B-Methods: Special Time-Integrators for Differential Equations with Blow-up Solutions

Mélanie Beck

Department of Mathematics and Statistics,
McGill University, Montréal
Québec, Canada

August, 2009

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements of the degree of
Doctorate of Philosophy

Abstract

Many nonlinear differential equations have solutions that cease to exist in finite time because their norm becomes infinite. We say that the solution blows up in finite time. In general, this phenomenon is especially important in the physical interpretation of the results, but unfortunately most of these differential equations can not be explicitly solved. Moreover numerically approximating blow-up phenomena is a delicate problem and most standard methods only yield poor results.

In this thesis we suggest ways to construct fixed-step numerical methods, specialized in the approximation of a blow-up solution, the so-called B-methods (in case of partial differential equations, we obtain semi-discretizations in time). Two approaches are presented in detail: one consists of a splitting method while the other comes from a variation of the constant. Both approaches are based on the same idea: to exploit the fact that the solution of a simplified equation (made up of the nonlinear part that is responsible for the blow-up) can be explicitly written.

We start by properly defining the problem and presenting an extensive literature review concerning both theoretical and numerical results. Then, after explaining the two methods of construction on an example, we apply them to different models and so we obtain numerous B-methods. All these methods are implemented and extensive numerical experiments illustrate the superiority of the performance of B-methods over standard methods. Finally a chapter is devoted to the theoretical study of some B-methods. Theorems which are proven reinforce the promising results of the numerical

tests.

Résumé

De nombreuses équations différentielles non-linéaires ont des solutions qui cessent d'exister en temps fini car leur norme devient infinie. On dit alors que la solution explose en temps fini. Ce phénomène revêt généralement une grande importance dans l'interprétation physique des résultats, malheureusement la plupart de ces équations différentielles ne peuvent pas être résolues explicitement. De plus l'approximation numérique du phénomène d'explosion est délicat et la plupart des méthodes standards ne donnent que des résultats médiocres.

Dans cette thèse nous proposons des façons de construire des méthodes numériques à pas de temps fixe, spécialisées dans l'approximation d'une solution qui explose, les B-méthodes (dans le cas d'équations aux dérivés partielles, nous obtenons des semi-discrétisations en temps). Deux approches sont présentées en détail : l'une consiste en une *splitting method* tandis que l'autre provient d'une variation de la constante. Toutes deux se basent sur la même idée : exploiter le fait que la solution d'une équation simplifiée (formée de la partie non-linéaire responsable de l'explosion) peut être écrite explicitement.

Nous commençons par bien définir le problème et présentons une revue étendue de la littérature consacrée au sujet, tant du point de vue théorique que du point de vue numérique. Puis, après avoir expliqué ces deux méthodes de construction sur un exemple, nous les appliquons à différents modèles et obtenons ainsi de nombreuses B-méthodes. Toutes ces méthodes sont ensuite programmées et des tests numériques

étendus viennent illustrer la supériorité des performances des B-méthodes sur celles des méthodes standards. Un chapitre est également consacré à l'étude théorique de quelques B-méthodes. Les théorèmes qui y sont prouvés viennent supporter les résultats prometteurs des tests numériques.

Acknowledgments

I would like to sincerely thank my supervisor Prof. Martin Gander for his ideas, his enthusiasm and his availability. He represents everything one could expect from a supervisor.

I also thank my co-supervisor Prof. Tony Humphries who accepted to co-supervise me when Martin left for Switzerland.

Merci à Olivier Dubois et à Felix Kwok d'avoir lu attentivement ma thèse et de m'avoir suggéré de nombreuses améliorations.

This work was supported by graduate scholarships from the Natural Sciences and Engineering Research Council of Canada (NSERC), McGill University and the Université de Genève where I was provided funding in the form of a teaching assistantship.

Finalement, merci à tous ceux qui m'ont soutenue et aidée dans cette longue entreprise, en particulier Charles qui a supporté ma charge de travail en plus de la sienne pendant des années, et Marion qui a eu la merveilleuse idée d'être un bébé si facile.

Table of Contents

Introduction

Many physical processes studied in applied sciences can be modelled using differential equations. For some of these applications, the process can be reproduced by means of linear (differential) equations. These equations have been deeply studied and their theory is well-understood. However many physical phenomena require the use of nonlinear models. Indeed some properties of nonlinear equations that are essential to properly reproduce real-world processes are absent from the theory of linear equations. Generally these new properties also represent new difficulties for the mathematical analysis. Moreover, most nonlinear equations can not be solved explicitly. Thus it is necessary to resort to approximations that can be obtained using numerical methods. These methods are well-developed. However some nonlinear equations have solutions that exhibit specific behaviors that may be hardly compatible with techniques of numerical methods. Hence the numerical approximations of these equations are less accurate than expected.

In this thesis, we are interested in finite-time singularities, a property that is specific to nonlinear equations. Indeed, even if the solution of a linear equation may develop a singularity in finite time, it is necessarily caused by some underlying singularity in the data of the problem (in the initial or boundary conditions) [59]. On the other hand, in the case of nonlinear equations, the singularity may arise from the nonlinearity itself, so that even with smooth initial and boundary conditions, the solution may develop a singularity in finite time.

The simplest form of singularity in nonlinear equations is known as blow-up: when a solution of an evolution equation grows without bound as time approaches some finite value T, we say that a blow-up occurs at time T. Essentially, the solution becomes infinite at one or more points of the domain so that we have

$$\lim_{t\to T} \sup_{x\in\Omega} |u(x,t)| = \infty,$$

where T is called the blow-up time. As we will explain in Chapter 1, it is generally very difficult to numerically simulate blow-up phenomena accurately, in particular when using a fixed timestep. The goal of our work is thus to provide appropriate numerical methods. We explain two different ways to construct semi-discretizations in time with fixed stepsize, designed to solve a specific problem whose solution is blowing up in finite time.

The first chapter gives a more detailed introduction to the subject of this thesis. We first show the importance of blow-up phenomena by presenting several of its domains of application. Then we present an overview of the most commonly studied models: the semilinear parabolic equation $u_t = \Delta u + f(u)$, the quasilinear equation $u_t = \Delta u^{\sigma+1} + \alpha u^{\beta+1}$, the wave equation $u_{tt} = \Delta u + f(u)$ and the Schrödinger equation $iu_t = \Delta u + F(|u|^2)u$. For each of these problems, we give a few criteria ensuring the occurrence of a finite-time blow-up. We then turn to the approximation of blow-up phenomena using numerical solutions. We explain the difficulties arising with the transition and in particular we address the problem of the definition of numerical blow-up and numerical blow-up time. Finally, we briefly explain the idea that underlies the constructions of B-methods that are presented in this work.

The second chapter consists of an historical review of the subject. In the first part, we present a chronological description of the development of the theory of equations with blow-up solutions. Articles suggesting proofs of blow-up solutions started to appear in the sixties. The first studies concentrated on conditions ensuring the

occurrence of blow-up and estimations on the blow-up time. Afterwards, other questions arose such as where and how the blow-up occurs. The number of papers devoted to the subject exploded in the eighties. Our presentation concentrates mainly on the earlier period, that is from the sixties to the eighties. In the second part, we turn ourselves to the different attempts to reproduce blow-up solutions numerically. It started mostly in the eighties on simple problems of the form $u_t = \Delta u + f(u)$ and the subject quickly took off in the nineties.

In the third chapter we present how to construct B-methods, which are numerical methods that are designed for a specific problem. Two different approaches are used and the constructions are explained on the semilinear parabolic equation $u_t = \Delta u + \delta F(u)$. The first approach consists of a splitting method whereas the second approach is based on the variation of a constant. Several schemes of each type are then derived for different problems, chosen among the most commonly studied. Most of these schemes are then put into practice in order to illustrate their performances; extensive numerical experiments complete the chapter.

Even though it is not possible to study every B-method, we present in the fourth chapter a theoretical study for a few schemes. We selected a method based on the backward Euler method for each type of construction, and applied these two methods on a semilinear parabolic problem and a quasilinear parabolic problem with power-type nonlinearities. The two methods were chosen because of their simplicity and their stability and the problems belong to the most-studied models. For each case, we prove the existence and uniqueness of a positive solution of the scheme over a finite-time interval. When necessary, an iterative method to compute the solution is given. A lower bound for the numerical blow-up time is derived for each method, and in two cases an upper bound is also given. Several results concerning the rate of growth of the numerical solution are also presented. All these results confirm the performance of these methods, as illustrated in Chapter 3.

Chapter 1: Presentation of the Problem

Blow-up solutions occur in various models coming from a large variety of physical backgrounds. The most well-known applications belong to combustion. The unknown function u then represents the temperature, or the excess of temperature of some substance (e.g. gas) in a recipient subject to a chemical reaction. The theory of thermal self-ignition of a chemically active mixture of gasses in a vessel was presented in particular by Gelfand in 1963 [61]. Thermal explosions are also discussed by Frank-Kamenetskii [50] and Joseph and Sparrow [79]. One can also see [41], [80] and [82].

The second important domain of application of blow-ups is fluid dynamics. Turbulent flows may be studied using nonlinear Schrödinger equations. These equations also model the temperature of a liquid flowing around a cylinder when the viscosity of the fluid decreases exponentially with temperature (see [93]).

In nonlinear optics, the cubic Schrödinger equation describes the propagation of light beams in nonlinear, dispersive media [36, 88, 151, 150, 44, 115, 103]. Several articles also stress the importance of the cubic Schrödinger equation in the domain of plasma physics, in particular in relation to Langmuir waves because the equation can be considered as a limit of Zakharov's model for these oscillations [67, 162, 163, 164].

Some applications appear in the field of biology. Indeed, in [141] Souplet suggested an interpretation in population dynamics for $u_t = \Delta u - \mu|\nabla u|^q + u^p$, where the

damping term has no obvious interpretation in terms of a thermal reaction-diffusion process. In the context of the population of a biological species, the unknown function u represents the spatial density of individuals. More models are presented in [71].

More applications of blow-up problems are found in areas like quantum mechanics [68], [144], colloid chemistry [159] and geometry [9, 84]. Some models can be used in chemotaxis [145], theory of gravitational equilibrium of polytropic stars [31, 78, 91] and Ohmic heating [94, 95].

The fields of application of blow-up solutions are varied, which explains the interest generated by these problems, among both mathematicians and applied scientists.

Equations admitting blow-up solutions are diverse and might be complex, however a few examples among the simplest models were more popular. The first problem that generated great interest is the following semilinear parabolic equation

$$\begin{cases} u_t &= \Delta u + f(u), & \Omega \times (0,T), \\ u &= 0, & \partial\Omega \times (0,T), \\ u(x,0) &= u_0(x), & \Omega, \end{cases} \tag{1.1}$$

where the domain $\Omega \subseteq \mathbb{R}^d$ can be bounded or not (including the Cauchy problem), the growth of f is superlinear at infinity and generally u_0 is taken to be a positive continuous function on $\bar{\Omega}$. This equation models a great variety of physical problems, from combustion to population dynamics. However, it most commonly represents nonlinear heat generation. If the source term f is positive, convex and grows fast enough at infinity, then diffusion can not prevent blow-up if the initial state u_0 is large enough. We present here in more detail some conditions ensuring that the solution blows up in finite time.

First, we assume that the function f is positive, strictly increasing and strictly convex on $(0,\infty)$, belongs to $C^2([0,\infty))$ and satisfies

$$\int_b^\infty \frac{ds}{f(s)} < \infty, \tag{1.2}$$

for $b > 0$. For the Cauchy problem, if we also assume that

$$\int_s^\infty \frac{d\sigma}{f(\sigma)} = o(s^{-2/d}), \text{ as } s \to 0,$$

then the solution blows up in finite time for all initial conditions $u_0 \geq 0$, $u_0 \not\equiv 0$. If the domain Ω is bounded, we need to introduce the first eigenvalue λ_1 of $-\Delta$,

$$-\Delta\varphi = \lambda_1\varphi, \qquad \text{in } \Omega,$$
$$\varphi = 0, \qquad \text{on } \partial\Omega.$$

The corresponding eigenfunction φ is chosen positive and normalized so that

$$\int_\Omega \varphi(x)\, dx = 1.$$

Then, if $u_0 \geq 0$ on Ω and

$$f(s) - \lambda_1 s > 0, \text{ for } s > \int_\Omega u_0\varphi\, dx, \tag{1.3}$$

the solution blows up in finite time. In particular, if we can replace $f(u)$ by $\delta F(u)$ where δ satisfies

$$\delta > \lambda_1 \sup_{s>0} \frac{s}{F(s)},$$

the solution of the problem blows up in finite time for any initial condition u_0. These results can be found in [56] and [92]. Two particular models for the source term f have been more specifically studied, in particular in the early works of Fujita [54, 55]. The exponential reaction model $f(u) = e^u$ is mostly known in combustion theory as the solid-fuel model or the Frank-Kamenetskii equation. In [18], Bebernes and Eberly explain in detail how this model is derived. The second classical choice for the source term is $f(u) = u^p$ with $p > 1$. Both examples satisfy the above conditions in the case of a bounded domain.

Studies quickly diversified to quasilinear equations $u_t = \Delta\varphi(u) + f(u)$, and in particular the porous-media equation with a power-type source term

$$u_t = \Delta u^{\sigma+1} + \alpha u^{\beta+1},$$

with $\sigma > 0$, $\beta > 0$, and $\alpha \geq 0$, on a bounded domain of $\mathbb{R}^d$. This problem is also a model for nonlinear heat propagation and has been used in plasma physics for the computation of the temperature in a fusion reaction plasma [133]. In the fast-diffusion case, $\sigma \in (-1, 0)$, the solution may vanish or blow up in finite time. In the slow-diffusion case, $\sigma > 0$, the solution either blows up in finite time or exists for all time. In case $\beta = \sigma > 0$, precise results are known. They can be found in particular in [134]: the solution blows up in finite time if and only if α is larger than the first eigenvalue of the problem $-\Delta\varphi = \lambda\varphi$ on Ω, with $\varphi = 0$ on $\partial\Omega$.

The third equation that has been deeply studied is the nonlinear wave equation $u_{tt} = \Delta u + f(u)$. In fact, it is among the first problems where mathematicians studied blow-up solutions (see Chapter 2), but this model is rarely related to real-life problems. The conditions ensuring the occurrence of finite-time blow-up are similar to those of the semilinear parabolic equation. The nonnegative, nondecreasing and convex function f must grow fast enough at infinity and satisfy (1.3). The initial conditions u_0 and u_{0t} must be nonnegative and nonidentically zero in the case of a bounded domain, and the initial conditions must be positive and satisfy $\Delta u_0 \geq 0$, for the Cauchy problem, in order for the solution to blow up in finite time. See for example [65].

Finally nonlinear Schrödinger equations

$$iu_t = \Delta u + F(|u|^2)u, \text{ in } \mathbb{R}^d, \ d \leq 3, \tag{1.4}$$

started to generate much interest in the eighties. Indeed these models (in particular the cubic Schrödinger equation, $F(s) = qs$) occur in various areas of mathematical physics. In one spatial dimension they arise in wave theory, in two dimensions, they appear in nonlinear optics and in three dimensions they are derived in plasma physics. Defining $G(u) = \int_0^u F(s)\,ds$ and $r = |x|$, if the functions F and u_0 satisfy

$$E_0 := \int_{\mathbb{R}^d} \left[|\nabla u_0|^2 - G(|u_0|^2) \right] dx \leq 0,$$

and

$$\text{Im } \left(\int r\bar{u_0} u_{0r} \, dx \right) > 0,$$

and there exists a constant $c > 1 + 2/d$ such that

$$sF(s) \geq cG(s), \quad \forall \, s \geq 0,$$

the solution of the equation blows up in finite time (see [66] or the monograph [148]). As only the Cauchy problem is studied for the Schrödinger equations, we will not address this example in this thesis.

These equations have been deeply studied and much is known about them. The conditions ensuring finite-time blow-up we gave above are chosen among the simplest and much more detailed and less restrictive conditions were developed (see Chapter 2). However, even though such conditions have been derived for many problems, it is still not possible to get precise information about the exact blow-up time. Blow-ups have an important physical interpretation. If the singularity is not caused by the use of unphysical initial (or boundary) conditions, it illustrates the collapse of some approximations used to derive the real-world model. It is important to be able to reproduce the blow-up as precisely as possible in order to be able to adapt the model in the most adequate manner according to both mathematical considerations and physical concerns. The blow-up time is of particular interest, since such an adjustment is only conceivable if the blow-up time can be properly predicted.

Since theoretical approaches do not provide appropriate results, it is natural to turn to numerical approximations to obtain more information. However this is a sensitive problem : most numerical methods lose accuracy as the solution becomes large, and in case of blow-up numerical data grow unboundedly as the blow-up time approaches. Close to the blow-up set, solutions vary quickly in time and the spatial gradients are very large, whereas the solution changes very slowly on the remainder of the domain.

To reproduce finite-time blow-ups numerically, a natural choice is to use methods that are adaptive in time. For these methods the definition of numerical blow-up follows easily from the definition of the blow-up for the exact solution. The blow-up time of the exact solution is defined to be the finite number T such that

$$\lim_{t \to T} \|u(x,t)\|_\infty = \infty.$$

For time-adaptive methods, the timestep naturally decreases as the blow-up time is approached so that the sequence $\{t_n\}$ generally tends to some finite value as n goes to infinity. We say that a numerical blow-up occurs in finite time if

$$T^* = \lim_{n \to \infty} t_n,$$

is finite with

$$\lim_{n \to \infty} \|u_n(x)\|_\infty = \infty.$$

The time T^* is called the numerical blow-up time. For time-adaptive methods, the strategy of time-stepping plays a key role, as is explained by Stuart and Floater in [146].

Nevertheless, methods with fixed timestep are interesting and we chose to concentrate our work on constructions leading to specialized fixed-step methods. If the definition of numerical blow-up naturally followed from the definition of the theoretical blow-up in the case of adaptative methods, such transition is not obvious in the case of fixed-step methods.

For any fixed timestep h, the numerical solution $u_n(x)$ approximates $u(t_n, x)$ where $t_n = nh$. If the function u blows up at finite time T, there exists n^* such that

$$t_{n^*} < T \leq t_{n^*+1},$$

and the numerical solution u_n should only be computed up to t_{n^*}. As a consequence, the solution, and its numerical approximation, can only reach a certain value K,

where $K \geq \|u(T-h,x)\|_\infty$, see Figure 1.1. Conversely, for any large number K, there exists h such that $\|u(t_n,x)\|_\infty \geq K$ for some $n < T/h$.

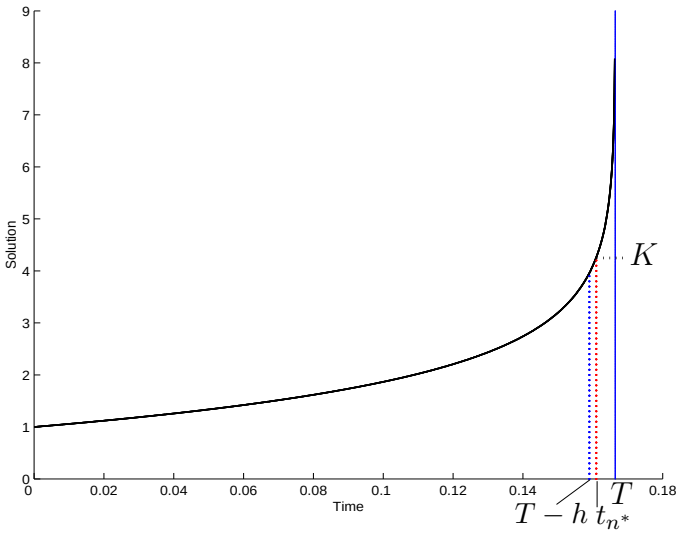

Figure 1.1: Numerical Blow-Up.

Hence the natural way to define the numerical blow-up time is the following. For a fixed large number K, we define $T_K^* = nh$ where n is such that $\|u_n(x)\|_\infty > K$. Then we say that the numerical solution blows up in finite time if the limit

$$T^* := \lim_{K\to\infty} T_K^*,$$

exists and is finite. The time T^* is called the numerical blow-up time. To prove that a numerical blow-up occurs and to give an upper bound on the blow-up time, we need to show the existence of an upper bound for $\{T_K^*\}$, that is we need to prove that there exists T_u such that for all $K > 0$ and h small enough, there exists $n < T_u/h$ such that

$$\|u_n(x)\|_\infty > K.$$

This will be done for one B-method in Section 4.2.3.

In this thesis, we present two different approaches that lead to the construction of semi-discretizations in time specialized for a specific blow-up problem. As we have seen, many blow-up models consist of two parts: a nonlinear part modelling the source (heat source, gas reaction...) and an other part (usually linear) modelling the diffusion process. It is clear that the nonlinear part is responsible for the blow-up and that the diffusion part only delays the occurrence of the blow-up. Moreover the larger the solution becomes, the more important the reaction part is. As we get closer to the blow-up time, the diffusion process plays a minor role. Many models are simple enough so that the differential equation obtained by removing the diffusion process part can be explicitly solved. It is thus interesting to exploit this information. In Chapter 3, we explain using an example how to use any standard method as a basis for new specialized methods. The first approach consists in constructing splitting methods and the second approach derives from the variation of the constant. Since the resulting methods are developed specially to properly reproduce blow-ups, we call them B-methods.

As a conclusion to this introductory chapter, we present an example of application of our methods. We apply standard first-order methods and the corresponding B-methods, to the classical problem

$$u_t = u_{xx} + 3e^u,$$

on $[-1, 1]$, with $u(-1, t) = u(1, t) = 0$ for all t and $u(x, 0) = \cos(\pi x/2)$. The spatial derivative u_{xx} is discretized using finite differences with 31 gridpoints. For the time derivative u_t we used the same fixed stepsize $h = 0.0001$ for all the methods we applied, that is the standard forward Euler (FE) and backward Euler (BE) methods, as well as six B-methods. Four of them are obtained using the first approach (splitting methods), they are labeled SpFE, SpFEA, SpBE and SpBEA. The two remaining-ones are obtained using the second approach (variation of the constant) and are labeled VCFE and VCBE. All six B-methods are derived in Chapter 3. For each of these

methods, we compute the solution u up to $T = 0.1663$. To get a clearer view of the results, instead of plotting the solutions at all gridpoints for all times, we chose to only plot the solutions at the midpoint of the domain $[-1, 1]$, so that we plot $u_n(0)$ for n between 0 and 1663. As the exact solution of the problem is not known, we use the adaptive method ode45 of Matlab to represent the exact solution.

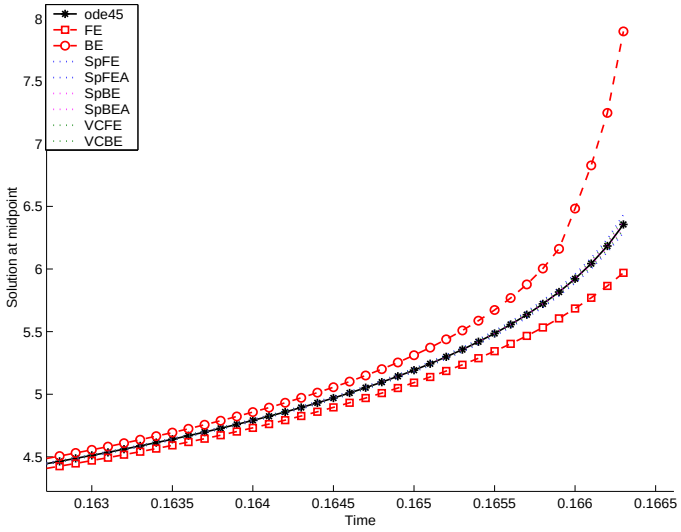

Figure 1.2: Application of first-order standard methods and B-methods.

On Figure 1.2, we observe that whereas the standard methods (circles and squares on the figure) go away from the exact solution (plain line), all B-methods (in dotted lines) stay close to it. We actually need to zoom (Figure 1.3) to properly distinguish them and we see that the approximations obtained by the B-methods really follow the trajectory of the exact solution. More detailed numerical examples are presented in Section 3.3, however one can already see on this simple example the potential of B-methods.

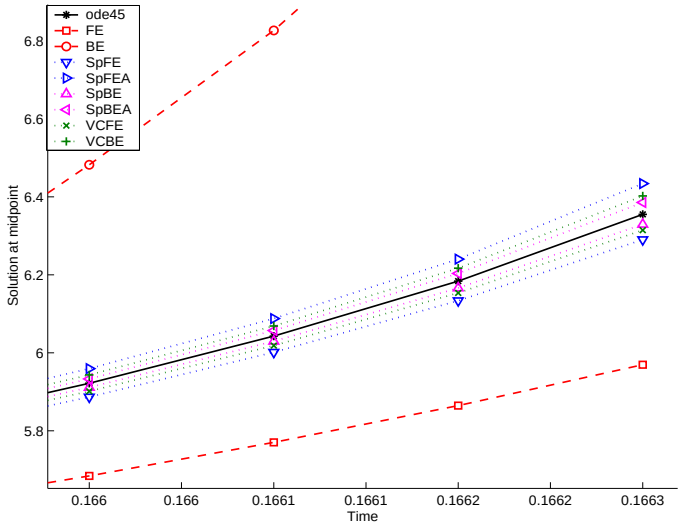

Figure 1.3: Zoom of the above application.

Chapter 2: Historical Review

2.1 Continuous Results

> *Another consequence of Theorem I is the existence, in any bounded domain, of a solution of $\Delta u = f(u)$ which becomes infinite everywhere on the boundary of the domain provided that f(u) is an increasing function.*
> Keller (1957) [86]
> *Consider the nonlinear wave equation $u_{tt} - c^2\Delta u = f(u)$. (...) We will show that for a certain class of functions $f(u)$ the solution u becomes infinite at a finite value of t, provided the initial data satisfy appropriate conditions.*
> Keller (1957) [87]
> *Finally, we consider the question of obtaining estimates which could be used to show that solutions of*
> $$\frac{\partial u}{\partial t} - \sum_{i,j=1}^{n} \frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial u}{\partial x_j}\right) = F(u),$$
> *blow up in some finite time interval, if F grows too rapidly as a function of u.*
> Kaplan (1963) [81]

Even though we can not really speak about blow-up phenomena for elliptic equations as time is absent from the equations, the existence of solutions which become infinite was first mentioned for these types of equations. It started as early as 1916 when Bieberbach published a first paper [24] about $\Delta u = e^u$ in two dimensions. In

1957, Keller generalized the known results to $\Delta u = f(u)$ in any dimension in [86] and decided to apply a similar procedure to nonlinear wave equations in a second article [87]. In this paper, Keller studied the general equation

$$u_{tt} = c^2 \Delta u + f(u), \tag{2.1}$$

with initial conditions $u(x,0) = \varphi(x)$, $u_t(x,0) = \psi(x)$, in a space of dimension $d = 1, 2$ or 3. Using a comparison theorem, he proved that under certain conditions on the function f and on the initial data, the solution becomes infinite in finite time. Moreover he generalized his result to the Euler-Poisson-Darboux equation

$$u_{tt} + \frac{k}{t} u_t - c^2 \Delta u = f(u), \tag{2.2}$$

for $d = 2$, $k > 1$, and $\psi \equiv 0$. Rosenbloom obtained separately, using the same method, equivalent results concerning (2.1) in the particular case $d = 1$ and $f(u) = u^2$. He stated his results in an abstract [132] published in 1954.

We now turn to the equations which are of most interest for us. Blow-up phenomena for semilinear parabolic equations in bounded domains have been studied for the first time in a paper by Kaplan [81] in 1963. This paper is rarely presented as the first reference on the subject, which may be explained by the fact that blow-up solutions do not represent an important part of it. This article contains the proof of a comparison theorem for a general parabolic equation in cylindrical domains under very general boundary conditions, followed by many applications. In particular, the following theorem was derived.

Theorem 2.1 (Kaplan)**.** *We suppose that Ω is bounded, and that $u(x,t) \in C^{2,1}$ in $Q_T := \Omega \times (0,T]$, and satisfies there*

$$\frac{\partial u}{\partial t} - L[u] \geq G(u,t), \tag{2.3}$$

where

$$L = \sum_{i,j=1}^{d} \left(\frac{\partial}{\partial x_i} a_{ij}(x) \frac{\partial}{\partial x_j} \right),$$

is a self-adjoint uniformly elliptic differential operator with smooth coefficients (in $C^3(\bar{\Omega})$*, say) and* $G(u,t)$ *is a convex function of* u *for each fixed* $t \geq 0$*. Let* $\phi(t)$ *satisfy*

$$\frac{d\phi(t)}{dt} = G(\phi(t),t) - \lambda_1(\phi(t) - k(t)), \qquad for \;\; 0 < t \leq T,$$

and

$$\phi(0) = \inf_{x\in\bar{\Omega}} u(x,0),$$

where $k(t) = \inf_{x\in\partial\Omega} u(x,t)$*, and where* λ_1 *is the first eigenvalue for the problem*

$$-L[\psi] = \lambda_1\psi, \qquad in\ \Omega,$$

with

$$\psi = 0, \qquad on\ \partial\Omega.$$

Then, we have

$$\sup_{x\in\bar{\Omega}} u(x,t) \geq \phi(t), \quad for \quad 0 \leq t \leq T.$$

If we impose homogeneous Dirichlet boundary conditions, the proof is very simple and since it will be of great importance in this thesis, we present this particular case here. Note that in that case, we have $k(t) = 0$ for all t.

Proof. (Case of homogeneous Dirichlet boundary condition) One key element of the proof is that by Courant's Theorem, the eigenfunction ψ corresponding to λ_1 does not change sign in Ω, so that we can take ψ to be nonnegative on Ω and normalized such that $\int_\Omega \psi\, dx = 1$. Multiplying the inequality (2.3) by ψ and integrating over Ω, we obtain

$$\int_\Omega \frac{\partial u}{\partial t}\psi\, dx - \int_\Omega L[u]\psi\, dx \geq \int_\Omega G(u,t)\psi\, dx.$$

Using the fact that u is continuously differentiable and applying Stoke's Theorem we obtain

$$\frac{dv}{dt} \geq \int_\Omega L[\psi]u\, dx + \int_\Omega G(u,t)\psi\, dx,$$

where $v(t)$ is the L^2-inner product of u and ψ, $v(t) = \int_\Omega u(x,t)\psi(x)\,dx$. By definition of ψ, we have $-L[\psi] = \lambda_1\psi$ and we can apply Jensen's inequality to the last integral since $G(u,t)$ is a convex function of u for each fixed t:

$$\frac{dv}{dt} \geq -\lambda_1 v(t) + G(v(t),t).$$

Since $v(0) \geq \phi(0)$, by standard comparison techniques for ordinary differential equations, $v(t) \geq \phi(t)$ for $0 \leq t \leq T$. Since $v(t) \leq \sup_{x\in\bar{\Omega}} u(x,t)$, the result follows. □

Using this theorem, it is easy to show that any solution of

$$\frac{\partial u}{\partial t} - L[u] = F(u), \tag{2.4}$$

where $F(u)$ is convex and positive for $u > u_0 > 0$, and $\int_{u_0}^{\infty} du/F(u) < +\infty$, can be made to blow up in any prescribed time interval by making initial values and/or boundary values large enough. Indeed, if $u_0 \leq m \leq u(x,0) \leq M$, $k \leq u(x,t) \leq K$ for $x \in \partial\Omega$ and $F(u) - \lambda_1(u-k) > 0$ for $u \geq m$, the function $\phi(t)$ satisfies $\phi'(t) = F(\phi(t)) - \lambda_1(\phi(t)-k)$ which gives

$$\int_{\phi(0)}^{\phi(t)} \frac{ds}{F(s) - \lambda_1(s-k)} = t.$$

Under the above assumptions, we have

$$\int_{u_0}^{\infty} \frac{ds}{F(s) - \lambda_1(s-k)} < \infty,$$

so ϕ and (by Theorem 2.1) any solution of (2.4) must become infinite as $t \to T_0$, for some T_0 such that

$$T_0 \leq \int_m^{\infty} \frac{du}{F(u) - \lambda_1(u-k)}.$$

Moreover, other results from the article imply that

$$T_0 \geq \int_{\max(M,K)}^{\infty} \frac{du}{F(u)},$$

so that estimates from above and below for the escape time for (2.4) are obtained. It is interesting to note that Kaplan's approach does not rely on the maximum principle, but rather on a study of the ordinary differential inequality satisfied by the scalar product of u and the lowest eigenfunction ψ.

The next cornerstone was the fundamental work of Fujita who published four papers in the late sixties. These articles are often referred to as the beginning of a deeper study of blow-up solutions. In the first paper [54], published in 1966, Fujita studied the Cauchy problem

$$u_t = \Delta u + u^{1+\alpha}, \quad \alpha > 0, \tag{2.5}$$

and nonnegative initial data (with some growth restriction as $|x| \to \infty$). Fujita noted that in the previous articles about blow-ups, one constant conclusion was that the solution blows up in finite time for large enough initial data. However, he proved that for the above mentioned Cauchy problem, if $0 < d\alpha < 2$ all nonnegative solutions apart from the null function blow up in finite time, no matter how small the initial data is, whereas if $2 < \alpha d$, there is a global solution for many initial data. To prove these results Fujita used the Green's function of the heat equation. The proof of the first theorem is similar in some way to the proof presented in Kaplan's paper [81]: using the Green's function and Jensen's inequality, Fujita obtained a differential inequality and used it to show that the solution cannot exist for all t. For the second theorem, Fujita used integral equations to construct a sequence of functions converging to the global solution.

