considering blow-up as a geometric property is a more recent area of research, and so far mostly ad hoc constructions have been used to obtain numerical schemes with good blow-up properties, see for example [39]. In a first paper [6], we have shown how one can systematically construct B-methods using the technique of variation of the constant. The goal of this paper is to present a second systematic way of obtaining B-methods, using splitting techniques.

2. B-Methods Based on Splitting

To fix ideas, we first show the construction of splitting B-methods for a quasi-linear parabolic problem. The construction for a few other scalar or systems of nonlinear partial differential equations can be found in Section 4.

2.1. Model Problem and Assumptions

We consider the quasi-linear parabolic partial differential equation

$$\begin{array}{lll} u_t &= \Delta u^m + \delta F(u), & \text{for } (x,t) \in \Omega \times (0,T), \\ u &= 0, & \text{for } (x,t) \in \partial\Omega \times (0,T), \\ u(x,0) &= u_0(x), & \text{for } x \in \Omega, \end{array} \tag{1}$$

where δ is a positive constant, Ω is a bounded domain of $\mathbb{R}^d$ and u_0 is a positive continuous function on $\bar{\Omega}$. In our analysis, we need the following

Assumption 1. *The function F is assumed to be positive, strictly increasing and strictly convex on $(0,\infty)$, belonging to $C^2([0,\infty))$ and satisfying*

$$\int_b^\infty \frac{ds}{F(s)} < \infty, \tag{2}$$

for $b>0$. Then the function $g(s)=\int_s^\infty \frac{1}{F(\sigma)}d\sigma$ is continuous and strictly decreasing on $(0,\infty)$. The function $G=g^{-1}$ is continuous and strictly decreasing on $(0,M)$, where $M=\lim_{s\to 0} g(s)\le\infty$. Note also that g and G are positive with $\lim_{s\to\infty} g(s)=0$ and $\lim_{s\to 0} G(s)=\infty$.

In order to be able to construct B-methods, we need to get an explicit form of g and often G. Examples of functions F which satisfy all these conditions are

- $F(u)=e^u,\ g(u)=e^{-u},\ G(u)=-\ln u,$
- $F(u)=(u+\alpha)^{p+1},\ \alpha\ge 0,\ p>0,\ g(u)=\frac{1}{p(u+\alpha)^p},\ G(u)=(pu)^{-1/p}-\alpha,$
- $F(u)=e^u-1,\ g(u)=\ln\left(\frac{e^u}{e^u-1}\right),\ G(u)=u-\ln(e^u-1),$
- $F(u)=(u+1)[\ln(u+1)]^{p+1},\ p>0,\ g(u)=\frac{1}{p[\ln(u+1)]^p},\ G(u)=e^{(pu)^{-1/p}}-1,$
- $F(u)=u^2+1,\ g(u)=\frac{\pi}{2}-\arctan(u),\ G(u)=\cot(u).$

In problem (1) the nonlinearity in F is responsible for the finite-time blow-up and becomes increasingly important as we approach the blow-up time. The conditions imposed on F allow us to write explicitly the solution of the nonlinear ordinary differential equation $y_t=\delta F(y)$. Indeed we get for any $S>0$, $\int_{y(t)}^{y(S)}\frac{ds}{F(s)}=\int_t^S\delta ds$, and then $g(y(t))=[g(y(S))+\delta S]-\delta t$, that is

$$y(t)=y(t,K)=G(K-\delta t), \tag{3}$$

where K is a constant, for all t satisfying $K-\delta t\in(0,M)$. It is then natural to seek integrators that exploit this information. In the following we present a new approach to obtain semi-discretizations in time for the semi-linear problem (1) from this exact solution. This approach allows us to derive many new B-methods which are different from the B-methods obtained using the variation of constants approach in [6].

2.2. Construction of B-Methods based on Splitting

As suggested in Hairer, Lubich and Wanner [28][1], one way to exploit the exact solution of the nonlinear part of the equation is by using splitting methods. We illustrate this construction on the quasi-linear scalar PDE (1); the construction for a system of semi-linear PDEs is given in Section 4.3.

If we decompose $u_t=\Delta u^m+\delta F(u)$ into $f^{[1]}(u)=\delta F(u)$ and $f^{[2]}(u)=\Delta u^m$, we can make good use of the fact that we know the exact flow $\varphi_t^{[1]}$ of $u_t=\delta F(u)$ (note that φ_t does not represent a time derivative) . Indeed, the exact flow of

[1]"It may happen that the differential equation $\dot y=f(y)$ can be split according to $\dot y=f^{[1]}(y)+f^{[2]}(y)$, such that only the flow of, say, $\dot y=f^{[1]}(y)$ can be computed exactly. If $f^{[1]}(y)$ constitutes the dominant part of the vector field, it is natural to search for integrators that exploit this information."

an equation $y_t = f(y)$ is the map defined by $\varphi_t(y_0) = y(t)$ if $y(0) = y_0$, so in this case, using (3), we have $\varphi_t^{[1]}(u_n) = G(g(u_n) - \delta t)$, for $t < g(u_n)/\delta$. Then we can choose any numerical integrator $\Phi_h^{[2]}$ for $u_t = \Delta u^m$, and by composing the exact flow and the numerical integrator, we obtain two new methods for $u_t = \Delta u^m + \delta F(u)$,

$$\Phi_h = \varphi_h^{[1]} \circ \Phi_h^{[2]} \qquad \text{and} \qquad \Phi_h^* = \Phi_h^{[2]*} \circ \varphi_h^{[1]}, \tag{4}$$

where $\Phi_h^{[2]*}$ is the adjoint of $\Phi_h^{[2]}$ (see Section II.3 in [28]). Note that the two original methods $\Phi_h^{[2]}$ and $\Phi_h^{[2]*}$ are consistent, that is

$$\Phi_h^{[2]}(z_0) = z_0 + hf^{[2]}(z_0) + O(h^p) \quad \text{and} \quad \Phi_h^{[2]*}(z_0) = z_0 + hf^{[2]}(z_0) + O(h^p),$$

with $p \geq 2$. Moreover, $\varphi_t^{[1]}$ is the exact flow of $u_t = \delta F(u)$, so that its Taylor expansion is

$$\varphi_h^{[1]}(y_0) = y(h) = y_0 + hf^{[1]}(y_0) + O(h^2).$$

Therefore, the resulting methods Φ_h and Φ_h^* are of first order. This construction can only lead to methods of first order, however as these two integrators are adjoint, we can use them as the basis of the composition method

$$\Phi_h = \Phi_{\alpha_s h} \circ \Phi_{\beta_s h}^* \circ \cdots \circ \Phi_{\beta_2 h}^* \circ \Phi_{\alpha_1 h} \circ \Phi_{\beta_1 h}^*,$$

to construct methods of any desired order (see [28]). In particular, by choosing $\alpha_1 = \beta_1 = 1/2$ for $s = 1$, we obtain a second-order symmetric method $\Psi_h = \Phi_{h/2} \circ \Phi_{h/2}^*$. It is interesting to note that if Φ_h (not $\Phi_h^{[2]}$) is the forward (respectively backward) Euler method, the resulting method Ψ_h corresponds to the midpoint (respectively trapezoidal) rule.

We saw that the exact flow of $u_t = \delta F(u)$ is given by $\varphi_t^{[1]}(u_n) = G(g(u_n) - \delta t)$, so we just have to choose a numerical integrator for the second part $u_t = \Delta u^m$. For example, even though this problem is stiff, we start with forward Euler $\Phi_h^{[2]}(u_n) = u_n + h\Delta u_n^m$, whose adjoint is backward Euler $\Phi_h^{[2]*}(u_n) = u_n + h\Delta u_{n+1}^m$. By composing these integrators with the exact flow $\varphi_t^{[1]}$, we get two B-methods. The first one is the *Splitting Forward Euler B-Method* (SpFE)

$$\Phi_h(u_n) = \varphi_h^{[1]} \circ \Phi_h^{[2]}(u_n), \tag{5}$$

which gives the explicit scheme

$$u_{n+1} = G(g(u_n + h\Delta u_n^m) - \delta h), \tag{6}$$

and requires the condition $g(u_n + h\Delta u_n^m) \in (0, M)$. The second one is the *Splitting Forward Euler Adjoint B-Method* (SpFE)*

$$\Phi_h^*(u_n) = \Phi_h^{[2]*} \circ \varphi_h^{[1]}(u_n), \tag{7}$$

which gives the implicit scheme

$$u_{n+1} = G(g(u_n) - \delta h) + h\Delta u_{n+1}^m, \tag{8}$$

and requires the condition $g(u_n) - \delta h \in (0, M)$. This scheme is studied in detail in Section 3 for the special semi-linear case, $m = 1$.

Instead of choosing $\Phi_h^{[2]}$ to be forward Euler in (4), we could choose it to be backward Euler; then the adjoint $\Phi_h^{[2]*}$ is forward Euler and the resulting schemes are the *Splitting Backward Euler B-Method* (SpBE)

$$\Phi_h(u_n) = u_{n+1} = G(g(v) - \delta h), \text{ with } v = u_n + h\Delta v^m, \tag{9}$$

and the *Splitting Backward Euler Adjoint B-Method* (SpBE)

$$\Phi_h^*(u_n) = u_{n+1} = G(g(u_n) - \delta h) + h\Delta(G(g(u_n) - \delta h))^m. \tag{10}$$

Another possibility would be to choose $\Phi_h^{[2]}$ to be a second-order method, like the symmetric midpoint rule (SpMid) or the trapezoidal rule (SpTrap). However, the scheme becomes more complicated without necessarily bringing more accuracy, as the resulting scheme is only first order. In order to get higher order methods, we need to compose first order methods. The simplest way to obtain a second-order method is thus to construct

$$\Psi_h = \Phi_{h/2} \circ \Phi_{h/2}^* = \varphi_{h/2}^{[1]} \circ \Phi_{h/2}^{[2]} \circ \Phi_{h/2}^{[2]*} \circ \varphi_{h/2}^{[1]}, \tag{11}$$

where $\Phi_h^{[2]}$ and $\Phi_h^{[2]*}$ are adjoint first-order methods.

If we choose $\Phi_h^{[2]}$ to be forward Euler, we obtain the *Second order Splitting Forward Euler B-Method* (SoSpFE)

$$\Psi_h(u_n) = G\left(g(v + \frac{h}{2}\Delta v^m) - \frac{\delta h}{2}\right), \text{ with } v - \frac{h}{2}\Delta v^m = G\left(g(u_n) - \frac{\delta h}{2}\right), \tag{12}$$

and if $\Phi_h^{[2]}$ is chosen to be backward Euler, we get the *Second order Splitting Backward Euler B-Method* (SoSpBE)

$$\Psi_h(u_n) = G(g(v) - \frac{\delta h}{2}), \text{ with } v - \frac{h}{2}\Delta v^m = G(g(u_n) - \frac{\delta h}{2}) + \frac{h}{2}\Delta(G(g(u_n) - \frac{\delta h}{2}))^m. \tag{13}$$

Similarly we can construct arbitrary high order splitting B-methods.

2.3. Truncation Error Analysis

In order to show that B-methods have the potential to be better than standard methods, we need to compare the local truncation errors of both types of methods. To start, we consider the problem $u_t = F(u) + \Upsilon(u)$, where Υ can be a function or an operator (like the Laplacian in our example). We denote by φ the function that satisfies

$$\varphi_t(t, v) = F(\varphi(t, v)), \quad \text{and} \quad \varphi(0, v) = v, \quad \forall v. \tag{14}$$

Keeping the notation introduced earlier, we have $\varphi(t,v) = G(g(v)-t)$. We also consider the numerical method Φ applied to $v_t = \Upsilon(v)$, with $v(0) = v_0$. If $v(t)$ solves this simplified problem, we have

$$\Phi(h, v_0) = v(h) + E(h), \tag{15}$$

where E represents the local truncation error of the standard method.