In 1969, Fujita published a short note [55] in which he stated several theorems concerning the relations between the boundary value problem $\Delta u + e^u = 0$ (with homogeneous Dirichlet boundary conditions), and the initial boundary value problem $u_t = \Delta u + e^u$ (with the same boundary conditions and a nonnegative initial condition) on a bounded domain Ω. The proofs of these theorems were only outlined in the note [55], however a more general version of the main result concerning blow-up

solutions was proved in a subsequent article published in 1970 [57]. In this paper, the function e^u was replaced by a general function f assumed to be increasing and strictly convex. Using the Green's function of $-\Delta$ in Ω with homogeneous Dirichlet boundary conditions, Fujita proved the following result: if the boundary value problem (BVP) has two distinct solutions $u_1 \leq u_2$, and if the initial data u_0 of the initial boundary value problem (IBVP) satisfies $u_0 \geq u_2$, $u_0 \not\equiv u_2$, then the solution of the IBVP blows up in finite time or diverges to infinity at infinity, whereas if $u_0 \leq u_2$, $u_0 \not\equiv u_2$, the solution of the IBVP is global. The same year, Fujita published another article [56] in which he proved that under certain conditions on f, the nonlinear parabolic problem

$$u_t = \Delta u + f(u), \tag{2.6}$$

with homogeneous Dirichlet boundary conditions have solutions which blow up in finite time. In this article, Fujita considered two cases: the Cauchy problem ($\Omega = \mathbb{R}^d$), and the case with bounded domain Ω. Concerning the Cauchy problem, Fujita first stated the results derived in his previous paper and presented a generalization thereof. Assuming that the initial data u_0 is nonnegative and non-identically zero, and that the function f satisfies the following conditions: f is locally Lipschitz continuous with $f(0) \geq 0$, $f(s) > 0$ for $s > 0$, $1/f$ integrable at infinity, f convex on $[0, \infty)$ and

$$\int_s^\infty \frac{d\lambda}{f(\lambda)} = o(s^{-2/d}), \quad \text{as } s \to 0^+,$$

then the solution of the initial value problem blows up in finite time. The proof uses the Green function of the heat equation and Jensen inequality. Fujita mentioned that this result could be generalized to the case where the Laplacian is replaced by some elliptic operator. Another theorem proving the existence of a global solution under certain conditions on the growth of f and on the size of the initial data was also given together with an outline of its proof. Concerning the problem on a bounded domain, Fujita's theorem is a particular case of Kaplan's result. It states that for nonnegative

initial data, if f is locally Lipschitz continuous with $f(0) \geq 0$, $f(s) > 0$ for $s > 0$, $1/f$ integrable at infinity, f convex on $[0, \infty)$ and

$$f(s) - \lambda s > 0, \quad \text{for } s > \int_\Omega u_0(x)\varphi(x)\,dx,$$

where λ and φ are the first eigenvalue and first eigenfunction of $-\Delta\varphi = \lambda\varphi$, with homogeneous Dirichlet conditions, φ normalized by $\int_\Omega \varphi(x)\,dx = 1$, then the solution blows up in finite time. In this case also, Fujita mentioned that the Laplacian could be replaced by a more general elliptic operator and that more general boundary conditions could be chosen. The proof of this result is the same as Kaplan's one, it uses the lowest eigenmode method. The critical case $\alpha = d/2$ that was not studied by Fujita in [54] and [56] was partly settled by Hayakawa in a short note [73] published in 1973. Hayakawa proved that the limit of a sequence of subsolutions blows up in finite time, so that the problem (2.5) has no global solution for any nontrivial initial data u_0 in case $d = 2$, $\alpha = 1$ or $d = 1$, $\alpha = 2$.

In the beginning of the seventies, Tsutsumi wrote several papers on blow-up solutions. In his first article of 1972 [154], Tsutsumi obtained results similar to those of Fujita [56], for solutions of

$$u_t = \sum_{i=1}^{d} \frac{\partial}{\partial x_i}\left(\left|\frac{\partial u}{\partial x_i}\right|^{p-2}\frac{\partial u}{\partial x_i}\right) + u^{1+\alpha},\ p \geq 2, \alpha \geq 0,$$

on a bounded set Ω with homogeneous Dirichlet boundary conditions. Using Galerkin's method, a compactness argument and an energy inequality, Tsutsumi proved that if $p > 2+\alpha$, the problem has a global solution for all nonnegative $u_0 \in \mathcal{W}_0^{1,p}(\Omega)$, whereas if $p < 2+\alpha$, the initial condition u_0 needs to be sufficiently small to get a global solution; otherwise, if u_0 is large, the solution blows up in finite time. In a second article published the same year [155], Tsutsumi considered the abstract Cauchy problems for

$$u_t + \mathcal{A}u + f(u) = 0 \qquad \text{and} \qquad u_{tt} + \mathcal{A}u + f(u) = 0,$$

where $\mathcal{A}$ is a nonnegative self-adjoint operator in a real Hilbert space H and f is a nonlinear operator mapping H into itself. For these two equations, Tsutsumi stated conditions which ensure the existence of a global solution and conditions which imply finite blow-up time. Finally, in 1974, Tsutsumi [156] looked at

$$u_t = \sum_{i=1}^{d} \frac{\partial}{\partial x_i} \left((1 + |u|^{p-2}) \frac{\partial u}{\partial x_i} \right) + u^{1+\alpha}, \; p \geq 2, \alpha > 0,$$

on a bounded set Ω with homogeneous Dirichlet boundary conditions and a nonnegative initial data u_0. After stating and proving some results concerning global existence and uniqueness, Tsutsumi stated the following: if $2 < p < \alpha$, and if the initial data u_0 satisfies

$$\frac{1}{2} \sum_{i=1}^{d} \int_{\Omega} (1 + (u_0(x))^{p-2}) (D_{x_i} u_0(x))^2 \, dx - \frac{1}{2+\alpha} \int_{\Omega} (u_0(x))^{2+\alpha} \, dx \leq 0,$$

which is satisfied if u_0 is large enough, then the solution blows up in finite time in the L^2-norm. Tsutsumi proved this theorem using a blow-up subsolution.

In 1973, Glassey [65] considered the semilinear wave equation (2.1) with $c = 1$ and showed that Kaplan's method can actually be applied to this equation on a bounded domain and thus proved that under certain conditions on the positivity, growth and convexity of f, and for some initial conditions, the solution of the equation must blow up in finite time. Glassey also considered the Cauchy problem and showed that for a positive and convex function f, the solution blows up in finite time for many initial conditions. He first proved his result for the case $d = 3$, using a combination of the methods of Kaplan and Keller, and then showed how this result can be extended to the general case. Finally Glassey considered the Cauchy problem for

$$u_{tt} = \Delta u + f(u_t), \tag{2.7}$$

for f positive and convex and he showed that the solution blows up at a rate greater than the one of the corresponding wave equation (2.1).

After the techniques involving the lowest eigenmode or the Green's function, a third type of well-used methods to prove that the solutions of some problems blow up are the concavity methods, also called energy methods. These were presented for the first time by Levine [104] in 1973. As Levine pointed out, the fact that no maximum principle, positive first eigenfunction or Green function is used, this method can be used for much more general problems, in particular for higher than second order parabolic equations. In his paper, the first of a long series, Levine concentrated on

$$Pu_t = -A(t)u + \mathcal{F}(u(t)),$$

where P and A are symmetric linear operators which satisfy certain positivity conditions and the nonlinear function $\mathcal{F}$ must satisfy the following growth condition

$$\int_\Omega x\mathcal{F}(x)\,dx \geq 2(\alpha+1)\int_0^1\int_\Omega x\mathcal{F}(\rho x)\,dx\,d\rho, \tag{2.8}$$

for all x in Ω. To prove that the solution blows up in finite time if the initial data is large enough, Levine considered the function

$$\Phi(t) = (T_0 - t)\int_\Omega u_0 P u_0\,dx + \beta(t+\tau)^2 + \int_0^t\int_\Omega uPu\,dx\,d\eta.$$

He proved that for some chosen values of $T_0 > 0$, $\beta > 0$, and $\tau > 0$, $\Phi^{-\alpha}$ is concave so that

$$\Phi(t) \geq \Phi^{(1+1/\alpha)}(0)[\Phi(0) - \alpha t\Phi'(0)]^{-1/\alpha},$$

and the solution u can not exist for $t \geq \Phi(0)/\alpha\Phi'(0)$. This result is applied to several examples in the remainder of the paper. The same method was used the following year in [105] for the nonlinear wave equation

$$Pu_{tt} = -A(t)u + \mathcal{F}(u(t)),$$

where A is a symmetric linear operator, P is a strictly positive symmetric operator and the nonlinear function $\mathcal{F}$ satisfies again (2.8). In this case, the function Φ is

defined by

$$\Phi(t) = Q^2 + \beta(t+\tau)^2 + \int_\Omega uPu\,dx,$$

and $\Phi^{-\alpha}$ is concave for some nonnegative constants Q, β and τ. As for the parabolic equation, Levine proved that if the initial potential energy of the nonlinearity is larger than the total initial energy of the linear problem, then the problem can not have a global solution. Some results of this paper were extended in a subsequent article [107], which studied in particular weak solutions. Levine also published a note [106] in which he extended some results of Keller about the Euler-Poisson-Darboux equation (2.2). Indeed Keller only considered the case $k > 1$ and Levine showed that Keller's results are valid in the case $0 < k \leq 1$ as well. In this case, independently of the space dimension, for certain functions f and certain initial conditions, every classical solution must blow up in finite time. The following articles, written with Payne, exploit the concavity method for different equations: the heat equation with nonlinear boundary conditions [109], more general classes of higher order equations [110], some abstract nonlinear equations (in a paper written by Knops, Levine and Payne [89]) and the abstract Cauchy problem [111].

The introduction of another paper [108] by Levine, published in 1975, is interesting as it lists different techniques that have been used to study finite-time blow-up, together with corresponding references. In the article, Levine studied weak solutions of $u_{tt} = \mathcal{L}u + \mathcal{F}_1(u)$ and $u_t = \mathcal{L}u + \mathcal{F}_2(u)$ on a bounded domain, where $\mathcal{L}$ is not necessarily elliptic but must have a positive eigenfunction ψ. Under some restrictions on $\mathcal{F}_1$ and $\mathcal{F}_2$, and for large enough initial conditions, all solutions must blow up in finite time. To prove his results, Levine obtained an ordinary differential equation for the corresponding Fourier coefficient $\int_\Omega \psi(x)u(x,t)dx$. This method was not original in the case of an elliptic operator $\mathcal{L}$, however Levine illustrated his results with several unusual examples. He then explained how one can extend these results to nonlinear equations in Banach spaces and concluded with a similar result for (2.7) in a bounded

domain.

Sugitani [147] extended the study of parabolic equations by replacing the Laplacian by the fractional power operator $-\left(-\frac{\Delta}{2}\right)^{\beta/2}$ $(0 < \beta \leq 2)$ in the semilinear parabolic equation (2.6). Using the fundamental solution of the equation with $f(u) \equiv 0$, and Jensen's inequality, he proved that if f is increasing, convex, and grows fast enough at infinity, the solution must blow up in finite time.

In 1977, Ball [11] illustrated the importance of combining the nonexistence arguments advanced by most authors with a continuation theorem. Indeed he presented an example of a semilinear parabolic equation (2.6) in a bounded domain of $\mathbb{R}^d$ whose solution ceases to exist but does not blow up in finite time.

Another kind of equation whose solutions might blow up in finite time is the nonlinear Schrödinger equations. This property was presented in the literature in 1977 [66], when Glassey studied $iu_t = \Delta u + F(|u|^2)u$ in $\mathbb{R}^d$. By first deriving a priori estimates, Glassey proved that under some conditions on the initial data and the function F, the solution blows up in finite time.

The results about semilinear parabolic equation (2.6) kept on becoming more precise as Kobayashi, Sirao and Tanaka [90] improved the known results for the Cauchy problem. By constructing the solution by iteration, using Green's functions, they proved that under some conditions on the function f, each positive solution of the problem blows up in finite time.

The article published in 1981 by Bebernes and Kassoy [20] presents a nice review of the known results about the semilinear parabolic equation in the specific case $f(u) = e^u$. A large part of this article consists in explaining the physics behind the equation so it is an interesting link between the math and the physics. Moreover numerical experiments comparing the actual blow-up time with the obtained bounds (the upper bound is the one derived by Kaplan in [81]) complete the paper.

The Russian school started to work on blow-up problems at the end of the sev-

enties. In particular, the Cauchy problem for $u_t = (u^\sigma)_{xx} + u^\beta$ was studied by Samarskii, Kudryumov and others. The corresponding boundary value problem on a bounded domain, with homogeneous Dirichlet boundary conditions, was first studied by Galaktionov in 1981 [58]. The author gave conditions under which the problem has no global solution and the solution blows up in some sense. An upper bound for the blow-up time is explicitly stated for both case $\beta = \sigma$ and $\beta > \sigma$. The same problem was also studied by Sacks in his Ph.D. thesis and one can find his results in [134]. Actually, the problem considered by Sacks is the "differential inclusion" $\beta(u)_t \ni \Delta u + q\nabla\gamma(u) + F(x,t,u)$ and the first section is devoted to the study of the classical solutions of the corresponding equation. A particular case of his work is of particular interest for this thesis: when $q = 0$, $F = F(u) = \alpha u$ and $\beta(u) = u^{1/m}$, where $m > 1$, that is the porous media equation. The equation written for $v = u^{1/m}$,

$$v_t = \Delta v^m + \alpha v^m,$$

is exactly the equation studied by Le Roux in [99] and will be studied in Section 4.4. Sacks's results imply in particular that if α is larger than the first eigenvalue of $-\Delta\rho = \lambda\rho$, with homogeneous Dirichlet boundary conditions, the solution must blow up in finite time.

In 1983 [92] and 1984 [93] Lacey too looked into the semilinear parabolic problem (2.6). He presented significant results and, by writing the problem in the form

$$u_t = \Delta u + \delta f(u),$$

he studied the connection between the existence and time of blow-up and the relationship between δ and the spectrum of the corresponding steady state problem. We denote by δ^* the critical value for which the steady state problem corresponding to the original equation has a positive classical solution. If $\delta > \delta^*$ under certain conditions on f, the solution of the IBVP blows up in finite time, whereas if $\delta \leq \delta^*$,

conditions of the size of the initial conditions are required in order to ensure finite-time blow-up. In case $\delta = \delta^*$, a lower bound for the blow-up time is derived and leads to an asymptotic approximation of the blow-up time for δ close to δ^*.

Ayeni [10] considered the more general equations

$$u_t = \Delta u + g(x,t)e^u \text{ and } u_t = \Delta u + g(x,t)(1+u^2),$$

with $g(x,t) \geq \lambda t^{n+\alpha-1} H(x,t)$, where H is the fundamental solution of the heat equation and λ and α are positive constants. Ayeni proved, using the maximum principle, that for certain classes of initial conditions the solution blows up in finite time, and he gave an upper bound for the blow-up time.

In the eighties, several authors started to broaden the questions about blow-up solutions. As we have seen, some equations had been deeply studied and the conditions leading to blow-up solutions were well-known, as well as approximations of the blow-up time. Once the question "when?" was answered, the next natural questions were "where and how does the blow-up occur"? The first to look at the blow-up set was Weissler [160] in 1984. He restricted his study to $u_t = u_{xx} + u^p$, on the domain $\Omega = [-R, R] \subset \mathbb{R}$, with homogeneous Dirichlet boundary conditions and nonnegative initial conditions, and proved that the blow-up occurs only at the point $x = 0$. Friedman and McLeod [53], and Giga and Kohn [62] studied the asymptotic behavior of the solution as the blow-up time is approached. In [62] Giga and Kohn gave, for the same equation as Weissler, a pointwise characterization of the asymptotic behavior of the solution near the blow-up point. Friedman and McLeod considered the more general equation $u_t = \Delta u + f(u)$ on a bounded domain of $\mathbb{R}^d$. They first studied the set of blow-up points in the symmetric case (when the domain is a ball and the initial condition is a radial function, decreasing in r) and the non-symmetric case (where the domain is assumed to be convex). In the symmetric case, the authors generalized the previous results by proving that the blow-up occurs at the single point $r = 0$. In the

non-symmetric case, the authors proved that the blow-up set lies in a compact subset of the domain. Then, for both cases, the asymptotic behavior of u at the blow-up points was studied in the cases where $f(u) = (u + \lambda)^p$ and $f(u) = e^u$ and precise estimates were given. This paper is of great importance for the development of the subject and is often cited in subsequent papers.

These articles and the following ones represent the beginning of a much more diversified study of equations with blow-up solutions. From that moment on, the studies on blow-up solutions also covered where [52, 126] and how the blow-up occurs (the asymptotic behavior is treated in [17, 33, 63, 161]), as well as what happens later (i.e. is the blow-up complete or is it possible to extend the solution after the blow-up time?) [14, 15, 96]. Quickly, the variety of the problems whose solution might blow up studied expanded (nonlinear wave equations [131, 138], nonlinear Schrödinger equations [120, 130], nonlinear parabolic equations including the gradient ∇u [37], with nonlinear memory [21], degenerate parabolic equations [112], parabolic equations degenerate in the time derivative [49], systems of parabolic equations [52] for example). So we will not attempt to expose an exhaustive overview of what has been done. We are only going to present a few more relevant articles concerning the methods used to prove that solutions blow up in finite time and approximate the blow-up time.

Indeed, another approach to prove that a solution is blowing up in finite time is based on the result stated by Sattinger [137]: if an upper solution $\bar{u}$ and an lower solution $\underline{u}$ exist, there is a solution u such that $\underline{u} \leq u \leq \bar{u}$. This statement applies to elliptic equations $Lu = f$ and to parabolic equations $Lu - u_t = f$. The first author to explore this path for parabolic equations was Meier in his thesis [116] in 1987. His results were announced in a short note [117] in 1986 and presented in an article [118] in 1988. He stated conditions on the functions $v(x,t)$ and $w(x,t)$ which ensure that the function $\underline{u}(x,t) = z(v(x,t); w(x,t))$ is a lower solution, where z is the solution of $z_v = f(z)$; $z(0) = w$. Meier's results can be applied to bounded and

unbounded domains and his article presents applications to several specific problems. In his thesis, Meier showed that the bounds obtained for the blow-up time are not directly comparable to those obtained by the method of Kaplan. Indeed the choice of the best estimate depends on the size of the domain, the size of the initial condition and the function f.

Another author who used a lower solution to study the blow-up time is Bellout in his article [22] published in 1987. In this paper, assuming the concavity of (f/f'), the author considered the problem $v_t = \Delta v + \delta a^2 t^2 f(v)$, for a certain constant a depending on δ and δ^*, where δ^* is the critical value for which the steady state problem corresponding to the original equation has a positive classical solution. By proving that there exists a point x_0 in the domain such that $v(x_0, t)$ tends to infinity as t tends to $T = 1/a$, Bellout obtained a sharp bound on the blow-up time of u.

At the same period, Kavian presented in [83] a new proof of Fujita's results by relating self-similar solutions of $u_t - Au = |u|^{p-1}u,\ (x \in \mathbb{R}^d)$, to stationary solutions of $v_s + Lv = |v|^{p-1}v + \lambda v,\ s > 0,\ y \in \mathbb{R}^d$, where $\lambda = (p-1)^{-1}$, $L = -K^{-1}\nabla \cdot K\nabla$, $K = \exp(|y|^2/4)$ and L^{-1} is compact self-adjoint and positive on the weighted space $L^2(\mathbb{R}^d; K)$. Sufficient conditions for global existence or for finite-time blow-up are given in terms of $E_\lambda(v_0)$, where E_λ is an energy functional for the second equation.

The technique explored by Meier in 1987, involving lower and upper solutions was not very often used, however in 1997 Souplet and Weissler [142] compared u with a self-similar subsolution that blows up in finite time. Assuming only a growth condition on the nonlinear function $F(u, \nabla u)$, it is possible to construct such a subsolution. This technique improved a large part of the known results on the existence of blow-up and allowed a unified treatment for problems that previously had to be handled by different methods.

Finally, one should mention the large book devoted to blow-up in quasilinear parabolic equations [135], written in 1995 by Samarskii, Galaktionov, Kurdyumov

and Mikhailov. The content includes in particular generalized solutions for degenerate equations, heat localization, self-similar solutions and their asymptotic stability, methods of generalized comparison of solutions of different equations and approximate self-similar solutions. Extensive bibliography and open problems are given for each topic.

For an overview of the progress of the study of blow-ups, we first refer to the third chapter of the early book by Berbernes and Eberly [18]. Motivated by the study of the solid fuel ignition model $\theta_t = \Delta\theta + \delta e^{\theta}$, Bebernes and Eberly presented an exhaustive summary of the results published about $u_t = \Delta u + \delta f(u)$. Their work covers the questions concerning the existence and uniqueness of the solution, the condition of the existence of a finite-time blow-up and when, where and how it occurs. Also, their Section 3.5 contains a nice overview of references. The survey article written in 1990 by Levine focuses mainly on the role of critical exponents, yet it is very interesting as it contains many references and covers different types of problems (nonlinear parabolic equations, nonlinear Schrödinger equations and nonlinear hyperbolic equations). The survey article published in 1998 by Bandle and Brunner follows a thematic approach: in Section 3 in particular, different approaches used for establishing blow-up are presented. Moreover important questions concerning numerical blow-up solutions are raised and Section 6 presents what had already been done in that topic. Besides, the problem of computing and reproducing the blow-up numerically started to arise in the mid-eighties. A historical review of the study of numerical solutions will be covered in the following section.

For a recent overview, we refer to the comprehensive survey by Galaktionov and Vazquez [59] published in 2002. This article focuses mainly on parabolic problems. It starts with a good introduction and a historical review, followed by a discussion of the main questions. These consist of the existence of the blow-up, when, where and how it occurs and what happens beyond it. The last question is discussed in more

detail. Different types of parabolic equations are studied, as well as systems and a brief section about other nonlinear equations. The numerical aspect is also raised, however they simply refer to the article by Bandle and Brunner [13]. This article contains an extensive list of references.

2.2 Numerical approximations of blow-up solutions

> *Our concern is directed to the numerical study of $u_t = u_{xx} + u^2$ with special emphasis on the case when the exact solution blows up with the blow-up time T_∞, where the blow-up is dealt with in the sense of L^2.*
> Nakagawa (1976) [127]
> *As [the blowup time] is approached, the discretization of the original problem results in a distortion of the blowup mechanism and, unless care is exercised, the numerical results can be misleading.*
> Tourigny and Sanz-Serna (1992) [153]

The pioneer article concerning approximations of blow-up solutions using numerical methods was written in 1976 by Nakagawa [127]. In this article, titled "Blowing up of a Finite Difference Solution to $u_t = u_{xx} + u^2$", Nakagawa restricted the study of this equation to the one-dimensional case $\Omega = (0, 1)$, with homogeneous Dirichlet boundary conditions and a nonnegative initial condition. He concentrated on the case when the exact solution blows up in finite time T_∞. The solution of the equation is approximated using the following difference scheme

$$D_t\, v^{n,i} = D_x\, D_{\bar{x}}\, v^{n,i} + (v^{n,i})^2, \tag{2.9}$$

where D_x and $D_{\bar{x}}$ represent respectively forward difference and backward difference in x, and D_t represents forward difference in t, with variable time-step $\Delta t_n = \lambda_n h^2$,

where h is the spatial mesh size. The parameter $\lambda = \max \lambda_n$ plays a crucial role. In particular, since the scheme (2.9) can be rewritten as

$$v^{n+1,i} = \lambda_n v^{n,i+1} + (1 - 2\lambda_n)v^{n,i} + \lambda_n v^{n,i-1} + \Delta t_n (v^{n,i})^2,$$

the condition $0 < \lambda \leq 1/2$ ensures the positivity of $v^{n,i}$ for all n and all i. The article contains two major results. First, if λ satisfies the above condition and if the maximal step-size τ is smaller than some value determined by the exact solution, we have for $1 \leq k \leq n-1$, where $0 \leq t_n \leq T < T_\infty$,

$$\max_i |v^{k,i} - u(t_k, x_i)| \leq c(T) \cdot h^2,$$

where u is the exact solution. The second important theorem states that the numerical blow-up time converges to T_∞ when τ tends to zero and $\Delta t_n = \tau \cdot \min\{1, \frac{1}{\|v_n\|}\}$. However these results are only valid in the case where the solution blows up in the sense of L^2, which is only true for some reaction functions. Indeed Friedman and McLeod proved several years later [53], that it is not true for many interesting functions.

A second article was published a year later by Nakagawa and Ushijima [129] to generalize the results obtained by Nakagawa. In this article, the problem studied is $u_t = \Delta u + f(u)$, where f is a locally Lipschitz-continuous convex function such that f is nonnegative on $\mathbb{R}$ and $f(u) \geq Cu^{1+\gamma}$, as u tends to infinity, for some positive constants γ and C. The space Ω is a bounded open set in $\mathbb{R}^d$ with smooth boundary. Moreover they assume homogeneous Dirichlet boundary conditions and the initial condition is taken to be continuous on $\bar{\Omega}$ and vanishing on the boundary. The finite element method of lumped mass type is used to discretize in space, whereas the time step is variable, its size being controlled by the size of the approximate solution u_h, or more precisely by the discretized analogue of $J(t) = \int_\Omega u(t,x)\varphi(x)\,dx$, where φ is the first eigenfunction of $-\Delta\varphi = \lambda\varphi$, with homogeneous Dirichlet boundary

conditions, $\varphi \geq 0$ and normalized ($\int_\Omega \varphi(x)\,dx = 1$). As in the previous article, the authors showed that the numerical blow-up time converges to the blow-up time of the continuous problem and proved the convergence of the approximate solution.

These results appeared again in a following paper by Nakagawa, Ikeda and Ushijima [128], put in a more abstract context and the method was applied not only to the heat equation but also to the wave equation $w_{tt} = \Delta w + f(w)$, where f is a convex polynomial of arbitrary degree if $n = 1$ or 2, or f is a convex quadratic function if $n = 3$, such that f is nonnegative on $\mathbb{R}$ and $f(w) \geq Cw^2$, as w tends to infinity, for some positive constant C. The authors introduced $v = w_t$ and the discretized analogue $I_h(t)$ of $I(t) = \int_\Omega w(t,x)\varphi(x)\,dx$ in addition to the functional $J_h(t)$ used in the previous article to control the time step. Here again the convergence of the numerical blow-up time of the approximate solution was proved.

Whereas these articles study the numerical approximation of problems that had already been deeply studied from a theoretical point of view, Chorin [38] used numerical methods to get some insight about equations for which no analysis of the blow-up properties had been done. His article focuses on the incompressible Euler equations in vorticity form,

$$\partial_t \xi + (u \cdot \nabla)\xi - (\xi \cdot \nabla)u = 0,$$

$$\xi = \nabla \times u, \quad \nabla \cdot u = 0,$$

where u is the velocity and ξ the vorticity, in a unit cube with periodic boundary conditions. The method used is quite sophisticated, it involves rescaling and mesh refinement: a maximum number of points is allowed and when it is exceeded, most parts of the domain are ignored (only the corner containing the singularity is kept) and the problem is rescaled and the boundary conditions are adjusted. The idea of such a study is quite different from the previous ones: whereas Nakagawa et al's articles put emphasis on the convergence of the numerical solution and the numerical blow-up time, the properties of the numerical solution are not proven in the article

of Chorin. The numerical results are used to described the nature of the solutions. A similar work for the nonlinear Schrödinger equation was presented by Sulem, Sulem and Patera [149] in 1984. The numerical blow-up was illustrated but not analytically studied.

The idea of Nakagawa was extended by Chen [34] in 1986. The equation considered is slightly more general than the one studied by Nakagawa: $u_t = u_{xx} + u^{1+\alpha}$, with $0 < x < 1$ and a positive constant α. The initial condition is assumed to be sufficiently smooth, nonnegative and zero on the boundary. Moreover Chen studied not only Dirichlet but also Neumann boundary conditions. Whereas the scheme proposed by Nakagawa was fully explicit, Chen changed it to be implicit in the linear part and explicit in the nonlinear part,

$$\frac{u_j^{n+1} - u_j^n}{\tau_n} = \frac{u_{j-1}^{n+1} - 2u_j^{n+1} + u_{j+1}^{n+1}}{h^2} + (u_j^n)^{1+\alpha},$$

and the size of the timestep is determined by the magnitude of the solution in a way that is a generalization of the method used by Nakagawa. In this article, the author was interested not only in the blow-up time but also in the blow-up set and the shape of the blow-up. In a first part, Chen proved the convergence of the numerical solution and the convergence of the numerical blow-up time in a way similar to Nakagawa's, then he proved that the blow-up concentrates to its maximum points. For this part, Chen assumed that the initial condition was non-constant, symmetric and increasing on $(0, 1/2)$. Assuming that the numerical solution blows up then if $\alpha \leq 1$ the blow-up is concentrated at its maximum point(s) (namely at $1/2$ and (if $\alpha = 1$) the two adjacent points), and if $\alpha \geq 1$ the value of the solutions apart from these points remains bounded. These conclusions are valid for Dirichlet conditions and Neumann conditions. While these results were restricted to the one-dimensional case, they will be extended to the general multi-dimensional case by Chen in 1992 [35], by means of a slight modification of the scheme.

Sanz-Serna and Verwer [136] chose another approach to add their contribution to the study of numerical blow-ups. They restricted their work to the numerical scheme obtained with the explicit Euler method for the simple ODE $y_t = y^m$. They derived sharp bounds for the error and asymptotic estimates of the numerical solution in order to illustrate the error propagation mechanism in nonlinear situations. The idea of restricting the subject of the study to a simple ODE was new and has been used on several occasions later. It allows for a better understanding of the semi-discretization in time of complex PDEs.

In their paper [23] published in 1988, Berger and Kohn suggested a sophisticated method for the numerical approximation of $u_t = u_{xx} + u^p$. This technique which was first applied to solve hyperbolic systems combines rescaling and mesh refinement in order to be able to keep accurate results over the entire physical interval up to very large magnitude of the solution. The solution is stepped forward until its maximum value reaches some predefined value. At that time the solution is rescaled, using the scale invariance of the equation, to make it small again and extra grid points are added in order to avoid the loss of accuracy caused by the stretch of the spatial variable. In this article, the method is presented in detail and some conjectures concerning the asymptotic shape are developed.

Stuart and Floater [146] used the same approach as Sanz-Serna and Verwer, that is the study of discretization of simple ODEs, to generate an important paper concerning the effect of time-discretization published in 1990. By studying the blow-up problem for ODEs, the authors evaluated various time-stepping strategies for partial differential equations that develop singularities in finite time. Using a natural continuous embedding of the numerical method, they explained why fixed-step methods are not a suitable choice and proposed a class of variable time-stepping strategies: they introduced a new time-like variable in order to transform the blow-up time to infinity and then showed that if the rescaling function is adequately defined the numerical

blow-up time converges to the true one. They illustrated their work by studying, theoretically and numerically, the parabolic equation $x^q u_t = u_{xx} + f(u)$ on $(0, 1)$ with Dirichlet boundary conditions.

Other authors pointed out the care required, in particular for the time-discretization, when trying to reproduce the blow-up numerically. Stewart and Geveci [143] used spectral and pseudospectral methods coupled with the well-known variable-stepsize Runge-Kutta method (RK45) in order to detect the blow-up of a nonlinear evolution equation involving a Hilbert transform, however this approach did not perform well and illustrated the fact that standard schemes are not well suited for solving their kind of problems and that special schemes have to be derived. Tourigny and Sanz-Serna [153] focused on the radial cubic nonlinear Schrödinger equation and showed that different discretizations lead to different rates of growth for the blow-up, which are thus irrelevant. They presented a procedure involving least squares fitting that ensures reliable conclusions concerning the growth rate of the blow-up, however their results are valid only asymptotically as the blow-up time is approached.

New equations, more and more complex, were studied. In 1992, Bona et al [25] suggested a complex algorithm to study the formation of singularity for the Korteweg-de Vries-Burges equations $u_t + u^p u_x - \delta u_{xx} + \epsilon u_{xxx} = 0$ with periodic initial values. Their method consists of a Galerkin finite element in space and an implicit Runge-Kutta method in time and it involves a local refinement of the spatial grid, a local selection of a temporal mesh size and a spatial translation of the solution. The proofs of the convergence results stated were only outlined.

We mentioned earlier that Stuart and Floater [146] explained why standard fixed-step methods were inappropriate when solving ODEs with blow-up solutions and indeed most of the numerical schemes developed to reproduce blow-up solutions numerically were based on variable time-steps governed by the amplitude of the solution. However it is possible to construct specific fixed-step methods designed to reproduce

the blow-up numerically. The first to explore this path was Le Roux in 1994 [99]. She considered the porous media equation

$$u_t = \Delta u^m + \alpha u^m,$$

with $m > 1$ on a bounded domain with Dirichlet boundary conditions. Using the exact solution of the differential equation $y_t = \alpha y^m$, she derived an implicit semi-discretization in time. She then proved existence and uniqueness of the solution of this scheme, showed that this solution has the same properties as the exact solution (i.e. blows up if α is larger than the first eigenvalue λ_1 of the Dirichlet problem $-\Delta v = \lambda v$) and finally she proved the convergence of the solution.

As our theoretical study (Chapter 4) was largely inspired by this article, we present shortly the strategy used by Le Roux to prove that the numerical solution blows up in finite time and define the numerical blow-up time. First we introduce the function $v = u^m$ and the constants $p = 1/m$ and $q = 1 - p$. The timestep is denoted by h. As a first step, Le Roux proves that the scheme has a unique positive solution:

Theorem 2.2 (Le Roux)**.** *If v_n satisfies*

$$\|v_n\|_\infty < K(h) := \left(\frac{p}{qh\alpha}\right)^{1/q}, \tag{2.10}$$

the scheme

$$\frac{p}{q} v_{n+1} v_n^{-q} - \frac{p}{q} v_{n+1}^p + h(-\Delta v_{n+1} - \alpha v_{n+1}) = 0, \tag{2.11}$$

has a unique solution v_{n+1}. This solution is positive and belongs to $H_0^1(\Omega) \cap C^2(\bar{\Omega})$.

Moreover if (2.10) is satisfied, the function v_{n+1} can be characterized using the functional

$$J_n(v) = \int_\Omega |\nabla v|^2 \, dx + \int_\Omega \left(\frac{p}{qh} v_n^{-q} - \alpha\right) v^2 \, dx. \tag{2.12}$$

Indeed, if we denote by ψ_n the non-negative function belonging to the space E defined by $E := \{v \in H_0^1(\Omega) \text{ such that } \int v^{p+1} dx = 1\}$ that satisfies $J_n(\psi_n) = \min_{v \in E} J_n(v)$,

we have

$$v_{n+1} = \left(\frac{p}{qhJ_n(\psi_n)}\right)^{1/q} \psi_n.$$

This expression and the related functional

$$F(v) = \frac{\int_\Omega (|\nabla v|^2 - \alpha v^2) dx}{\|v\|_{p+1}^2} \tag{2.13}$$

lead in particular to the following inequality, that holds as long as (2.10) is satisfied,

$$\frac{q}{p} hF(v_{n+1}) \leq \|v_{n+1}\|_{p+1}^{-q} - \|v_n\|_{p+1}^{-q} \leq \frac{q}{p} hF(v_n).$$

This implies, in case $\alpha > \lambda_1$ and $F(v_0) < 0$,

$$\|v_{n+1}\|_{p+1}^{-q} \leq \|v_0\|_{p+1}^{-q} + \frac{q}{p} nhF(v_0).$$

The right-hand side of this inequality is negative if

$$t_n := nh > T_{\max} := \frac{p}{q} \frac{\|v_0\|_{p+1}^{-q}}{(-F(v_0))},$$

hence there must be $t_{\tilde{n}} < T_{\max}$ for which (2.10) is not satisfied anymore, in other words we would have

$$\|v_{\tilde{n}}\|_\infty \geq K(h).$$

As $K(h)$ can be made as large as desired by decreasing h, Le Roux refers to that time $t_{\tilde{n}}$ as the numerical blow-up time and says that the numerical solution becomes infinite at $t_{\tilde{n}}$. Note that the numerical solution should not be computed further as the numerical result may become irrelevant, since there are no theorems concerning the case where the condition (2.10) is not satisfied. Hence the scheme may have no solution at all or it may have one or more solutions, and these would not necessarily be positive.

This definition of numerical blow-up time is slighlty different from the one we introduced in Chapter 1 as in this case the numerical blow-up time $t_{\tilde{n}}$ depends on h and K (it actually corresponds to what we called T_K^*), however as $T_{\max}$ only depends

on v_0, this result proves the existence of a numerical blow-up in the sense defined in Chapter 1. The advantage of Le Roux's analysis is not only the fact that $T_{\max}$ does not depend on h nor K, but also the fact that for a fixed timestep h, we are sure to reach the bound $K(h)$ before $T_{\max}$. This makes the result particularly interesting, however the technique to obtain this result relies on the fact that the solution has a variational characterization (2.12). As such a characterization was not available for most of the schemes we analyze in Chapter 4, we defined the numerical blow-up time differently.

We will see later that scheme (2.11) can be obtained by applying one of the constructions suggested in Chapter 3, so more details about the results proved in Le Roux's paper are presented in Section 4.4.1. This scheme and the corresponding results were extended to the more general case $u_t = \Delta u^m + \alpha u^p$, for $m > 0$ and $p \geq m$ in Le Roux's subsequent articles [100, 101] and a paper co-authored with Mainge [102]. This time discretization developed in 1999 was recently combined by Mainge with a suitable finite-dimensional space to derive a full discretization for the Cauchy problem for the fast-diffusion equation $u_t = (u^m)_{xx} + \alpha u^p$, with $m \in (0, 1)$, $\alpha > 0$ and $p > 1$ [114]. Recently also, M-N. Le Roux constructed with A-Y. Le Roux [97, 98] a full discretization involving variable timestep for the Cauchy problem of $u_t = (u^m u_x)_x + u^p$, with $m > 0$ and $p \geq m + 1$.

Tourigny and Grinfeld [152] presented a new approach based on discrete methods employed in the complex domain: they discretized the governing equation and "timestepped" in the complex domain. Their approach combines classical discretization (Runge-Kutta methods) and methods of Taylor series (using Lyness's algorithm to compute approximate Taylor coefficients using Fast Fourier Transform). It is first explained on the scalar Cauchy problem $x_t = f(x, t)$, then illustrated using several examples, including the semilinear parabolic problem $u_t = u_{xx} + u^2$ on $(0, 1)$ with Dirichlet boundary conditions.

In [12], Bandle and Brunner discussed the choice of the time-step sequence when the discretization in time is made by collocation using piecewise linear functions and standard finite difference are used in space, for the semilinear parabolic problem $u_t = \Delta u + f(u)$ in a bounded domain with Dirichlet boundary conditions. The criterion for the size of the time-step is a function of the collocation parameters so that the implicit collocation schemes are well-defined and can be used for computing the blow-up. Error estimates for the solutions are given but not for the blow-up time. Their work can be generalized to more general second-order elliptic operators.

A natural approach to tackle the numerical reproduction of a blow-up is to use moving-mesh methods. It is only in 1996 with the papers of Budd, Huang and Russell [30] and of Budd, Chen, Huang and Russell [27] that this technique was analyzed in the context of blow-up solution. Their work is based on the moving mesh partial differential equations developed by Huang, Ren and Rusell in 1994 [76, 75], it uses the scaling invariance of the solution and involves a spatial mesh that is modified as time goes forward. They presented some analysis concerning the semilinear parabolic equation $u_t = u_{xx} + f(u)$ in one dimension. This approach has been further developed for more complicated equations in [29, 28] and more recently in [74, 140].

Also in 1996, Abia, López-Marcos and Martínez studied a semi-discretization in space based on a uniform mesh for the semilinear parabolic equation in one dimension [1]. This article, and the following [2], focus on the approximation of the blow-up time. The authors gave conditions for the semi-discrete solution to blow up and bounds on the blow-up time. They proved the convergence of the blow-up time of the semi-discrete problem to the theoretical one. The main contribution of the second article is that a strong hypothesis (that the solution achieves blow-up in some $L^q(0,1)$ norm, with $1 \leq q < \infty$) is removed under some conditions on f. Groisman and Rossi [70] extended later this study by studying the blow-up rate and the blow-up set in the special case $f(u) = u^p$. In the continuity of [1, 2], Abia et al completed this

semi-discretization using the forward Euler method for time discretization which, in some sense, comes down to generalizing Nakagawa and Chen's results.

In 1997, Meyer-Spasche and Düchs [124] emphasized the relevance of nonstandard difference schemes applied to ODEs with blow-up solutions. Indeed in the previous studies of nonstandard schemes, see for example Mickens [125] and Agarwal [6], the emphasis was put on avoiding numerical instabilities but the correct reproduction of a blow-up was not explicitly set out as an asset of such schemes. In their paper, Meyer-Spasche and Düchs studied the relation between several examples of nonstandard schemes taken from the literature and the linearized trapezoidal rule, a time-centered scheme coupled with the first step of a Newton iteration, that actually is a Rosenbrock-type scheme. They showed that on their first example, $\dot{u} = \lambda u^2$, their scheme is exact and on their second example, the logistic equation, they showed that the discrete solution exhibits a finite-time blow-up in case the solution of the continuous problem blows up in finite time and they compared the continuous and the discrete blow-up times. The linearized trapezoidal rule (Lintrap) was studied further in a subsequent article by Meyer-Spasche [121]. It is shown on an example that even if the discrete blow-up occurs, the rate of growth is not correctly reproduced, so that the scheme is not suitable. Even if in special cases the a priori bound for the blow-up time is good, it is not necessarily true in most cases. In this article and the following ones [121, 122, 123], Meyer-Spasche also gave examples of situations where exact schemes for simpler equations led to efficient schemes for more complicated equations. One of the examples presented is the scheme developed by Le Roux [99]: as we already mentioned, she started from the exact scheme for $y_t = \alpha y^m$ to construct her specific scheme for the nonlinear equation $u_t = \Delta u^m + \alpha u^m$.

It is in 1998 that the first (and only) survey about the numerical study of blow-ups was published. Indeed the survey by Bandle and Brunner [13] that we already mentioned in Section 2.1 contains an excellent review of the beginning of the numerical

study of parabolic equations of type $u_t = \Delta u + f(x, t, u, \nabla u)$. In their introduction they enumerated the different methods used for spatial discretization and emphasized the importance of the choice of the time integrator. A more detailed review of timestepping strategies is given at the end of the paper (Section 6). In particular several of the papers cited above are referred to in more detail.

A new and quite different approach to prove the convergence of the numerical blow-up time to the original equation's blow-up time in case of a semi-discretization in space was presented by Ushijima [157] in 2000. His approach is based on functionals. He assumed there exists a functional J such that $J[u](t)$ or $\frac{d}{dt}J[u](t)$ tends to infinity as t tends to the blow-up time and a corresponding discrete functional. Ushijima showed that if the semi-discrete solution u_h converges to the solution in the sense of functional then, under certain assumptions on J and J_h, the numerical solution u_h blows up in finite time T_h that converges to the exact blow-up time. He illustrated his theory by applying it to different problems, including the semilinear parabolic equation $u_t = \Delta u + f(u)$.

As one can see, the subject really took off during the nineties. In the new decade, the most commonly used semi-discretization in space remained the piecewise linear finite element (which coincides with the classical central finite difference second order scheme in one dimension) with mass lumping, that was already used in the pioneering work of Nakagawa and Ushijima [129]. Duran, Etchevery and Rossi [42] applied it to the heat equation with nonlinear flux condition and their work was generalized by Acosta and al [5]. It was also applied to a system of semilinear heat equations [69] and used to study the blow-up sets of the heat equation with nonlinear boundary conditions [46]. However Ferreira, Groisman and Rossi explained the limits of this uniform mesh when it comes to reproducing the numerical blow-up sets or rate of growth [47] and suggested two ways to adapt the spatial mesh, either by adding mesh points or by moving mesh points [48].

Most of the fully-discrete schemes suggested also used the same spatial-discretization as the one mentioned above, with or without mesh refinement, and they involved a time-stepping strategy: some used a control of the time increment [77, 3], others used a rescaling in time ([39] for a nonlinear Schrödinger equation, and [4] for a heat equation with nonlinear flux boundary conditions).

Finally, very few of the authors suggested fixed time-step discretizations. Actually among those who did, we mostly find people who took up the time discretization suggested by Le Roux [99]. Barro et al [16] modified it for the reaction-diffusion equation $u_t - \Delta u^{1+\delta} + \gamma \overrightarrow{V} \cdot \nabla u^{1+\delta} = \alpha u^p$ and combined it with a finite difference in space to produce a numerical simulation. In [43] Duvnjak and Eberl studied a reaction-diffusion equation arising in biofilm modelling $u_t = \Delta \Phi(u) + ku$, where $\Phi(u) = \int_0^u \frac{s^b}{(1-s)^a} ds$, $a, b \geq 1$. By applying the change of variables $v = \Phi(u)$, the non-standard diffusion effects were removed from the spatial operator: v satisfies $\beta(v) = \Delta v + k\beta(v)$, where $\beta = \Phi^{-1}$. This approach can be related to Le Roux's work [99]; as in this paper, Duvnjak and Eberl applied the implicit Euler discretization which leads to a scheme equivalent to Le Roux's. The authors then followed her way to study the properties of their scheme.

Chapter 3: Construction of Specialized Methods

In this chapter, we explain how to construct B-methods. For this purpose, we chose to use the semilinear parabolic equation which has been widely studied (see Chapters 1 and 2). We explain in detail the two types of construction of schemes for this problem. The idea is then applied to several other examples and a section devoted to numerical experiments concludes the chapter.

3.1 Methods of construction of schemes

The construction of the methods will be illustrated using the semilinear parabolic problem presented in Chapter 1

$$\begin{cases} u_t &= \Delta u + \delta F(u), & \text{for } (x,t) \in \Omega \times (0,T), \\ u &= 0, & \text{for } (x,t) \in \partial\Omega \times (0,T), \\ u(x,0) &= u_0(x), & \text{for } x \in \Omega, \end{cases} \tag{3.1}$$

where δ is a positive constant, Ω is a bounded domain of $\mathbb{R}^d$ and u_0 is a positive continuous function on $\bar{\Omega}$. We recall that the function F is supposed to be positive, strictly increasing and strictly convex on $(0,\infty)$, to belong to $C^2([0,\infty))$ and to satisfy

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

for some finite b, so that the function

$$g(s) = \int_s^\infty \frac{1}{F(\sigma)} d\sigma$$

is well-defined on $(0, \infty)$. We have $\lim_{s\to\infty} g(s) = 0$ and we denote $M = \lim_{s\to 0} g(s)$, so that if we allow $F(0) = 0$, then M can be finite or infinite. Moreover, since F is continous and positive on $(0, \infty)$, g is continuous and strictly decreasing on $\mathbb{R}_+^*$. Hence g is invertible on $\mathbb{R}_+^*$ and $G = g^{-1}$ is defined on $(0, M)$ and satisfies

$$\lim_{s\to 0} G(s) = \infty, \quad \text{and} \quad \lim_{s\to M} G(s) = 0.$$

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

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

In problem (3.1) 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

$$y(t) = y(t, K) = G(K - \delta t), \tag{3.2}$$

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 two different ways to obtain semi-discretizations in time for the semilinear problem (3.1) from this exact solution. We obtain two different types of B-methods.

3.1.1 Splitting Methods

> *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.*
> Hairer, Lubich, Wanner (2002) [72]

As suggested in Hairer, Wanner and Lubich [72], one way to exploit the exact solution of the nonlinear part of the equation is by using splitting methods. If we decompose $u_t = \Delta u + \delta F(u)$ into

$$f^{[1]}(u) = \delta F(u) \quad \text{and} \quad f^{[2]}(u) = \Delta u,$$

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 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 (3.2), we have

$$\varphi_t^{[1]}(u_n) = G(g(u_n) - \delta t), \text{ for } t < g(u_n)/\delta.$$

Note that the notation φ_t represents a map, not a time derivative. Then we can choose any numerical integrator $\Phi_h^{[2]}$ for $u_t = \Delta u$, and by composing the exact flow and the numerical integrator, we obtain two new methods for $u_t = \Delta u + \delta F(u)$,

$$\Phi_h = \varphi_h^{[1]} \circ \Phi_h^{[2]} \qquad \text{and} \qquad \Phi_h^* = \Phi_h^{[2]*} \circ \varphi_h^{[1]}, \tag{3.3}$$

where $\Phi_h^{[2]*}$ is the adjoint of $\Phi_h^{[2]}$ (see Section II.3 in [72]). As 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 $p \geq 2$, and $\varphi_t^{[1]}$ is the exact flow of $u_t = \delta F(u)$, so that its Taylor expansion is

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

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 [72]). 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 $u_t = \Delta u$. For example, even though this problem is stiff, we start with forward Euler

$$\Phi_h^{[2]}(u_n) = u_n + h\Delta u_n,$$

whose adjoint is backward Euler

$$\Phi_h^{[2]*}(u_n) = u_n + h\Delta u_{n+1}.$$

By composing these integrators with the exact flow $\varphi_t^{[1]}$, we get two B-methods. The first one is

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

which gives the explicit scheme

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

and requires the condition $g(u_n + h\Delta u_n) \in (0, M)$, and the second method is

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

which gives the implicit scheme

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

and requires the condition $g(u_n) - \delta h \in (0, M)$. This scheme is studied in detail in Section 4.2. We derive more schemes for problem (3.1) in Section 3.2.1.

Local truncation error. We mentioned above that the B-methods obtained using this construction are of first-order if the standard methods used in the construction are consistent. 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.

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

$$\varphi_t(t, v) = F(\varphi(t, v)), \quad \text{and} \quad \varphi(0, v) = v, \quad \forall v. \tag{3.4}$$

Keeping the notation introduced at the beginning of the chapter, 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 $v(t)$ is this simplified problem, we have

$$\Phi(h, v_0) = v(h) + E(h), \tag{3.5}$$

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: 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. We have

$$u_1'(h) = \varphi_t + \varphi_v \left(v'(h) + E'(h)\right),$$

and

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

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

From the definition of φ given in (3.4) (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

$$u(h) = u_0 + h(\Upsilon(u_0) + F(u_0)) + \frac{h^2}{2}(\Upsilon'(u_0) + F'(u_0))(\Upsilon(u_0) + F(u_0)) + \cdots, \quad (3.6)$$

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

$$\tau_B := u_1 - u(h) = \frac{h^2}{2}\left(F'(u_0)\Upsilon(u_0) - \Upsilon'(u_0)F(u_0) + E''(0)\right) + O(h^3), \quad (3.7)$$

if a first-order standard method is used, while it becomes

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

if a higher-order standard method is used.

Before comparing these results with the local truncation error of standard methods, let us derive the error of the adjoint of these B-methods in order to verify that we obtain

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

as is expected by definition of adjoint methods.

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 (3.4), and we compute $u_1 = \Phi(h, v_0)$, where Φ is defined by (3.5) (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

$$\Phi(0,\xi) = \xi + E(0), \quad \Phi_t(0,\xi) = \Upsilon(\xi) + E'(0), \quad \Phi_{tt}(0,\xi) = \Upsilon'(\xi)\Upsilon(\xi) + E''(0), \tag{3.8}$$

and

$$\Phi_v(0,\xi) = 1, \quad \Phi_{vv}(0,\xi) = 0, \quad \text{and} \quad \Phi_{tv}(0,\xi) = \Upsilon'(\xi). \tag{3.9}$$

We now expand

$$u_1 = \Phi(h, \varphi(h, u_0))$$

in a series of h. We have $u_1(0) = \Phi(0, \varphi(0, u_0)) = \varphi(0, u_0) = u_0$, and then

$$u_1'(h) = \Phi_t(h, \varphi(h, u_0)) + \Phi_v(h, \varphi(h, u_0)) \cdot \varphi_t(h, u_0),$$

and

$$u_1''(h) = \Phi_{tt}(h,\varphi) + 2\Phi_{tv}(h,\varphi)\varphi_t(h,\varphi) + \Phi_{vv}(h,\varphi)\varphi_t(h,\varphi)^2 + \Phi_v(h,\varphi)\varphi_{tt}(h,\varphi).$$

Using the properties of Φ stated in (3.8) and (3.9) and the definition of φ given in (3.4), the derivatives of u evaluated at $h = 0$ become

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

and

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

As the Taylor expansion of the exact solution u is given by (3.6), the local truncation error of these B-methods are, as expected,

$$\tau_{B^*} = \frac{h^2}{2}\left(\Upsilon'(u_0)F(u_0) + E''(0) - F'(u_0)\Upsilon(u_0)\right) + O(h^3), \tag{3.10}$$

in case a first-order standard method is used, while they become

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

if a higher-order standard method is used.

We now need to show that in case of finite-time blow-up, the local truncation error of B-methods is smaller than the one of the corresponding standard methods. We can not get a general result but we will consider the two most used first-order methods: forward and backward Euler.

The local truncation errors of the forward and backward Euler methods applied to the general equation $y_t = f(t, y)$ are given by

$$\tau := y_1 - y(h) = \mp\frac{h^2}{2}(f_t + f_y f) + O(h^3), \tag{3.11}$$

respectively, which means that if we apply these methods to $u_t = F(u) + \Upsilon(u)$, we obtain

$$\tau_s = \mp\frac{h^2}{2}(\Upsilon'(u_0) + F'(u_0))(\Upsilon(u_0) + F(u_0)) + O(h^3). \tag{3.12}$$

On the other hand, if we apply forward or backward Euler to $v_t = \Upsilon(v)$, we obtain respectively

$$E(h) = \mp\frac{h^2}{2}\left[\Upsilon'(v_0)\Upsilon(v_0)\right] + O(h^3),$$

which gives $E''(0) = \mp\Upsilon'(v_0)\Upsilon(v_0)$. Going back to (3.7) and (3.10) we obtain the truncation error of the corresponding B-methods,

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

and

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

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. Let's 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 errors can then be written as

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

for the standard methods and

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

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

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

$$\tau_s = \mp\frac{h^2}{2}\left(\Delta(\Delta u_0 + 3e^{u_0}) + 3e^{u_0}\Delta u_0 + 9e^{2u_0}\right) + O(h^3),$$

and

$$\tau_B = \pm\frac{h^2}{2}\left(3e^{u_0}\Delta u_0 - \Delta(3e^{u_0}) \mp \Delta(\Delta u_0)\right) + O(h^3).$$

In this case also the term e^{2u_0} of τ_s is absent from τ_B, however 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

$$\|\Delta(\Delta u_0)\|_2 = 342,439, \quad \|\Delta(3e^{u_0})\|_2 = 1,466,377, \quad \|3e^{u_0}(\Delta u_0)\|_2 = 1,542,768,$$

and

$$\|(3e^{u_0})^2\|_2 = 16,544,121.$$

So we see that removing this last term from the local error might greatly improve the results.

3.1.2 Variation of the Constant

A more original approach to construct schemes for $u_t = \Delta u + \delta F(u)$ from the exact solution of the nonlinear part is to apply an idea of variation of the constant to the exact solution. The solution of $y' = \delta F(y)$ being $y(t) = G(K - \delta t)$, we introduce the variation of the constant $K = K(x,t)$ and we look for a solution in the form

$$u(x,t) = G(K(x,t) - \delta t), \tag{3.13}$$

which is possible since G is onto. Using the fact that $G'(s) = 1/(g'(G(s))) = -F(G(s))$, we obtain

$$u_t(x,t) = G'(K - \delta t)(K_t - \delta) = -F(G(K - \delta t))K_t + \delta F(G(K - \delta t)),$$

so that u is a solution of (3.1) if

$$\Delta u + \delta F(u) = -F(G(K - \delta t))K_t + \delta F(u),$$

that is

$$\Delta u = -F(G(K - \delta t))K_t.$$

Hence we get a differential equation for K

$$K_t(x,t) = \frac{-1}{F(G(K - \delta t))}\Delta(G(K - \delta t)). \tag{3.14}$$

To solve this differential equation, one can use several methods, each leading to a different semi-dicretization in time for the original partial differential equation (3.1). Note also that from (3.13), we get

$$K(x,t) = g(u(x,t)) + \delta t. \tag{3.15}$$

As first example, we solve the differential equation (3.14) using the backward Euler method with timestep h to get

$$K_{n+1} - K_n = \frac{-h}{F(G(K_{n+1} - \delta t_{n+1}))} \Delta(G(K_{n+1} - \delta t_{n+1})),$$

(where $K_n = K_n(x)$). Then, introducing the definition of K given in (3.15), we obtain

$$g(u_{n+1}) - g(u_n) + \delta h = \frac{-h}{F(u_{n+1})} \Delta u_{n+1},$$

where u_n and u_{n+1} are functions of x. This scheme is studied in detail in Section 4.1.

Consistency of the methods and local truncation error. We need to show that the B-methods constructed using this approach are consistent and (as for the B-methods obtained using splitting methods) we want to show that they have the potential to be better than standard methods by comparing the local truncation errors of both types of methods.

As in the previous section, we consider the general problem $u_t = F(u) + \Upsilon(u)$, where Υ can be a function or an operator. The function φ satisfies $\varphi_t = F(\varphi)$, that is, with the notation introduced at the beginning of the chapter, $\varphi(t, K) = G(K - t)$. We then consider $u(x, t) = \varphi(t, K(t))$ so that

$$u_t = \varphi_t + \varphi_K K' = F(\varphi) + \Upsilon(\varphi), \tag{3.16}$$

and thus K must satisfy

$$K' = \frac{1}{\varphi_K} \Upsilon(\varphi). \tag{3.17}$$

To construct B-methods by variation of the constant, we define K_0 so that $u_0 = \varphi(0, K_0)$ and then we apply a standard method to equation (3.17) to obtain K_1, so that we have

$$K_1 = K(h) + E(h),$$

where E represents the local truncation error of the standard method. Finally we define

$$u_1 = \varphi(h, K_1) = \varphi(h, K(h) + E(h)).$$

To consider the local truncation error of the resulting B-method, we need to expand u_1 in a series in h:

$$u_1(h) = u_1(0) + u_1'(0)h + u_1''(0)\frac{h^2}{2} + \cdots .$$

Differentiating $u_1 = \varphi(h, K(h) + E(h))$ with respect to t we obtain

$$u_1'(h) = \varphi_t(h, K(h) + E(h)) + \varphi_K(h, K(h) + E(h)) \cdot (K' + E'(h)).$$

We saw in (3.16) that the exact solution u satisfies

$$u_t(h) = \varphi_t(h, K) + \varphi_K(h, K) \cdot K'.$$

In other words, the derivatives coincide except that K and its derivatives are replaced by $(K + E)$ and the corresponding derivatives. For example the second derivative is

$$\begin{aligned} u_1''(h) = {} & \varphi_{tt}(h, K + E) + 2\varphi_{tK}(h, K + E) \cdot (K' + E') \\ & + \varphi_{KK}(h, K + E) \cdot (K' + E')^2 + \varphi_K(h, K + E) \cdot (K'' + E''), \end{aligned}$$

while

$$\begin{aligned} u_{tt}(h) &= \varphi_{tt}(h, K) + 2\varphi_{tK}(h, K) \cdot K' + \varphi_{KK}(h, K) \cdot K'^2 + \varphi_K(h, K) \cdot K'' \\ &= [F'(\varphi) + \Upsilon'(\varphi)][F(\varphi) + \Upsilon(\varphi)]. \end{aligned}$$

Moreover we observe that in the p-th derivative of u_1, the highest-order derivative of E, which is $E^{(p)}$, appears only once, in the term $\varphi_K(h, K + E) \cdot E^{(p)}$.

Using these observations, we obtain that if the standard method is of order p (that is $E(h) = O(h^{p+1})$ so that $E(0) = E'(0) = \cdots = E^{(p)}(0) = 0$), the local truncation error of the resulting B-method is given by

$$\tau_B = \frac{h^{p+1}}{(p+1)!}\varphi_K(0, K_0)E^{(p+1)}(0) + O(h^{p+2}). \tag{3.18}$$

So the B-method is of same order as the original standard method, and in particular it is consistent. In order for the B-method to be more accurate than the standard

method, the local truncation error of the former needs to be smaller than the error of the latter. From (3.18) we can not obtain a general result, however we illustrate below the cases of forward and backward Euler, as we did for the splitting B-methods.

First, we rewrite (3.17) using $\varphi(t,K) = G(K-t)$ and $G' = -F(G)$, that is $\varphi_K = -F(\varphi)$, to get

$$K'(t) = \frac{-1}{F(\varphi(t,K))}\Upsilon(\varphi(t,K)),$$

and since the local truncation errors of forward and backward Euler are given by (3.11), we obtain in this case

$$\begin{aligned} E(h) &= \mp\frac{h^2}{2}\left[\frac{F'(\varphi)}{F(\varphi)^2}\Upsilon(\varphi) - \frac{1}{F(\varphi)}\Upsilon'(\varphi)\right][\varphi_t + \varphi_K K'] + O(h^3) \\ &= \mp\frac{h^2}{2}\left[\frac{F'(\varphi)}{F(\varphi)^2}\Upsilon(\varphi) - \frac{1}{F(\varphi)}\Upsilon'(\varphi)\right][F(\varphi) + \Upsilon(\varphi)] + O(h^3). \end{aligned}$$

Replacing $E''(0)$ in (3.18) we obtain the truncation error of the corresponding B-methods,

$$\tau_B = \pm\frac{h^2}{2}\left[\frac{F'(u_0)}{F(u_0)}\Upsilon(u_0) - \Upsilon'(u_0)\right][F(u_0) + \Upsilon(u_0)] + O(h^3).$$

Comparing τ_B with τ_s given in (3.12), we see that the term $F'(u_0)[F(u_0)+\Upsilon(u_0)]$ in τ_s is replaced by $-\frac{F'(u_0)}{F(u_0)}\Upsilon(u_0)[F(u_0)+\Upsilon(u_0)]$ in τ_B.

As for splitting B-methods, we first consider the case $F(u) = e^u$ and $\Upsilon(u) = \sin(u)$. We have

$$\begin{aligned} \tau_B &= \pm\frac{h^2}{2}\left[\frac{e^{u_0}}{e^{u_0}}\sin(u_0) - \cos(u_0)\right][e^{u_0} + \sin(u_0)] + O(h^3) \\ &= \pm\frac{h^2}{2}\left[e^{u_0}(\sin(u_0) - \cos(u_0)) + \sin^2(u_0) - \cos(u_0)\sin(u_0)\right] + O(h^3), \end{aligned}$$

while

$$\tau_s = \mp\frac{h^2}{2}\left[e^{2u_0} + e^{u_0}(\sin(u_0) + \cos(u_0)) + \cos(u_0)\sin(u_0)\right] + O(h^3).$$

We see that the highly-weighted term e^{2u_0} that appears in τ_s is replaced in τ_B by $\sin^2(u_0)$ which remains bounded by 1 for any u_0, and thus τ_B should be much smaller than τ_s.

We then consider the case $\Upsilon(u) = \Delta u$ and $F(u) = 3e^u$. The local truncation errors are

$$\tau_s = \mp\frac{h^2}{2}\left(\Delta(\Delta u_0 + 3e^{u_0}) + 3e^{u_0}(\Delta u_0 + 3e^{u_0})\right) + O(h^3),$$

and

$$\tau_B = \pm\frac{h^2}{2}\left[\Delta u_0(\Delta u_0 + 3e^{u_0}) - \Delta\left(\Delta u_0 + 3e^{u_0}\right)\right] + O(h^3).$$

As these terms are not easily evaluated theoretically, we use again numerical experiments to compare them. We use the same example as in the previous section: the solution of $u_t = 3e^u + \Delta u$ on $[-1, 1]$, with $u_0(x) = \cos(\pi x/2)$ is computed using the apdaptive method ode45 of Matlab and we evaluate the norm of the different terms of τ_s and τ_B at $= 0.1660$. The common term

$$\|\Delta\left(\Delta u_0 + 3e^{u_0}\right)\|_2 = 1,145,556,$$

is of same order than the remaining terms of τ_B

$$\|\Delta u_0(\Delta u_0 + 3e^{u_0})\|_2 = 1,391,072,$$

while the second term of τ_s is considerably larger

$$\|3e^{u_0}(\Delta u_0 + 3e^{u_0})\|_2 = 15,062,542.$$

Hence we expect the error of the B-methods to be significantly smaller than the error of the corresponding standard methods.

3.2 More Equations, More Schemes

For many problems with blow-up solutions, the two types of construction presented in Section 3.1 lead to examples of B-methods. We chose several examples, among the most-studied ones, to illustrate further the derivation of B-methods. For each of these problems, we derive several B-methods of each type.

3.2.1 Semilinear Parabolic Equation

We already derived three B-methods for problem (3.1) in the previous section. We now present several other methods, of first and second order.

For the splitting methods, instead of choosing $\Phi_h^{[2]}$ to be forward Euler in (3.3), we could choose it to be backward Euler; then $\Phi_h^{[2]*}$ is forward Euler and the resulting schemes are

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

and

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

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

$$\Psi_h = \Phi_{h/2} \circ \Phi_{h/2}^* = \varphi_{h/2}^{[1]} \circ \Phi_{h/2}^{[2]} \circ \Phi_{h/2}^{[2]*} \circ \varphi_{h/2}^{[1]}, \tag{3.19}$$

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

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

and if $\Phi_h^{[2]}$ is chosen to be backward Euler, we get

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

Concerning the approach by variation of the constant, we obtain more schemes by applying different methods to solve the differential equation (3.14): using forward

Euler we would obtain the explicit scheme

$$g(u_{n+1}) = g(u_n) - \delta h - \frac{h}{F(u_n)}\Delta u_n;$$

using the trapezoidal rule, we get

$$g(u_{n+1}) - g(u_n) + \delta h + \frac{h}{2F(u_n)}\Delta u_n + \frac{h}{2F(u_{n+1})}\Delta u_{n+1} = 0,$$

and the midpoint rule leads to

$$[g(u_{n+1}) - g(u_n) + \delta h]F\left(G\left(\frac{g(u_n) + g(u_{n+1})}{2}\right)\right) + h\Delta\left(G\left(\frac{g(u_n) + g(u_{n+1})}{2}\right)\right) = 0.$$

More generally, if a general s-stage Runge-Kutta method given by Table 3.1 is applied,

c	$A = (a_{ij})$
	b^T

Table 3.1: s-stage Runge-Kutta method

we obtain

$$g(u_{n+1}) = g(u_n) - \delta h + \sum_{j=1}^{s} b_j k_j,$$

with, for $i = 1..s$,

$$k_i = \frac{-h}{F\left(G(g(u_n) + \sum_{j=1}^{s} a_{ij}(k_j - \delta h))\right)}\Delta\left(G\left(g(u_n) + \sum_{j=1}^{s} a_{ij}(k_j - \delta h)\right)\right).$$

Similar schemes can be obtained for the slightly more general equation (results concerning blow-up of the exact solutions of this equation can be found in [10], [118] and [119])

$$u_t = \Delta u + \delta q(x)\psi(t)F(u),$$

where q is bounded on $\bar{\Omega}$ with $q(x) > 0$ and ψ is continuous on $[0, \infty)$, with $\psi(t) > 0$. We also assume that the function F satisfies the conditions stated in Section 3.1 and we only consider the case where the function

$$\varphi(t) = \int_0^t \psi(s)\, ds,$$

can be explicitly computed so that the solution of $y_t = \delta q \psi(t) F(y)$ is

$$y(t) = G(g(s) - \delta q \varphi(h)).$$

Using this result, we notice that we can easily modify the above schemes for this new equation. In all schemes obtained using the splitting method, it is enough to replace each $g(s) - \delta h$ by $g(s) - \delta q(x)\varphi(h)$, and in the schemes obtained by variation of the constant, the term δh must be replaced by $\delta q(x)(\varphi(t_{n+1}) - \varphi(t_n))$. Moreover the scheme obtained using the midpoint rule becomes much more complicated,

$$\begin{array}{l} [g(u_{n+1}) - g(u_n) + \delta q \varphi(h)] \cdot \\ F\left(G\left(\frac{g(u_n)+g(u_{n+1})}{2} + \delta q \left[\frac{\varphi(t_n)+\varphi(t_{n+1})}{2} - \varphi\left(\frac{t_n+t_{n+1}}{2}\right)\right]\right)\right) \\ +\Delta\left(G\left(\frac{g(u_n)+g(u_{n+1})}{2} + \delta q \left[\frac{\varphi(t_n)+\varphi(t_{n+1})}{2} - \varphi\left(\frac{t_n+t_{n+1}}{2}\right)\right]\right)\right) = 0. \end{array}$$

Similarly, the schemes obtained by applying general Runge-Kutta methods would be quite complicated.