We first consider the B-methods obtained by applying the numerical method first and use the result in the exact scheme (like SpFE or SpBE): starting with u_0, we define $v_0 = u_0$ and we apply the numerical method Φ to get $v_1 = v(h)+E(h)$, then we set $u_1(h) := \varphi(h, v_1) = \varphi(h, v(h)+E(h))$. To expand u_1 as a series of h, we need to compute its derivatives, $u_1'(h) = \varphi_t + \varphi_v\,(v'(h)+E'(h))$ and

$$u_1''(h) = \varphi_{tt} + 2\varphi_{tv}\,(v'+E') + \varphi_{vv}\,(v'+E')^2 + \varphi_v\,(v''+E''),$$

where the derivatives of φ are evaluated at $(h, v(h)+E(h))$.

From the definition of φ given in (14) (or using $\varphi(t,v) = G(g(v)-t)$), we obtain $u_1(0) = \varphi(0, v(0)+E(0)) = \varphi(0,u_0) = u_0$, $\varphi_t = F(\varphi)$, $\varphi_v(0,v) = 1$, $\varphi_{tt} = F'(\varphi)\varphi_t$, $\varphi_{tv} = F'(\varphi)\varphi_v$ and $\varphi_{vv}(0,v) = 0$. Moreover we have $v'(h) = \Upsilon(v)$ and $v''(h) = \Upsilon'(v)\Upsilon(v)$. Hence the derivatives of u_1 evaluated at $h = 0$ are $u_1'(0) = F(u_0) + \Upsilon(u_0) + E'(0)$, and

$$u_1''(0) = F'(u_0)F(u_0) + 2F'(u_0)(\Upsilon(u_0)+E'(0)) + \Upsilon'(u_0)\Upsilon(u_0) + E''(0).$$

The values of $E'(0)$ and $E''(0)$ depend on the standard method used, in particular for any consistent method, we have $E'(0) = 0$ and if the method is of second or higher order, we also have $E''(0) = 0$.

The Taylor expansion of the exact solution u is

$$u(h) = u_0 + h(\Upsilon(u_0)+F(u_0)) + \frac{h^2}{2}(\Upsilon'(u_0)+F'(u_0))(\Upsilon(u_0)+F(u_0)) + \cdots, \tag{16}$$

where the derivative $\Upsilon'(u_0)$ can be an operator, so the local truncation error of the B-methods is given by

$$\tau_B := u_1 - u(h) = \frac{h^2}{2}\Big(F'(u_0)\Upsilon(u_0) - \Upsilon'(u_0)F(u_0) + E''(0)\Big) + O(h^3), \tag{17}$$

if a first-order standard method is used, and for higher order standard methods we get

$$\tau_B = \frac{h^2}{2}\Big(F'(u_0)\Upsilon(u_0) - \Upsilon'(u_0)F(u_0)\Big) + O(h^3).$$

To construct the adjoint methods, we first use the exact scheme and then apply a numerical methods on the result. In other words, starting with the initial condition u_0, we define $v_0 = \varphi(h, u_0)$, where φ satisfies condition (14), and we compute $u_1 = \Phi(h, v_0)$, where Φ is defined by (15) (to get a simpler notation, we denote the numerical method by Φ instead of Φ^*). The definition of Φ implies in particular that for all ξ, we have

$$\Phi(0,\xi) = \xi + E(0), \quad \Phi_t(0,\xi) = \Upsilon(\xi) + E'(0), \quad \Phi_{tt}(0,\xi) = \Upsilon'(\xi)\Upsilon(\xi) + E''(0), \tag{18}$$

and

$$\Phi_v(0,\xi) = 1, \quad \Phi_{vv}(0,\xi) = 0, \quad \text{and} \quad \Phi_{tv}(0,\xi) = \Upsilon'(\xi). \tag{19}$$

We now expand $u_1 = \Phi(h, \varphi(h, u_0))$ in a series of h. The derivatives of u_1 are

$$\begin{aligned} u_1'(h) &= \Phi_t(h, \varphi(h, u_0)) + \Phi_v(h, \varphi(h, u_0)) \cdot \varphi_t(h, u_0), \\ u_1''(h) &= \Phi_{tt}(h,\varphi) + 2\Phi_{tv}(h,\varphi)\varphi_t(h,\varphi) + \Phi_{vv}(h,\varphi)\varphi_t(h,\varphi)^2 + \Phi_v(h,\varphi)\varphi_{tt}(h,\varphi). \end{aligned}$$

Noting that $u_1(0) = \Phi(0, \varphi(0, u_0)) = \varphi(0, u_0) = u_0$, we evaluate u_1, u_1' and u_1'' at $h = 0$ and get

$$\begin{aligned} u_1'(0) &= \Upsilon(u_0) + E'(0) + F(u_0), \\ u_1''(0) &= \Upsilon'(u_0)\Upsilon(u_0) + E''(0) + 2\Upsilon'(u_0)F(u_0) + F'(u_0)F(u_0), \end{aligned}$$

where we used the properties of Φ in (18) and (19) and the definition of φ given in (14). As the Taylor expansion of the exact solution u is given by (16), the local truncation errors of these B-methods are, as expected,

$$\tau_{B^*} = \frac{h^2}{2}\Big(\Upsilon'(u_0)F(u_0) + E''(0) - F'(u_0)\Upsilon(u_0)\Big) + O(h^3), \tag{20}$$

for first-order standard methods, and for higher-order standard methods we get

$$\tau_{B^*} = \frac{h^2}{2}\Big(\Upsilon'(u_0)F(u_0) - F'(u_0)\Upsilon(u_0)\Big) + O(h^3).$$

We now need to show that in case of finite-time blow-up, the local truncation error of B-methods is smaller than that of the corresponding standard methods. We illustrate the difference in truncation errors by considering the forward Euler method.

The local truncation error of the forward Euler method applied to the general equation $y_t = f(t, y)$ is given by

$$\tau := y_1 - y(h) = -\frac{h^2}{2}(f_t + f_y f) + O(h^3), \tag{21}$$

which means that if we apply this method to $u_t = F(u) + \Upsilon(u)$, we obtain

$$\tau_s = -\frac{h^2}{2}(\Upsilon'(u_0) + F'(u_0))(\Upsilon(u_0) + F(u_0)) + O(h^3). \tag{22}$$

On the other hand, if we apply forward Euler to $v_t = \Upsilon(v)$, we obtain $E(h) = -\frac{h^2}{2}\left[\Upsilon'(v_0)\Upsilon(v_0)\right] + O(h^3)$, which gives $E''(0) = -\Upsilon'(v_0)\Upsilon(v_0)$. Going back to

(17) and (20) we obtain the truncation error of the corresponding B-methods,

$$\tau_B = \frac{h^2}{2}\Big(F'(u_0)\Upsilon(u_0) - \Upsilon'(u_0)F(u_0) - \Upsilon'(u_0)\Upsilon(u_0)\Big) + O(h^3)$$
$$\tau_{B^*} = -\frac{h^2}{2}\Big(F'(u_0)\Upsilon(u_0) - \Upsilon'(u_0)F(u_0) + \Upsilon'(u_0)\Upsilon(u_0)\Big) + O(h^3).$$

In order for the function F to be responsible for the finite-time blow-up, it needs to be superlinear at infinity, while the remaining part $\Upsilon(u)$ becomes less important as u becomes large. We therefore expect the term $F'(u_0)F(u_0)$, which is present in τ_s but absent in τ_B and τ_{B^*}, to be large relative to the other terms. As an example, let us first consider the case where $\Upsilon(u)$ is a bounded function of u. We define $F(u) := e^u$ and $\Upsilon(u) := \sin(u)$. The local truncation error can then be written as

$$\begin{aligned}\tau_s &= -\frac{h^2}{2}(\cos(u_0) + e^{u_0})(\sin(u_0) + e^{u_0}) + O(h^3),\\ &= -\frac{h^2}{2}\Big(e^{2u_0} + e^{u_0}(\sin(u_0) + \cos(u_0)) + \cos(u_0)\sin(u_0)\Big) + O(h^3)\end{aligned}$$

for the standard method and

$$\tau_B = \frac{h^2}{2}\Big(e^{u_0}(\sin(u_0) - \cos(u_0)) - \cos(u_0)\sin(u_0)\Big) + O(h^3),$$

for the specialized SpFE method. We see that the fastest growing term $(e^{u_0})^2$ in τ_s does not appear in τ_B, while the other terms are of similar order. Given the size of this term compared to the remaining terms, τ_B is considerably smaller than τ_s.

Going back to the case $\Upsilon(u) = \Delta u$, we observe numerically the same phenomenon. Indeed, with $F(u) = 3e^u$ and $\Upsilon(u) = \Delta u$, the local truncation errors are

$$\begin{aligned}\tau_s &= -\frac{h^2}{2}\Big(\Delta(\Delta u_0 + 3e^{u_0}) + 3e^{u_0}\Delta u_0 + 9e^{2u_0}\Big) + O(h^3),\\ \tau_B &= \frac{h^2}{2}\Big(3e^{u_0}\Delta u_0 - \Delta(3e^{u_0}) - \Delta(\Delta u_0)\Big) + O(h^3)\end{aligned}$$

for the SpFE method, whereas for the (SpFE)* method, we have

$$\tau_{B^*} = -\frac{h^2}{2}\Big(3e^{u_0}\Delta u_0 - \Delta(3e^{u_0}) + \Delta(\Delta u_0)\Big) + O(h^3).$$

In this case, the term e^{2u_0} of τ_s is also absent from τ_B and τ_{B^*}, but it is not obvious that this term is much larger than the remaining terms. Some numerical experiments using Matlab show that the difference between e^{2u} and the other terms is considerable and increases as u gets larger. Using the built-in adaptive method ode45 we computed the solution of $u_t = 3e^u + \Delta u$ on $[-1,1]$ with $u_0(x) = \cos(\pi x/2)$, we then evaluated each of the four terms that appear in τ_s. When $t = 0.1660$ (the blow-up occurs approximately at t=0.1664), the norm of the different terms are $\|\Delta(\Delta u_0)\|_2 = 342\,439$, $\|\Delta(3e^{u_0})\|_2 = 1\,466\,377$, $\|3e^{u_0}(\Delta u_0)\|_2 = 1\,542\,768$, and $\|(3e^{u_0})^2\|_2 = 16\,544\,121$. So removing this last term from the local error greatly improves the results in this example.

3. Analysis of (SpFE)* for semi-linear parabolic problems

We now analyze the properties of the (SpFE)* scheme applied to the model problem (1) for the special case of $m = 1$, i.e., when the problem is semi-linear. An explicit formula for the method is given by (8). By letting $A := -\Delta$, the scheme (8) for $m = 1$ can be written in the form

$$Au_{n+1} = f(x, u_{n+1}) = -\frac{1}{h}u_{n+1} + \frac{1}{h}G(g(u_n) - \delta h). \tag{23}$$

3.1. Existence and Uniqueness of the Solution

Since the scheme (23) is linear, it has a unique solution if and only if $G(g(u_n) - \delta h)$ is well defined, i.e., $g(u_n) \in (\delta h, M + \delta h)$. Since g is decreasing, $M = \lim_{s\to 0} g(s)$ and $u_n > 0$, we have $g(u_n) < M + \delta h$, so the only condition is $\|u_n\|_\infty < G(\delta h)$. We will need the following theorem due to Amann [1]:

Theorem 1 (Amann). *Let $f \in C^\alpha(\bar{\Omega} \times \mathbb{R}_+)$ be given, with $\alpha \in (0,1)$, and assume that $f(x,0) \geq 0$. Then a necessary and sufficient condition for the existence of a non-negative solution $u \in C^{2+\alpha}(\Omega)$ of the BVP*

$$Au := -\Delta u = f(x,u) \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega, \tag{24}$$

is the existence of a non-negative function $v \in C^{2+\alpha}(\bar{\Omega})$ satisfying

$$Av \geq f(x,v) \text{ in } \Omega, \quad v \geq 0 \text{ on } \partial\Omega.$$

Moreover, if this condition is satisfied, there exist a maximal non-negative solution $\hat{u} \leq v$ and a minimal non-negative solution $\bar{u} \leq v$ in the sense that, for every non-negative solution $u \leq v$ of (24), the inequality $\bar{u} \leq u \leq \hat{u}$ holds.