In some specific cases, it is possible to apply these methods for more general problems. We consider for example

$$u_t = \Delta u + m \int_0^t e^{u(x,\tau)}\, d\tau + g(x), \tag{3.20}$$

whose nonlinear part

$$y'(t) = m \int_0^t e^{y(\tau)}\, d\tau,$$

can be solved explicitly. Indeed the solution of this equation is given by

$$y(t) = 2 \ln\left[K \sec\left(\alpha t K\right)\right], \tag{3.21}$$

where K is the constant of integration and $\alpha = \sqrt{2m}/2$. If we apply the method of variation of the constant, two difficulties arise. The first one is due to the integral term: if we let $u(x,t) = 2\ln\left[K(x,t)\sec\left(\alpha t K(x,t)\right)\right]$, the term

$$m\int_0^t e^{u(x,\tau)}d\tau = m\int_0^t K(x,\tau)^2\sec^2(\alpha\tau K(x,\tau))d\tau,$$

is not equal to $2\alpha K(x,t)\tan(\alpha t K(x,t))$ therefore this simplification which is the main advantage of this method cannot be applied. The differential equation for K would be

$$K_t = \frac{K}{1+\alpha t K\tan(\alpha t K)}\Bigg[\Delta(\ln(K\sec(\alpha t K))) + \frac{g}{2} + \frac{m}{2}\int_0^t K^2\sec(\alpha\tau K)d\tau - \alpha K\tan(\alpha t K)\Bigg].$$

The second difficulty comes from the fact that it is not possible to invert formula (3.21) analytically to express K as a function of y. A numerical inversion would then be required. The interest of the resulting method is clearly weakened by these two complications. Yet, such problems do not arise when we apply the splitting method. The exact flow of the nonlinear part of equation (3.20) is

$$\varphi_t^{[1]}(u_n) = 2\ln\left[e^{u_n/2}\sec(\alpha t e^{u_n/2})\right],$$

so if we apply the forward Euler method to the linear part of the equation, we obtain two schemes

$$\Phi_h(u_n) = u_{n+1} = 2\ln\left[e^{(u_n+h\Delta u_n+hg)/2}\sec(\alpha h e^{(u_n+h\Delta u_n+hg)/2})\right],$$

and

$$\Phi_h^*(u_n) = u_{n+1} = 2\ln\left[e^{u_n/2}\sec(\alpha h e^{u_n/2})\right] + h\Delta u_{n+1} + hg.$$

If we use the backward Euler method, the resulting schemes are

$$\Phi_h(u_n) = 2\ln\left(e^{v/2}\sec(\alpha h e^{h/2})\right), \text{ with } v - h\Delta v = u_n + hg,$$

and

$$\Phi_h^*(u_n) = 2\ln\left(e^{u_n/2}\sec(\alpha h e^{u_n/2})\right) + 2h\Delta\ln\left(e^{u_n/2}\sec(\alpha h e^{u_n/2})\right) + hg.$$

One can compose these methods as in (3.19) to obtain second-order methods.

3.2.2 Quasilinear Parabolic Equation

As we said in Chapter 1, another model that has been deeply studied is the quasilinear equations of the type

$$u_t = \Delta\phi(u) + Q(u),$$

and more specifically, the case of power-type nonlinearities

$$u_t = \Delta u^{\sigma+1} + \alpha u^{\beta+1}, \tag{3.22}$$

with $\beta > 0$, $\sigma > 0$ and $\alpha \geq 0$. We now use the two constructions presented in Section 3.1 to derive B-methods for this problem.

To derive the schemes, the first step is to consider the nonlinear part

$$y_t = \alpha y^{\beta+1},$$

whose solution is given by

$$y(t) = \left(\frac{1}{K - \alpha\beta t}\right)^{1/\beta}. \tag{3.23}$$

We can now use this explicit solution to construct B-methods.

Splitting method. We recall that the first construction consists in splitting the equation (3.22) into $f^{[1]}(u) = \alpha u^{\beta+1}$ and $f^{[2]}(u) = \Delta u^{\sigma+1}$. The exact flow of the first part is derived from (3.23):

$$\varphi_t^{[1]}(u_n) = \left[u_n^{-\beta} - \alpha\beta t\right]^{-1/\beta}.$$

By choosing $\Phi_h^{[2]}$ to be the forward Euler method, so that $\Phi_h^{[2]*}$ is the backward Euler method we obtain

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

and its adjoint

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

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

$$\Phi_h(u_n) = \left[v^{-\beta} - \alpha\beta h\right]^{-1/\beta}, \text{ where } v \text{ is solution of } v - h\Delta(v^{\sigma+1}) = u_n,$$

and

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

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

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

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 given implicitly by

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

where

$$v = \left[u_n^{-\beta} - \frac{\alpha\beta h}{2}\right]^{-1/\beta} + \frac{h}{2}\Delta\left[\left(u_n^{-\beta} - \frac{\alpha\beta h}{2}\right)^{-(\sigma+1)/\beta}\right].$$

Variation of the constant. The second type of B-methods is obtained by introducing the variation of the constant $K = K(x,t)$ and let

$$u(x,t) = \left(\frac{1}{K(x,t) - \alpha\beta t}\right)^{1/\beta}.$$

Differentiating with respect to t and going back to (3.22), we obtain

$$\frac{-1}{\beta}\left(\frac{1}{K - \alpha\beta t}\right)^{\frac{1}{\beta}+1} K_t = \Delta u^{\sigma+1},$$

so that the differential equation for K is

$$K_t = -\beta\Delta\left(\frac{1}{K(x,t) - \alpha\beta t}\right)^{\frac{\sigma+1}{\beta}} (K - \alpha\beta t)^{\frac{\beta+1}{\beta}},$$

and we can express K as a function of u and t

$$K(x,t) = u(x,t)^{-\beta} + \alpha\beta t.$$

Solving the differential equation for K with different methods, we obtain the following schemes: using the forward Euler method,

$$u_{n+1}^{-\beta} u_n^{\beta+1} - u_n + h\beta[\alpha u_n^{\beta+1} + \Delta u_n^{\sigma+1}] = 0,$$

using the backward Euler method,

$$u_{n+1} - u_n^{-\beta} u_{n+1}^{\beta+1} + h\beta[\alpha u_{n+1}^{\beta+1} + \Delta u_{n+1}^{\sigma+1}] = 0,$$

using the trapezoidal rule,

$$u_{n+1}^{-\beta} - u_n^{-\beta} + \alpha\beta h + \frac{\beta h}{2}\left[v_n^{-(\beta+1)}\Delta u_n^{\sigma+1} + u_{n+1}^{-(\beta+1)}\Delta u_{n+1}^{\sigma+1}\right] = 0,$$

and using the midpoint rule,

$$u_{n+1}^{-\beta} - u_n^{-\beta} + \alpha\beta h + \beta h\left(\frac{u_n^{-\beta} + u_{n+1}^{-\beta}}{2}\right)^{\frac{\beta+1}{\beta}} \Delta\left[\left(\frac{u_n^{-\beta} + u_{n+1}^{-\beta}}{2}\right)^{-\frac{\sigma+1}{\beta}}\right] = 0.$$

More generally, if we consider the general s-stage Runge-Kutta method given by Table 3.1, we obtain

$$u_{n+1}^{-\beta} = u_n^{-\beta} - \alpha\beta h + \sum_{j=1}^{s} b_j k_j,$$

where for $i = 1..s$,

$$k_i = -h\beta \left(u_n^{-\beta} + \sum_{j=1}^{s} a_{ij}(k_j - \alpha\beta h) \right)^{\frac{\beta+1}{\beta}} \Delta \left(u_n^{-\beta} + \sum_{j=1}^{s} a_{ij}(k_j - \alpha\beta h) \right)^{-\frac{\sigma+1}{\beta}}.$$

3.2.3 Systems

In [52] and [51], 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 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 [19], Gang and Sleeman [60] and Chen [32]. In this section, we derive several specialized methods for the simple case

$$\begin{cases} u_t &= \Delta u + \delta e^v, \\ v_t &= \Delta v + \gamma e^u. \end{cases} \tag{3.24}$$

We first solve the nonlinear system of ordinary differential equations

$$\begin{cases} y'(t) &= \delta e^{z(t)}, \\ z'(t) &= \gamma e^{y(t)}, \end{cases} \tag{3.25}$$

to get

$$\begin{cases} y(t) &= \ln K - \ln[1 - \delta e^{Kt+D}] - \ln \gamma, \\ z(t) &= \ln K - \ln[1 - \delta e^{Kt+D}] + Kt + D, \end{cases} \tag{3.26}$$

where K and D are constants of integration.

Variation of the constant. To derive specialized methods using variation of the constants, we set

$$\begin{cases} u(x,t) &= \ln K(x,t) - \ln[1-\delta e^{K(x,t)t+D(x,t)}] - \ln\gamma, \\ v(x,t) &= \ln K(x,t) - \ln[1-\delta e^{K(x,t)t+D(x,t)}] + K(x,t)t + D(x,t), \end{cases} \tag{3.27}$$

and compute the derivatives

$$u_t = \frac{K_t}{K} + \frac{\delta e^{D+tK}(D_t + K + tK_t)}{1-\delta e^{d+tK}} = \frac{K_t}{K} + \frac{\delta e^{D+tK}(D_t + tK_t)}{1-\delta e^{d+tK}} + \delta e^v,$$

and

$$\begin{aligned} v_t &= \frac{K_t}{K} + \frac{\delta e^{D+tK}(D_t + K + tK_t)}{1-\delta e^{d+tK}} + D_t + K + tK_t, \\ &= \frac{K_t}{K} + \frac{\delta e^{D+tK}(D_t + tK_t)}{1-\delta e^{d+tK}} + D_t + tK_t + \frac{K}{1-\delta e^{d+tK}}, \\ &= \frac{K_t}{K} + \frac{\delta e^{D+tK}(D_t + tK_t)}{1-\delta e^{d+tK}} + D_t + tK_t + \gamma e^u. \end{aligned}$$

So for u and v to satisfy the system (3.24), we need

$$\begin{cases} \Delta u = \frac{K_t}{K} + \frac{\delta e^{Kt+D}}{1-\delta e^{Kt+D}}(tK_t + D_t), \\ \Delta v = \Delta u + tK_t + D_t, \end{cases}$$

which lead to the following system

$$\begin{cases} K_t &= \frac{K}{1-\delta e^{Kt+D}}\left(\Delta u - \delta e^{Kt+D}\Delta v\right), \\ D_t &= \Delta v - \Delta u - \frac{tK}{1-\delta e^{Kt+D}}\left(\Delta u - \delta e^{Kt+D}\Delta v\right), \end{cases} \tag{3.28}$$

where u and v are given by (3.27). We also need to invert the system (3.27) to obtain K and D as functions of u and v:

$$\begin{cases} K &= \gamma e^u - \delta e^v, \\ D &= v - u - \ln\gamma + \delta t e^v - \gamma t e^u. \end{cases}$$

Solving the system (3.28) using the forward Euler method and simplifying, we obtain

$$\gamma(e^{u_{n+1}} - e^{u_n}) - \delta(e^{v_{n+1}} - e^{v_n}) = h(\gamma e^{u_n}\Delta u_n - \delta e^{v_n}\Delta v_n),$$

and

$$(v_{n+1} - v_n + \delta t_{n+1} e^{v_{n+1}} - \delta t_n e^{v_n}) - (u_{n+1} - u_n + \gamma t_{n+1} e^{u_{n+1}} - \gamma t_n e^{u_n}) \\ = h(\Delta v_n - \Delta u_n) - h t_n (\gamma e^{u_n} \Delta u_n - \delta e^{v_n} \Delta v_n),$$

and solving it using the backward Euler method, we obtain

$$\gamma(e^{u_{n+1}} - e^{u_n}) - \delta(e^{v_{n+1}} - e^{v_n}) = h(\gamma e^{u_{n+1}} \Delta u_{n+1} - \delta e^{v_{n+1}} \Delta v_{n+1}),$$

and

$$(v_{n+1} - v_n + \delta t_{n+1} e^{v_{n+1}} - \delta t_n e^{v_n}) - (u_{n+1} - u_n + \gamma t_{n+1} e^{u_{n+1}} - \gamma t_n e^{u_n}) \\ = h(\Delta v_{n+1} - \Delta u_{n+1}) - h t_{n+1} (\gamma e^{u_{n+1}} \Delta u_{n+1} - \delta e^{v_{n+1}} \Delta v_{n+1}).$$

If we solve system (3.28) using the midpoint rule, the resulting scheme is quite complex to write. First we define

$$\tilde{K} = \frac{K_{n+1} + K_n}{2} = \frac{1}{2}(\gamma e^{u_{n+1}} - \delta e^{v_{n+1}}) + \frac{1}{2}(\gamma e^{u_n} - \delta e^{v_n}),$$

and

$$\tilde{D} = \frac{D_{n+1} + D_n}{2} = \frac{1}{2}(v_{n+1} - u_{n+1} - \ln \gamma + \delta t_{n+1} e^{v_{n+1}} - \gamma t_{n+1} e^{u_{n+1}}) \\ + \frac{1}{2}(v_n - u_n - \ln \gamma + \delta t_n e^{v_n} - \gamma t_n e^{u_n}),$$

and for simplicity

$$\tilde{E} = 1 - \delta \exp(\tilde{K}(t_n + h/2) + \tilde{D}).$$

The first equation can now be written as

$$(\gamma e^{u_{n+1}} - \delta e^{v_{n+1}}) - (\gamma e^{u_n} - \delta e^{v_n}) \\ = \frac{h\tilde{K}}{\tilde{E}} \left(\Delta \left(\ln \tilde{K} - \ln[\tilde{E}] \right) + (\tilde{E} - 1) \Delta \left(\ln \tilde{K} - \ln[\tilde{E}] + (\tilde{K}(t_n + h/2) + \tilde{D}) \right) \right),$$

and the second equation is

$$(v_{n+1} - u_{n+1} + \delta t_{n+1} e^{v_{n+1}} - \gamma t_{n+1} e^{u_{n+1}}) - (v_n - u_n + \delta t_n e^{v_n} - \gamma t_n e^{u_n}) \\ = h\Delta(\tilde{K}(t_n + h/2) + \tilde{D}) - (t_n + \frac{h}{2}) \frac{h\tilde{K}}{\tilde{E}} \left(\Delta \left(\ln \tilde{K} - \ln[\tilde{E}] \right) \right. \\ \left. + (\tilde{E} - 1) \Delta \left(\ln \tilde{K} - \ln[\tilde{E}] + (\tilde{K}(t_n + h/2) + \tilde{D}) \right) \right).$$

For the method obtained using the trapezoidal rule, we first define

$$E_1 = 1 - \delta e^{K_n t_n + D_n} = \frac{\gamma e^{u_n} - \delta e^{v_n}}{\gamma e^{u_n}} \quad \text{and} \quad E_2 = \frac{\gamma e^{u_{n+1}} - \delta e^{v_{n+1}}}{\gamma e^{u_{n+1}}},$$

and

$$K_1^t = \gamma e^{u_n}(\Delta u_n + (E_1 - 1)\Delta v_n) = \gamma e^{u_n}\Delta u_n - \delta e^{v_n}\Delta v_n,$$

$$K_2^t = \gamma e^{u_{n+1}}(\Delta u_{n+1} + (E_2 - 1)\Delta v_{n+1}) = \gamma e^{u_{n+1}}\Delta u_{n+1} - \delta e^{v_{n+1}}\Delta v_{n+1}.$$

The first part of the scheme can then be written as

$$(\gamma e^{u_{n+1}} - \delta e^{v_{n+1}}) - (\gamma e^{u_n} - \delta e^{v_n}) = \frac{h}{2}(K_1^t + K_2^t),$$

and the second

$$\begin{aligned}&(v_{n+1} - u_{n+1} + \delta t_{n+1} e^{v_{n+1}} - \gamma t_{n+1} e^{u_{n+1}}) - (v_n - u_n + \delta t_n e^{v_n} - \gamma t_n e^{u_n}) \\ &= \frac{h}{2}(\Delta(v_n - u_n) - t_n K_1^t - \Delta(v_{n+1} - u_{n+1}) + t_{n+1} K_2^t).\end{aligned}$$

Splitting method. To construct B-methods using the splitting method, we first note that the exact solution of the system of ODEs (3.25) is given by (3.26) with $K = \gamma e^{y(0)} - \delta e^{z(0)}$ and $D = z(0) - y(0) - \ln \gamma$. Then for each choice of numerical integrator $\Phi_h^{[2]}$ applied to

$$\begin{cases} u_t = \Delta u, \\ v_t = \Delta v, \end{cases}$$

we obtain two adjoint schemes. The forward Euler method leads to the explicit 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

$$\begin{cases} K_n = \gamma e^{u_n + h\Delta u_n} - \delta e^{v_n + h\Delta v_n}, \\ D_n = v_n + h\Delta v_n - u_n - h\Delta u_n - \ln \gamma, \end{cases}$$

and the implicit scheme Φ_h^* given by

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

where

$$\begin{cases} K_n = \gamma e^{u_n} - \delta e^{v_n}, \\ D_n = v_n - u_n - \ln\gamma. \end{cases} \tag{3.29}$$

If we choose instead the backward Euler method, we obtain

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

where

$$\begin{cases} K_n = \gamma e^{w_1} - \delta e^{w_2}, \\ D_n = w_2 - w_1 - \ln\gamma, \end{cases}$$

and w_1 and w_2 are solutions of

$$w_1 = u_n + h\Delta w_1 \text{ and } w_2 = v_n + h\Delta w_2.$$

For its adjoint method, we first define

$$\begin{cases} K_n = \gamma e^{u_n} - \delta e^{v_n}, \\ D_n = v_n - u_n - \ln\gamma, \end{cases}$$

and

$$\begin{cases} w_1 = \ln K_n - \ln[1 - \delta e^{D_n + hK_n}] - \ln\gamma, \\ w_2 = \ln K_n - \ln[1 - \delta e^{D_n + hK_n}] + D_n + hK_n. \end{cases}$$

Then the 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 specialized methods. For these, we first define K_n and D_n as in (3.29). 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

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

and we define

$$\begin{cases} \tilde{K} &= \gamma \exp(w_1 + \frac{h}{2}\Delta w_1) - \delta \exp(w_2 + \frac{h}{2}\Delta w_2), \\ \tilde{D} &= w_2 + \frac{h}{2}\Delta w_2 - w_1 - \frac{h}{2}\Delta w_1 - \ln\gamma, \end{cases}$$

to finally get

$$\begin{cases} u_{n+1} &= \ln \tilde{K} - \ln[1 - \delta e^{\tilde{D} + \frac{h}{2}\tilde{K}}] - \ln\gamma, \\ v_{n+1} &= \ln \tilde{K} - \ln[1 - \delta e^{\tilde{D} + \frac{h}{2}\tilde{K}}] + \tilde{D} + \frac{h}{2}\tilde{K}. \end{cases} \tag{3.30}$$

If we choose to use the backward Euler method as $\Phi_h^{[2]}$, we need to first define

$$\begin{cases} \tilde{u} &= \ln K_n - \ln[1 - \delta e^{D_n + \frac{h}{2}K_n}] - \ln\gamma, \\ \tilde{v} &= \ln K_n - \ln[1 - \delta e^{D_n + \frac{h}{2}K_n}] + D_n + \frac{h}{2}K_n, \end{cases}$$

then w_1 and w_2 are the solutions of

$$\begin{cases} w_1 - \frac{h}{2}\Delta w_1 &= \tilde{u} + \frac{h}{2}\Delta \tilde{u}, \\ w_2 - \frac{h}{2}\Delta w_2 &= \tilde{v} + \frac{h}{2}\Delta \tilde{v}, \end{cases}$$

and we define

$$\begin{cases} \tilde{K} &= \gamma \exp(w_1) - \delta \exp(w_2), \\ \tilde{D} &= w_2 - w_1 - \ln\gamma, \end{cases}$$

to finally get u_{n+1} and v_{n+1} by (3.30).

3.2.4 Wave Equation

As mentioned in Chapter 2, blow-up phenomena were first studied by Keller [87] for nonlinear wave equations of the form $u_{tt} = c^2\Delta u + f(u)$, in a space of dimension 1, 2

or 3. In [65], Glassey considered the same problem,

$$\begin{cases} u_{tt} &= \Delta u + \delta F(u), & \text{for } (x,t) \in \Omega \times (0,T), \\ u &= 0, & \text{for } (x,t) \in \partial\Omega \times (0,T), \\ u(x,0) &= u_0(x), & \text{for } x \in \Omega, \\ u_t(x,0) &= u_{t0}(x), & \text{for } x \in \Omega, \end{cases} \tag{3.31}$$

on a bounded domain. We present here the case where $F(u) = e^u$. The general solution of the ordinary differential equation

$$y''(t) = \delta e^{y(t)},$$

is

$$y(t) = \ln\left[\frac{2K^2}{\delta}(1 + \tan^2(D + tK))\right],$$

which we rewrite as

$$y(t) = \ln 2 - \ln\delta + 2\ln K - 2\ln[\cos(D + tK)].$$

Variation of the constant. First we apply the variation of the constant and look for a solution $u(x,t)$ of the form

$$u(x,t) = \ln 2 - \ln\delta + 2\ln K(x,t) - 2\ln[\cos(D(x,t) + tK(x,t))]. \tag{3.32}$$

We have

$$u_t(x,t) = 2\frac{K_t}{K} - 2\frac{-\sin(D + tK)(D_t + K + tK_t)}{\cos(D + tK)},$$

so the first condition we set is

$$\frac{K_t}{K} + \frac{\sin(D + tK)(D_t + tK_t)}{\cos(D + tK)} = 0, \tag{3.33}$$

so that

$$u_t(x,t) = 2K\tan(D + tK).$$

If we differentiate again, we obtain

$$\begin{aligned} u_{tt}(x,t) &= 2K_t \tan(D+tK) + 2K(D_t + tK_t + K)\sec^2(D+tK), \\ &= 2K_t \tan(D+tK) - \frac{2K_t}{\tan(D+tK)}\sec^2(D+tK) + 2K^2\sec^2(D+tK), \end{aligned}$$

where we used (3.33) which gives

$$(D_t + tK_t) = \frac{-K_t}{K\tan(D+tK)}.$$

Since u is given by (3.32), it becomes

$$\begin{aligned} u_{tt}(x,t) &= 2K_t \frac{\tan^2(D+tK) - \sec^2(D+tK)}{\tan(D+tK)} + \delta e^u, \\ &= \delta e^u - 2K_t \frac{1}{\tan(D+tK)}. \end{aligned}$$

So in order for u to satisfy the PDE (3.31), we need

$$(\Delta u =) - 2\Delta\left[\ln(\cos(D+tK)) - \ln K\right] = \frac{-2K_t}{\tan(D+tK)},$$

that is

$$K_t = \tan(D+tK)\Delta\left[\ln(\cos(D+tK)) - \ln K\right].$$

Hence we obtained a system of equations for K and D,

$$\left\{ \begin{array}{rcl} K_t &=& \tan(D+tK)\Delta\left[\ln(\cos(D+tK)) - \ln K\right], \\ D_t &=& -\Delta\left[\ln(\cos(D+tK)) - \ln K\right]\left(t\,\tan(D+tK) + \frac{1}{K}\right). \end{array} \right. \tag{3.34}$$

Remember that we have

$$\left\{ \begin{array}{rcl} u &=& \ln 2 - \ln\delta + 2\ln K - 2\ln[\cos(D+tK)], \\ v &=& u_t = 2K\tan(D+tK), \end{array} \right.$$

from which we obtain

$$\left\{ \begin{array}{rcl} K &=& \frac{1}{2}\sqrt{2\delta e^u - v^2}, \\ D &=& \arctan\frac{v}{\sqrt{2\delta e^u - v^2}} - \frac{t}{2}\sqrt{2\delta e^u - v^2}. \end{array} \right.$$

Note that we need $2\delta e^u \geq v^2$.

Now we can apply different numerical schemes to system (3.34), for example using the backward Euler method, we get

$$K_{n+1} - K_n = h \tan(D_{n+1} + t_{n+1}K_{n+1})\Delta \left[\ln(\cos(D_{n+1} + t_{n+1}K_{n+1})) - \ln K_{n+1}\right],$$

and

$$\begin{aligned}
&D_{n+1} - D_n \\
&= -\Delta \left[\ln(\cos(D_{n+1} + t_{n+1}K_{n+1})) - \ln K_{n+1}\right] \left(t_{n+1}\, \tan(D_{n+1} + t_{n+1}K_{n+1}) + \frac{1}{K_{n+1}}\right),
\end{aligned}$$

so going back to u and v it becomes

$$\begin{aligned}
&\sqrt{2\delta e^{u_{n+1}} - v_{n+1}^2} - \sqrt{2\delta e^{u_n} - v_n^2} \\
&\quad = 2h \frac{v_{n+1}}{\sqrt{2\delta e^{u_{n+1}} - v_{n+1}^2}} \Delta \left[\ln\left(\frac{\sqrt{2\delta e^{u_{n+1}} - v_{n+1}^2}}{\sqrt{2\delta e^{u_{n+1}}}}\right) - \ln\left(\frac{1}{2}\sqrt{2\delta e^{u_{n+1}} - v_{n+1}^2}\right)\right], \\
&\quad = -h \frac{v_{n+1}}{\sqrt{2\delta e^{u_{n+1}} - v_{n+1}^2}} \Delta u_{n+1},
\end{aligned}$$

and

$$\begin{aligned}
&\arctan \frac{v_{n+1}}{\sqrt{2\delta e^{u_{n+1}} - v_{n+1}^2}} - \frac{t_{n+1}}{2}\sqrt{2\delta e^{u_{n+1}} - v_{n+1}^2} \\
&\qquad\qquad - \arctan \frac{v_n}{\sqrt{2\delta e^{u_n} - v_n^2}} + \frac{t_n}{2}\sqrt{2\delta e^{u_n} - v_n^2} \\
&\quad = \frac{h}{2}\Delta u_{n+1} \left(t_{n+1} \frac{v_{n+1}}{\sqrt{2\delta e^{u_{n+1}} - v_{n+1}^2}} + \frac{2}{\sqrt{2\delta e^{u_{n+1}} - v_{n+1}^2}}\right).
\end{aligned}$$

If we apply the forward Euler method instead, we get the following scheme

$$\begin{cases}
\sqrt{2\delta e^{u_{n+1}} - v_{n+1}^2} - \sqrt{2\delta e^{u_n} - v_n^2} = h \frac{v_n}{\sqrt{2\delta e^{u_n} - v_n^2}} \Delta u_n, \\
\arctan \frac{v_{n+1}}{\sqrt{2\delta e^{u_{n+1}} - v_{n+1}^2}} - \frac{t_{n+1}}{2}\sqrt{2\delta e^{u_{n+1}} - v_{n+1}^2} \\
\qquad\qquad - \arctan \frac{v_n}{\sqrt{2\delta e^{u_n} - v_n^2}} + \frac{t_n}{2}\sqrt{2\delta e^{u_n} - v_n^2} \\
\qquad\qquad = \frac{h}{2}\Delta u_{n+1} \left(\frac{t_{n+1}v_{n+1} + 2}{\sqrt{2\delta e^{u_{n+1}} - v_{n+1}^2}}\right).
\end{cases}$$

In order to write the scheme obtained using the midpoint rule, we need to introduce

$$S_1 = \sqrt{2\delta e^{u_n} - v_n^2} \quad \text{and} \quad S_2 = \sqrt{2\delta e^{u_{n+1}} - v_{n+1}^2},$$

so that $\tilde{K} = \dfrac{S_1 + S_2}{4}$ and

$$\tilde{D} = \frac{1}{2}\left(\arctan\left(\frac{v_n}{S_1}\right) - \frac{t_n}{2}S_1 + \arctan\left(\frac{v_{n+1}}{S_2}\right) - \frac{t_{n+1}}{2}S_2\right).$$

The scheme is now given by

$$\frac{S_2 - S_1}{2} = h\tan\left(\tilde{K}\left(t_n + \frac{h}{2}\right) + \tilde{D}\right)\Delta\left[\ln\left(\cos\left(\tilde{K}(t_n + h/2) + \tilde{D}\right)\right) - \ln\tilde{K}\right],$$

and

$$\begin{aligned}&\left(\arctan\left(\frac{v_{n+1}}{S_2}\right) - \frac{t_{n+1}}{2}S_2\right) - \left(\arctan\left(\frac{v_n}{S_1}\right) - \frac{t_n}{2}S_1\right)\\ &= -h\Delta\left[\ln\frac{\cos\left(\tilde{K}(t_n + \frac{h}{2}) + \tilde{D}\right)}{\tilde{K}}\right]\left((t_n + \frac{h}{2})\tan\left(\tilde{K}(t_n + \frac{h}{2}) + \tilde{D}\right) + \frac{1}{\tilde{K}}\right).\end{aligned}$$

For the scheme obtained by the trapezoidal rule, using the same notation as above, we can write

$$\frac{S_2 - S_1}{2} = -\frac{h}{4}\left[\frac{v_{n+1}}{S_2}\Delta u_{n+1} + \frac{v_n}{S_1}\Delta u_n\right]$$

and

$$\begin{aligned}&\left(\arctan\left(\frac{v_{n+1}}{S_2}\right) - \frac{t_{n+1}}{2}S_2\right) - \left(\arctan\left(\frac{v_n}{S_1}\right) - \frac{t_n}{2}S_1\right)\\ &\qquad = -\frac{h}{2}\left[\Delta u_n\left(\frac{t_n v_n + 2}{S_1}\right) + \Delta u_{n+1}\left(\frac{t_{n+1}v_{n+1} + 2}{S_2}\right)\right].\end{aligned}$$

Splitting method. In order to get schemes by the splitting method, it is more convenient to write the second order equation (3.31) as a first order system:

$$\begin{cases} u_t &= v, \\ v_t &= \Delta u + \delta e^u. \end{cases} \tag{3.35}$$

The general solution of the simplified system

$$\begin{cases} y_t &= z, \\ z_t &= \delta e^y, \end{cases} \tag{3.36}$$

is given by

$$\begin{cases} y(t) &= 2\ln[K\sec(\alpha K t + D)] = -2\ln[\frac{1}{K}\cos(\alpha K t + D)], \\ z(t) &= 2\alpha K \tan(\alpha K t + D), \end{cases}$$

where $\alpha = \sqrt{2\delta}/2$. In order to get the exact flow of the simplified system, we need to express the constants K and D as functions of the initial conditions y_0 and z_0. From

$$y(0) = 2\ln(K\sec D) \quad \text{and} \quad z(0) = 2\alpha K \tan D,$$

we obtain

$$\sec D = \frac{1}{K} e^{y_0/2} \quad \text{and} \quad \tan D = \frac{z_0}{2\alpha K}, \tag{3.37}$$

and then

$$\frac{e^{y_0}}{K^2} - \frac{z_0^2}{4\alpha^2 K^2} = 1,$$

from which we obtain

$$K^2 = e^{y_0} - \frac{z_0^2}{4\alpha^2}.$$

This gives K explicitly, however for D we only have (3.37), so we need to isolate $\tan D$ and $\cos D$ in the expression of $y(t)$ and $z(t)$. For this we use the following identities

$$\tan(x+y) = \frac{\tan x + \tan y}{1 - \tan x \tan y},$$

and

$$\cos(x+y) = \cos x \cos y - \sin x \sin y = \cos y[\cos x - \sin x \tan y].$$

The exact flow of (3.36) is given by

$$\begin{aligned} y(t) &= -2\ln\left[\frac{1}{K}\cos D[\cos(\alpha K t) - \sin(\alpha K t)\tan D]\right] \\ &= -2\ln[e^{-y_0/2}[\cos(\alpha K t) - \frac{z_0}{2\alpha K}\sin(\alpha K t)]] \\ &= y_0 - 2\ln[\cos(\alpha K t) - \frac{z_0}{2\alpha K}\sin(\alpha K t)], \end{aligned}$$

and

$$z(t) = 2\alpha K \left(\frac{\tan(\alpha K t) + \tan D}{1 - \tan(\alpha K t)\tan D} \right) = 2\alpha K \left(\frac{z_0 + 2\alpha K \tan(\alpha K t)}{2\alpha K - z_0 \tan(\alpha K t)} \right),$$

with $K = (e^{y_0} - \frac{z_0^2}{4\alpha^2})^{1/2}$.

In order to get specialized methods for the original system (3.35), we simply choose a numerical method for

$$\begin{cases} u_t = 0, \\ v_t = \Delta u, \end{cases}$$

and compose it with the above exact flow. Since by the first equation we have $u_{n+1} = u_n$, choosing $\Phi_h^{[2]}$ to be forward Euler or backward Euler leads to the same scheme

$$\begin{cases} u_{n+1} &= u_n - 2\ln[\cos(\alpha K_n h) - \frac{v_n + h\Delta u_n}{2\alpha K_n}\sin(\alpha K_n h)], \\ v_{n+1} &= 2\alpha K_n \left(\frac{v_n + h\Delta u_n + 2\alpha K_n \tan(\alpha K_n h)}{2\alpha K_n - (v_n + h\Delta u_n)\tan(\alpha K_n h)} \right), \end{cases}$$

with $K_n = \sqrt{e^{u_n} - \frac{(v_n + h\Delta u_n)^2}{4\alpha^2}}$. The adjoint of this method is given by

$$\begin{cases} u_{n+1} &= u_n - 2\ln[\cos(\alpha K_n h) - \frac{v_n}{2\alpha K_n}\sin(\alpha K_n h)], \\ v_{n+1} &= 2\alpha K_n \left(\frac{v_n + 2\alpha K_n \tan(\alpha K_n h)}{2\alpha K_n - v_n \tan(\alpha K_n h)} \right) + h\Delta u_{n+1}, \end{cases}$$

with $K_n = \sqrt{e^{u_n} - \frac{v_n^2}{4\alpha^2}}$. To write the second-order method obtained by composing theses two, we first set $K_n = \sqrt{e^{u_n} - \frac{v_n^2}{4\alpha^2}}$ and

$$\begin{cases} \tilde{u} &= u_n - 2\ln[\cos(\alpha K_n \frac{h}{2}) - \frac{v_n}{2\alpha K_n}\sin(\alpha K_n \frac{h}{2})], \\ \tilde{v} &= 2\alpha K_n \left(\frac{v_n + 2\alpha K_n \tan(\alpha K_n \frac{h}{2})}{2\alpha K_n - v_n \tan(\alpha K_n \frac{h}{2})} \right) + \frac{h}{2}\Delta u_{n+1}, \end{cases}$$

then the scheme is given by

$$\begin{cases} u_{n+1} &= \tilde{u} - 2\ln[\cos(\alpha K_n \frac{h}{2}) - \frac{(\tilde{v} + \frac{h}{2}\Delta\tilde{u})}{2\alpha K_n}\sin(\alpha K_n \frac{h}{2})], \\ v_{n+1} &= 2\alpha K_n \left(\frac{(\tilde{v} + \frac{h}{2}\Delta\tilde{u}) + 2\alpha K_n \tan(\alpha K_n \frac{h}{2})}{2\alpha K_n - (\tilde{v} + \frac{h}{2}\Delta\tilde{u})\tan(\alpha K_n \frac{h}{2})} \right), \end{cases}$$

with $\tilde{K} = \sqrt{e^{\tilde{u}} - \frac{(\tilde{v} + \frac{h}{2}\Delta\tilde{u})^2}{4\alpha^2}}$.

3.2.5 An "Accretive" Equation

Another type of second order equation that exhibits blow-up behavior, the initial-value problem $u_{tt} = \Delta u + F(u_t)$, was first studied by Glassey in [65]. The corresponding initial-boundary value problem was studied by Levine [108]. In this section, we derive specialized methods for the following problem:

$$\begin{cases} u_{tt} &= \Delta u + \delta e^{u_t}, & \text{for } (x,t) \in \Omega \times (0,T), \\ u &= 0, & \text{for } (x,t) \in \partial\Omega \times (0,T), \\ u(x,0) &= u_0(x), & \text{for } x \in \Omega, \\ u_t(x,0) &= u_{t0}(x), & \text{for } x \in \Omega. \end{cases} \tag{3.38}$$

As for the nonlinear wave equation, we rewrite the equation as a system

$$\begin{cases} u_t &= v, \\ v_t &= \Delta u + \delta e^v. \end{cases}$$

The general solution of the simplified system

$$\begin{cases} y_t &= z, \\ z_t &= \delta e^z, \end{cases}$$

is given by

$$\begin{cases} y(t) &= D + \frac{K-\delta t}{\delta}[\ln(K-\delta t) - 1], \\ z(t) &= -\ln(K-\delta t). \end{cases}$$

Splitting method. In order to use the splitting methods approach, we need to get the exact flow of the simplified system, so first, we express the constants K and D as functions of the initial conditions y_0 and z_0. Since $z(0) = -\ln(K)$, we obtain $K = e^{-z_0}$ and then $y(0) = D + \frac{K}{\delta}[\ln(K) - 1]$, which gives $D = y_0 + \frac{K}{\delta}(z_0 + 1)$. So we get

$$\begin{cases} y(t) &= y_0 + \frac{e^{-z_0}}{\delta}(z_0+1) + \frac{e^{-z_0}-\delta t}{\delta}[\ln(e^{-z_0} - \delta t) - 1], \\ z(t) &= -\ln(e^{-z_0} - \delta t), \end{cases}$$

and the exact flow is given by

$$\left\{\begin{array}{ll} \varphi_h(u_n,v_n) & = u_n + \frac{e^{-v_n}}{\delta}(v_n+1) + \frac{e^{-v_n}-\delta h}{\delta}[\ln(e^{-v_n}-\delta h)-1], \\ \psi_h(u_n,v_n) & = -\ln(e^{-v_n}-\delta h). \end{array}\right.$$

The numerical method $\Phi_h^{[2]}$ is only applied to the very simple system

$$\left\{\begin{array}{ll} u_t & = 0, \\ v_t & = \Delta u. \end{array}\right.$$

As for the wave equation, the first equation implies that $u_{n+1} = u_n$, so that choosing $\Phi_h^{[2]}$ to be forward Euler or backward Euler leads to the same scheme

$$\Phi_h(u_n,v_n) = \left(\begin{array}{l} u_n + \frac{e^{-v_n}}{\delta}(v_n+1) + \frac{e^{-v_n}-\delta h}{\delta}[\ln(e^{-v_n}-\delta h)-1] \\ -\ln(e^{-v_n-h\Delta u_n}-\delta h) \end{array}\right),$$

whose adjoint Φ_h^* is given by

$$\left\{\begin{array}{l} u_{n+1} = u_n + \frac{e^{-v_n}}{\delta}(v_n+1) + \frac{e^{-v_n}-\delta h}{\delta}[\ln(e^{-v_n}-\delta h)-1], \\ v_{n+1} = -\ln(e^{-v_n}-\delta h) + h\Delta\left(u_n + \frac{e^{-v_n}}{\delta}(v_n+1) + \frac{e^{-v_n}-\delta h}{\delta}[\ln(e^{-v_n}-\delta h)-1]\right). \end{array}\right.$$

We can obtain a method of second order by composing these two schemes: $\Phi_{h/2}\circ\Phi_{h/2}^*$, which gives the explicit second-order method

$$\left\{\begin{array}{ll} u_{n+1} & = \varphi_{h/2}(\varphi_{h/2}(u_n,v_n),\psi_{h/2}(u_n,v_n)+h\Delta\varphi_{h/2}(u_n,v_n)), \\ v_{n+1} & = \psi_{h/2}(\varphi_{h/2}(u_n,v_n),\psi_{h/2}(u_n,v_n)+h\Delta\varphi_{h/2}(u_n,v_n)). \end{array}\right.$$

Variation of the constant. We now turn to the method by variation of the constant: we look for a solution of the form

$$u(x,t) = D(x,t) + (K(x,t)-t)[\ln\delta - 1 + \ln(K(x,t)-t)].$$

The first partial derivative in time is

$$\begin{aligned} u_t(x,t) &= D_t + (K_t-1)[\ln\delta - 1 + \ln(K-t)] + (K-t)\left[\frac{K_t-1}{K-t}\right] \\ &= D_t + (K_t-1)(\ln\delta + \ln(K-t)), \end{aligned}$$

so we get the first condition $D_t + K_t(\ln\delta + \ln(K-t)) = 0$. Differentiating again $u_t = -(\ln\delta + \ln(K-t))$, we obtain the second partial derivative

$$u_{tt} = -\frac{K_t - 1}{K - t}.$$

Since we have

$$\Delta u + \delta e^{u_t} = \Delta u + \delta\frac{1}{\delta K - \delta t} = \Delta u + \frac{1}{K-t},$$

we need

$$\Delta u = \frac{-K_t}{K-t},$$

for u to be solution of (3.38). Hence we end up with the following system for K and D

$$\left\{\begin{array}{rcl} K_t &=& -(K-t)\Delta[D + (K-t)[\ln\delta - 1 + \ln(K-t)]], \\ D_t &=& [\ln\delta + \ln(K-t)](K-t)\Delta[D + (K-t)[\ln\delta - 1 + \ln(K-t)]]. \end{array}\right. \tag{3.39}$$

From the definitions of u and v, $u = D + (K-t)[\ln\delta - 1 + \ln(K-t)]$ and $v = -[\ln\delta + \ln(K-t)]$, we obtain

$$\left\{\begin{array}{rcl} K &=& t + \frac{1}{\delta e^v}, \\ D &=& u + \frac{v+1}{\delta e^v}. \end{array}\right.$$

We can now apply different standard methods to the system (3.39) to obtain several specialized schemes for our original equation. Applying forward Euler, we obtain

$$\left\{\begin{array}{l} e^{-v_{n+1}} - e^{-v_n} + \delta h + h e^{-v_n}\Delta u_n = 0, \\ \delta(u_{n+1} - u_n) + (v_{n+1}+1)e^{-v_{n+1}} - (v_n+1)e^{-v_n} + h v_n e^{-v_n}\Delta u_n = 0, \end{array}\right.$$

which can be written in explicit form

$$\left\{\begin{array}{l} v_{n+1} = -\ln(e^{-v_n} - \delta h - h e^{-v_n}\Delta u_n), \\ u_{n+1} = \frac{v_n+1}{\delta}e^{-v_n} + u_n - \frac{v_{n+1}+1}{\delta}e^{-v_{n+1}} - v_n\frac{h}{\delta}e^{-v_n}\Delta u_n, \end{array}\right.$$

and applying backward Euler, we obtain

$$\left\{\begin{array}{l} e^{-v_{n+1}} - e^{-v_n} + \delta h + h e^{-v_{n+1}}\Delta u_{n+1} = 0, \\ \delta(u_{n+1} - u_n) + (v_{n+1}+1)e^{-v_{n+1}} - (v_n+1)e^{-v_n} + h v_{n+1} e^{-v_{n+1}}\Delta u_{n+1} = 0. \end{array}\right.$$

In order to write the scheme obtained with the midpoint rule, we first introduce

$$\tilde{E} = \frac{K_n + K_{n+1}}{2} - \frac{t_n + t_{n+1}}{2} = \frac{e^{-v_n} + e^{-v_{n+1}}}{2\delta},$$

and

$$\tilde{D} = \frac{D_n + D_{n+1}}{2} = \frac{u_n + u_{n+1}}{2} + \frac{v_n + 1}{2\delta e^{v_n}} + \frac{v_{n+1} + 1}{2\delta e^{v_{n+1}}},$$

then the schemes are

$$\begin{cases} e^{-v_{n+1}} - e^{-v_n} + \delta h = -\delta h \tilde{E} \Delta \left[\tilde{D} + \tilde{E} \left(\ln(\delta \tilde{E}) - 1 \right) \right], \\ \delta(u_{n+1} - u_n) + \frac{v_{n+1}+1}{e^{v_{n+1}}} - \frac{v_n+1}{e^{v_n}} = \delta h \tilde{E} \ln(\delta \tilde{E}) \Delta \left[\tilde{D} + \tilde{E} \left(\ln(\delta \tilde{E}) - 1 \right) \right]. \end{cases}$$

Finally the schemes corresponding to the trapezoidal rule are

$$\begin{cases} e^{-v_{n+1}} - e^{-v_n} + \delta h = -\frac{h}{2} \left(e^{-v_n} \Delta u_n + e^{-v_{n+1}} \Delta u_{n+1} \right), \\ \delta(u_{n+1} - u_n) + \frac{v_{n+1}+1}{e^{v_{n+1}}} - \frac{v_n+1}{e^{v_n}} = -\frac{h}{2} (v_n e^{-v_n} \Delta u_n + v_{n+1} e^{-v_{n+1}} \Delta u_{n+1}). \end{cases}$$

3.3 Numerical Experiments

3.3.1 Implementation of the methods

Most of the B-methods are implicit. In order to implement them, we need to solve a nonlinear system at each step. One way to do so is to use a simple fixed-point iteration, however it has been shown (see [113]) that a better approach is to use Newton's method. It is also possible to use the simplified Newton's iterations presented by Hairer et al in [72]. Our extensive numerical experiments showed that in most cases simplified iterations lead to the same results as classical ones. The classical Newton iteration for $F(v) = 0$ is given by

$$J_F(v_n)(v_{n+1} - v_n) = -F(v_n),$$

where J_F represents the Jacobian matrix of F. In order to simplify this, Hairer et al. [72] approximate $J_F(v_n)$ by $J_F(v_0)$, which allows them to compute only once the

LU-decomposition of the Jacobian matrix. Hence, to get u_{n+1} from u_n, we would solve iteratively

$$J_F(u_n)(v_{k+1} - v_k) = -F(v_k),$$

with $v_0 = u_n$, and then we would define $u_{n+1} = v_{k+1}$. In other words, the procedure would be

`Compute` $J_F(u_n)$

`Set` $v_0 = u_n$

`Newton's Iteration:` `Compute` $F(v_k)$

`Solve` $J_F(u_n)dv = -F(v_k)$

`Define` $v_{k+1} = v_k + dv$

`Set` $u_{n+1} = v_{k+1}$.

3.3.2 Numerical Experiments

In this section, we illustrate with numerical experiments the improvement brought by the B-methods derived in Section 3.2 for computing blow-up solutions accurately. The schemes we derived are only semi-discretizations in time. To complete them, we chose to apply finite-differences to discretize the Laplacian in space. The mesh was chosen in order to lead to stable solutions.

We apply the fourteen methods listed in Table 3.2 to each problem, except for the second-order equations for which several of these methods are identical (see Sections 3.2.4 and 3.2.5). Hereafter, we use the abbreviations listed in Table 3.2 in legends and tables of values.

We present two types of results for each problem. The first two figures, together with the corresponding tables of values, illustrate the general improvement in the accuracy of numerical solutions obtained with B-methods. They also show the order of each method. To obtain these figures, all methods are applied to the problem on

Methods	Abbreviations
Standard Forward Euler	FE
Standard Backward Euler	BE
Splitting Forward Euler	SpFE
Adjoint Splitting Forward Euler	SpFEA
Splitting Backward Euler	SpBE
Adjoint Splitting Backward Euler	SpBEA
Variation of the Constant with Forward Euler	VCFE
Variation of the Constant with Backward Euler	VCBE
Standard Midpoint Rule	MR
Standard Trapezoidal Rule	TR
Second-Order splitting method, with forward Euler	SoSpFE
Second-Order splitting method, with backward Euler	SoSpBE
Variation of the constant with midpoint rule	VCMR
Variation of the constant with trapezoidal rule	VCTR

Table 3.2: List of methods and their abbreviations.

$[0, T_f]$, where T_f is smaller than, yet quite close to the blow-up time, with several different timesteps. For each stepsize, we compute the infinity norm of the error at time T_f. Since we do not have the exact solution, an adaptive method is used as a reference to compute the error: the function `ode45` in Matlab, which is based on an explicit Runge-Kutta (4,5) formula, the Dormand-Prince pair. The tolerance is set to 10^{-12}. The errors obtained are then plotted on two loglog graphs, one for first-order methods and the other for second-order methods.

The second set of figures illustrates how the accuracy of the numerical solutions evolves as we approach blow-up time. To obtain these figures, we apply the methods with a fixed stepsize on $[0, T_f]$, where T_f is very close to the blow-up time, and we compute and plot the infinity norm of the error at each timestep. The adaptive method `ode45` is again used to represent the exact solution. The figures are then refocused on the timesteps close to the blow-up time as the potential of the B-methods is more noticeable there. As for the first set of figures, we separate the first-order and second-order methods.

For the most important problem, the semilinear parabolic equation (3.1), we treated several examples of functions F. An example of the case $F(u) = e^u$ is presented below, whereas more examples with different initial conditions and parameter choices as well as the results concerning $F(u) = (u+\alpha)^p$ and $F(u) = (u+1)\ln(u+1)^{p+1}$ are presented in the Appendix. We also present one example of each type of problem: the quasilinear problem (3.22) and the second-order equation (3.31) that are widespread models (see Chapter 1) are presented below, whereas the system (3.24) and the second-order equation (3.38) are put into the Appendix.

Semilinear parabolic equation $u_t = \Delta u + \delta e^u$. For the first example, we use the function $F(u) = e^u$ in the semilinear equation (3.1), on the interval $\Omega = [-1, 1]$. We set $\delta = 3$ and $u_0(x) = \cos(\pi x/2)$, which is concave on the whole interval. (Different

initial conditions are presented in the Appendix.) Using adaptive methods, we can evaluate the blow-up time at $T_b \approx 0.1664$.

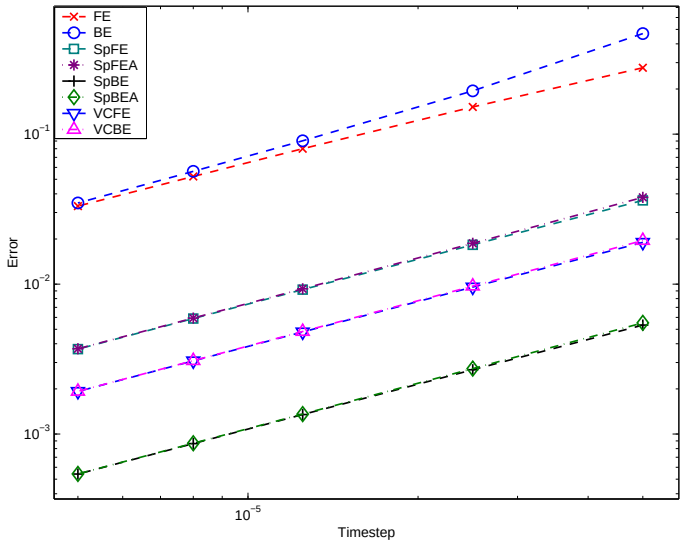

Figure 3.1: Error at $T_f = 0.1660$ for first-order methods applied to the semilinear equation with $F(u) = e^u$, with different values of h.

For Figures 3.1 and 3.2, we computed the solution up to $T_f = 0.1660$ with different stepsizes. For first-order methods, we used $h = 0.00005$, 0.000025, 0.0000125, 0.000008 and 0.000005. For second-order methods, we used $h = 0.0002$, 0.000125, 0.0001, 0.00005 and 0.000025. 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. The values used to generate these figures are listed in Tables 3.3 and 3.4. 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.

For Figures 3.3 and 3.4, we used $h = 0.0001$ and computed the solution up to

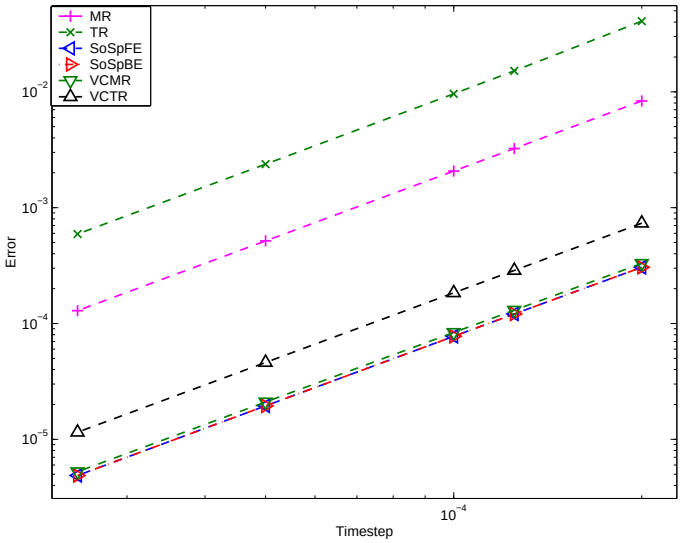

Figure 3.2: Error at $T_f = 0.1660$ for second-order methods applied to the semilinear equation with $F(u) = e^u$, with different values of h.

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
VCBE	0.0195	0.0097	0.00483	0.00309	0.00193

Table 3.3: Error at $T_f = 0.1660$ for first-order methods applied to the semilinear equation with $F(u) = e^u$.

Timestep	0.0002	0.000125	0.0001	5e-005	2.5e-005
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 3.4: Error at $T_f = 0.1660$ for second-order methods applied to the semilinear equation with $F(u) = e^u$.

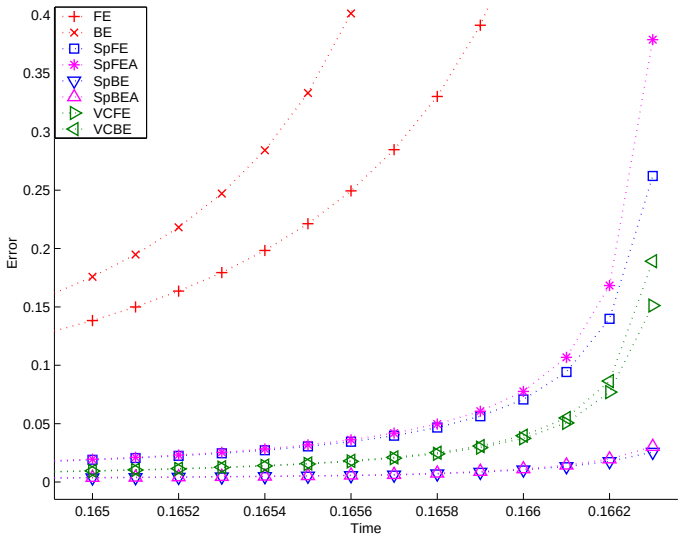

Figure 3.3: Error for first-order methods applied to the semilinear equation with $F(u) = e^u$, for timesteps close to $T_f = 0.1663$.

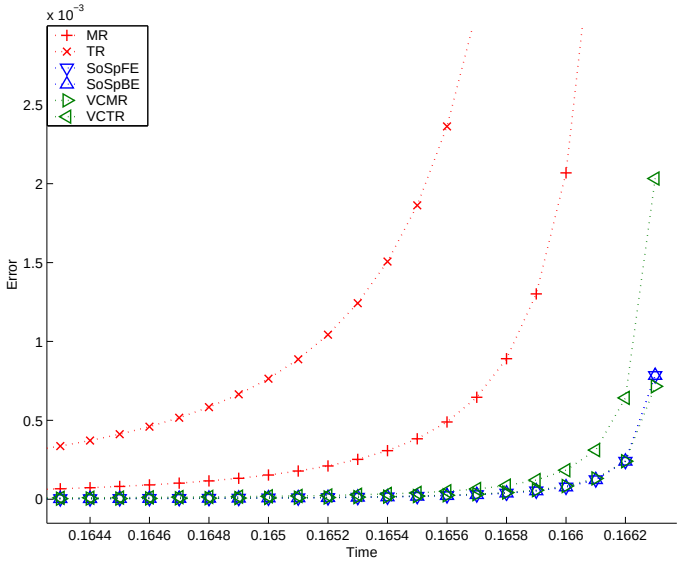

Figure 3.4: Error for second-order methods applied to the semilinear equation with $F(u) = e^u$, for timesteps close to $T_f = 0.1663$.

$T_f = 0.1663$.

Quasilinear Equation For the quasilinear equation (3.22), we consider

$$u_t = \Delta u^2 + 8u^3, \tag{3.40}$$

on $\Omega = [-1, 1]$ with the same initial condition as above: $u_0(x) = \cos(\pi x/2)$. The blow-up time is approximately $T_b \approx 0.1128$.

For Figures 3.5 and 3.6 we computed the solution up to $T_f = 0.1000$, using the stepsizes $h = 0.000125$, 0.00008, 0.00005, 0.000025 and 0.0000125 for first-order methods and $h = 0.0005$, 0.00025, 0.000125, 0.00008 and 0.00005 for second-order methods. The errors are listed in Tables 3.5 and 3.6. 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

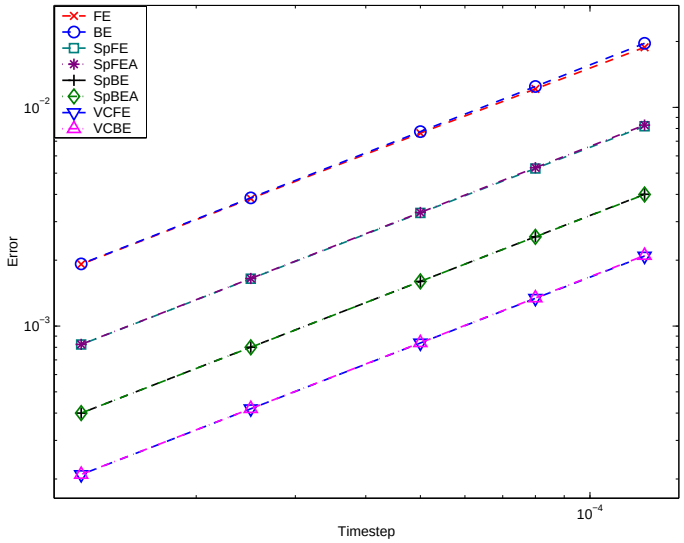

Figure 3.5: Error at $T_f = 0.1000$ for first-order methods applied to the quasilinear equation (3.40), with different values of h.

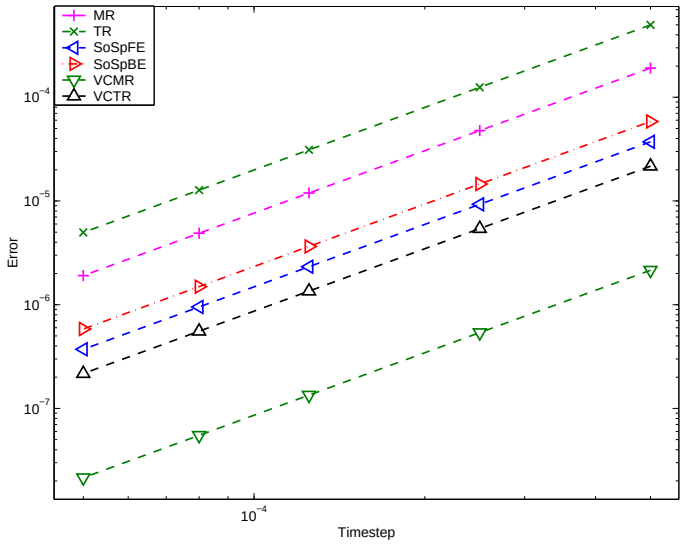

Figure 3.6: Error at $T_f = 0.1000$ for second-order methods applied to the quasilinear equation (3.40), with different values of h.

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

Table 3.5: Error at $T_f = 0.1000$ for first-order methods applied to the quasilinear equation (3.40).

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 3.6: Error at $T_f = 0.1000$ for second-order methods applied to the quasilinear equation (3.40).

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.

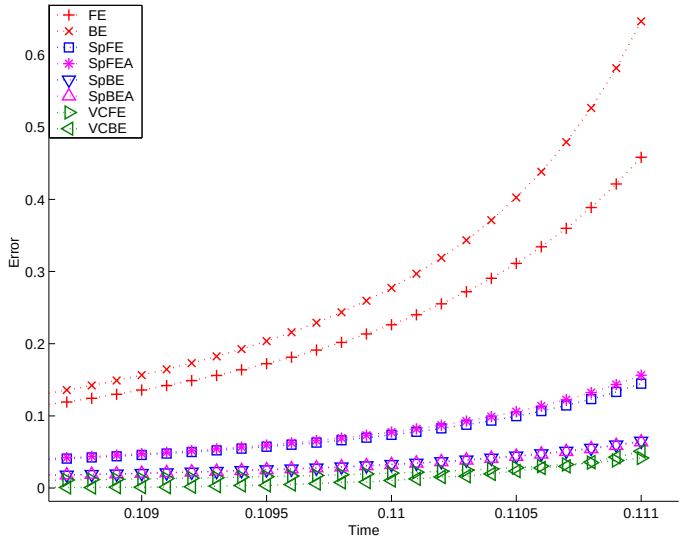

Figure 3.7: Error for first-order methods applied to the quasilinear equation (3.40), for timesteps close to $T_f = 0.1110$.

The step-by-step errors is plotted in Figures 3.7 and 3.8 up to $T_f = 0.1110$, when the solutions are computed using the timestep $h = 0.0001$.

Wave Equation As a last example (more examples are presented in the Appendix), we present the wave equation

$$u_{tt} = \Delta u + 5e^u, \tag{3.41}$$

on $\Omega = [-1, 1]$. The initial conditions are $u_0(x) = \cos(\pi x/2)$ and $u_{t0} = 0.1$. The blow-up time can be approximated by $T_b = 0.643$.

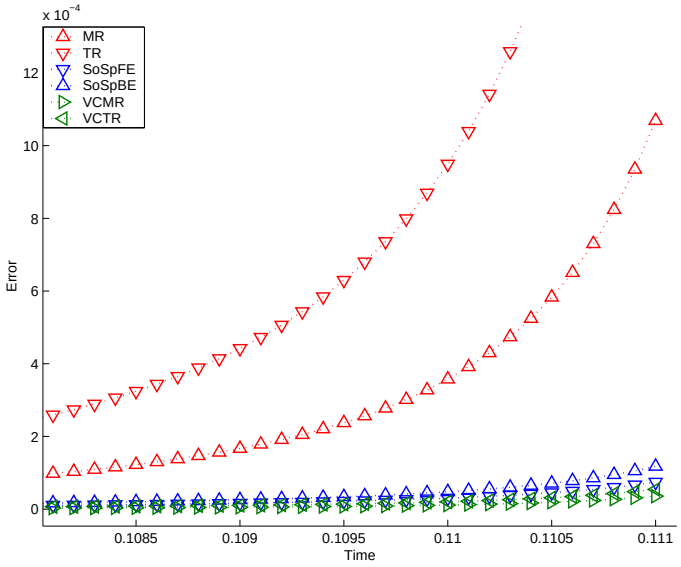

Figure 3.8: Error for second-order methods applied to the quasilinear equation (3.40), for timesteps close to $T_f = 0.1110$.

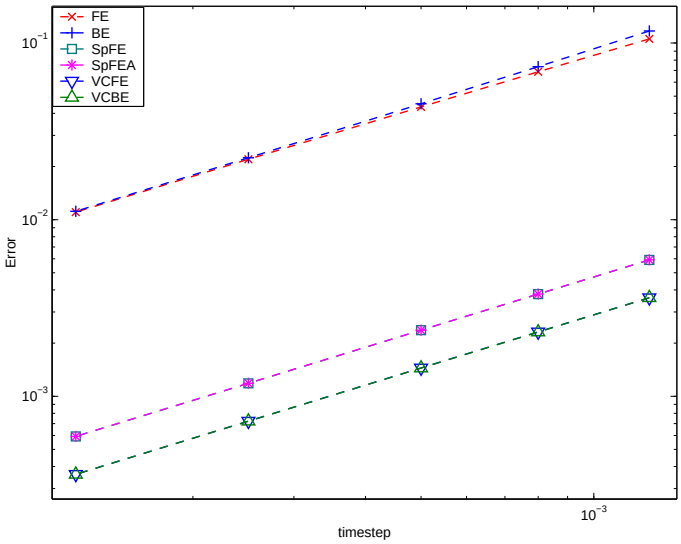

Figure 3.9: Error at $T_f = 0.600$ for first-order methods applied to the wave equation (3.41) with different values of h.

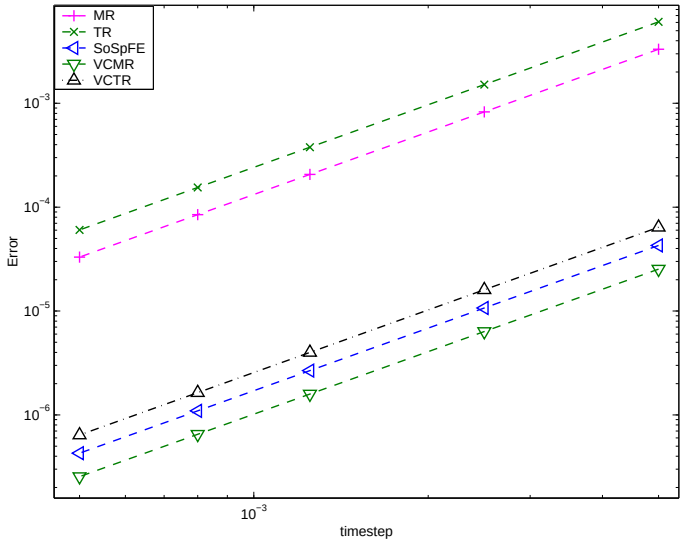

Figure 3.10: Error at $T_f = 0.600$ for second-order methods applied to the wave equation (3.41) with different values of h.

Timestep	0.00125	0.0008	0.0005	0.00025	0.000125
FE	0.106	0.0688	0.0435	0.022	0.011
BE	0.117	0.0735	0.0453	0.0224	0.0112
SpFE	0.00591	0.00379	0.00237	0.00118	0.000592
SpFEA	0.00593	0.00379	0.00237	0.00118	0.000592
VCFE	0.00362	0.00231	0.00145	0.000723	0.000361
VCBE	0.00361	0.00231	0.00145	0.000723	0.000361

Table 3.7: Error at $T_f = 0.600$ for first-order methods applied to the wave equation (3.41).