Theorem 2. *If the function u_n is positive in Ω, continuous in $\bar{\Omega}$, and satisfies*

$$\|u_n\|_\infty < G(\delta h), \tag{25}$$

then the scheme (23) has a maximal non-negative solution $\hat{u} \leq C_n = G(g(\|u_n\|_\infty) - \delta h)$, a minimal solution $\bar{u} \geq 0$, and if u is solution, then $u \in C^2(\bar{\Omega})$ and $\bar{u} \leq u \leq \hat{u}$.

Remark 1. *We can make the bound in condition (25) on the right-hand side as large as desired by choosing h small enough. This condition is necessary for the scheme (23) to be well-defined.*

PROOF. The constant C_n is a supersolution of the scheme, if it satisfies $-\frac{1}{h}C_n + \frac{1}{h}G(g(u_n) - \delta h) \leq 0 \; (= AC_n)$, that is $C_n \geq G(g(u_n) - \delta h)$. Hence the constant $C_n = G(g(\|u_n\|_\infty) - \delta h)$, which is well-defined if condition (25) is satisfied and positive by definition of G (see Assumption 1), is a supersolution. Moreover, since $f(x, 0) = \frac{1}{h}G(g(u_n) - \delta h) > 0$, we conclude using Theorem 1 from Amann.

Since $u_n \equiv 0$ is not a solution of the scheme, this result implies that there exists a non-zero nonnegative solution. Moreover the strong maximum principle applies (see for example [51]) and any nonnegative solution is positive on Ω. Uniqueness of the positive solution can also be obtained using the following result of Keller [34] with $m = 0$ and $M = C_n$.

Theorem 3 (Keller). *If there exist two constants m and M such that for all $x \in \Omega$ and all u_1, u_2 such that $m \leq u_1 < u_2 \leq M$, we have $f(x, u_1) \geq f(x, u_2)$, then problem (24) has at most one solution $u \in C^2$ satisfying $m \leq u \leq M$.*

Since $f(x, u)$ defined in (23) is decreasing in u, we get the uniqueness of the solution, and we can show the same minimal time of existence as for the VBE scheme in [6]:

Theorem 4. *Scheme (23) has a unique positive solution u_n for n such that $t_n = nh < T_1$, where $T_1 = \frac{1}{\delta}\, g(\|u_0\|_\infty) = \int_{\|u_0\|_\infty}^{\infty} \frac{ds}{\delta F(s)}$.*

Since we know from Theorem 2 that

$$\text{if} \quad \|u_n\|_\infty < G(\delta h), \quad \text{then } \|u_{n+1}\|_\infty \leq C_n = G(g(\|u_n\|_\infty) - \delta h),$$

the proof is exactly the same as the proof of [6, Theorem 3.11].

Finally, we recall that the scheme (23) is linear, so that no specialized nonlinear solver is required to solve for u_{n+1}.

3.2. Rate of Growth

We now prove some growth rate estimates for the scheme (23). Note that we will do this on a case-by-case basis for the functions $F(u)$ listed in the introduction, since the estimate depends on the particular function at hand. We first consider the function $F(u) = e^u$, before turning our attention to the case of $F(u) = (u + \alpha)^{p+1}$.

Theorem 5. *Let C_0 be a constant such that*

$$C_0 \geq \delta e^{\|u_0\|_\infty} \quad \text{and} \quad Au_0 - \delta e^{u_0} + C_0 \geq 0. \tag{26}$$

If $t_{n+1} < T_2 := \frac{1}{C_0}$, the function u_{n+1} given by

$$u_{n+1} + hAu_{n+1} = -\ln(e^{-u_n} - \delta h) \tag{27}$$

satisfies for all x

$$u_{n+1}(x) \leq u_n(x) + \ln\left(\frac{T_2 - t_n}{T_2 - t_{n+1}}\right).$$

Remark 2. *Note that if $Au_0 \geq 0$, we can take $C_0 = \delta e^{\|u_0\|_\infty}$, so that $T_2 = \frac{1}{C_0} = \frac{1}{\delta} g(\|u_0\|_\infty) = T_1$, as defined in Theorem 4.*

Proof. We prove this result by induction, using a supersolution approach. First, let us prove that if $t_1 = h < T_2$, we have

$$u_1 \leq u_0 + \ln\left(\frac{T_2}{T_2 - h}\right). \tag{28}$$

The function $u_0 + \ln\left(\frac{T_2}{T_2-h}\right) = u_0 - \ln\left(1 - \frac{h}{T_2}\right) = u_0 - \ln(1 - hC_0)$ is a supersolution of (27) with $n = 0$ if $u_0 - \ln(1 - hC_0) + hA(u_0 - \ln(1 - hC_0)) \geq -\ln(e^{-u_0} - \delta h)$, which simplifies to

$$Au_0 \geq \frac{1}{h}\ln(1 - hC_0) - \frac{1}{h}\ln(1 - \delta h e^{u_0}). \tag{29}$$

Since $\ln(1-x) = -\sum_{k\geq 1} \frac{x^k}{k}$ for x smaller than 1, we have

$$\begin{aligned}\beta(h) := \frac{1}{h}\ln(1 - hC_0) - \frac{1}{h}\ln(1 - \delta h e^{u_0}) &= \frac{-1}{h}\sum_{k=1}^{\infty}\frac{(hC_0)^k}{k} + \frac{1}{h}\sum_{k=1}^{\infty}\frac{(\delta h e^{u_0})^k}{k} \\ &= \sum_{k=0}^{\infty}\frac{h^k}{k+1}[(\delta e^{u_0})^{k+1} - C_0^{k+1}].\end{aligned}$$

Since $\frac{1}{T_2} = C_0 \geq \delta e^{\|u_0\|} \geq \delta e^{u_0}$, the bracket is negative and β is decreasing in h so inequality (29) holds for all $h \in (0, T_2)$ if

$$Au_0 \geq \lim_{h\to 0}\left(\frac{1}{h}\ln(1 - hC_0) - \frac{1}{h}\ln(1 - \delta h e^{u_0})\right) = \delta e^{u_0} - C_0,$$

which is exactly condition (26), and we get (28). To complete the induction we assume that

$$u_n \leq u_{n-1} + \ln\left(\frac{T_2 - t_{n-1}}{T_2 - t_n}\right), \tag{30}$$

and we show that $u_n + \ln\left(\frac{T_2 - t_n}{T_2 - t_{n+1}}\right)$ is a supersolution of (27), that is

$$u_n + \ln\left(\frac{T_2 - t_n}{T_2 - t_{n+1}}\right) + hAu_n + \ln(e^{-u_n} - \delta h) \geq 0. \tag{31}$$

First, we note that since $\frac{1}{T_2} = C_0 \geq \delta e^{\|u_0\|}$, we have $T_2 \leq \frac{1}{\delta e^{\|u_0\|}}$, and $u_0 \leq \|u_0\| \leq \ln(\frac{1}{\delta T_2})$, and by induction

$$\begin{aligned}u_n &\leq u_{n-1} + \ln\left(\frac{T_2 - t_{n-1}}{T_2 - t_n}\right) \\ &\leq \ln\left(\frac{1}{\delta(T_2 - t_{n-1})}\right) + \ln\left(\frac{T_2 - t_{n-1}}{T_2 - t_n}\right) = \ln\left(\frac{1}{\delta(T_2 - t_n)}\right).\end{aligned} \tag{32}$$

By definition of u_n, we have $u_n + hAu_n = -\ln(e^{-u_{n-1}} - \delta h)$, and from the induction hypothesis (30), we obtain $-\ln(e^{-u_{n-1}} - \delta h) > -\ln[e^{-u_n}(\frac{T_2 - t_{n-1}}{T_2 - t_n}) - \delta h]$, so that inequality (31) is satisfied if

$$\ln\left(\frac{T_2 - t_n}{T_2 - t_{n+1}}\right) - \ln\left[e^{-u_n}\left(\frac{T_2 - t_{n-1}}{T_2 - t_n}\right) - \delta h\right] + \ln(e^{-u_n} - \delta h) \geq 0.$$

which simplifies to $(T_2 - t_n)\delta \leq e^{-u_n}$, which is exactly (32).

Next, we consider the (SpFE)* scheme for $F(u) = (u + \alpha)^{p+1}$, $p > 0$.

Theorem 6. *Let $p > 0$. Suppose there exists a constant C_0 that satisfies*

$$C_0 \geq p\delta(\|u_0\|_\infty + \alpha)^p \quad \textit{and} \quad Au_0 \geq 0. \tag{33}$$

If $t_{n+1} < T_2 := \frac{1}{C_0}$, the function u_{n+1} given by

$$u_{n+1} + hAu_{n+1} = [(u_n + \alpha)^{-p} - p\delta h]^{-1/p} - \alpha, \tag{34}$$

satisfies for all x

$$u_{n+1} + \alpha \leq \left(\frac{T_2 - t_n}{T_2 - t_{n+1}}\right)^{1/p}(u_n + \alpha).$$

Proof. Throughout this proof, we will write $w_n = u_n + \alpha$ for all n. The recurrence can then be written as

$$w_{n+1} + hAw_{n+1} = (w_n^{-p} - p\delta h)^{-1/p}, \tag{35}$$

where we have $Au_{n+1} = Aw_{n+1}$, since $A = -\Delta$ annihilates the constant α. For the initial step, we want to show that $v_1 = (\frac{T_2 - t_0}{T_2 - t_1})^{1/p} w_0$ satisfies $v_1 + hAv_1 \geq [w_0^{-p} - p\delta h]^{-1/p}$. We calculate

$$\begin{aligned}
v_1 + hAv_1 &= \left(\frac{T_2}{T_2 - h}\right)^{1/p}(w_0 + hAw_0) \geq \left(\frac{T_2}{T_2 - h}\right)^{1/p} w_0 && (\text{since } Aw_0 = Au_0 \geq 0) \\
&= \left(\frac{w_0^p}{1 - hC_0}\right)^{1/p} && (\text{since } T_2 = 1/C_0) \\
&= (w_0^{-p} - hC_0 w_0^{-p})^{-1/p} \geq (w_0^{-p} - p\delta h)^{-1/p} && (\text{since } C_0 w_0^{-p} \geq p\delta).
\end{aligned}$$

v_1 is therefore a supersolution, so by Theorem 1, there exists a solution w_1 of (35) that satisfies $\alpha \leq w_1 \leq (\frac{T_2}{T_2 - h})^{1/p} w_0$ and $w_1^{-p} \geq \frac{T_2 - h}{T_2} w_0^{-p} \geq (T_2 - h)p\delta$. This completes the base case.

For the induction step, suppose the solution w_n satisfies

$$\alpha \leq w_n \leq \left(\frac{T_2 - t_{n-1}}{T_2 - t_n}\right)^{1/p} w_{n-1} \qquad \text{and} \qquad w_n^{-p} \geq (T_2 - t_n)p\delta.$$

We now need to show that there exists a solution w_{n+1} such that

$$\alpha \leq w_{n+1} \leq \left(\frac{T_2 - t_n}{T_2 - t_{n+1}}\right)^{1/p} w_n \qquad \text{and} \qquad w_{n+1}^{-p} \geq (T_2 - t_{n+1})p\delta.$$

We start by showing that $v_{n+1} := (\frac{T_2 - t_n}{T_2 - t_{n+1}})^{1/p} w_n$ is a supersolution, i.e., we have

$$v_{n+1} + hAv_{n+1} - (w_n^{-p} - p\delta h)^{-1/p} \geq 0. \tag{36}$$

It is clear from the definition that $v_{n+1} \geq \alpha \geq 0$, since $T_2 - t_n > T_2 - t_{n+1}$. Substituting into (36) gives $(\frac{T_2 - t_n}{T_2 - t_{n+1}})^{1/p}(w_n + hAw_n) - (w_n^{-p} - p\delta h)^{-1/p} \geq 0$. Since w_n satisfies $w_n + hAw_n = (w_{n-1}^{-p} - p\delta h)^{-1/p}$, this criterion is equivalent to $(\frac{T_2 - t_n}{T_2 - t_{n+1}})^{1/p}(w_{n-1}^{-p} - p\delta h)^{-1/p} - (w_n^{-p} - p\delta h)^{-1/p} \geq 0$. In other words, we need to show that $(\frac{T_2 - t_n}{T_2 - t_{n+1}})^{1/p}(w_{n-1}^{-p} - p\delta h)^{-1/p} \geq (w_n^{-p} - p\delta h)^{-1/p}$, which is equivalent to showing that

$$w_{n-1}^{-p} - p\delta h \leq \left(\frac{T_2 - t_n}{T_2 - t_{n+1}}\right)(w_n^{-p} - p\delta h). \tag{37}$$

To prove the above inequality, we use the induction hypothesis: we have $w_{n-1}^{-p} \leq (\frac{T_2 - t_{n-1}}{T_2 - t_n}) w_n^{-p}$ which implies

$$\begin{aligned} w_{n-1}^{-p} - p\delta h &\leq \left(\frac{T_2 - t_{n-1}}{T_2 - t_n}\right) w_n^{-p} - p\delta h \\ &= \left(\frac{T_2 - t_n}{T_2 - t_{n+1}}\right)\left[\frac{(T_2 - t_{n-1})(T_2 - t_{n+1})}{(T_2 - t_n)^2} w_n^{-p} - \frac{(T_2 - t_{n+1})p\delta h}{T_2 - t_n}\right]. \end{aligned}$$

Therefore, the inequality (37) is true if

$$\frac{(T_2 - t_{n-1})(T_2 - t_{n+1})}{(T_2 - t_n)^2} w_n^{-p} - \frac{(T_2 - t_{n+1})p\delta h}{T_2 - t_n} \leq w_n^{-p} - p\delta h,$$

or

$$p\delta h \left(1 - \frac{T_2 - t_{n+1}}{T_2 - t_n}\right) \leq w_n^{-p}\left(1 - \frac{(T_2 - t_{n-1})(T_2 - t_{n+1})}{(T_2 - t_n)^2}\right),$$

which simplifies to $(T_2 - t_n)p\delta \leq w_n^{-p}$, which we know is true by the induction hypothesis. Thus, we have shown that v_{n+1} is a supersolution; by Theorem 1, a solution w_{n+1} exists and satisfies $\alpha \leq w_{n+1} \leq (\frac{T_2 - t_n}{T_2 - t_{n+1}})^{1/p} w_n$, so that $w_{n+1}^{-p} \geq (\frac{T_2 - t_{n+1}}{T_2 - t_n}) w_n^{-p} \geq (T_2 - t_{n+1})p\delta$, which completes the induction step.

3.3. Numerical Blow-up

In this section, we want to prove that for values of δ large enough, the (SpFE)* method will blow up before a certain time $T^* \leq \int_0^\infty \frac{ds}{\delta F(s) - \lambda s} < \infty$, with λ being the smallest positive eigenvalue of $-\Delta$. The existence of such a blow-up time in the continuous case has been shown by Kaplan in [32]. Since we already proved that $T^* > T_1 = \frac{1}{\delta}(g(\|u_0\|_\infty)$, proving this result leads to exactly the same bounds as Kaplan for the discrete case.

To do so, we first need to define what we mean by numerical blow-up time. Suppose we use a numerical method of fixed time step size h to integrate the model problem (1). We define the numerical blow-up time T_h^* to be the smallest multiple of h such that the numerical solution ceases to exist. To estimate T_h^*, we adapt the approach used by Kaplan for the continuous problem to our semi-discretization: we show that there exists a finite time T^* such that for all $K > 0$ and h small enough, there exists $n < T^*/h$ such that $\|u_n\|_\infty > K$, so that $T_h^* \leq T^*$ for all h small enough.[2] We now state our main result.

Theorem 7. *Suppose that δ satisfies*

$$\delta F(u) - \lambda u > 0, \qquad \forall\, u \geq 0, \tag{38}$$

where λ is the first eigenvalue of $-\Delta\varphi = \lambda\varphi$, $\varphi = 0$ on the boundary. We fix some large positive constant K and choose $\varepsilon \in (0, g(K))$. Then there exists $h^* > 0$ such that for all $h < \min(h^*, \frac{g(K)-\varepsilon}{\delta})$, the numerical scheme

$$u_{n+1} + hAu_{n+1} = G(g(u_n) - \delta h), \tag{39}$$

has a numerical blow-up time $T^* \leq \int_0^\infty \frac{ds}{\delta F(s) - \lambda s}$, in the sense that there exists $n^* < \frac{T^*}{h}$ such that $\|u_{n^*}\|_\infty > K$.

Note that the proof presented in this section is constructive so that one can compute an explicit bound h^*. We suppose thereafter that K and ε are fixed.

Remark 3. *The assumption $h < \frac{g(K)-\varepsilon}{\delta}$ implies that $K < G(\delta h + \varepsilon) < G(\delta h)$ so that as long as $\|u_n\|_\infty \leq K$, condition* (25) *is satisfied and scheme* (39) *has a unique positive solution.*

Remark 4. *Condition* (38) *imposed on δ is identical to the one given by Kaplan in [32]. It cannot be satisfied at $u = 0$ if $F(0) = 0$, however, if $F(0) > 0$, since F satisfies* (2)*, we have $\lim_{u\to 0} \frac{u}{F(u)} = 0$ and $\lim_{u\to\infty} \frac{u}{F(u)} = 0$, and condition* (38) *is satisfied for all δ large enough. For example, if we consider $F(u) = e^u$, condition* (38) *becomes $\delta > \frac{\lambda u}{e^u}$, for all $u \geq 0$, that is $\delta > \frac{\lambda}{e}$. If we consider $F(u) = (u+\alpha)^p$, with $\alpha > 0$, since the derivative of the function $\beta(u) := u/(u+\alpha)^p$ satisfies $\beta'(u) = \frac{(u+\alpha)^p - p(u+\alpha)^{p-1}u}{(u+\alpha)^{2p}} > 0 \iff u < \frac{\alpha}{p-1}$, we have $\beta(\frac{\alpha}{p-1}) = \frac{\alpha}{(p-1)(\alpha p)^p}$, and condition* (38) *becomes $\delta > \frac{\lambda\alpha}{(p-1)(\alpha p)^p}$.*

[2] While most of our previous results were following Le Roux's approach in [39], we could not use the same method as hers to prove this result. Indeed, a key element of Le Roux's approach is the use specific functionals and no equivalent functionals could be found for this scheme.

3.3.1. Outline of the Proof

We need to show that there exists $n^* < T^*/h$ such that $\|u_{n^*}\|_\infty > K$, where K is a fixed large constant. Following the eigenfunction methods, we introduce the sequence (a_n), defined by

$$a_n = \int_\Omega \varphi\, u_n dx, \tag{40}$$

where φ is the eigenfunction corresponding to the first eigenvalue λ of $-\Delta\varphi = \lambda\varphi$, $\varphi = 0$ on the boundary, with $\lambda > 0$, $\varphi \geq 0$ and $\int_\Omega \varphi\, dx = 1$ (we can assume $\varphi \geq 0$ since by Courant's theorem, the eigenfunction φ does not change sign in Ω). Our approach consists of finding n^* such that $a_{n^*} > K$. Indeed we have

$$a_n \leq \int_\Omega \varphi\, \|u_n\|_\infty \, dx = \|u_n\|_\infty \int_\Omega \varphi \, dx = \|u_n\|_\infty.$$

We divide our proof into the following steps: 1. We prove that (a_n) is increasing. 2. We define $a(t)$, solution of $a'(t) = \delta F(a(t)) - \lambda a(t),\ \ a(0) = a^* \in (0, a_0)$, which blows up in finite time at $T = \int_{a^*}^{\infty} \frac{ds}{\delta F(s) - \lambda s}$ if δ satisfies condition (38). Defining $D_n = a_n - a(nh)$, we need to bound D_n from below in order to prove that for h small enough, D_n is positive for all n for which a_n and $a(t_n)$ are well-defined.

3.3.2. Growth of the sequence (a_n)

To prove that (a_n) is increasing, we need the following lemma.

Lemma 1. *As long as u_n satisfies $\|u_n\|_\infty < G(\delta h)$, the sequence (a_n) defined in (40) satisfies $a_{n+1} \geq \frac{1}{1+h\lambda} G(g(a_n) - \delta h)$. The condition is satisfied in particular if $h < \frac{g(K)-\varepsilon}{\delta}$ and $\|u_n\|_\infty \leq K$.*

Proof. Since $\|u_n\|_\infty < G(\delta h)$, scheme (39) is well-defined. We multiply each side by φ and integrate over Ω to get $\int_\Omega \varphi\, u_{n+1} - h\varphi\Delta u_{n+1} dx = \int_\Omega \varphi\, G(g(u_n) - \delta h) dx$. Using the fact that u_n and φ vanish on the boundary, the left-hand side can be rewritten as $a_{n+1} - h\int_\Omega u_{n+1}\Delta\varphi\, dx = (1 + h\lambda)a_{n+1}$, and we obtain $a_{n+1} = \frac{1}{1+h\lambda} \int_\Omega \varphi\, G(g(u_n) - \delta h) dx$. We now prove that the function $f(x) := G(g(x) - \delta h)$ is convex for $x \geq 0$. We have

$$f'(x) = G'(g(x) - \delta h) g'(x) = -F(G(g(x) - \delta h)) \frac{-1}{F(x)} = \frac{1}{F(x)} F(G(g(x) - \delta h)),$$

since $G'(s) = -F(G(s))$ and $g'(s) = \frac{-1}{F(s)}$, and then

$$\begin{aligned} f''(x) &= \frac{1}{F(x)^2} \left[F'(G(g(x) - \delta h))G'(g(x) - \delta h)g'(x)F(x) - F'(x)F(G(g(x) - \delta h))\right] \\ &= \frac{1}{F(x)^2} \left[F'(G(g(x) - \delta h))F(G(g(x) - \delta h)) - F'(x)F(G(g(x) - \delta h))\right] \\ &= \frac{F(G(g(x) - \delta h))}{F(x)^2} \left(F'(G(g(x) - \delta h)) - F'(x)\right), \end{aligned}$$

which is positive since F being strictly convex implies that F' is increasing and we have $G(g(x)-\delta h) \geq x$. Hence f is convex and we apply Jensen's inequality to get

$$\int \varphi(x) f(u_n(x))\, dx \geq f\left(\int \varphi(x) u_n(x)\, dx\right) = f(a_n),$$

which completes the proof.

Lemma 2. *If δ satisfies condition* (38)*, the sequence* (a_n) *defined in* (40) *is increasing as long as u_n satisfies $\|u_n\|_\infty < G(\delta h)$ (this is satisfied in particular if $h < \frac{g(K)-\varepsilon}{\delta}$ and $\|u_n\|_\infty \leq K$).*

Proof. To prove this result, we show that for all $x \in (0, G(\delta h))$, we have

$$\frac{1}{1+h\lambda} G(g(x)-\delta h) > x, \tag{41}$$

that is $g(x) - g((1+h\lambda)x) < \delta h$. Since g is continuously differentiable, we can apply the Mean Value Theorem on the interval $(x,(1+h\lambda)x)$, so there exists $\xi \in (x,(1+h\lambda)x)$, such that $g(x)-g((1+h\lambda)x) = g'(\xi)(x-(1+h\lambda)x) = -g'(\xi)h\lambda x$, which becomes $g(x) - g((1+h\lambda)x) = \frac{1}{F(\xi)} h\lambda x$. So we need $\frac{1}{F(\xi)} h\lambda x < \delta h$, i.e. $F(\xi) > \frac{\lambda x}{\delta}$, $\quad \forall\, \xi \in (x,(1+h\lambda)x)$. Since F is increasing and δ satisfies condition (38), we have $F(\xi) > F(x) > \frac{\lambda x}{\delta}$. Hence inequality (41) holds for all $x \in (0, G(\delta h))$ and Lemma 1 completes the proof.