Timestep	0.005	0.0025	0.00125	0.0008	0.0005
MR	0.00331	0.000827	0.000207	8.46e-005	3.31e-005
TR	0.00608	0.00151	0.000378	0.000155	6.04e-005
SoSpFE	4.28e-005	1.07e-005	2.67e-006	1.09e-006	4.28e-007
VCMR	2.53e-005	6.35e-006	1.59e-006	6.51e-007	2.54e-007
VCTR	6.41e-005	1.6e-005	4e-006	1.64e-006	6.4e-007

Table 3.8: Error at $T_f = 0.600$ for second-order methods applied to the wave equation (3.41).

For Figures 3.9 and 3.10 we computed the solution up to $T_f = 0.6$, using stepsizes $h = 0.00125$, 0.0008, 0.0005, 0.00025 and 0.000125 for first-order methods and $h = 0.005$, 0.0025, 0.00125, 0.0008 and 0.0005 for second-order methods. The errors are listed in Tables 3.7 and 3.8. For second-order methods, the error is between 80 and 125 times smaller and for first-order methods it is 20 or 30 times smaller.

We used $h = 0.0001$ to compute the solutions of (3.41) up to $T_f = 0.630$. The errors for the last steps are plotted in Figures 3.11 and 3.12.

Concluding remarks As a conclusion, we note that on all examples, B-methods bring a clear improvement for the numerical approximation of blow-up solutions. This improvement is generally increasing as we approach the blow-up time. This allows us to obtain a better approximation of the blow-up time by fixed-step methods.

These methods can also be used to construct specialized adaptive methods. For example, the function ode12, which involves backward Euler and the midpoint rule, can be improved to a B-ode12 using VCBE and VCME. To compare the performance of ode12 and B-ode12, we applied both methods to the wave equation (3.41), with the

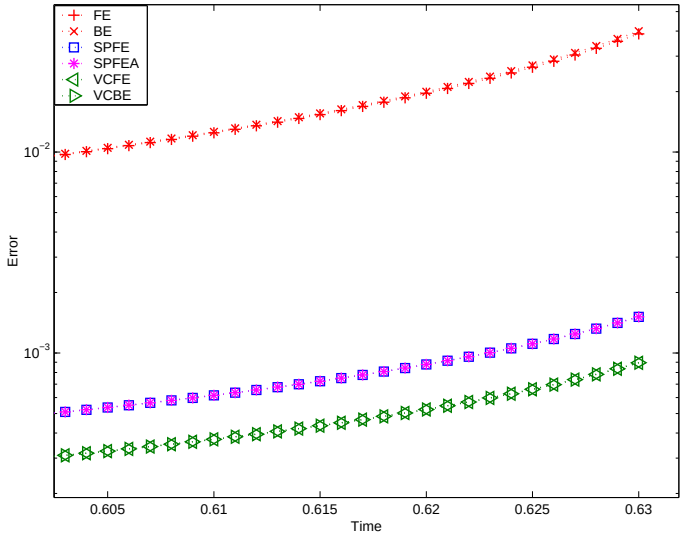

Figure 3.11: Error for first-order methods applied to the wave equation (3.41), for timesteps close to $T_f = 0.630$.

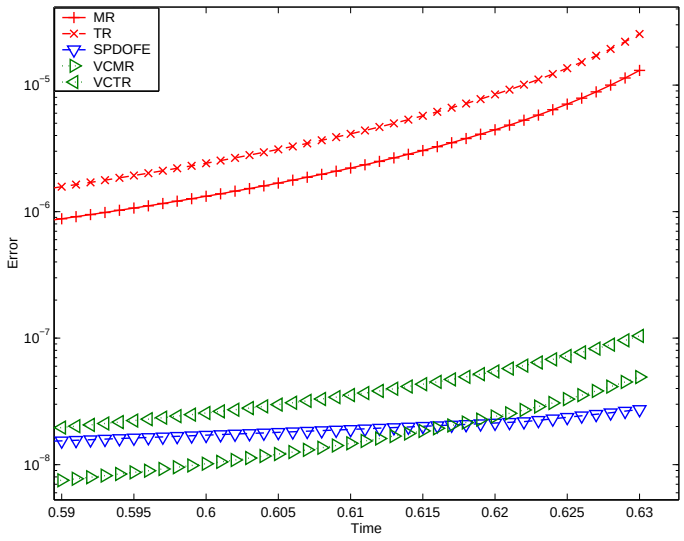

Figure 3.12: Error for second-order methods applied to the wave equation (3.41), for timesteps close to $T_f = 0.630$.

same initial conditions as above. To compute the solution up to $T_f = 0.64$, the function ode12 requires $32,333$ timesteps (with 87 rejected steps), whereas the function B-ode12 requires only $13,184$ timesteps (with 26 rejected steps). As a consequence, it is approximately 5 times faster to compute the solution using B-ode12. The difference is even more obvious as we get closer to the blow-up time. Indeed computing the solution up to $T_f = 0.6434$ requires $41,530$ timesteps (with 127 rejected steps) for ode12 and $13,414$ timesteps (with 49 rejected steps) for B-ode12, and it is close to 7 times faster to obtain the results with B-ode12.

Chapter 4: Theoretical Study of Selected Schemes

As B-methods are numerous and different for every model, it is not possible to study them as a whole and one needs to study each one separately. Thus we chose to concentrate on a few selected schemes, chosen among those derived in Chapter 3. Since the problem is stiff and we wanted to consider schemes as simple as possible, we chose two methods based on the backward Euler method. The first B-method, referred to as VCBE, is obtained using the backward Euler method in the construction by variation of the constant. The second B-method, referred to as SpFEA, was obtained by composing the backward Euler method (which is the adjoint of Forward Euler) and the exact flow of a simpler equation.

Since we only constructed semi-discretizations in time, B-methods are partial differential equations. Thus we first need to prove the existence and uniqueness of a positive solution. Of course this solution should only exist for a finite time. In this chapter we prove that the numerical solution u_n exists as long as it is smaller than a certain constant (which depends on the timestep h). We also give a minimal time T_1 that does not depend on h, for which u_n is small enough so that the solution exists. This value corresponds to a lower bound for the numerical blow-up time. For some B-methods, we were also able to find an upper bound for the numerical blow-up time. For some specific problems, the behaviour of the exact solution close to the blow-

up time has been studied, so we prove that the numerical solution follows the same rate of growth as the exact solution. While these results do not prove by themselves that B-methods are better than standard methods, they do show that they lead to solutions that exhibit the expected behavior.

In Chapter 3, we derived several B-methods to solve the semilinear parabolic problem

$$\begin{cases} u_t &= \Delta u + \delta F(u), & \Omega \times (0,T), \\ u &= 0, & \partial\Omega \times (0,T), \\ u(x,0) &= u_0(x), & \Omega, \end{cases}$$

where Ω is a bounded domain in $\mathbb{R}^d$, u_0 is a positive continuous function on $\bar{\Omega}$ and δ is a positive constant. In Sections 4.1 and 4.2, we present several results concerning two of these B-methods (namely, VCBE and SpFEA). As in Section 3.1, we introduce $g(s) = \int_s^\infty \frac{1}{F(\sigma)} d\sigma$ and $G = g^{-1}$. As the behaviour of these functions and the function F will be used many times throughout the chapter, it may be convenient to summarize most information in a lemma for future references.

Assumption 4.1. *The function F is assumed to be positive, strictly increasing and strictly convex on $(0,\infty)$, belonging to $C^2([0,\infty))$ and satisfying*

$$\int_b^\infty \frac{ds}{F(s)} < \infty, \tag{4.1}$$

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 $M = \lim_{s\to 0} g(s) \leq \infty$. Note also that g and G are positive with

$$\lim_{s\to\infty} g(s) = 0 \quad \text{and} \quad \lim_{s\to 0} G(s) = \infty.$$

Several examples of such functions are given in Section 3.1, including the most studied examples $F(u) = e^u$ and $F(u) = (u+\alpha)^p$, with $\alpha \geq 0$ and $p > 1$ (see Chapters 1 and 2).

For the first B-method (VCBE), some additional assumptions on the function F are necessary to prove the existence and uniqueness of a positive solution. Moreover, as the resulting differential equation is nonlinear, we present an iterative method leading to the solution. As the second B-method (SpFEA) is linear, there is no need for such a solver, and the existence and uniqueness of a positive solution are easily obtained. We prove for both methods that the solution exists as long as $\|u_n\|_\infty < G(\delta h)$ and we prove that this condition is satisfied at least until $T_1 = g(\|u_0\|)/\delta$. This lower bound for the numerical blow-up time is the same as the one given by Kaplan [81]. Kaplan also gives an upper bound for the exact blow-up time and we prove that the solution obtained by the second B-method blows up in a finite time that is smaller than this bound. Finally we also prove for both methods that if $F(u) = e^u$ or $F(u) = (u + \alpha)^2$, the rate of growth of the numerical solution follows the rate of growth of the exact solution.

In Section 4.3 we show how the results of existence and uniqueness of a positive solution need to be modified when the two B-methods are adapted to the more general equation

$$u_t = \Delta u + \delta q(x)\psi(t)F(u),$$

where q is bounded on $\bar{\Omega}$ with $q(x) > 0$, ψ is continuous on $[0, \infty)$, with $\psi(t) > 0$. Some conditions on the function ψ are necessary and the condition $\|u_n\|_\infty < G(\delta h)$ must be adapted and leads to a different condition for each method. Accordingly, two different lower bounds T_1 for the blow-up time are derived.

Finally we present in Section 4.4 several results about two schemes for the quasi-linear parabolic equation with power-like nonlinearities. The first method (VCBE) has been deeply studied by Le Roux [99] so we quickly present some of her results. In her paper, the existence and uniqueness of a positive solution were proven and a lower bound T_1 for the numerical blow-up time was derived. As the scheme is nonlinear a specific solver was presented. Moreover Le Roux proved that the upper bound for

the numerical blow-up time is the same as the bound for the exact one. The rate of growth and the existence of a subsolution were also studied. Concerning the second B-method, we only prove a few results: the existence and uniqueness of a positive solution and the minimal time of existence T_1.

4.1 B-Method Obtained by Variation of the Constant and Backward Euler (VCBE)

In this section, we study the scheme obtained by using the backward Euler method when applying the variation of the constant construction. This scheme was given in Section 3.1.2 in the form

$$g(u_{n+1}) - g(u_n) + \delta h = \frac{-h}{F(u_{n+1})}\Delta u_{n+1}. \tag{4.2}$$

In order to study that scheme, we introduce $Au = -\Delta u$ and write it as $Au_{n+1} = f(x, u_{n+1})$ with

$$f(x,u) = \frac{1}{h}\, F(u)\,(g(u) - g(u_n(x)) + \delta h). \tag{4.3}$$

For our purposes, we need f to be defined and continuous at $u = 0$. This is clearly the case if $F(0) > 0$ since $g(s) = \int_s^\infty \frac{1}{F(\sigma)} d\sigma$, however if $F(0) = 0$, the function f may not be defined at $u = 0$ as $g(0) = \infty$. We prove below that $\lim_{u\to 0^+} f(x,u) = 0$, so that f can be continuously extended by setting $f(x,0) = 0$ for all x in Ω.

By definition, for each $c > 0$, we have $g(c) = \int_c^\infty \frac{ds}{F(s)}$, so that

$$F(c)g(c) = \int_c^\infty \frac{F(c)}{F(s)}\, ds = \int_c^a \frac{F(c)}{F(s)}\, ds + \int_a^\infty \frac{F(c)}{F(s)}\, ds.$$

for any fixed $a \geq c$. Then, since F is increasing and $s \geq c$,

$$F(c)g(c) \leq \int_c^a 1\, ds + \int_a^\infty \frac{F(c)}{F(s)}\, ds = (a-c) + F(c)\int_a^\infty \frac{ds}{F(s)}.$$

The last integral is finite by condition (4.1), we call it I_a. Then we let c tend to zero. We get

$$\lim_{c\to 0^+} F(c)g(c) \leq a + F(0)I_a = a,$$

since $F(0) = 0$. So for any fixed $a > 0$, we get $\lim_{c\to 0^+} F(c)g(c) \leq a$, hence

$$\lim_{c\to 0^+} F(c)g(c) = 0,$$

and $\lim_{u\to 0^+} f(x,u) = 0$ for all $x \in \Omega$.

By abuse of notation, we shall refer to f as its continuous extension on $[0,\infty)$.

4.1.1 Existence and Uniqueness of the solution

Existence

Amann proved in [8] that in case $f(x,0) \geq 0$, a necessary and sufficient condition for the existence of a non-negative solution of problem (4.4) is the existence of a non-negative supersolution.

Theorem 4.2 (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*

$$\begin{aligned} Au := -\Delta u = f(x,u), \qquad & \text{in } \Omega, \\ u = 0, \qquad & \text{on } \partial\Omega, \end{aligned} \tag{4.4}$$

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

$$\begin{aligned} Av \geq f(x,v), \qquad & \text{in } \Omega, \\ v \geq 0, \qquad & \text{on } \partial\Omega. \end{aligned}$$

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 (4.4), the inequality

$$\bar{u} \leq u \leq \hat{u}$$

holds.

We use this result to prove the existence of a non-negative solution of the scheme.

Theorem 4.3. *If the function u_n is positive in Ω, continuous in $\bar{\Omega}$, and satisfies*

$$\|u_n\|_\infty < G(\delta h), \tag{4.5}$$

then scheme (4.2) *has a maximal nonnegative solution $\hat{u} \leq C_n$ with*

$$C_n = G(g(\|u_n\|_\infty) - \delta h),$$

and a minimal solution $\bar{u} \geq 0$ and if u is a solution, then $u \in C^2(\bar{\Omega})$ and satisfies $\bar{u} \leq u \leq \hat{u}$.

Note that by definition of G we have $\lim_{h\to 0} G(\delta h) = \infty$ (see Assumption 4.1), so that by choosing h small, the bound on the right-hand side of (4.5) can be made as large as desired. All following results need this condition to be satified, hence even when the solution can be computed further, the numerical result can become incorrect once this bound is reached.

Proof. As stated at the beginning of the section, if $F(0) = 0$, we have $f(x, 0) = 0$ for all $x \in \Omega$. If $F(0) > 0$, since g is decreasing, we have $g(u_n) < g(0) + \delta h$ and we get $f(x, 0) > 0$. So to apply Theorem 4.2, we need to show that the constant C_n is a supersolution, that is

$$\frac{1}{h} F(C_n) \left(g(C_n) - g(u_n) + \delta h\right) \leq 0 \; (= AC_n),$$

and since $F(C_n) > 0$ and G is decreasing, it becomes

$$C_n \geq G(g(u_n) - \delta h).$$

Hence, the constant

$$C_n = G(g(\|u_n\|_\infty) - \delta h),$$

which is well-defined if condition (4.5) is satisfied and positive by definition of G, is a supersolution. □

If $F(0) = 0$, the identically zero function is solution of scheme (4.2). In this case we need to use a stronger result, proved in [26] by Brezis and Oswald, to prove the existence of a non-identically zero solution.

We consider a problem of the form

$$\begin{cases} -\Delta u &= f(x,u), & \text{in } \Omega, \\ u \geq 0, & u \not\equiv 0 & \text{in } \Omega, \\ u &= 0, & \text{on } \partial\Omega, \end{cases} \tag{4.6}$$

where $\Omega \subset \mathbb{R}^d$ is a bounded domain with smooth boundary and $f(x,u) : \Omega \times [0,\infty) \to \mathbb{R}$. Set

$$a_0(x) = \lim_{u\to 0} \frac{f(x,u)}{u},$$

and

$$a_\infty(x) = \lim_{u\to\infty} \frac{f(x,u)}{u}.$$

For $a(x) = a_0(x), a_\infty(x)$, we denote by $\lambda_1(-\Delta - a(x))$ the first eigenvalue of $-\Delta - a(x)$ with zero Dirichlet boundary condition. Brezis and Oswald proved the following result.

Theorem 4.4 (Brezis and Oswald)**.** *We suppose that the function f satisfies the following properties*

a. *for almost every $x \in \Omega$, the function $u \mapsto f(x,u)$ is continuous on $[0,\infty)$;*

b. *for each $u \geq 0$, the function $x \mapsto f(x,u)$ belongs to $L^\infty(\Omega)$;*

c. *there exists $C_1 > 0$ such that $f(x,u) \leq C_1(u+1)$ for almost every $x \in \Omega$, and for all $u \geq 0$;*

d. for each $\delta > 0$, there exists $C_\delta \geq 0$ such that $f(x,u) \geq -C_\delta u$ for all $u \in [0,\delta]$, and almost every $x \in \Omega$;

e. we have $\lambda_1(-\Delta - a_0(x)) < 0$ and $\lambda_1(-\Delta - a_\infty(x)) > 0$. Note that in the special case where $a_0(x)$ and $a_\infty(x)$ are independent of x, this is equivalent to

$$a_\infty < \lambda_1(-\Delta) < a_0.$$

Then problem (4.6) *has a solution $u \in H_0^1(\Omega) \cap L^\infty(\Omega)$.*

As we only need this stronger result for the case where $F(0) = 0$, the following theorem is only proved for that case.

Theorem 4.5. *If the function u_n is positive in Ω, continuous in $\bar{\Omega}$, and satisfies condition* (4.5)*, the following hold.*

a. If $F(0) = 0$ and $F'(0) > 0$, scheme (4.2) *has a non-identically zero nonnegative solution.*

b. If $F(0) = 0$, $F'(0) = 0$ and

$$L := \lim_{u \to 0} \frac{-F(u)}{F(u) - uF'(u)},$$

is positive, then scheme (4.2) *has a non-identically zero nonnegative solution if*

$$h < \frac{1}{\lambda_1(-\Delta)} L.$$

Proof. We already proved in the introduction of the section that the function $u \to f(x,u)$ is continuous on $[0,\infty)$, and since Ω is bounded, condition (b) of Theorem 4.4 is satisfied for all $u \geq 0$.

Since $F(0) = 0$ by hypothesis, we have $f(x,0) = 0$ and condition (d) of Theorem 4.4 is satisfied for $u = 0$. For $u > 0$, condition (d) of Theorem 4.4 requires that there exists C_δ such that for all $u \in (0,\delta]$ and all $x \in \Omega$,

$$f(x,u) \geq -C_\delta u \iff g(u) \geq c(x) - \frac{hC_\delta u}{F(u)}, \text{ where } c(x) = g(u_n) - \delta h.$$

Since g is positive, it is enough that the right-hand side be negative, that is

$$C_\delta > \frac{F(u)}{u}\frac{c(x)}{h}.$$

By definition of M and because u_n satisfies condition (4.5), we have $c(x) \in (0, M)$. By hypothesis $F(0) = 0$ so that

$$\lim_{u\to 0} \frac{F(u)}{u} = F'(0) < \infty,$$

(since $F \in C^2([0,\infty))$), hence such a constant C_δ exists for all δ and condition (d) of Theorem 4.4 is satisfied.

Since u_n satisfies condition (4.5), we have $g(\|u_n\|_\infty) > \delta h$, so that there exists $\varepsilon > 0$ such that $g(\|u_n\|_\infty) > \delta h + \varepsilon$, and then

$$c(x) > \varepsilon > 0,$$

for all $x \in \Omega$. Hence

$$f(x,u) \le \frac{F(u)}{h}(g(u) - \varepsilon),$$

and since $\lim_{u\to\infty} g(u) = 0$ and $\lim_{u\to\infty} F(u) = \infty$ (see Assumption 4.1), we have

$$\lim_{u\to\infty} \frac{F(u)}{h}(g(u) - \varepsilon) = -\infty,$$

and condition (c) of Theorem 4.4 is satisfied with

$$C_1 = \max_{u\ge 0} \frac{F(u)}{h}(g(u) - \varepsilon).$$

Because of condition (4.1), we have

$$\lim_{u\to\infty} \frac{F(u)}{u} = \infty,$$

and since $\lim_{u\to\infty} g(u) = 0$, we obtain

$$a_\infty(x) = \lim_{u\to\infty} \frac{F(u)}{hu}(g(u) - c(x)) = -\infty,$$

for all x. Finally if $\lim_{u\to 0}\frac{F(u)}{u} = F'(0) > 0$,

$$a_0(x) = \lim_{u\to 0}\frac{F(u)}{hu}(g(u) - c(x)) = \infty,$$

for all x, whereas if $F'(0) = 0$, we need to use l'Hôpital's Rule to compute

$$\begin{aligned} a_0(x) &= \lim_{u\to 0}\frac{F(u)}{hu}(g(u) - c(x)) = \lim_{u\to 0}\frac{F(u)}{hu}g(u) = \frac{1}{h}\lim_{u\to 0}\frac{g(u)}{\frac{u}{F(u)}} \\ &= \frac{1}{h}\lim_{u\to 0}\frac{g'(u)}{\frac{F(u)-uF'(u)}{F^2(u)}} = \frac{1}{h}\lim_{u\to 0}\frac{-F(u)}{F(u) - uF'(u)}, \end{aligned}$$

where we used that $g' = -1/F$. So a_0 and a_∞ are both independent of x, and condition (e) of Theorem 4.4 becomes

$$-\infty < \lambda_1(-\Delta) < a_0,$$

where $a_0 = \infty$ if $F'(0) > 0$, and $a_0 = \frac{1}{h}L$ if $F(0) = 0$. □

Corollary 4.1. *If $F(u) = u^{p+1}$ or $F(u) = (u+1)[\ln(u+1)]^{p+1}$ with $p > 0$ and $h < \frac{1}{p\lambda_1(-\Delta)}$ or if $F(u) = e^u - 1$, scheme (4.2) has a non-identically zero nonnegative solution.*

Proof. If $F(u) = e^u - 1$, we have $F(0) = 0$ and $F'(0) = 1$, so we can apply part (a) of Theorem 4.5.

If $F(u) = u^{p+1}$, we have $F'(u) = (p+1)u^p$ and $F'(0) = 0$, so to obtain the existence of a non-identically zero solution, we need

$$h\lambda_1(-\Delta) < \lim_{u\to 0}\frac{-F(u)}{F(u) - uF'(u)} = \lim_{u\to 0}\frac{-u^{p+1}}{u^{p+1} - (p+1)u^{p+1}} = \frac{1}{p}.$$

Similarly, if $F(u) = (u+1)[\ln(u+1)]^{p+1}$, we have $F'(u) = [\ln(u+1)]^p[(p+1) + \ln(u+1)]$ and $F'(0) = 0$, so to obtain the existence of a non-identically zero solution, we need

$$\begin{aligned} h\lambda_1(-\Delta) &< \lim_{u\to 0}\frac{-F(u)}{F(u) - uF'(u)} \\ &= \lim_{u\to 0}\frac{-(u+1)[\ln(u+1)]^{p+1}}{(u+1)[\ln(u+1)]^{p+1} - u[\ln(u+1)]^p[(p+1) + \ln(u+1)]} \\ &= \lim_{u\to 0}\frac{-(u+1)\ln(u+1)}{\ln(u+1) - (p+1)u} = \frac{1}{p}. \end{aligned}$$

□

Uniqueness

A second result of Amann [8] could be used to prove the uniqueness of positive solutions in case the function f is decreasing in u, as will be done in Section 4.2, however as the function f defined in (4.3) is not generally decreasing in u, we need to use a more general result by Brezis and Oswald [26].

Theorem 4.6 (Brezis and Oswald)**.** *Consider system* (4.6)*. If the function f satisfies the following properties*

a. for almost every $x \in \Omega$, the function $u \mapsto f(x,u)$ is continuous on $[0,\infty)$;

b. for each $u \geq 0$, the function $x \mapsto f(x,u)$ belongs to $L^\infty(\Omega)$;

c. for almost every $x \in \Omega$, the function $u \mapsto \varphi(u) := \frac{f(x,u)}{u}$ is decreasing on $(0,\infty)$;

then problem (4.6) *has at most one solution and this solution is positive.*

We apply this result to problem (4.2).

Theorem 4.7. *We suppose that the function u_n is positive in Ω, continuous in $\bar{\Omega}$, and satisfies condition* (4.5)*. If the function F satisfies the following property*

$$\left(\frac{F(u)}{u}\right)' \left(\int_u^\infty \frac{1}{F(s)}\, ds - c\right) < \frac{1}{u}, \ \ \forall\, u > 0,\ \forall\, c \in (0, M - \delta h), \tag{4.7}$$

then scheme (4.2) *has at most one solution and this solution is positive.*

Proof. We already proved that the first two conditions of Theorem 4.6 are satisfied, so it only remains to show that the function $\varphi(u) = f(x,u)/u$ is decreasing on $(0,\infty)$ for all x. From

$$\varphi(u) = \frac{1}{h}\frac{F(u)}{u}(g(u) - g(u_n) + \delta h),$$

we get

$$\varphi'(u) = \frac{1}{h}\left(\frac{F(u)}{u}\right)'(g(u) - g(u_n) + \delta h) - \frac{1}{h}\frac{1}{u}.$$

Since $g(u_n(x)) - \delta h \in (0, M - \delta h)$, φ is decreasing if (4.7) is satisfied. □

Condition (4.7) is satisfied by many functions F of interest; two important examples are given in the following corollary.

Corollary 4.2. *We suppose that the function u_n is positive in Ω, continuous in $\bar{\Omega}$, and satisfies condition (4.5). If $F(u) = e^u$ or $F(u) = (u+\alpha)^{p+1}$, $\alpha \geq 0$, $p > 0$, the scheme has a unique non-identically zero solution which is positive.*

Proof. If $F(u) = e^u$, condition (4.7) becomes

$$\frac{e^u(u-1)}{u^2}(e^{-u} - c) < \frac{1}{u}, \qquad \forall\, u > 0,\ \forall\, c \in (0, 1 - \delta h),$$

that is,

$$-ce^u(u-1) < 1.$$

Since the function on the left-hand side is decreasing in u and is equal to c if $u = 0$, this condition is satisfied for all $c \in (0,1)$ and $u > 0$ and Theorem 4.7 applies.

Similarly, if $F(u) = (u+\alpha)^{p+1}$, with $\alpha \geq 0$ and $p > 0$, condition (4.7) can be written as

$$\frac{[(p+1)u - (u+\alpha)](u+\alpha)^p}{u^2}\left[\frac{(u+\alpha)^{-(p)}}{p} - c\right] < \frac{1}{u},$$

for all $u > 0$ and $c \in (0, \frac{1}{p\alpha^p} - \delta h)$ (or $c > 0$ if $\alpha = 0$), which becomes after simplifications

$$-c[pu - \alpha](u+\alpha)^p < \frac{\alpha}{p}.$$

If $\alpha = 0$, this condition is clearly satisfied for all $u > 0$, $c > 0$. If $\alpha > 0$, the function on the left-hand side is again decreasing in u and is equal to $c\alpha^{p+1}$ when $u = 0$, so that we end up with the condition $c < 1/(p\,\alpha^p)$. In both cases, Theorem 4.7 applies. □

Minimal Time of Existence of the Solution

If the function F satisfies the hypothesis of Theorem 4.7, and either $F(0) \neq 0$ or F satisfies the hypothesis of Theorem 4.5, then it remains to show that the condition (4.5) is satisfied for a positive number of steps.

Theorem 4.8. *If the function F satisfies the hypothesis of Theorem 4.7, and either $F(0) \neq 0$ or F satisfies the hypothesis of Theorem 4.5, the scheme* (4.2) *has a positive solution u_n for n such that $t_n = nh < T_1$, where*

$$T_1 = \frac{1}{\delta} g(\|u_0\|_\infty) = \int_{\|u_0\|_\infty}^{\infty} \frac{ds}{\delta F(s)}.$$

This theorem gives a lower bound on the numerical blow-up time equal to the one given by Kaplan in [81] for the exact solution.

Proof. We want to prove that if $t_n < T_1$, that is

$$\|u_0\|_\infty < G(\delta t_n),$$

we have $\|u_{n-1}\|_\infty < G(\delta h)$ so that u_n is well-defined. To obtain this result, we prove by induction that if $\|u_0\|_\infty < G(\delta t_n)$, then u_n is well-defined and satisfies

$$\|u_n\|_\infty \leq G(g(\|u_0\|_\infty) - \delta t_n). \tag{4.8}$$

For this, we will need in particular the following result which comes from Theorem 4.3:

$$\text{if} \quad \|u_n\|_\infty < G(\delta h), \quad \text{then } \|u_{n+1}\|_\infty \leq C_n = G(g(\|u_n\|_\infty) - \delta h). \tag{4.9}$$

By choosing $n = 0$ in (4.9), we obtain the initial step of the induction, in particular (4.8) for $n = 1$. We suppose now that for some fixed n, if $\|u_0\|_\infty < G(\delta t_n)$, then u_n is well-defined and satisfies (4.8), and we also suppose that

$$\|u_0\|_\infty < G(\delta t_{n+1}). \tag{4.10}$$

Since G is decreasing, (4.10) implies $\|u_0\|_\infty < G(\delta t_n)$ and then by induction hypothesis, we get

$$\|u_n\|_\infty \leq G(g(\|u_0\|_\infty) - \delta t_n). \tag{4.11}$$

Moreover from (4.10) we also get

$$g(\|u_0\|_\infty) > \delta t_{n+1},$$

that we write as

$$g(\|u_0\|_\infty) - \delta t_n > \delta h.$$

Inserting this estimate into (4.11), and using that G is decreasing, we obtain

$$\|u_n\|_\infty < G(\delta h),$$

which implies that u_{n+1} is well-defined. Moreover using Theorem 4.3, we have

$$\|u_{n+1}\|_\infty \leq G(g(\|u_n\|_\infty) - \delta h).$$

Since from (4.11), we get that

$$g(\|u_n\|_\infty) - \delta h \geq g(\|u_0\|_\infty) - \delta t_{n+1},$$

we obtain

$$\|u_{n+1}\|_\infty \leq G(g(\|u_0\|_\infty) - \delta t_{n+1})$$

and the induction is proved. □

Computation of the numerical solutions

In this section, we introduce a specific fixed-point iteration method to solve $Au = f(x,u)$, namely

$$\begin{cases} -\Delta v_k - \varphi(x) v_k &= f(x, v_{k-1}) - \varphi(x) v_{k-1}, & \text{in } \Omega, \\ v_k &= 0, & \text{on } \partial\Omega. \end{cases} \tag{4.12}$$

where the preconditioning function $\varphi \in C^\gamma(\bar{\Omega})$ is non-positive and satisfies (4.14).

The iteration scheme was first presented by Courant and Hilbert in [40]; however we present the results in the form given by Keller in [85]. The proof of the following theorem follows the proofs of Theorems 4.1 and 4.2 in [85].

Theorem 4.9 (Keller)**.** *Consider the problem*

$$\begin{cases} -\Delta u &= f(x,u), \quad \text{in } \Omega, \\ u &= 0, \qquad\quad \text{on } \partial\Omega. \end{cases} \tag{4.13}$$

Suppose there exist two constants $M \geq 0$ and $m \leq 0$ and a non-positive function $\varphi(x) \in C^\gamma(\bar{\Omega})$ such that the function f satisfies $f(x,u) \in C^\gamma(\bar{\Omega}\times[m,M])$, $f(x,m) \geq 0$ and $f(x,M) \leq 0$ on Ω, and

$$\frac{f(x,z_1) - f(x,z_2)}{z_1 - z_2} \geq \varphi(x) \quad \text{on } \Omega, \quad \text{for all } z_1, z_2 \in [m,M]. \tag{4.14}$$

If $v_0(x) \in C^\gamma(\bar{\Omega})$ is a supersolution (resp. subsolution) of problem (4.13), with $m \leq v_0(x) \leq M$, then problem (4.13) has at least one solution $u(x) \in C^{2+\gamma}(\bar{\Omega})$, with $m \leq u(x) \leq M$ and given by

$$u(x) = \lim_{n\to\infty} v_n(x),$$

where the monotone non-increasing (resp. non-decreasing) sequence $\{v_n(x)\}$ is defined by (4.12).

Proof. The operator L defined by $Lu = \Delta u + \varphi(x)u$ is elliptic, with $\varphi(x) \leq 0$, and $\varphi \in C^\gamma(\bar{\Omega})$, $f \in C^\gamma(\bar{\Omega} \times [m,M])$, so from Schauder's theory (see Theorem 6.14 in [64]), problem (4.12) has a unique solution lying in $C^{2+\gamma}(\bar{\Omega})$. Hence the sequence $\{v_k\}$ is well-defined and $v_k \in C^{2+\gamma}(\bar{\Omega})$ for all $k \geq 1$.

We only prove the monotonicity of the sequence for the case where v_0 is a supersolution so that the sequence $\{v_k\}$ satisfies

$$m \leq \cdots \leq v_k(x) \leq \cdots \leq v_0(x) \leq M, \tag{4.15}$$

because the proof is similar if v_0 is a subsolution. First, we show that $v_1 \leq u_0$ on $\bar{\Omega}$. We have from (4.12),

$$-\Delta(v_1 - v_0) - \varphi(x)(v_1 - v_0) = f(x, v_0) - \varphi(x)v_0 + \Delta v_0 + \varphi(x)v_0 = f(x, v_0) + \Delta v_0,$$

which is negative since v_0 is a supersolution of problem (4.13). Thus we have $L(v_1 - v_0) \geq 0$ and from the strong maximum principle (see Theorem 2.1 of [85]) we obtain $v_1 - v_0 \leq 0$ on $\bar{\Omega}$.