3.3.3. Definition of $a(t)$ and D_n

From now on, we assume that condition (38) is satisfied and $h < \frac{g(K)-\varepsilon}{\delta}$ and $\|u_n\|_\infty \leq K$. This implies that u_{n+1} is well-defined, thus so are a_{n+1} and D_{n+1} defined below.

Definition of $a(t)$. From Lemma 1, we have $\frac{a_{n+1}-a_n}{h} \geq \frac{1}{h}(\frac{1}{1+h\lambda} G(g(a_n)-\delta h) - a_n)$, hence we will compare (a_n) with $(a(t_n))$ where $t_n = nh$ and $a(t)$ is the solution of

$$a'(t) = \lim_{h\to 0} \frac{1}{h}\left(\frac{1}{1+h\lambda} G(g(a(t))-\delta h) - a(t)\right), \quad a(0) = a^*,$$

where a^* can be any fixed number in $[0, a_0)$. This limit simplifies to

$$\begin{aligned}
&\lim_{h\to 0} \frac{1}{h}\left(\frac{1}{1+h\lambda} G(g(a)-\delta h) - a\right) \\
&\quad = \lim_{h\to 0} \frac{1}{h}\left[\left(\frac{1}{1+h\lambda} - 1\right) G(g(a)-\delta h) + G(g(a)-\delta h) - G(g(a))\right] \\
&\quad = \lim_{h\to 0} \frac{1}{h}\left[\left(\frac{-h\lambda}{1+h\lambda}\right) G(g(a)-\delta h) - \delta\, \frac{G(g(a)-\delta h) - G(g(a))}{-\delta}\right] \\
&\quad = \lim_{h\to 0} \left[\left(\frac{-\lambda}{1+h\lambda}\right) G(g(a)-\delta h)\right] - \delta\, G'(g(a)) \\
&\quad = -\lambda G(g(a)) + \delta\, F(G(g(a))) = \delta F(a) - \lambda a.
\end{aligned}$$

So $a(t)$ is the solution of

$$a'(t) = \delta F(a(t)) - \lambda a(t), \quad a(0) = a^* < a_0.$$

By integrating this equation, we note that $a(t)$ is defined on $[0, T_{a^*})$, where $T_{a^*} = \int_{a^*}^{\infty} \frac{1}{\delta F(s) - \lambda s} \, ds < \infty$, so that $a(t)$ blows up at finite time T_{a^*}. Our goal is to show that a_n is larger than $a(t_n)$.

Definition of D_n. For all n such that a_n and $a(t_n)$ are well-defined, we define $D_n := a_n - a(t_n)$. To prove Theorem 7, we will prove by induction that there exists h^* such that $\forall\, h \le h^*$, $\forall n$ such that $\|u_n\|_\infty \le K$, we have $D_{n+1} > 0$. The initial condition a^* was chosen such that D_0 is positive, so assuming that D_n is positive, we prove that D_{n+1} is also positive. First, we need to verify that $a(t_{n+1})$ exists so that D_{n+1} is well-defined and $t_{n+1} < T_{a^*}$.

Lemma 3. *If $D_n > 0$, the function $a(t_n+\xi)$, with $\xi \in [0, h]$, is bounded above by $a(t_n+\xi) < G(\varepsilon)$, where ε is a fixed number belonging to $(0, g(K))$ (see Theorem 7).*

Proof. We introduce for $t \ge t_n$ the function $b(t)$, solution of

$$b'(t) = \delta F(b(t)) > \delta F(b(t)) - \lambda b(t), \quad b(t_n) = a(t_n).$$

This function can be written explicitly, $b(t) = G(g(a(t_n)) + \delta t_n - \delta t)$, and we have $a(t) \le b(t)$, $\forall t \ge t_n$. Moreover since δ satisfies condition (38), $a(t)$ is increasing and we have $a(t_n + \xi) \le a(t_n + h) \le b(t_{n+1}) = G(g(a(t_n)) - \delta h)$, and since $a(t_n) < a_n \le K$ and $h < \frac{g(K)-\varepsilon}{\delta}$, we get as required $a(t_n + \xi) \le G(g(a(t_n)) - \delta h) \le G(g(K) - \delta h) < G(\varepsilon)$.

Hence D_{n+1} is well-defined and we first bound it using Lemma 1,

$$D_{n+1} \ge \frac{1}{1+h\lambda} G(g(a_n) - \delta h) - a(t_n + h) =: \eta(h).$$

We then take a Taylor expansion of the right hand side function $\eta(h)$ around $h = 0$,

$$\begin{aligned}
\eta(0) &= a_n - a(t_n) = D_n, \\
\eta(h) &= \frac{-\lambda G(g(a_n) - \delta h)}{(1+h\lambda)^2} + \delta \frac{F(G(g(a_n) - \delta h))}{1+h\lambda} - a'(t_n + h), \\
\eta'(0) &= \delta F(a_n) - \lambda a_n - (\delta F(a(t)) - \lambda a(t)).
\end{aligned}$$

Thus, we have

$$D_{n+1} \ge D_n + h(\psi(a_n) - \psi(a(t_n))) + \frac{h^2}{2} \eta''(\xi) \tag{42}$$

for some $\xi \in (0, h)$, with $\psi(x) = \delta F(x) - \lambda x$.

Since $\eta(h)$ is twice continuously differentiable for all $0 \le h \le \frac{g(K)-\epsilon}{\delta}$, $\eta''(h)$ is continuous on the same interval, so there exists a (possibly negative) constant C_2 (which depends on δ, K and ϵ, but not on h) such that $\eta''(h) \ge C_2$ for all $0 < h < \frac{g(K)-\epsilon}{\delta}$. We are now able to prove Theorem 7.

3.3.4. Proof of Theorem 7

We suppose that $\|u_n\|_\infty \leq K$ and $D_n > 0$. We now show that $D_{n+1} > 0$. Indeed, since $a(t)$ blows up at time T_{a^*} with $T_{a^*} \leq T_0 = \int_0^\infty \frac{ds}{\delta F(s)-\lambda s}$, there exists $\tilde{n} < T_{a^*}/h$, such that $a(t_{\tilde{n}}) \leq K$ and either $t_{\tilde{n}+1} \geq T_{a^*}$ or $a(t_{\tilde{n}+1}) > K$. The first case implies that $\|u_n\|_\infty > K$ for some $n \leq \tilde{n}$, and in the second case, by the positivity of D_{n+1}, we have $\|u_{\tilde{n}+1}\|_\infty > a(t_{\tilde{n}+1}) > K$ with $t_{\tilde{n}+1} < T_{a^*}$. Hence there exists $n^* < T_0/h$ such that $\|u_{n^*}\|_\infty > K$.

We assume that $D_n > 0$ and we go back to (42) to write

$$\begin{aligned} D_{n+1} &\geq D_n + h[\psi(a_n) - \psi(a(t_n))] + \frac{h^2}{2}\eta(\xi) \\ &\geq D_n + h[\psi(a(t_n) + D_n) - \psi(a(t_n))] + \frac{h^2}{2}C_2 \\ &\geq D_n + hD_n\psi'(\zeta) + \frac{h^2}{2}C_2, \end{aligned}$$

with $\zeta \in (a(t_n), a(t_n) + D_n)$, by the Mean Value Theorem. The derivative $\psi'(x) = \delta F'(x) - \lambda$ is increasing and $\zeta > a(t_n) \geq a(0) = a^*$ so we get

$$D_{n+1} \geq D_n(1 + h\psi'(a^*)) + \frac{h^2}{2}C_2. \tag{43}$$

By induction, we obtain $D_{n+1} \geq (1+h\psi'(a^*))^{n+1}D_0 + \frac{h^2}{2}C_2\sum_{k=0}^{n}(1+h\psi'(a^*))^k$. We assume that $1 + h\psi'(a^*) > 0$, so if $\psi'(a^*) < 0$, we need h to be smaller than $1/(-\psi'(a^*))$, that is: if $F'(a^*) < \frac{\lambda}{\delta}$, then $h < \frac{1}{\lambda-\delta F'(a^*)}$. If C_2 is positive, the positivity of D_{n+1} follows from (43). We now study the case $C_2 < 0$. We obtain different bounds on h depending on the sign of $\psi'(a^*)$:

- if $\psi'(a^*) = 0$, we get $D_{n+1} \geq D_0 + (n+1)\frac{h^2}{2}C_2$, so that since $C_2 < 0$ and $t_{n+1} < T_{a^*}$, D_{n+1} is positive if $h < \frac{2D_0}{(-C_2)T_{a^*}}$.

- if $\psi'(a^*) > 0$, we get $D_{n+1} \geq (1+h\psi'(a^*))^{n+1}D_0 + \frac{h^2}{2}C_2\left(\frac{(1+h\psi'(a^*))^{n+1}-1}{h\psi'(a^*)}\right)$, so we need $\frac{h^2}{2}C_2 \geq -\underbrace{\frac{(1+h\psi'(a^*))^{n+1}}{(1+h\psi'(a^*))^{n+1}-1}}\, h\psi'(a^*)D_0$. The underbraced term is greater than 1 since $\psi'(a^*) > 0$, so we need $h < \frac{2\psi'(a^*)D_0}{(-C_2)}$.

- if $\psi'(a^*) < 0$ we also get $D_{n+1} \geq (1+h\psi'(a^*))^{n+1}D_0 + \frac{h^2}{2}C_2\left(\frac{(1+h\psi'(a^*))^{n+1}-1}{h\psi'(a^*)}\right)$, so we need $(1 + h\psi'(a^*))^{n+1}D_0 + \frac{h}{2}\frac{C_2}{\psi'(a^*)}[(1 + h\psi'(a^*))^{n+1} - 1] > 0$, which simplifies to $\frac{h}{(1+h\psi'(a^*))^{n+1}} < \frac{2D_0}{(-C_2)}(-\psi'(a^*)) + h$. Since $h > 0$, it is enough to satisfy $\frac{h}{(1+h\psi'(a^*))^{n+1}} \leq \frac{2D_0}{(-C_2)}(-\psi'(a^*))$. Since $t_{n+1} = (n+1)h < T_{a^*}$, i.e. $(n+1) < T_{a^*}/h$, and $(1+h\psi'(a^*)) \in (0,1)$, we have $\beta(h) := \frac{h}{(1+h\psi'(a^*))^{T_{a^*}/h}} >$

$\frac{h}{(1+h\psi'(a^*))^{n+1}}$. To prove that $\beta(h)$ is strictly increasing for $h > 0$, we consider

$$\begin{aligned}\beta'(h) &= \frac{1}{(1+h\psi'(a^*))^{T_{a^*}/h}} - \frac{h}{[(1+h\psi'(a^*))^{T_{a^*}/h}]^2}(1+h\psi'(a^*))^{T_{a^*}/h} \\ &\qquad \cdot \left[\frac{-T_{a^*}}{h^2}\ln(1+h\psi'(a^*)) + \frac{T_{a^*}}{h}\frac{\psi'(a^*)}{1+h\psi'(a^*)}\right] \\ &= \frac{1}{(1+h\psi'(a^*))^{T_{a^*}/h}}\left[1 + T_{a^*}\left(\frac{1}{h}\ln(1+h\psi'(a^*)) - \frac{\psi'(a^*)}{1+h\psi'(a^*)}\right)\right],\end{aligned}$$

which is clearly positive if $\ln(1+h\psi'(a^*)) > h\frac{\psi'(a^*)}{1+h\psi'(a^*)}$. Since $x - \ln x > 1$ for $x > 1$, and $(1+h\psi'(a^*))^{-1} \in (1,\infty)$, the above inequality is satisfied and $\beta(h)$ is strictly increasing. Moreover $\beta(0) = 0$ and $\lim_{h\to\frac{-1}{\psi'(a^*)}} \beta(h) = +\infty$, so that the equation $\frac{h}{(1+h\psi'(a^*))^{T_{a^*}/h}} = \frac{2D_0(-\psi'(a^*))}{(-C_2)}$ has exactly one solution $\tilde{h}$ and if $h < \tilde{h}$ we have $D_{n+1} > 0$.

The existence and uniqueness results of this section can be generalized to quasi-linear parabolic equations with power-like nonlinearities

$$\begin{aligned} u_t &= \alpha u^m + \Delta u^m, && \text{in } \Omega \times (0,T), \\ u &= 0, && \text{on } \partial\Omega \times (0,T), \\ u(x,0) &= u_0(x), && \text{in } \Omega, \end{aligned}$$

where Ω is a bounded domain in $\mathbb{R}^d$, $m > 1$ and $\alpha \geq 0$, see [5], but the upper bound blow up estimate remains currently open.

4. Numerical Results

We now test the new splitting B-methods on several non-linear partial differential equations, and also compare them to the B-methods based on the variation of constants approach from [6] called VCFE, VCBE, VCMR and VCTR.

4.1. A Semi-Linear Parabolic Equation

For the first example, we study the semi-linear parabolic equation $u_t = \Delta u + \delta e^u$ on the interval $\Omega = [-1,1]$. We discretize the Laplacian operator in space using a fourth order finite difference method with a five point stencil and a mesh size of $\Delta x = 2/30$. We set $\delta = 3$ and $u_0(x) = \cos(\pi x/2)$, which is concave on the whole interval. Using adaptive methods, we can estimate the blow-up time at $T_b \approx 0.1664$. We show in Table 1 the errors in the computed solutions up to $T_f = 0.1660$ with different step sizes. We observe that the error of B-methods is approximately 10 times smaller for first-order methods (and even more for SpBE and SpBEA) and 30 times smaller for second-order B-methods compared to standard methods. In Figure 1, we show these results graphically. As expected, the slopes of the lines corresponding to first-order methods are approximately one, whereas the slopes of the lines corresponding to second-order methods are close to two. In Figure 2 we show the behavior in time of the methods as blow-up is approached, using $h = 0.0001$ and computing the solution up to $T_f = 0.1663$.

Timestep	5e-005	2.5e-005	1.25e-005	8e-006	5e-006
FE	0.277	0.152	0.08	0.0522	0.0331
BE	0.468	0.194	0.0904	0.0565	0.0347
SpFE	0.0361	0.0183	0.00919	0.00589	0.00369
SpFEA	0.0379	0.0187	0.0093	0.00594	0.00371
SpBE	0.00533	0.00269	0.00135	0.000864	0.000541
SpBEA	0.00551	0.00273	0.00136	0.000869	0.000543
VCFE	0.019	0.00956	0.0048	0.00307	0.00192
Timestep	0.0002	0.000125	0.0001	5e-005	2.5e-005
VCBE	0.0195	0.0097	0.00483	0.00309	0.00193
MR	0.00833	0.00324	0.00207	0.000516	0.000129
TR	0.0407	0.0152	0.00961	0.00237	0.000591
SoSpFE	0.000305	0.000121	7.75e-005	1.94e-005	4.87e-006
SoSpBE	0.000305	0.000121	7.75e-005	1.94e-005	4.87e-006
VCMR	0.00033	0.00013	8.36e-005	2.1e-005	5.25e-006
VCTR	0.000733	0.000287	0.000184	4.6e-005	1.15e-005

Table 1: Error at $T_f = 0.1660$ for first-order methods (top) and second-order methods (bottom) applied to the semi-linear equation with $\delta F(u) = 3e^u$.

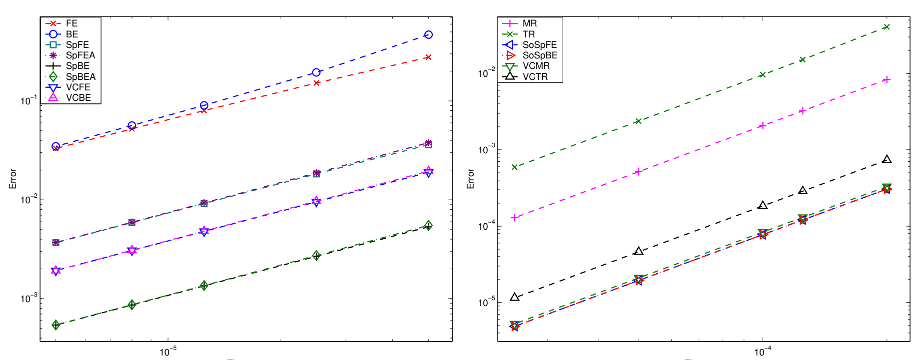

Figure 1: Error at $T_f = 0.1660$ for first-order methods (left) and second-order methods (right) applied to the semi-linear equation with $\delta F(u) = 3e^u$, with different values of h.

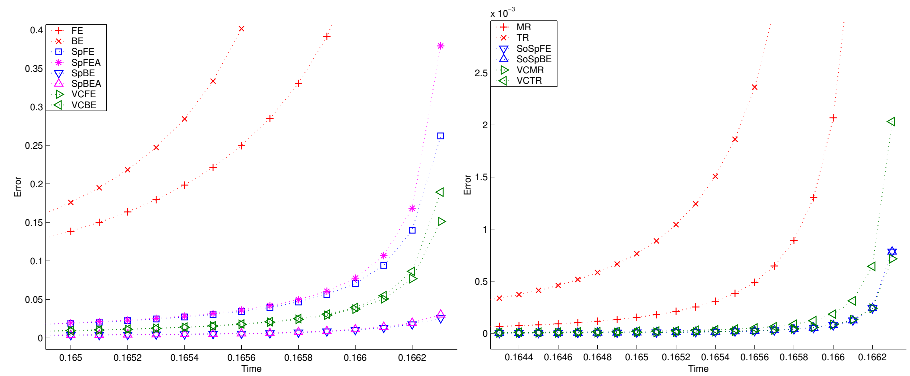

Figure 2: Error for first-order methods (left) and second-order methods (right) applied to the semi-linear equation with $\delta F(u) = 3e^u$, for time steps close to $T_f = 0.1663$.

4.2. A Quasi-Linear Parabolic Equation

Splitting B-methods can also be constructed for more general non-linear PDEs. We illustrate this now with the quasi-linear equation with power-type nonlinearities, $u_t = \Delta u^{\sigma+1} + \alpha u^{\beta+1}$, with $\beta > 0$, $\sigma > 0$ and $\alpha \geq 0$. We split the PDE right hand side into $f^{[1]}(u) = \alpha u^{\beta+1}$ and $f^{[2]}(u) = \Delta u^{\sigma+1}$, and using that the nonlinear part $y_t = \alpha y^{\beta+1}$ is solved by $y(t) = (\frac{1}{K-\alpha\beta t})^{1/\beta}$, the exact flow of the first part is $\varphi_t^{[1]}(u_n) = \left[u_n^{-\beta} - \alpha\beta t\right]^{-1/\beta}$. By choosing $\Phi_h^{[2]}$ to be the forward Euler method, so that $\Phi_h^{[2]*}$ is the backward Euler method we obtain the corresponding SpFE method

$$\Phi_h(u_n) = u_{n+1} = \left[(u_n + h\Delta u_n^{\sigma+1})^{-\beta} - \alpha\beta h\right]^{-1/\beta}, \tag{44}$$

and its adjoint, the (SpFE)*,

$$\Phi_h^*(u_n) = u_{n+1} = \left(u_n^{-\beta} - \alpha\beta h\right)^{-1/\beta} + h\Delta u_{n+1}^{\sigma+1}. \tag{45}$$

If we choose $\Phi_h^{[2]}$ to be backward Euler we get SpBE,

$$\Phi_h(u_n) = \left[v^{-\beta} - \alpha\beta h\right]^{-1/\beta}, \tag{46}$$

where v is solution of $v - h\Delta(v^{\sigma+1}) = u_n$, and its adjoint (SpBE)*

$$\Phi_h^*(u_n) = \left[u_n^{-\beta} - \alpha\beta h\right]^{-1/\beta} + h\Delta\left(\left[u_n^{-\beta} - \alpha\beta h\right]^{-(\sigma+1)/\beta}\right). \tag{47}$$

The second-order methods obtained by composing these methods are quite simple. If $\Phi_h^{[2]}$ is the forward Euler method, the composed method is SoSpFE

$$\Psi_h(u_n) = \left(\left(v + \frac{h}{2}\Delta(v^{\sigma+1})\right)^{-\beta} - \alpha\beta\frac{h}{2}\right)^{-1/\beta}, \tag{48}$$

Timestep	0.000125	8e-005	5e-005	2.5e-005	1.25e-005
FE	0.0188	0.0121	0.00762	0.00383	0.00192
BE	0.0196	0.0125	0.00774	0.00386	0.00192
SpFE	0.0082	0.00526	0.00329	0.00165	0.000824
SpFEA	0.00829	0.0053	0.00331	0.00165	0.000825
SpBE	0.004	0.00256	0.0016	0.0008	0.0004
SpBEA	0.004	0.00256	0.0016	0.0008	0.0004
VCFE	0.00209	0.00134	0.000837	0.000419	0.000209
VCBE	0.0021	0.00134	0.000839	0.000419	0.00021
Timestep	0.0005	0.00025	0.000125	8e-005	5e-005
MR	0.000191	4.78e-005	1.19e-005	4.89e-006	1.91e-006
TR	0.000499	0.000125	3.11e-005	1.28e-005	4.98e-006
SoSpFE	3.72e-005	9.29e-006	2.32e-006	9.52e-007	3.72e-007
SoSpBE	5.84e-005	1.46e-005	3.65e-006	1.49e-006	5.84e-007
VCMR	2.15e-006	5.37e-007	1.34e-007	5.5e-008	2.15e-008
VCTR	2.17e-005	5.42e-006	1.35e-006	5.55e-007	2.17e-007

Table 2: Error at $T_f = 0.1000$ for first-order methods (top) and second-order methods (bottom) applied to the quasi-linear equation $u_t = \Delta u^2 + 8u^3$.

where v is the solution of $v - \frac{h}{2}\Delta(v^{\sigma+1}) = (u_n^{-\beta} - \alpha\beta\frac{h}{2})^{-1/\beta}$. Similarly, the second-order method obtained using the backward Euler method for $\Phi_h^{[2]}$ is SoSpBE, given implicitly by

$$\Psi_h(u_n) = u_{n+1} = \left(\left(v + \frac{h}{2}\Delta(u_{n+1}^{\sigma+1})\right)^{-\beta} - \frac{\alpha\beta h}{2}\right)^{-1/\beta}, \tag{49}$$

where $v = [u_n^{-\beta} - \frac{\alpha\beta h}{2}]^{-1/\beta} + \frac{h}{2}\Delta[(u_n^{-\beta} - \frac{\alpha\beta h}{2})^{-(\sigma+1)/\beta}]$. We show a numerical example for the quasi-linear equation $u_t = \Delta u^2 + 8u^3$, on $\Omega = (-1, 1)$ with the same initial condition as above: $u_0(x) = \cos(\pi x/2)$. The blow-up time is approximately $T_b \approx 0.1128$. We list in Table 2 the errors we obtained. We observe that the B-methods obtained by variation of the constant are more accurate than those obtained by splitting methods. Compared with standard methods, the errors are 10 times smaller for first-order methods of the first type and between 2 and 7 times smaller for first-order methods of the second type. Among second-order methods, the method obtained by variation of the constant and the midpoint rule (VCMR) is remarkably better than the others, as its error is more than fifty times smaller that the error of the standard midpoint rule. In Figure 3 we show the corresponding data graphically. The step-by-step errors are plotted in Figure 4 up to $T_f = 0.1110$, when the solutions are computed using the timestep $h = 0.0001$.