By hypothesis, we have $m \leq v_0 \leq M$. We then suppose that $m \leq v_k(x) \leq M$ and we show that $v_{k+1} \geq m$ on $\bar{\Omega}$. Choosing $z_1 = v_k$ and $z_2 = m$ in (4.14), we obtain

$$\frac{f(x, v_k) - f(x, m)}{v_k - m} \geq \varphi(x),$$

which gives, since $v_k \geq m$,

$$f(x, v_k) - \varphi(x)v_k \geq f(x, m) - \varphi(x)m.$$

Using (4.12), it becomes

$$-\Delta v_{k+1} \geq f(x, m) + \varphi(x)(v_{k+1} - m).$$

Since $f(x, m) \geq 0$ and $\varphi(x) \leq 0$ by hypothesis, we have $\Delta v_{k+1} \leq 0$ on $\Omega \cap \{x \in \bar{\Omega} \,|\, v_{k+1}(x) \leq m\}$. Using Theorem 2.2 of [85], which is a consequence of the maximum principle, we obtain $v_{k+1} \geq m$ on $\bar{\Omega}$.

Finally, we need to show that if $m \leq v_k(x) \leq v_{k-1}(x) \leq M$ on $\bar{\Omega}$, we have $v_{k+1} \leq v_k$ on $\bar{\Omega}$. We consider

$$-\Delta(v_{k+1} - v_k) - \varphi(x)(v_{k+1} - v_k) = f(x, v_k) - f(x, v_{k-1}) - \varphi(x)(v_k - v_{k-1}).$$

Since $v_k - v_{k-1} \leq 0$, choosing $z_1 = v_k$ and $z_2 = v_{k-1}$ in (4.12) leads to

$$-\Delta(v_{k+1} - v_k) - \varphi(x)(v_{k+1} - v_k) \leq 0,$$

and using as above Theorem 2.1 of [85], we obtain $v_{k+1} \leq v_k$ on $\bar{\Omega}$.

Hence the monotonicity of the sequence $\{v_k(x)\}$ is established, together with (4.15). As the sequence is monotone and uniformly bounded, it converges to some function $\hat{u}$ defined by

$$\hat{u}(x) = \lim_{k\to\infty} v_k(x).$$

To show that $\hat{u}$ belongs to $C^{2+\gamma}(\bar{\Omega})$ and is a solution of (4.13), we will use the Compactness Theorem 12.2 in [7], however we first need to show that the convergence is uniform on $\bar{\Omega}$.

From Morrey's inequality (see Section 5.6.2 in [45]), we have

$$\max_{x,\xi\in\bar{\Omega}} \frac{|v_k(x) - v_k(\xi)|}{|x-\xi|^\alpha} \leq K_0 \|v_k\|_{1,p}\,, \tag{4.16}$$

for some constant K_0 independent of v_k, and $\alpha = 1 - \frac{d}{p}$, for any $p \geq d$ (recall that $\Omega \subset \mathbb{R}^d$). Moreover, the following estimate, taken from [139],

$$\|u\|_{s,p} \leq K_1(\|Au\|_{s-2,p} + \|u\|_{s-2,p}),$$

leads to, using (4.12) and letting $s = 2$,

$$\|v_k\|_{2,p} \leq K_1(\|f(x, v_{k-1}) - \varphi(x)v_{k-1}\|_{0,p} + \|v_k\|_{0,p}).$$

Since $f \in C^\gamma(\bar{\Omega} \times [m, M])$, $\varphi \in C^\gamma(\bar{\Omega})$ and $u_k(x) \leq M$ on $\bar{\Omega}$ for all k, there exists a constant K_2 such that

$$\|v_k\|_{2,p} \leq K_2, \qquad \text{for all } k \geq 0.$$

Hence inequality (4.16) becomes

$$\max_{x,\xi\in\bar{\Omega}} \frac{|v_k(x) - v_k(\xi)|}{|x-\xi|^\alpha} \leq K_0 K_2,$$

and the v_k are uniformly Hölder continuous, from which the equicontinuity and the uniform convergence of $\{v_k(x)\}$ to $\hat{u}(x)$ follow.

Thus we can apply the Compactness Theorem 12.2 of [7] with $L_i = L$ for all i and $F_i = -f(x, v_{i-1}) + \varphi(x)v_{i-1}$ which converges uniformly to $-f(x, \hat{u}) + \varphi(x)\hat{u}$. Hence a subsequence of $\{v_k\}$ converges to a solution of (4.12) that belongs to $C^{2+\gamma}(\bar{\Omega})$ and this solution must be $\hat{u}$ by monotonicity of the sequence $\{v_k\}$.

□

We consider functions F that satisfy the hypothesis of Theorem 4.8 so that the scheme has a unique positive solution $0 < u(x) < C_n$. In order to apply Theorem 4.9 with $m = 0$ and $M = C_n$ to scheme (4.2), we need to find a function φ that satisfies (4.14). Indeed, the conditions $f(x, 0) \geq 0$ and $f(x, C_n) \leq 0$ on Ω were proved in Theorem 4.3. We proved in Corollaries 4.1 and 4.2 that two important functions F satisfy the hypothesis of Theorem 4.8. We now show that they also satisfy the hypothesis of Theorem 4.9. Note that since the constant C_n is a supersolution, one can use $v_0 = C_n$ as the initial iterate.

Corollary 4.3. *If $F(u) = e^u$, the solution of the scheme is given by iteration* (4.12) *with*

$$\varphi(x) = -\frac{1}{h}(g(u_n) - \delta h)e^{C_n},$$

where $C_n = G(g(\|u_n\|_\infty) - \delta h)$, with $g(s) = e^{-s}$ and $G(s) = -\ln(s)$.

Proof. We define $c(x) = g(u_n) - \delta h > 0$ (by condition (4.5)). In order to apply Theorem 4.9, we need to prove that

$$\frac{f(x, u) - f(x, v)}{u - v} \geq -\frac{c(x)}{h}e^{C_n}, \quad \forall\, v \leq u \in [0, C_n]. \tag{4.17}$$

Recall that

$$f(x, u) = \frac{1}{h}e^u(e^{-u} - c(x)) = \frac{1}{h}(1 - e^u c(x)).$$

Suppose first that $v < u$. After simplifications inequality (4.17) becomes

$$\frac{e^u - e^v}{u - v} \leq e^{C_n}.$$

We rewrite it as

$$e^u \frac{1-e^{-(u-v)}}{u-v} \leq e^{C_n},$$

so all we need is to check that for $x > 0$,

$$\frac{1-e^{-x}}{x} \leq 1,$$

which is obtained using the inequality

$$1 - x \leq e^{-x}.$$

Finally, we consider the case $u = v$. We have for u fixed,

$$\lim_{v\to u} \frac{f(x,u)-f(x,v)}{u-v} = -\frac{c(x)}{h} \lim_{v\to u} \frac{e^u - e^v}{u-v} = -\frac{c(x)}{h} e^u \geq -\frac{c(x)}{h} e^{C_n}.$$

Hence Theorem 4.9 applies with $\varphi(x) = -\frac{c(x)}{h}\, e^{C_n} < 0$ and the proof is complete. □

Corollary 4.4. *If $F(u) = (u+\alpha)^{p+1}$, $\alpha \geq 0$, $p > 0$, the solution of the scheme is given by iteration* (4.12) *with*

$$\varphi(x) = \frac{1}{h}\left[\frac{1}{p} - (p+1)(C_n+\alpha)^p (g(u_n) - \delta h)\right],$$

where $C_n = G(g(\|u_n\|_\infty) - \delta h)$, with $g(s) = \frac{1}{p}(s+\alpha)^{-p}$ and $G(s) = (ps)^{-1/p} - \alpha$.

Proof. We define $c(x) = g(u_n) - \delta h > 0$ (by condition (4.5)). In order to apply Theorem 4.9, we need to prove that

$$\frac{f(x,u)-f(x,v)}{u-v} \geq \frac{1}{h}\left[\frac{1}{p} - (p+1)(C_n+\alpha)^p c(x)\right], \quad \forall\, u \leq v \in [0, C_n]. \tag{4.18}$$

Recall that $f(x,u) = \frac{1}{h}F(u)(g(u) - c(x))$, that is

$$f(x,u) = \frac{1}{h}(u+\alpha)^{p+1}\left(\frac{1}{p(u+\alpha)^p} - c(x)\right) = \frac{1}{h}\left(\frac{u+\alpha}{p} - c(x)(u+\alpha)^{p+1}\right).$$

We suppose first that $u < v$, then

$$\begin{aligned}\frac{f(x,u)-f(x,v)}{u-v} &= \frac{1}{h(u-v)}\left[\frac{u+\alpha}{p} - (u+\alpha)^{p+1}c(x) - \frac{v+\alpha}{p} + (v+\alpha)^{p+1}c(x)\right] \\ &= \frac{1}{h}\left[\frac{1}{p} - \frac{(u+\alpha)^{p+1} - (v+\alpha)^{p+1}}{u-v}\, c(x)\right],\end{aligned}$$

so the inequality (4.18) is satisfied if and only if

$$\frac{(u+\alpha)^{p+1}-(v+\alpha)^{p+1}}{u-v} \leq (p+1)\,(C_n+\alpha)^p.$$

We define $U = u+\alpha$ and $V = v+\alpha$, so that $0 \leq \alpha \leq U < V \leq C_n+\alpha$ and we need

$$\frac{U^{p+1}-V^{p+1}}{U-V} \leq (p+1)\,(C_n+\alpha)^p.$$

If $V = 0$, the inequality is satisfied. If $V \neq 0$, we define $\xi := U/V \in (0,1)$ and get

$$V^p\,\frac{\xi^{p+1}-1}{\xi-1} \leq (p+1)(C_n+\alpha)^p,$$

so all we need is the sufficient condition

$$\frac{1-\xi^{p+1}}{1-\xi} \leq p+1,$$

which is satisfied for $p > 0$ and $\xi \in (0,1)$. Now for the case $u = v$, we have

$$\lim_{v\to u}\frac{f(x,u)-f(x,v)}{u-v} = \frac{1}{h}\left[\frac{1}{p} - c(x)(p+1)(u+\alpha)^p\right].$$

Hence Theorem 4.9 applies with $\varphi(x) = \frac{1}{h}\left[\frac{1}{p} - (p+1)(C_n+\alpha)^p c(x)\right]$, if $\varphi(x)$ is negative for all x. Since $C_n = G(g(\|u_n\|_\infty) - \delta h)$ and G and g are both decreasing, we have

$$C_n \geq G(g(u_n) - \delta h) = [p\,c(x)]^{\frac{-1}{p}} - \alpha,$$

and then

$$\varphi(x) = \frac{1}{h}\left[\frac{1}{p} - (p+1)(C_n+\alpha)^p c(x)\right] \leq \frac{1}{h}\left[\frac{1}{p} - (p+1)[p\,c(x)]^{-1}c(x)\right] = -\frac{1}{h}.$$

□

In Section 3.3, we chose not to use this fixed-point method to implement the nonlinear schemes and to use Newton's method instead. We prove here the convergence of Newton's method under certain conditions on F. Using another result of Keller [85], we prove that if f satisfies the hypotheses of Theorem 4.9 and f_u is decreasing and if the first iterate w_0 is a supersolution, then Newton's method converges monotonically to the solution of the problem.

Theorem 4.10 (Keller)**.** *Suppose that f satisfies the hypothesis of Theorem 4.9 and*

$$f_u(x,u) \in C^\gamma(\bar{\Omega} \times [m,M]),$$

and

$$f_u(x,z_1) \geq f_u(x,z_2), \quad \forall\, x \in \Omega,\ 0 \leq z_1 \leq z_2 \leq M.$$

Then the unique solution $u(x) \in [m,M]$ of problem (4.13) *is given by*

$$u(x) = \lim_{n\to\infty} w_n(x),$$

where $\{w_n(x)\}$, the Newton iterates, form a monotone non-increasing sequence defined by

$$\begin{aligned} Aw_{n+1} - f_u(x,w_n)w_{n+1} &= f(x,w_n) - f_u(x,w_n)w_n, \qquad \text{in } \Omega, \\ w_{n+1} &= 0 \qquad\qquad\qquad\qquad\quad \text{on } \partial\Omega, \end{aligned} \tag{4.19}$$

with an initial iterate w_0 satisfying $Aw_0 \geq f(w_0)$ and $f_u(x,w_0) \leq 0$ in Ω and $w_0 \geq 0$ on $\partial\Omega$.

We proved in Corollaries 4.1 and 4.2 that two important examples of functions F satisfy the hypotheses of Theorem 4.8. We now show that they also satisfy the hypotheses of Theorem 4.10 with $m = 0$ and $M = C_n$.

Corollary 4.5. *If $F(u) = e^u$ or $F(u) = (u+\alpha)^{p+1}$, the solution of* (4.13) *can be obtained by Newton's iteration* (4.19)*, with $w_0 = C_n$.*

Proof. If $F(u) = e^u$, we have $f_u(x,u) = -\frac{1}{h}(g(u_n) - \delta h)e^u \in C^\gamma(\bar{\Omega} \times [0, C_n])$. Since $g(u_n) - \delta h$ is positive, f_u is negative and decreasing in u, so Theorem 4.10 applies with $w_0 = C_n$.

Similarly, if $F(u) = (u+\alpha)^{p+1}$, we have

$$f_u(x,u) = \frac{1}{ph} - \frac{1}{h}(p+1)(g(u_n) - \delta h)(u+\alpha)^p \in C^\gamma(\bar{\Omega} \times [0, C_n]),$$

which is decreasing in u and since C_n satisfies $g(C_n) \leq g(u_n) - \delta h$, that is

$$(g(u_n) - \delta h)(C_n + \alpha)^p \geq \frac{1}{p},$$

we have $f_u(x, C_n) \leq -1/h$. Hence we can apply Theorem 4.10 with $w_0 = C_n$. □

4.1.2 Rate of Growth

For specific functions F, the rate of growth of the exact solution close to the blow-up has been approximated. In this section we derive similar results for the solution of scheme (4.2).

Since the solution of $y_t = \delta F(y)$ is given by $y(t) = G(\delta(T-t))$, where T is the blow-up time, we expect

$$u(t) \sim G(\delta(T-t)),$$

close to the blow-up. In [53], Friedman and McLeod proved that if $F(u) = u^p$ (and then $G(u) = [(p-1)u]^{-1/(p-1)}$) and $\delta = 1$, solutions $u(x,t)$ with suitable initial-boundary conditions satisfy

$$(T-t)^{\frac{1}{p-1}} u(x,t) \to \frac{1}{p-1}^{\frac{1}{p-1}}, \qquad \text{as } t \to T^-,$$

provided $|x| \leq C(T-t)^{1/2}$, for some $C > 0$. For $F(u) = e^u$ (and $G(u) = 1/\ln u$), $\delta = 1$ and $n = 1$ or 2, Bebernes and al [17] proved that the solutions $u(x,t)$ satisfy

$$u(x,t) - \ln \frac{1}{T-t} \to 0, \qquad \text{as } t \to T^-,$$

uniformly on $|x| \leq C(T-t)^{1/2}$, $C \geq 0$.

Similarly, if $F(u) = e^u$, we expect that the numerical solution satisfies

$$u_n(x) \sim \ln\left(\frac{1}{\delta(T^* - t_n)}\right),$$

for some T^* and for values of x close to the blow-up point, and then

$$u_{n+1}(x) - u_n(x) \sim \ln\left(\frac{1}{\delta(T^* - t_{n+1})}\right) - \ln\left(\frac{1}{\delta(T^* - t_n)}\right) = \ln\left(\frac{T^* - t_n}{T^* - t_{n+1}}\right).$$

This motivates the following theorem.

Theorem 4.11. *Let C_0 be a constant such that*

$$C_0 \geq \delta e^{\|u_0\|_\infty} \quad \text{and} \quad Au_0 - \delta e^{u_0} + C_0 \geq 0. \tag{4.20}$$

Note that if $Au_0 \geq 0$, we can take $C_0 = \delta e^{\|u_0\|_\infty}$.

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

$$Au_{n+1} - \delta e^{u_{n+1}} + \frac{1}{h}(e^{u_{n+1}-u_n} - 1) = 0, \tag{4.21}$$

satisfies for all x

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

Proof. First, let's prove that if $t_1 = h < T_2$, then

$$u_1 \leq u_0 + \ln\left(\frac{T_2}{T_2 - h}\right). \tag{4.22}$$

The function $u_0 + \ln\left(\frac{T_2}{T_2-h}\right)$ is a supersolution of (4.21) for $n = 0$ if

$$Au_0 \geq \delta e^{u_0}\left(\frac{T_2}{T_2 - h}\right) + \frac{1}{h}\left(\frac{T_2}{T_2 - h} - 1\right) = \frac{1}{T_2 - h}(\delta e^{u_0} T_2 - 1).$$

Since $\frac{1}{T_2} = C_0 \geq \delta e^{\|u_0\|}$, the right-hand side is decreasing in h so in order to get a bound valid for all $h \in (0, T_2)$, we need

$$Au_0 \geq \lim_{h\to 0} \frac{1}{T_2 - h}(\delta e^{u_0} T_2 - 1) = \delta e^{u_0} - \frac{1}{T_2},$$

which is exactly condition (4.20), hence we get (4.22).

We now assume that

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

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

$$\begin{aligned} u_n &\leq u_{n-1} + \ln\left(\frac{T_2 - t_{n-1}}{T_2 - t_n}\right) \\ &\leq \ln\left(\frac{1}{\delta(T_2 - t_{n-1})}\right) + \ln\left(\frac{T_2 - t_{n-1}}{T_2 - t_n}\right) = \ln\left(\frac{1}{\delta(T_2 - t_n)}\right). \end{aligned} \tag{4.23}$$

The function $u_n + \ln\left(\frac{T_2-t_n}{T_2-t_{n+1}}\right)$ is a supersolution of the scheme (4.21) if

$$Au_n - \delta e^{u_n}\left(\frac{T_2-t_n}{T_2-t_{n+1}}\right) + \frac{1}{h}\left(\frac{T_2-t_n}{T_2-t_{n+1}} - 1\right) \geq 0. \tag{4.24}$$

Since u_n is solution of

$$Au_n = \delta e^{u_n} - \frac{1}{h}(e^{u_n-u_{n-1}} - 1),$$

and the induction hypothesis gives

$$e^{u_n-u_{n-1}} \leq \frac{T_2-t_{n-1}}{T_2-t_n},$$

condition (4.24) is satisfied if

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

which simplifies to

$$\delta e^{u_n} \leq \frac{1}{T_2-t_n},$$

which is exactly (4.23). □

If $F(u) = (u+\alpha)^2$, we expect the numerical solution to satisfy

$$u_n \sim \frac{1}{\delta(T^*-t_n)},$$

for some T^*, so that

$$\frac{u_{n+1}}{u_n} \sim \frac{\delta(T^*-t_n)}{\delta(T^*-t_{n+1})} = \frac{T^*-t_n}{T^*-t_{n+1}}.$$

Theorem 4.12. *Suppose these exists a constant C_0 that satisfies*

$$C_0 \geq \delta(\|u_0\|_\infty + \alpha), \quad \text{and} \quad Au_0 - \delta(u_0+\alpha)^2 + C_0u_0 \geq 0. \tag{4.25}$$

Note that if $Au_0 \geq 0$, we can take $C_0 = \delta(\|u_0\|_\infty + \alpha)$.

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

$$hAu_{n+1} = (u_{n+1}+\alpha)^2\left[\frac{1}{u_{n+1}+\alpha} - \frac{1}{u_n+\alpha} + \delta h\right], \tag{4.26}$$

satisfies for all x

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

Proof. First, we need to show that $\frac{T_2}{T_2-h}u_0$ is a supersolution if $h < T_2$ to prove that (4.12) for $n = 1$. Letting $\lambda = (T_2 - h)/T_2$, the condition to satisfy is

$$\frac{hAu_0}{\lambda} \geq \left[\frac{u_0}{\lambda} + \alpha\right]^2 \left[\frac{1}{\frac{u_0}{\lambda} + \alpha} - \frac{1}{u_0 + \alpha} + \delta h\right]. \tag{4.27}$$

It becomes

$$hAu_0 \geq \frac{1}{\lambda}\left(\frac{u_0 + \alpha\lambda}{u_0 + \alpha}\right)\left[(\lambda - 1)u_0 + \delta h(u_0 + \alpha\lambda)(u_0 + \alpha)\right],$$

and substituting back $\lambda = (T_2 - h)/T_2$, we obtain

$$Au_0 \geq \frac{T_2}{T_2 - h}\left(1 - \frac{\alpha h}{T_2(u_0 + \alpha)}\right)\left[\delta(u_0 + \alpha)^2 - \frac{1}{T_2}u_0 - \frac{\alpha\delta}{T_2}h(u_0 + \alpha)\right].$$

We denote by $\beta(h)$ the function of h on the right-hand side and we prove that it is decreasing. The derivative

$$\begin{aligned}\beta'(h) = {} & \frac{T_2}{(T_2 - h)^2}\left(1 - \frac{\alpha h}{T_2(u_0 + \alpha)}\right)\left[\delta(u_0 + \alpha)^2 - \frac{1}{T_2}u_0 - \frac{\alpha\delta}{T_2}h(u_0 + \alpha)\right] \\ & + \frac{T_2}{T_2 - h}\left(-\frac{\alpha}{T_2(u_0 + \alpha)}\right)\left[\delta(u_0 + \alpha)^2 - \frac{1}{T_2}u_0 - \frac{\alpha\delta}{T_2}h(u_0 + \alpha)\right] \\ & + \frac{T_2}{T_2 - h}\left(1 - \frac{\alpha h}{T_2(u_0 + \alpha)}\right)\left(-\frac{\alpha\delta}{T_2}(u_0 + \alpha)\right),\end{aligned}$$

is negative if

$$\begin{aligned}&\frac{1}{T_2 - h}\left(1 - \frac{\alpha h}{T_2(u_0 + \alpha)}\right)\left[\delta(u_0 + \alpha)^2 - \frac{1}{T_2}u_0 - \frac{\alpha\delta}{T_2}h(u_0 + \alpha)\right] - \left(\frac{\alpha}{T_2(u_0 + \alpha)}\right) \\ &\left[\delta(u_0 + \alpha)^2 - \frac{1}{T_2}u_0 - \frac{\alpha\delta}{T_2}h(u_0 + \alpha)\right] - \left(1 - \frac{\alpha h}{T_2(u_0 + \alpha)}\right)\left(\frac{\alpha\delta}{T_2}(u_0 + \alpha)\right) < 0.\end{aligned}$$

Expanding and simplifying, we obtain

$$-\alpha^2\delta(u_0 + \alpha)h^2 + 2T_2\alpha^2\delta(u_0 + \alpha)h + (\delta T_2^2(u_0 + \alpha)^2(u_0 - \alpha) - T_2u_0^2) < 0.$$

Since the leading coefficient is negative, this quadratic polynomial $P(h)$ attains its maximum at

$$h = \frac{2T_2\alpha^2\delta(u_0 + \alpha)}{2\alpha^2\delta(u_0 + \alpha)} = T_2,$$

so we only need to show that $P(T_2)$ is negative. We have

$$\begin{aligned} P(T_2) &= -\alpha^2\delta(u_0+\alpha)T_2^2 + 2T_2\alpha^2\delta(u_0+\alpha)T_2 + \delta T_2^2(u_0+\alpha)^2(u_0-\alpha) - T_2u_0^2 \\ &= T_2[\alpha^2\delta(u_0+\alpha)T_2 + \delta T_2(u_0^2-\alpha^2)(u_0+\alpha) - u_0^2] \\ &= T_2u_0^2\left[\delta\, T_2\,(u_0+\alpha) - 1\right], \end{aligned}$$

and since $T_2 \leq \frac{1}{\delta(\|u_0\|+\alpha)}$, $P(T_2)$ is negative and the function $\beta(h)$ is decreasing. Hence condition (4.27) becomes

$$Au_0 \geq \lim_{h\to 0}\beta(h) = \delta(u_0+\alpha)^2 - \frac{1}{T_2}u_0,$$

which is exactly (4.25).

We now assume that $\lambda_n u_n \leq u_{n-1}$, with

$$\lambda_n = \frac{T_2 - t_n}{T_2 - t_{n-1}},$$

and we prove that u_n/λ_{n+1} is a supersolution of (4.26).

First we note that we have

$$u_0 < \frac{1}{\delta T_2} - \alpha,$$

and by induction

$$\begin{aligned} u_n &\leq \frac{T_2 - t_{n-1}}{T_2 - t_n}u_{n-1} \\ &< \frac{T_2 - t_{n-1}}{T_2 - t_n}\left(\frac{1}{\delta(T_2 - t_{n-1})} - \alpha\right) \\ &< \frac{1}{\delta(T_2 - t_n)} - \alpha, \end{aligned}$$

where we used that

$$\frac{T_2 - t_{n-1}}{T_2 - t_n} > 1.$$

The function u_n/λ_{n+1} is a supersolution if

$$\frac{hAu_n}{\lambda_{n+1}} \geq \left(\frac{u_n}{\lambda_{n+1}} + \alpha\right)^2\left[\frac{\lambda_{n+1}}{u_n + \alpha\lambda_{n+1}} - \frac{1}{u_n+\alpha} + \delta h\right]. \tag{4.28}$$

Since u_n satisfies

$$hAu_n = (u_n + \alpha)^2 \left[\frac{1}{u_n + \alpha} - \frac{1}{u_{n-1} + \alpha} + \delta h \right],$$

and the induction hypothesis gives

$$\frac{-1}{\lambda_n u_n + \alpha} \leq \frac{-1}{u_{n-1} + \alpha},$$

condition (4.28) is satisfied if

$$\begin{aligned}(u_n + \alpha)^2 &\left[\frac{1}{u_n + \alpha} - \frac{1}{\lambda_n u_n + \alpha} + \delta h \right] \\ &\geq \lambda_{n+1} \left(\frac{u_n + \alpha\lambda_{n+1}}{\lambda_{n+1}} \right)^2 \left[\frac{\lambda_{n+1}}{u_n + \alpha\lambda_{n+1}} - \frac{1}{u_n + \alpha} + \delta h \right].\end{aligned}$$

Expanding, simplifying and collecting the terms in u_n, we obtain

$$\begin{aligned}&- \delta h \lambda_n (1 - \lambda_{n+1}) u_n^4 + [-\alpha\delta h (1 + \lambda_n)(1 - \lambda_{n+1}) + (\lambda_n - \lambda_{n+1})] u_n^3 \\ &+ \alpha[-\alpha\delta h (1 - \lambda_{n+1})(1 - \lambda_n\lambda_{n+1}) + (1 - 3\lambda_{n+1} + 3\lambda_{n+1}\lambda_n - \lambda_{n+1}^2\lambda_n)] u_n^2 \\ &+ \alpha^2 \lambda_{n+1} [\alpha\delta h (1 - \lambda_{n+1})(1 + \lambda_n) + (\lambda_n - \lambda_{n+1})] u_n \\ &+ \alpha^4 \delta h \lambda_{n+1} (1 - \lambda_{n+1}) \geq 0.\end{aligned}$$

Substituting back λ_n and λ_{n+1}, we simplify further

$$\begin{aligned}&- \delta(T_2 - t_n) u_n^4 + (1 - \alpha\delta(T_2 - t_n + T_2 - t_{n-1})) u_n^3 + 2\alpha(1 - \alpha\delta h) u_n^2 \\ &\quad + \alpha^2 \lambda_{n+1} (1 + \alpha\delta(T_2 - t_{n-1} + T_2 - t_n)) u_n + \alpha^4 \delta \lambda_{n+1} (T_2 - t_{n-1}) \geq 0.\end{aligned}$$

We denote by P this polynomial of degree 4 in u_n. We note that the leading coefficient is negative, the second coefficient may be positive or negative and the remaining ones are all positive. If we consider the second derivative of P, we remark that it is a quadratic polynomial, which is concave (as the leading coefficient of P (and thus P'') is negative) and that $P''(0) = 4\alpha(1 - \alpha\delta h) > 0$, so that P'' admits two roots, one positive X_+ and one negative X_-. Moreover, since $P'(0) > 0$, we have the following analysis

		X_-		0		X_+	
P''	-	0	+		+	0	-
P'	$\searrow$		$\nearrow$	+	$\nearrow$		$\searrow$

Hence there exists $\tilde{x}$ such that P' is positive for all $x \in [0, \tilde{x}]$ and negative for $x > \tilde{x}$. We conclude that since $P(0) > 0$, there exists $\underline{x}$ such that P is positive on $[0, \underline{x})$ and negative on $(\underline{x}, \infty)$. Since we need to prove that P is positive for $u_n \in (0, \frac{1}{\delta(T_2 - t_n)} - \alpha)$, it is enough to check that

$$P\left(\frac{1}{\delta(T_2 - t_n)} - \alpha\right) \geq 0.$$

One can compute

$$P\left(\frac{1}{\delta(T_2 - t_n)} - \alpha\right) = \frac{\alpha}{\delta^2(T_2 - t_n)^3}\left[\alpha^2\delta^2 h(T_2 - t_n)(T_2 - t_{n+1}) + (T_2 - t_{n+1})\right]$$

which is positive for $t_{n+1} < T_2$, so that u_n/λ_{n+1} is a supersolution of (4.26). □

There is no obvious way to generalize this proof for $F(u) = (u+\alpha)^{p+1}$, with $p > 0$, however we can suggest which form the theorem should take. Since we expect

$$u_n \sim \left(\frac{1}{\delta(T^* - t_n)}\right)^{1/p},$$

for some T^*, we would have

$$\frac{u_{n+1}}{u_n} \sim \left(\frac{\delta(T^* - t_n)}{\delta(T^* - t_{n+1})}\right)^{1/p} = \left(\frac{T^* - t_n}{T^* - t_{n+1}}\right)^{1/p}.$$

Conjecture 4.13. *Suppose there exists a constant C_0 that satisfies*

$$C_0 \geq \delta p\,(\|u_0\|_\infty + \alpha)^p, \qquad \text{and} \qquad Au_0 - \delta(u_0 + \alpha)^{p+1} + \frac{C_0}{p}\,u_0 \geq 0. \tag{4.29}$$

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

$$hAu_{n+1} = (u_{n+1} + \alpha)^{p+1}\left[\frac{1}{p}(u_{n+1} + \alpha)^{-p} - \frac{1}{p}(u_n + \alpha)^{-p} + \delta h\right],$$

satisfies for all x

$$u_{n+1}(x) \leq \left(\frac{T_2 - t_n}{T_2 - t_{n+1}}\right)^{1/p} u_n(x), \qquad \text{if } t_{n+1} \leq T_2.$$

Justification of condition (4.29)**:** We obtain the condition on u_0 by following an approach close to the one used in the case $p = 1$.

For the initial step, we would need to prove that $\left(\frac{T_2}{T_2-h}\right)^{1/p} u_0$ is a supersolution, that is

$$Au_0 \geq \frac{1}{h\lambda}\left(\lambda u_0 + \alpha\right)^{p+1}\left[\frac{1}{p}\left(\lambda u_0 + \alpha\right)^{-p} - \frac{1}{p}(u_0 + \alpha)^{-p} + \delta h\right].$$

with $\lambda = \left(\frac{T_2}{T_2-h}\right)^{1/p}$. Though we are not able to prove that the function of the right-hand side is non-increasing in h, it seems to be the case, so that this inequality is satisfied if

$$Au_0 \geq \lim_{h\to 0} \frac{1}{h\lambda}\left(\lambda u_0 + \alpha\right)^{p+1}\left[\frac{1}{p}\left(\lambda u_0 + \alpha\right)^{-p} - \frac{1}{p}(u_0 + \alpha)^{-p} + \delta h\right],$$

that is

$$Au_0 \geq \delta(u_0 + \alpha)^{p+1} - \frac{1}{pT_2}u_0.$$

4.2 B-Method Obtained by Splitting Methods and Backward Euler (SpFEA)

In this section, we prove results analogous to the ones derived in Section 4.1, for the scheme obtained by using the backward Euler method in the composition method. This scheme was derived in Section 3.1.1 and can be written in the form

$$Au_{n+1} = f(x, u_{n+1}) = -\frac{1}{h}u_{n+1} + \frac{1}{h}G(g(u_n) - \delta h). \tag{4.30}$$

As we mentioned at the beginning of the chapter, the proofs of the existence and uniqueness of a positive solution are easily obtained: they directly follow from Amann's results [8]. The lower bound for the numerical blow-up time and the rate of growth of the solution are identical to the ones derived is Section 4.1 for the first method and

the upper bound for the numerical blow-up time derived in Section 4.2.3 is the same as the one obtained by Kaplan [81] for the exact solution. Since the scheme is linear, the numerical solution can be computed without the use of a specific solver.

4.2.1 Existence and Uniqueness of the Solution

Since scheme (4.30) is linear, it has a unique solution if and only if $G(g(u_n) - \delta h)$ is well-defined, that is

$$g(u_n) \in (\delta h, M + \delta h).$$

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 $\|u_n\|_\infty < G(\delta h)$. This result can also be obtained using Theorem 4.2, which leads to the following theorem, identical to Theorem 4.3.

Theorem 4.14. *If the function u_n is positive in Ω, continuous in $\bar{\Omega}$, and satisfies*

$$\|u_n\|_\infty < G(\delta h), \tag{4.31}$$

then scheme (4.30) *has a maximal nonnegative solution*

$$\hat{u} \leq C_n = G(g(\|u_n\|_\infty) - \delta h),$$

and a minimal solution $\bar{u} \geq 0$ and if u is a solution, then $u \in C^2(\bar{\Omega})$ and satisfies $\bar{u} \leq u \leq \hat{u}$.