4.3. A Semi-Linear System

In [18] and [19] , Friedman and Giga considered parabolic systems of the form $u_t - u_{xx} = f(v)$, $v_t - v_{xx} = g(u)$, where f and g are positive, increasing

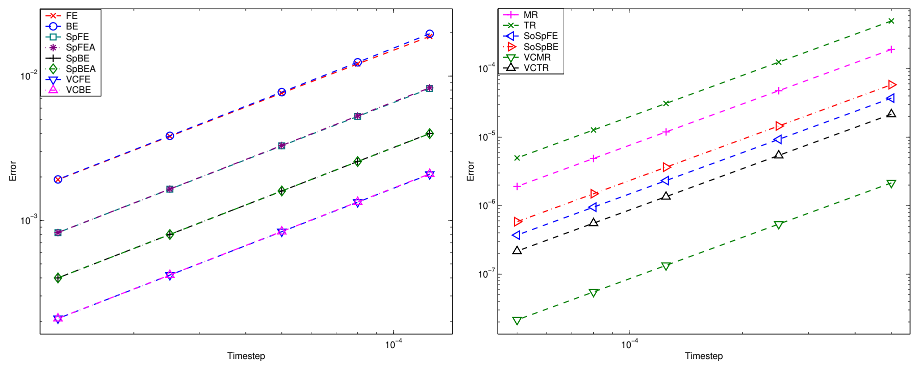

Figure 3: Error at $T_f = 0.1000$ for first-order methods (left) and second-order methods (right) applied to the quasi-linear equation $u_t = \Delta u^2 + 8u^3$, with different values of h.

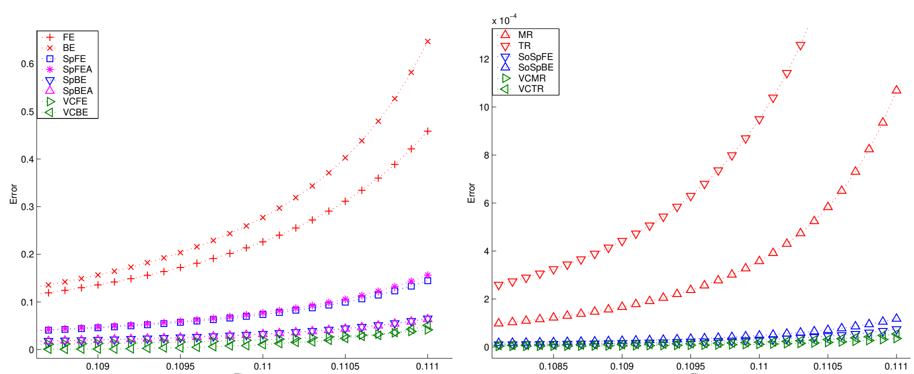

Figure 4: Error for first-order methods (left) and second-order methods (right) applied to the quasi-linear equation $u_t = \Delta u^2 + 8u^3$, for timesteps close to $T_f = 0.1110$.

and superlinear. They showed that the solutions exhibit a single-point blow-up. More complex systems of the form $(u_i)_t = \Delta u_i + f_i(u_1, \ldots, u_m)$ were studied by Bebernes and Lacey [4], Gang and Sleeman [22] and Chen [12]. In this subsection, we derive several specialized methods for the simple case

$$u_t = \Delta u + \delta e^v, \quad v_t = \Delta v + \gamma e^u. \tag{50}$$

We first solve the nonlinear system of ordinary differential equations $y'(t) = \delta e^{z(t)}$, $z'(t) = \gamma e^{y(t)}$, to get $y(t) = \ln K - \ln[1 - \delta e^{Kt+D}] - \ln \gamma$, $z(t) = \ln K - \ln[1-\delta e^{Kt+D}] + Kt + D$, where K and D are constants of integration, determined by the initial conditions, $K = \gamma e^{y(0)} - \delta e^{z(0)}$ and $D = z(0) - y(0) - \ln \gamma$. Then for each choice of numerical integrator $\Phi_h^{[2]}$ applied to $u_t = \Delta u$, $v_t = \Delta v$, we obtain two schemes that are adjoint to each other. The forward Euler method leads to the explicit SpFE scheme

$$\Phi_h(u_n, v_n) = \begin{pmatrix} \ln K_n - \ln[1 - \delta e^{D_n + hK_n}] - \ln \gamma \\ \ln K_n - \ln[1 - \delta e^{D_n + hK_n}] + D_n + hK_n \end{pmatrix},$$

where $K_n = \gamma e^{u_n + h\Delta u_n} - \delta e^{v_n + h\Delta v_n}$, $D_n = v_n + h\Delta v_n - u_n - h\Delta u_n - \ln \gamma$, and the adjoint scheme Φ_h^* we call (SpFE)* is given by

$$\begin{aligned} u_{n+1} &= \ln K_n - \ln[1 - \delta e^{D_n + hK_n}] - \ln \gamma + h\Delta u_{n+1}, \\ v_{n+1} &= \ln K_n - \ln[1 - \delta e^{D_n + hK_n}] + D_n + hK_n + h\Delta v_{n+1}, \end{aligned}$$

where $K_n = \gamma e^{u_n} - \delta e^{v_n}$, $D_n = v_n - u_n - \ln \gamma$. If we choose instead the backward Euler method, we obtain the SpBE scheme

$$\Phi_h(u_n, v_n) = \begin{pmatrix} \ln K_n - \ln[1 - \delta e^{D_n + hK_n}] - \ln \gamma \\ \ln K_n - \ln[1 - \delta e^{D_n + hK_n}] + D_n + hK_n \end{pmatrix},$$

where $K_n = \gamma e^{w_1} - \delta e^{w_2}$, $D_n = w_2 - w_1 - \ln \gamma$, and w_1 and w_2 are solutions of $w_1 = u_n + h\Delta w_1$ and $w_2 = v_n + h\Delta w_2$. For its adjoint method (SpBE)*, we first define $K_n := \gamma e^{u_n} - \delta e^{v_n}$, $D_n := v_n - u_n - \ln \gamma$, and $w_1 := \ln K_n - \ln[1 - \delta e^{D_n + hK_n}] - \ln \gamma$, $w_2 := \ln K_n - \ln[1 - \delta e^{D_n + hK_n}] + D_n + hK_n$. Then the (SpBE)* scheme can be written as

$$\Phi_h^*(u_n, v_n) = \begin{pmatrix} w_1 + h\Delta w_1 \\ w_2 + h\Delta w_2 \end{pmatrix}.$$

We can also compose these methods to construct second-order splitting B-methods. For these, we first define $K_n := \gamma e^{u_n} - \delta e^{v_n}$, $D_n := v_n - u_n - \ln \gamma$ as before. Then, if we choose $\Phi_h^{[2]}$ to be the forward Euler method, we define w_1 and w_2 to be the solutions of

$$\begin{aligned} w_1 - \tfrac{h}{2}\Delta w_1 &= \ln K_n - \ln[1 - \delta e^{D_n + \frac{h}{2}K_n}] - \ln \gamma, \\ w_2 - \tfrac{h}{2}\Delta w_2 &= \ln K_n - \ln[1 - \delta e^{D_n + \frac{h}{2}K_n}] + D_n + \tfrac{h}{2}K_n, \end{aligned}$$

and we define $\tilde{K} := \gamma \exp(w_1 + \frac{h}{2}\Delta w_1) - \delta \exp(w_2 + \frac{h}{2}\Delta w_2)$, $\tilde{D} := w_2 + \frac{h}{2}\Delta w_2 - w_1 - \frac{h}{2}\Delta w_1 - \ln \gamma$, to finally get the SoSpFE scheme

$$\begin{aligned} u_{n+1} &= \ln \tilde{K} - \ln[1 - \delta e^{\tilde{D} + \frac{h}{2}\tilde{K}}] - \ln \gamma, \\ v_{n+1} &= \ln \tilde{K} - \ln[1 - \delta e^{\tilde{D} + \frac{h}{2}\tilde{K}}] + \tilde{D} + \tfrac{h}{2}\tilde{K}. \end{aligned} \tag{51}$$

Timestep	0.0001	5e-005	2.5e-005	1.25e-005	8e-006
FE	0.0146	0.00736	0.00369	0.00185	0.00118
BE	0.015	0.00747	0.00372	0.00186	0.00119
SpFE	0.00146	0.00073	0.000365	0.000183	0.000117
SpFEA	0.000679	0.000339	0.000169	8.47e-005	5.42e-005
SpBE	0.000675	0.000338	0.000169	8.46e-005	5.42e-005
SpBEA	0.00146	0.000731	0.000365	0.000183	0.000117
VCFE	0.00118	0.00059	0.000295	0.000148	9.45e-005
VCBE	0.00118	0.000591	0.000295	0.000148	9.45e-005

Timestep	0.0004	0.0002	0.0001	5e-005	2.5e-005
MR	5.91e-005	1.48e-005	3.69e-006	9.23e-007	2.31e-007
TR	0.000339	8.48e-005	2.12e-005	5.3e-006	1.32e-006
SoSpFE	4.85e-006	1.21e-006	3.03e-007	7.57e-008	1.89e-008
SoSpBE	4.85e-006	1.21e-006	3.03e-007	7.57e-008	1.89e-008
VCMR	5.82e-006	1.46e-006	3.64e-007	9.1e-008	2.28e-008
VCTR	6.4e-006	1.6e-006	4e-007	1e-007	2.5e-008

Table 3: Error at $T_f = 0.1100$ for first-order methods (top) and second-order methods (bottom) applied to the system of semi-linear equations.

If we choose to use the backward Euler method as $\Phi_h^{[2]}$, we need to first define $\tilde{u} := \ln K_n - \ln[1-\delta e^{D_n+\frac{h}{2}K_n}] - \ln\gamma$, $\tilde{v} := \ln K_n - \ln[1-\delta e^{D_n+\frac{h}{2}K_n}] + D_n + \frac{h}{2}K_n$, and then w_1 and w_2 are the solutions of

$$w_1 - \frac{h}{2}\Delta w_1 = \tilde{u} + \frac{h}{2}\Delta\tilde{u}, \quad w_2 - \frac{h}{2}\Delta w_2 = \tilde{v} + \frac{h}{2}\Delta\tilde{v},$$

and we define $\tilde{K} := \gamma\exp(w_1) - \delta\exp(w_2)$, $\tilde{D} := w_2 - w_1 - \ln\gamma$, to finally get u_{n+1} and v_{n+1} by (51) for the SoSpBE scheme.

We now present the results of numerical experiments for the system of semi-linear parabolic equations (50) with $\delta = 3$ and $\gamma = 5$. The initial conditions are $u_0(x) = \cos(\pi x/2)$ and $v_0(x) = \cos(\pi x/2)$ on $\Omega = [-1, 1]$. The blow-up time is approximately $T_b \approx 0.1181$. The errors are listed in Table 3. and shown graphically in Figure 5.

In Figure 6, we show again the evolution of the solution as we approach blow up. We used $h = 0.0001$ and computed the solutions up to $T_f = 0.1170$. Further examples can be found in [5, Appendix A].

5. Conclusions

We presented in this paper a systematic approach for deriving numerical integrators which are very accurate for semi- and quasi-linear parabolic and hyperbolic partial differential equations exhibiting blow-up in finite time. We call this new class of geometric integration methods B-methods, where B stands for blow-up. Our construction is completely general, and can lead to B-methods for many other nonlinear partial differential equations that were not considered

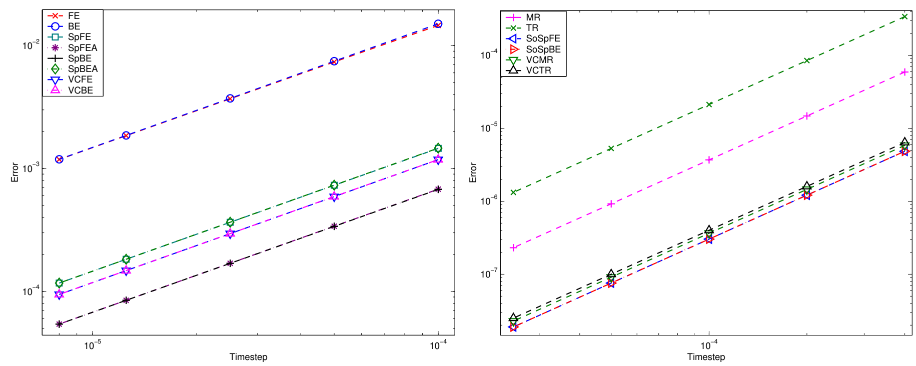

Figure 5: Error at $T_f = 0.1100$ for first-order methods (left) and second-order methods (right) applied to the system of semi-linear equations with different values of h.

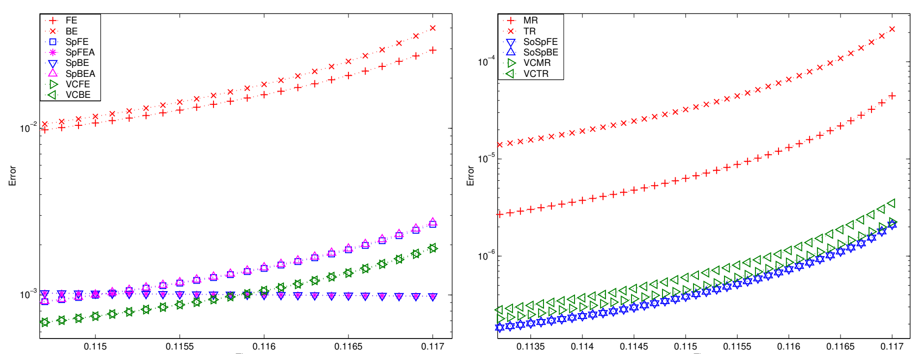

Figure 6: Error for first-order methods (left) and second order methods (right) applied to the system of semi-linear equations, for timesteps close to $T_f = 0.1170$.

in this paper. Because of their construction, which takes the blow-up behavior into account, all these methods will behave substantially better close to blow-up, while their behavior before blow-up is similar to classical time stepping schemes.

References

[1] H. Amann. On the existence of positive solutions of nonlinear elliptic boundary value problems. *Indiana Univ. Math. J.*, 21:125–146, 1971/72.

[2] R. O. Ayeni. On the blow up problem for semilinear heat equations. *SIAM J. Math. Anal.*, 14(1):138–141, 1983.

[3] C. Bandle and H. Brunner. Blowup in diffusion equations: a survey. *J. Comput. Appl. Math.*, 97(1-2):3–22, 1998.

[4] J. Bebernes and A. Lacey. Finite-time blowup for a particular parabolic system. *SIAM Journal on Mathematical Analysis*, 21(6):1415–1425, 1990.

[5] M. Beck. *B-methods: Special Time-integrators for Differential Equations with Blow-up Solutions.* PhD thesis, McGill University, 2009.

[6] M. Beck, M. J. Gander, and F. Kwok. B-methods for the numerical solution of evolution problems with blow-up solutions part I: Variation of the constant. *SIAM Journal on Scientific Computing*, 37(6):A2998–A3029, 2015.

[7] N. Bou-Rabee and J. M. Sanz-Serna. Geometric integrators and the hamiltonian monte carlo method. *Acta Numerica*, 27:113–206, 2018.

[8] H. Brezis and J. L. Vázquez. Blow-up solutions of some nonlinear elliptic problems. *Rev. Mat. Univ. Complut. Madrid*, 10(2):443–469, 1997.

[9] C. J. Budd, B. Dold, and A. Stuart. Blowup in a partial differential equation with conserved first integral. *SIAM Journal on Applied Mathematics*, 53(3):718–742, 1993.

[10] C. J. Budd, W. Huang, and R. D. Russell. Moving mesh methods for problems with blow-up. *SIAM Journal on Scientific Computing*, 17(2):305–327, 1996.

[11] C. J. Budd, J. Williams, and V. A. Galaktionov. Self-similar blow-up in higher-order semilinear parabolic equations. *SIAM Journal on Applied Mathematics*, 64(5):1775–1809, 2004.

[12] S. Chen. A sufficient condition for blowup solutions of nonlinear heat equations. *Journal of mathematical analysis and applications*, 293(1):227–236, 2004.

[13] S. Chen, H. Li, J. Li, M. Mei, and K. Zhang. Global and blow-up solutions for compressible euler equations with time-dependent damping. *Journal of Differential Equations*, 268(9):5035–5077, 2020.

[14] G. K. Duong, N. Nouaili, and H. Zaag. Refined asymptotic for the blow-up solution of the complex Ginzburg-Landau equation in the subcritical case. In *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, volume 39(1), pages 41–85, 2021.

[15] E. Faou. *Geometric numerical integration and Schrödinger equations*, volume 15. European Mathematical Society, 2012.

[16] M. S. Floater. Blow-up at the boundary for degenerate semilinear parabolic equations. *Arch. Rational Mech. Anal.*, 114(1):57–77, 1991.

[17] D. Frank-Kamenetskii. *Diffusion and heat exchange in chemical kinetics.* Princeton, N.J. : Princeton University Press, 1955.

[18] A. Friedman. Blow-up of solutions of nonlinear parabolic equations. In *Nonlinear diffusion equations and their equilibrium states, I (Berkeley, CA, 1986)*, volume 12 of *Math. Sci. Res. Inst. Publ.*, pages 301–318. Springer, New York, 1988.

[19] A. Friedman and Y. Giga. A single point blow-up for solutions of semilinear parabolic systems. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.*, 34(1):65–79, 1987.

[20] V. A. Galaktionov. A boundary value problem for the nonlinear parabolic equation $u_t = \Delta u^{\sigma+1} + u^\beta$ {English translation: Differential Equations 17 (1981), no. 5, 551-556.}. *Differentsial′nye Uravneniya*, 17(5):836–842, 956, 1981.

[21] V. A. Galaktionov and J. L. Vázquez. The problem of blow-up in nonlinear parabolic equations. *Discrete Contin. Dyn. Syst.*, 8(2):399–433, 2002.

[22] L. Gang and B. Sleeman. *Non Existence of Global Solutions to Systems of Semi Linear Parabolic Equations*. University of Dundee. Department of Mathematical Sciences, 1990.

[23] I. M. Gel'fand. Some problems in the theory of quasilinear equations. *Amer. Math. Soc. Transl. (2)*, 29:295–381, 1963.

[24] R. T. Glassey. Blow-up theorems for nonlinear wave equations. *Math. Z.*, 132:183–203, 1973.

[25] R. T. Glassey. On the blowing up of solutions to the Cauchy problem for nonlinear Schrödinger equations. *J. Math. Phys.*, 18(9):1794–1797, 1977.

[26] M. E. Gurtin and R. C. MacCamy. On the diffusion of biological populations. *Mathematical biosciences*, 33(1):35–49, 1977.

[27] E. Hairer, M. Hochbruck, A. Iserles, and C. Lubich. Geometric numerical integration. *Oberwolfach Reports*, 3(1):805–882, 2006.

[28] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration*. Springer-Verlag. Berlin, 2002.

[29] M. A. Hamza and H. Zaag. The blow-up rate for a non-scaling invariant semilinear wave equations. *Journal of Mathematical Analysis and Applications*, 483(2):123652, 2020.

[30] C. Hirota and K. Ozawa. Numerical method of estimating the blow-up time and rate of the solution of ordinary differential equations—an application to the blow-up problems of partial differential equations. *Journal of computational and applied mathematics*, 193(2):614–637, 2006.

[31] A. Kapila. Reactive-diffusive system with Arrhenius kinetics: dynamics of ignition. *SIAM Journal on Applied Mathematics*, 39(1):21–36, 1980.

[32] S. Kaplan. On the growth of solutions of quasi-linear parabolic equations. *Comm. Pure Appl. Math.*, 16:305–330, 1963.

[33] D. Kassoy and J. Poland. The thermal explosion confined by a constant temperature boundary: I-the induction period solution. *SIAM Journal on Applied Mathematics*, 39(3):412–430, 1980.

[34] H. B. Keller. Elliptic boundary value problems suggested by nonlinear diffusion processes. *Arch. Rational Mech. Anal.*, 35:363–381, 1969.

[35] J. B. Keller. On solutions of nonlinear wave equations. *Comm. Pure Appl. Math.*, 10:523–530, 1957.

[36] P. Kelley. Self-focusing of optical beams. *Physical Review Letters*, 15(26):1005–1008, 1965.

[37] C. E. Kenig, T. Duyckaerts, and F. Merle. Universality of blow-up profile for small radial type II blow-up solutions of the energy-critical wave equation. *Journal of the European Mathematical Society*, 13(3):533–599, 2011.

[38] A. A. Lacey. The form of blow-up for nonlinear parabolic equations. *Proc. Roy. Soc. Edinburgh Sect. A*, 98(1-2):183–202, 1984.

[39] M.-N. Le Roux. Semidiscretization in time of nonlinear parabolic equations with blowup of the solution. *SIAM Journal on Numerical Analysis*, 31(1):170–195, 1994.

[40] B. Leimkuhler and S. Reich. *Simulating Hamiltonian dynamics*. Number 14. Cambridge university press, 2004.

[41] B. LeMesurier, G. Papanicolaou, C. Sulem, and P.-L. Sulem. The focusing singularity of the nonlinear Schrödinger equation. In *Directions in partial differential equations (Madison, WI, 1985)*, volume 54 of *Publ. Math. Res. Center Univ. Wisconsin*, pages 159–201. Academic Press, Boston, MA, 1987.

[42] D. McLaughlin, G. Papanicolaou, C. Sulem, and P. Sulem. Focusing singularity of the cubic Schrödinger equation. *Physical Review A*, 34(2):1200, 1986.

[43] F. Merle and H. Zaag. Determination of the blow-up rate for a critical semilinear wave equation. *Mathematische Annalen*, 331(2):395–416, 2005.

[44] F. Merle and H. Zaag. Existence and classification of characteristic points at blow-up for a semilinear wave equation in one space dimension. *American Journal of Mathematics*, 134(3):581–648, 2012.

[45] N. Saito and T. Sasaki. Blow-up of finite-difference solutions to nonlinear wave equations. *J. Math. Sci. Univ. Tokyo*, 23(1):349–380, 2016.

[46] P. Souplet. Finite time blow-up for a non-linear parabolic equation with a gradient term and applications. *Math. Methods Appl. Sci.*, 19(16):1317–1333, 1996.

[47] Y. Su and D. Zhang. Construction of minimal mass blow-up solutions to rough nonlinear schrödinger equations. *Journal of Functional Analysis*, 284(5):109796, 2023.

[48] A. Takayasu, K. Matsue, T. Sasaki, K. Tanaka, M. Mizuguchi, and S. Oishi. Numerical validation of blow-up solutions of ordinary differential equations. *Journal of computational and applied mathematics*, 314:10–29, 2017.

[49] M. Tsutsumi. Existence and nonexistence of global solutions for nonlinear parabolic equations. *Publ. Res. Inst. Math. Sci.*, 8:211–229, 1972.

[50] M. Tsutsumi. Existence and nonexistence of global solutions of the first boundary value problem for a certain quasilinear parabolic equation. *Funkcial. Ekvac.*, 17:13–24, 1974.

[51] J. L. Vázquez. A strong maximum principle for some quasilinear elliptic equations. *Appl. Math. Optim.*, 12(3):191–202, 1984.