Note that condition (4.31) is the same as condition (4.5), so that we can make the bound on the right-hand side as large as desired by choosing h small enough. This condition is necessary for scheme (4.30) 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 (4.31) is satisfied and positive by definition of G (see Assumption 4.1), is a supersolution. Moreover, since

$$f(x,0) = \frac{1}{h} G(g(u_n) - \delta h) > 0,$$

we can apply Theorem 4.2 and we get the result. □

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 [158]) and any nonnegative solution is positive on Ω. Uniqueness of the positive solution can also be obtained using the following result of Keller [85] with $m = 0$ and $M = C_n$.

Theorem 4.15 (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 (4.4) *has at most one solution $u \in C^2$ satisfying $m \leq u \leq M$.*

Since $f(x,u)$ defined in (4.30) is decreasing in u, we get the uniqueness of the solution. Hence Theorem 4.8 applies to scheme (4.30) and we obtain the same minimal time of existence for the solution:

Theorem 4.16. *Scheme* (4.30) *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 4.14 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 Theorem (4.8).

Finally, we recall that scheme (4.30) is linear so that no particular solver is required.

4.2.2 Rate of Growth

In this section we prove that Theorems 4.11 and 4.12 of Section 4.1.2 are also valid when scheme (4.30) is used.

Theorem 4.17. *Let C_0 be a constant such that*

$$C_0 \geq \delta e^{\|u_0\|_\infty} \quad \text{and} \quad Au_0 - \delta e^{u_0} + C_0 \geq 0. \tag{4.32}$$

Note that if $Au_0 \geq 0$, we can take $C_0 = \delta e^{\|u_0\|_\infty}$.

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

$$u_{n+1} + hAu_{n+1} = -\ln(e^{-u_n} - \delta h). \tag{4.33}$$

satisfies for all x

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

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

$$u_1 \leq u_0 + \ln\left(\frac{T_2}{T_2 - h}\right). \tag{4.34}$$

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 (4.33) 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

$$Au_0 \geq \frac{1}{h}\ln(1 - hC_0) - \frac{1}{h}\ln(1 - \delta h e^{u_0}). \tag{4.35}$$

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 (4.35) 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 (4.32), and we get (4.34).

To complete the induction we assume that

$$u_n \leq u_{n-1} + \ln\left(\frac{T_2 - t_{n-1}}{T_2 - t_n}\right), \tag{4.36}$$

and we show that $u_n + \ln\left(\frac{T_2-t_n}{T_2-t_{n+1}}\right)$ is a supersolution of (4.33), that is

$$u_n + \ln\left(\frac{T_2 - t_n}{T_2 - t_{n+1}}\right) + hAu_n + \ln(e^{-u_n} - \delta h) \geq 0. \tag{4.37}$$

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

$$\begin{aligned}u_n &\leq u_{n-1} + \ln\left(\frac{T_2 - t_{n-1}}{T_2 - t_n}\right) \\ &\leq \ln\left(\frac{1}{\delta(T_2 - t_{n-1})}\right) + \ln\left(\frac{T_2 - t_{n-1}}{T_2 - t_n}\right) = \ln\left(\frac{1}{\delta(T_2 - t_n)}\right).\end{aligned} \tag{4.38}$$

By definition of u_n, we have

$$u_n + hAu_n = -\ln(e^{-u_{n-1}} - \delta h),$$

and from the induction hypothesis (4.36), we obtain

$$-\ln(e^{-u_{n-1}} - \delta h) > -\ln\left[e^{-u_n}\left(\frac{T_2 - t_{n-1}}{T_2 - t_n}\right) - \delta h\right],$$

so that inequality (4.37) 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 (4.38). □

Theorem 4.18. *Suppose these exists a constant C_0 that satisfies*

$$C_0 \geq \delta(\|u_0\|_\infty + \alpha), \quad \text{and} \quad Au_0 - \delta(u_0 + \alpha)^2 + C_0 u_0 \geq 0. \tag{4.39}$$

Note that if $Au_0 \geq 0$, we can take $C_0 = \delta(\|u_0\|_\infty + \alpha)$.

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

$$u_{n+1} + hAu_{n+1} = \frac{1}{\frac{1}{u_n + \alpha} - \delta h} - \alpha, \tag{4.40}$$

satisfies for all x

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

Note that if $Au_0 \geq 0$, we can take $C_0 = \delta(\|u_0\|_\infty + \alpha)$.

Proof. For the initial step, we need to prove that if $h < T_2$, then $\frac{T_2}{T_2 - h}u_0$ is a supersolution of (4.40) with $n = 0$, that is

$$u_0 + hAu_0 \geq \left(\frac{T_2 - h}{T_2}\right)\left[\frac{u_0 + \alpha}{1 - \delta h(u_0 + \alpha)} - \alpha\right], \tag{4.41}$$

which simplifies to

$$Au_0 \geq \frac{1}{T_2}\frac{1}{1-\delta h(u_0+\alpha)}\left[\delta T_2(u_0+\alpha)^2 - u_0 - \delta h\alpha(u_0+\alpha)\right].$$

Let $x = \delta h(u_0+\alpha) < 1$, so that the function on the right-hand side is

$$f(x) = \frac{1}{T_2(1-x)}[\delta T_2(u_0+\alpha)^2 - u_0 - \alpha x].$$

Its derivative

$$f'(x) = \frac{-\alpha(1-x) + \delta T_2(u_0+\alpha)^2 - u_0 - \alpha x}{(1-x)^2},$$

is negative if

$$\delta T_2(u_0+\alpha)^2 - (u_0+\alpha) < 0,$$

that is

$$T_2 < \frac{1}{\delta(u_0+\alpha)}.$$

As $1/T_2 \geq \delta(\|u_0\|_\infty + \alpha)$, this condition is satisfied and f is decreasing. So condition (4.41) is satisfied if

$$Au_0 \geq \lim_{x\to 0}\frac{1}{T_2(1-x)}[\delta T_2(u_0+\alpha)^2 - u_0 - x] = \frac{1}{T_2}[\delta T_2(u_0+\alpha)^2 - u_0],$$

which is exactly (4.39).

We define

$$\lambda_n = \frac{T_2 - t_n}{T_2 - t_{n-1}} \in (0,1),$$

and we assume by induction hypothesis that $\lambda_n u_n \leq u_{n-1}$. First we note that we have $u_0 < \frac{1}{\delta T_2} - \alpha$, and by induction

$$\begin{aligned} u_n &\leq \frac{T_2 - t_{n-1}}{T_2 - t_n} u_{n-1} \\ &< \frac{T_2 - t_{n-1}}{T_2 - t_n}\left(\frac{1}{\delta(T_2 - t_{n-1})} - \alpha\right) \\ &< \frac{1}{\delta(T_2 - t_n)} - \alpha, \end{aligned}$$

where we used that

$$\frac{T_2 - t_{n-1}}{T_2 - t_n} > 1.$$

The function u_n/λ_{n+1} is a supersolution of (4.40) if

$$\frac{1}{\lambda_{n+1}}(u_n + hAu_n) \geq [(u_n + \alpha)^{-1} - \delta h]^{-1} - \alpha. \tag{4.42}$$

Since u_n satisfies

$$u_n + hAu_n = \frac{1}{\frac{1}{u_{n-1}+\alpha} - \delta h} - \alpha,$$

and the induction hypothesis gives

$$\frac{1}{\lambda_n u_n + \alpha} \geq \frac{1}{u_{n-1} + \alpha},$$

condition (4.42) is satisfied if

$$[(\lambda_n u_n + \alpha)^{-1} - \delta h]^{-1} - \alpha \geq \lambda_{n+1}[(u_n + \alpha)^{-1} - \delta h]^{-1} - \lambda_{n+1}\alpha.$$

Expanding, simplifying and collecting the terms in u_n, we get

$$\begin{aligned} &- \delta h \lambda_n (1 - \lambda_{n+1})(1 + \delta h \alpha) u_n^2 + [\lambda_n - \lambda_{n+1} - (\delta h \alpha)^2 (1 + \lambda_n)(1 - \lambda_{n+1})] u_n \\ &\quad + \delta h \alpha^2 (1 - \delta h \alpha)(1 - \lambda_{n+1}) \geq 0. \end{aligned}$$

We denote by $P(u_n)$ this quadratic polynomial. Its constant term is positive, so $P(0) \geq 0$ and since its leading coefficient is negative, if P is positive at $u_n = \frac{1}{\delta(T_2 - t_n)} - \alpha$ then P is positive on the whole interval $[0, \frac{1}{\delta(T_2 - t_n)} - \alpha]$. Algebraic manipulations lead to

$$P\left(\frac{1}{\delta(T_2 - t_n)} - \alpha\right) = \frac{\alpha h^2 (T_2 - t_{n+1})(1 + \delta h \alpha)}{(T_2 - t_n)^2 (T_2 - t_{n-1})} \geq 0.$$

Since P is positive on $[0, \frac{1}{\delta(T_2 - t_n)} - \alpha]$, u_n/λ_{n+1} is a supersolution and the theorem is proved. □

Similarly, for the case $F(u) = (u + \alpha)^{p+1}$, with $p > 0$, we obtain the following conjecture.

Conjecture 4.19. *Suppose there exists a constant C_0 that satisfies*

$$C_0 \geq \delta p \left(\|u_0\|_\infty + \alpha\right)^p \quad \text{and} \quad Au_0 - \delta(u_0+\alpha)^{p+1} + \frac{C_0}{p}\, u_0 \geq 0. \tag{4.43}$$

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

$$u_{n+1} + hAu_{n+1} = \left[(u_n+\alpha)^{-p} - p\,\delta h\right]^{-1/p} - \alpha, \tag{4.44}$$

satisfies for all x

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

Justification of condition (4.43)**:** The justification given for Conjecture 4.13 holds for scheme (4.44):

For the initial step, we would need to prove that $\left(\frac{T_2}{T_2-h}\right)^{1/p} u_0$, is a supersolution, that is

$$\left(\frac{T_2}{T_2-h}\right)^{1/p} (u_0 + hAu_0) \geq \left[(u_0+\alpha)^{-p} - \delta hp\right]^{1/p} - \alpha,$$

that we rewrite

$$Au_0 \geq -\frac{u_0}{h} + \frac{1}{h}\left(1 - \frac{h}{T_2}\right)^{1/p} \left(\frac{(u_0+\alpha)}{[1-\delta hp\,(u_0+\alpha)^p]^{1/p}}\right) - \frac{\alpha}{h}\left(1-\frac{h}{T_2}\right)^{1/p}.$$

Though we were not able to prove that the function of the right-hand side is decreasing in h, it seems to be the case, so that this inequality is satisfied if

$$Au_0 \geq \lim_{h\to 0}\left[-\frac{u_0}{h} + \frac{1}{h}\left(1 - \frac{h}{T_2}\right)^{1/p} \left(\frac{(u_0+\alpha)}{[1-\delta hp\,(u_0+\alpha)^p]^{1/p}}\right) - \frac{\alpha}{h}\left(1-\frac{h}{T_2}\right)^{1/p}\right],$$

that is

$$Au_0 \geq \delta(u_0+\alpha)^{p+1} - \frac{1}{T_2 p} u_0.$$

4.2.3 Numerical Blow-up

In this section we want to prove that for values of δ large enough, the numerical blow-up time T^* satisfies

$$T^* \leq \int_0^\infty \frac{ds}{\delta F(s) - \lambda s} < \infty.$$

Since we already proved that $T^* > T_1 = \frac{1}{\delta}\left(g(\|u_0\|_\infty)\right.$, we obtain exactly the same bounds as Kaplan in [81].

While most of our previous results were following Le Roux's approach in [99], 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 of the functionals J_n and F (defined in (2.12) and (2.13) in Chapter 2) and no equivalent functionals could be found for this scheme. Hence we chose to adapt the approach used by Kaplan for the continuous problem to our semi-discretization. This lead to the definition of numerical blow-up time given in Chapter 1. As we explained there, we need to prove that for all $K > 0$ and h small enough, there exists $n < T^*/h$ such that $\|u_n\|_\infty > K$. We now state it in our main result.

Theorem 4.20. *Suppose that δ satisfies*

$$\delta F(u) - \lambda u > 0, \qquad \forall\, u \geq 0, \tag{4.45}$$

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

$$u_{n+1} + hAu_{n+1} = G(g(u_n) - \delta h), \tag{4.46}$$

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 $\|u_{n^*}\|_\infty > K$.

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 4.1. *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* (4.31) *is satisfied and scheme* (4.46) *has a unique positive solution.*

Remark 4.2. *Condition* (4.45) *imposed on* δ *is identical to the one given by Kaplan in [81] (see Chapter 2). It can not be satisfied at* $u = 0$ *if* $F(0) = 0$, *however, if* $F(0) > 0$, *since* F *satisfies* (4.1), *we have*

$$\lim_{u\to 0} \frac{u}{F(u)} = 0 \quad \textit{and} \quad \lim_{u\to\infty} \frac{u}{F(u)} = 0,$$

and condition (4.45) *is satisfied for all* δ *large enough. For example, if we consider* $F(u) = e^u$, *condition* (4.45) *becomes* $\delta > \frac{\lambda u}{e^u}$, *for all* $u \geq 0$, *that is*

$$\delta > \frac{\lambda}{e},$$

and if we consider $F(u) = (u+\alpha)^p$, *with* $\alpha > 0$, *since the derivative of the function* $\beta(u) := u/(u+\alpha)^p$ *satisfies*

$$\beta'(u) = \frac{(u+\alpha)^p - p(u+\alpha)^{p-1}u}{(u+\alpha)^{2p}} > 0 \quad \Leftrightarrow \quad u < \frac{\alpha}{p-1},$$

and we have

$$\beta\left(\frac{\alpha}{p-1}\right) = \frac{\alpha}{(p-1)(\alpha p)^p},$$

condition (4.45) *becomes*

$$\delta > \frac{\lambda\alpha}{(p-1)(\alpha p)^p}.$$

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

$$a_n = \int_\Omega \varphi\, u_n dx, \tag{4.47}$$

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 in 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 work into the following steps:

a. We prove that (a_n) is increasing.

b. We define $a(t)$, solution of

$$\begin{cases} a'(t) &= \delta F(a(t)) - \lambda a(t), \\ a(0) &= a^* \in (0, a_0), \end{cases}$$

which blows up in finite time at $T = \int_{a^*}^{\infty} \frac{ds}{\delta F(s) - \lambda s}$ if δ satisfies condition (4.45). 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.

Jensen's Inequality

An important tool in proving finite-time blow-up is Jensen's inequality. As mentioned in Chapter 2 it has been used under different forms in several articles. We prove below the version we will need.

Lemma 4.1. *If a function $f \in C^2([0,\infty))$ is convex, and φ satisfies $\varphi \geq 0$ and $\int_\Omega \varphi(x)\, dx = 1$, we have for all functions $u(x) \geq 0$*

$$f\left(\int_\Omega \varphi\, u\, dx\right) \leq \int_\Omega f(u)\varphi\, dx. \tag{4.48}$$

Proof. First, we note that f is convex if and only if $g := f - \alpha x$ is convex, for $\alpha \in \mathbb{R}$. For some fixed $c \geq 0$, we let $\alpha = f'(c)$, so that $g'(c) = f'(c) - \alpha = 0$, and since g is convex, c is a minimum of g and $g(s) \geq g(c)$, $\forall s \geq 0$, that is

$$f(s) \geq \alpha(s - c) + f(c).$$

So, for $s = u(x)$, we get

$$f(u(x)) \geq \alpha(u(x) - c) + f(c),$$

and multiplying by $\varphi(x) \geq 0$ and integrating over Ω,

$$\int_\Omega \varphi(x) f(u(x)) dx \geq \alpha \left(\int_\Omega \varphi(x) u(x) dx - c \int_\Omega \varphi(x) dx \right) + f(c) \int_\Omega \varphi(x) dx.$$

Now, we let $c = \int_\Omega u\varphi \, dx$, then since $\int \varphi(x) dx = 1$, we obtain the Jensen's inequality (4.48). □

Growth of the sequence (a_n)

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

Lemma 4.2. *As long as u_n satisfies $\|u_n\|_\infty < G(\delta h)$, the sequence (a_n) defined in (4.47) 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 (4.46) 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 (4.48) to complete the proof. □

Lemma 4.3. *If δ satisfies condition (4.45), the sequence (a_n) defined in (4.47) 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

$$\frac{1}{1+h\lambda}G(g(x) - \delta h) > x, \tag{4.49}$$

that is

$$g(x) - g((1+h\lambda)x) < \delta h.$$

Since g is continuous, 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

$$g(x) - g((1+h\lambda)x) = g'(\xi)(x - (1+h\lambda)x),$$

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.

$$F(\xi) > \frac{\lambda x}{\delta}, \qquad \forall\, \xi \in (x, (1 + h\lambda)x).$$

Since F is increasing and δ satisfies condition (4.45), we have

$$F(\xi) > F(x) > \frac{\lambda x}{\delta}.$$

Hence inequality (4.49) holds for all $x \in (0, G(\delta h))$ and Lemma 4.2 completes the proof. □

Definition of $a(t)$ and D_n

From now on, we assume that condition (4.45) 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 4.2, we have

$$\frac{a_{n+1} - a_n}{h} \geq \frac{1}{h} \left(\frac{1}{1 + h\lambda} G(g(a_n) - \delta h) - a_n \right),$$

hence we will compare (a_n) with $(a(t_n))$ where $t_n = nh$ and $a(t)$ is the solution of

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

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

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

So $a(t)$ is the solution of

$$\begin{cases} a'(t) &= \delta F(a(t)) - \lambda a(t), \\ a(0) &= a^* < a_0. \end{cases}$$

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 4.20, 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 4.4. *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 fix number belonging to $(0, g(K))$ (see Theorem 4.20).

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

$$\begin{cases} b'(t) &= \delta F(b(t)) > \delta F(b(t)) - \lambda b(t), \\ b(t_n) &= a(t_n). \end{cases}$$

This function can be written explicitly,

$$b(t) = G(g(a(t_n)) + \delta t_n - \delta t),$$

and we have $a(t) \leq b(t)$, $\forall t \geq t_n$. Moreover since δ satisfies condition (4.45), $a(t)$ is increasing and we have

$$a(t_n + \xi) \leq a(t_n + h) \leq b(t_{n+1}) = G(g(a(t_n)) - \delta h),$$

and since $a(t_n) < a_n \leq K$ and $h < \frac{g(K)-\varepsilon}{\delta}$, we get

$$a(t_n + \xi) \leq G(g(a(t_n)) - \delta h) \leq G(g(K) - \delta h) < G(\varepsilon).$$

□

Hence D_{n+1} is well-defined and we first bound it using Lemma 4.2

$$D_{n+1} \geq \frac{1}{1 + h\lambda} G(g(a_n) - \delta h) - a(t_n + h),$$

then we consider the right-hand side as a function of h and we take a Taylor expansion around $h = 0$. We get

$$D_{n+1} \geq D_n + h(\psi(a_n) - \psi(a(t_n))) + \frac{h^2}{2}\eta(\xi), \tag{4.50}$$

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

$$\begin{aligned}\eta(\xi) = &\frac{2\lambda^2}{(1+\xi\lambda)^3} G(g(a_n) - \delta\xi) - \frac{2\delta\lambda}{(1+\xi\lambda)^2} F(G(g(a_n) - \delta\xi)) \\ &+ \frac{\delta^2}{1+\xi\lambda} F'(G(g(a_n) - \delta\xi)) F(G(g(a_n) - \delta\xi)) \\ &- [(\delta F'(a(t_n+\xi)) - \lambda)(\delta F(a(t_n+\xi)) - \lambda a(t_n+\xi))].\end{aligned} \tag{4.51}$$

To be able to bound D_{n+1} further, we need another lemma.

Lemma 4.5. *The function $\eta(\xi)$ defined in* (4.51) *satisfies $\eta(\xi) \geq C_2$ for all $\xi \in (0,h)$, with*

$$C_2 = \frac{2\lambda^2\delta^3 a_0}{(\delta + \lambda(g(K) - \varepsilon))^3} - 2\delta\lambda F(G(\varepsilon)) - \beta(G(\varepsilon)),$$

where $\beta(x) := (\delta F'(x) - \lambda)(\delta F(x) - \lambda x) = \psi'(x)\psi(x)$, and ε is a fix number belonging to $(0, g(K))$ (see Theorem 4.20).

Proof. In order to bound below $\eta(\xi)$, we will try to bound each term separately. Recall that we suppose $h < \frac{g(K)-\varepsilon}{\delta}$ and $a_n \leq K$ so that $G(g(a_n) - \delta h) < G(\varepsilon)$.

- To bound $\frac{2\lambda^2}{(1+\xi\lambda)^3} G(g(a_n) - \delta\xi)$, we use the fact that $\xi \in (0,h)$: since

$$\xi < h < \frac{g(K) - \varepsilon}{\delta},$$

we have

$$\frac{1}{(1+\xi\lambda)^3} > \frac{\delta^3}{(\delta + \lambda(g(K) - \varepsilon))^3},$$

and since $\xi > 0$ and (a_n) is increasing, we have

$$G(g(a_n) - \delta\xi) > G(g(a_n)) = a_n > a_0.$$

Hence

$$\frac{2\lambda^2}{(1+\xi\lambda)^3} G(g(a_n) - \delta\xi) > \frac{2\lambda^2\delta^3 a_0}{(\delta + \lambda(g(K) - \varepsilon))^3}.$$

- To bound $\frac{-2\delta\lambda}{(1+\xi\lambda)^2}F(G(g(a_n)-\delta\xi))$, we first observe that the minimum is reached on the boundary of the interval. Indeed, let

$$\alpha(x) = \frac{1}{(1+x\lambda)^2}F(G(g(a_n)-\delta x)),$$

then its derivative

$$\alpha'(x) = \frac{-2\lambda}{(1+x\lambda)^3}F(G(\dots)) + \frac{\delta}{(1+x\lambda)^2}F'(G(\dots))F(G(\dots))$$

is positive if

$$F'(G(g(a_n)-\delta x)) \geq \frac{2\lambda}{\delta(1+x\lambda)}.$$

Since F' is increasing, and G is decreasing, the function $(1+x\lambda)F'(G(g(a_n)-\delta x))$ is increasing in x. Hence α has no maximum in the interval and we have

$$\begin{aligned}\max_{0\leq\xi\leq h}\alpha(\xi) &= \max\{\alpha(0),\alpha(h)\} = \max\{F(a_n); \frac{1}{(1+h\lambda)^2}F(G(g(a_n)-\delta h))\}\\ &\leq \max\{F(K); F(G(\varepsilon))\}.\end{aligned}$$

Since $g(K)-\delta h > \varepsilon$, we have $K < G(\varepsilon)$, so

$$\max_{0\leq\xi\leq h}\alpha(\xi) \leq F(G(\varepsilon)),$$

and

$$\frac{-2\delta\lambda}{(1+\xi\lambda)^2}F(G(g(a_n)-\delta\xi)) \geq -2\delta\lambda F(G(\varepsilon)).$$

- The next term to bound is $\frac{\delta^2}{1+\xi\lambda}F'(G(g(a_n)-\delta\xi))F(G(g(a_n)-\delta\xi))$.

Let $\alpha(x) = \frac{1}{1+x\lambda}F'(G(g(a_n)-\delta x))F(G(g(a_n)-\delta x))$, then

$$\begin{aligned}\alpha'(x) &= \frac{-\lambda}{(1+x\lambda)^2}F'(G)F(G) + \frac{-\delta}{1+x\lambda}F''(G)G'F(G) + \frac{-\delta}{1+x\lambda}F'(G)F'(G)G'\\ &= \frac{-\lambda}{(1+x\lambda)^2}F'(G)F(G) + \frac{\delta}{1+x\lambda}F''(G)(F(G))^2 + \frac{\delta}{1+x\lambda}(F'(G))^2F(G)\end{aligned}$$

is positive if $F''(G)F(G) + [F'(G)]^2 \geq \frac{\lambda}{\delta(1+x\lambda)}F'(G)$, i.e.

$$(1+x\lambda)\left(\frac{F''(G(g(a_n)-\delta x))F(G(g(a_n)-\delta x))}{F'(G(g(a_n)-\delta x))} + F'(G(g(a_n)-\delta x))\right) \geq \frac{\lambda}{\delta}.$$

In many cases, in particular if $(F''F/F')' \geq 0$, the function of x on the left-hand side is increasing and α may have a minimum at $x^* \in (0,h)$. This would lead to

$$\min \alpha(x) = \min\{\alpha(0), \alpha(h), \alpha(x^*)\}.$$

So for a general function F, it is not possible to evaluate this minimum and since the term to bound is positive, we simply bound it by zero.

- For the last part, we let $\beta(x) = (\delta F'(x) - \lambda)(\delta F(x) - \lambda x)$, for $x \geq 0$. Since $F''(x) \geq 0$ and δ satisfies condition (4.45), we have

$$\beta'(x) = \delta F''(x)(\delta F(x) - \lambda x) + (\delta F'(x) - \lambda)^2 \geq 0.$$

Thus β is non-decreasing and in order to bound $\beta(a(t_n + \xi))$, we use Lemma 4.4. □

We are now able to prove Theorem 4.20.

Proof of Theorem 4.20

We suppose that $\|u_n\|_\infty \leq K$ and $D_n > 0$ and we 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

$$\text{either} \quad t_{\tilde{n}+1} \geq T_{a^*} \quad \text{or} \quad a(t_{\tilde{n}+1}) > K.$$

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 (4.50) 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

$$D_{n+1} \geq D_n(1 + h\psi'(a^*)) + \frac{h^2}{2}C_2. \tag{4.52}$$

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

$$\boxed{\text{if } F'(a^*) < \frac{\lambda}{\delta}, \quad h < \frac{1}{\lambda - \delta F'(a^*)}.}$$

If C_2 is positive, the positivity of D_{n+1} follows from (4.52). 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 $D_{n+1} > 0$.

Numerical Example

We present an example of computation of h^* in the case where $\Omega = [-1,1]$, so that

$$\lambda = \frac{\pi^2}{4} \quad \text{and} \quad \varphi = \frac{\pi}{4}\cos\left(\frac{\pi}{2}x\right).$$

We consider the case where $F(x) = (x+\alpha)^{p+1}$, with $\alpha = 2$ and $p = 1$, and $\delta = 3$. The initial condition is given by $u_0(x) = \cos(\pi x/2)$, so

$$a_0 = \int_{-1}^{1} \varphi(x)u_0(x)\,dx = \frac{\pi}{4}$$

and we can choose $a^* \in [0, \pi/4)$. We now fix $K = 500$, so that

$$g(K) = \frac{1}{K+\alpha} \approx 0.001992031872,$$

and we choose $\varepsilon = 0.0019$.

A first bound on h is given by

$$\frac{g(K) - \varepsilon}{\delta} \approx 0.00003067729067.$$

We now need to compute C_2. We have

$$\begin{aligned}\beta(G(\varepsilon)) &= (\delta(p+1)(G(\varepsilon)+\alpha)^p - \lambda)(\delta(G(\varepsilon)+\alpha)^{p+1} - \lambda x) \\ &= \left(\frac{\delta(p+1)}{pu} - \lambda\right)(\delta(pu)^{-(p+1)/p} - \lambda x) \\ &\approx 2.618156616 \cdot 10^9,\end{aligned}$$

and then

$$\begin{aligned}C_2 &= \frac{2\lambda^2\delta^3 a_0}{(\delta + \lambda(g(K) - \varepsilon))^3} - 2\delta\lambda((p\,\varepsilon)^{-(p+1)/p}) - \beta(G(\varepsilon)) \\ &\approx -2.622257550 \cdot 10^9.\end{aligned}$$

Since C_2 is negative, we need to derive a second bound on h to ensure the positivity of D_{n+1}. As we said above, we can choose $a^* \in [0, \pi/4)$. For all these values, $\psi'(a^*) = \delta(p+1)(a^*+\alpha)^p - \lambda$, is positive in which case the second bound for h is given by

$$h^* = \frac{2\psi'(a^*)D_0}{(-C_2)} = \frac{2(\delta(p+1)(a^*+\alpha)^p - \lambda)(a_0 - a^*)}{(-C_2)}.$$

The greatest value of h^* is obtained by choosing $a^* = 0$, for which we obtain

$$h^* \approx 5.710259596 \cdot 10^{-9}$$

So if we choose h smaller than h^*, there exists $n^* < \frac{T}{h}$ such that $\|u_{n^*}\|_\infty > 500$.

4.3 More General Equation

In this section, we present how several of the previous results of existence and uniqueness of a positive solution can be generalized to the case of the more general equation

$$u_t = \Delta u + \delta q(x)\psi(t)F(u),$$

where q is bounded on $\bar{\Omega}$ with $q(x) > 0$, ψ is continuous on $[0, \infty)$, with $\psi(t) > 0$ and F satisfies the conditions mentioned in Assumption 4.1. We explained in Section 3.2.1 how the previous schemes should be modified for this equation in the case where φ defined by

$$\varphi(t) = \int_0^t \psi(s)\, ds,$$

can be explicitly computed. Recall that φ is strictly increasing and $\varphi(0) = 0$. Since δh in the original scheme is replaced by $\delta q(x)(\varphi(t_{n+1}) - \varphi(t_n))$ in the first method and by $\delta q(x)\varphi(h)$ in the second method, the condition $\|u_n\|_\infty < G(\delta h)$ of Theorems 4.3 and 4.14 leads to a different condition for each method. Indeed the first scheme has a solution if

$$\|u_n\|_\infty < G\Big(\delta \|q\|_\infty (\varphi(t_{n+1}) - \varphi(t_n))\Big),$$

whereas the second scheme has a solution if

$$\|u_n\|_\infty < G\Big(\delta \|q\|_\infty \varphi(h))\Big).$$

We present in detail how these conditions affect the minimal time of existence T_1 of the solution.

4.3.1 Variation of the Constant and Backward Euler

The scheme obtained using the backward Euler method and the variation of the constant construction can be written as $Au_{n+1} = f(x, u_{n+1})$, with

$$f(x, v) = \frac{1}{h} F(v)[g(v) - g(u_n) + \delta q(\varphi(t_{n+1}) - \varphi(t_n))].$$

As in Section 4.1, if $F(0) = 0$, we have $\lim_{v \to 0^+} f(x, v) = 0$ for all $x \in \Omega$ and by abuse of notation, we shall refer to f as its continuous extension on $[0, \infty)$. For this scheme, Theorem 4.3 becomes

Theorem 4.21. *If the function u_n is positive in Ω, continuous in $\bar{\Omega}$, and satisfies*

$$\|u_n\|_\infty < G\bigg(\delta\|q\|_\infty(\varphi(t_{n+1}) - \varphi(t_n))\bigg), \tag{4.53}$$

then our scheme has a maximal nonnegative solution

$$\hat{u} \leq C_n = G\bigg(g(\|u_n\|) - \delta\|q\|\,[\varphi(t_{n+1}) - \varphi(t_n)]\bigg),$$

and a minimal solution $\bar{u} \geq 0$ and if u is a solution, then $u \in C^2(\bar{\Omega})$ and satisfies $\bar{u} \leq u \leq \hat{u}$.

Note that, in case M is finite, we need to take h small enough so that

$$\delta\|q\|_\infty\,[\varphi(t_{n+1}) - \varphi(t_n)] \in (0, M),$$

that is

$$\varphi(t_{n+1}) - \varphi(t_n) < \frac{M}{\delta\|q\|_\infty}.$$

Proof. As stated above, if $F(0) = 0$, we have $f(x, 0) = 0$ for all $x \in \Omega$. If $F(0) > 0$, since g is decreasing, we have

$$g(u_n) < g(0) + \delta\|q\|\,[\varphi(t_{n+1}) - \varphi(t_n)],$$

and we get $f(x, 0) > 0$. Moreover, $f(x, C_n) \leq 0$ if

$$g(C_n) - g(u_n) + \delta q(\varphi(t_{n+1}) - \varphi(t_n)) \leq 0,$$

that is

$$C_n \geq G(g(u_n) - \delta q(\varphi(t_{n+1}) - \varphi(t_n)),$$

so that under condition (4.53), we get the constant supersolution

$$C_n = G\bigg(g(\|u_n\|) - \delta\|q\|\,[\varphi(t_{n+1}) - \varphi(t_n)]\bigg).$$

□

Theorem 4.5 is valid if we replace condition (4.5) with condition (4.53), thus so is Corollary 4.1. In Theorem 4.7 condition (4.5) must be replaced by condition (4.53) and condition (4.7) must be satisfied for all $c \in (0, M)$. Thus Corollary 4.2 is also satisfied.

In order to show that condition (4.53) is satisfied for a positive number of steps, we need to adapt the induction of Theorem 4.8. The idea of that proof is to use a result of the form

$$\|u_n\| < G(\beta(h)) \qquad \Rightarrow \qquad \|u_{n+1}\| \leq G(g(u_n) - \beta(h)),$$

where β depends on h but not on n, to prove by induction that

$$\|u_0\| < G(\beta(t_n)) \qquad \Rightarrow \qquad \|u_{n+1}\| \leq G(g(u_0) - \beta(t_n)).$$

However the result we get from Theorem 4.21 is

$$\begin{aligned} &\|u_n\| < G(\delta\|q\| \left[\varphi(t_{n+1}) - \varphi(t_n)\right]) \\ &\qquad\qquad \Rightarrow \ \|u_{n+1}\| \leq G(g(\|u_n\|) - \delta\|q\| \left[\varphi(t_{n+1}) - \varphi(t_n)\right]), \end{aligned} \tag{4.54}$$

so that unless φ is linear, β does depend on n. So we need to find how to adapt the induction hypothesis. We are looking for a result of the form

$$\|u_0\| < G(\delta\|q\| f_1(n)) \quad \Rightarrow \quad \|u_n\| \leq G(g(\|u_0\|) - \delta\|q\| f_1(n)), \tag{4.55}$$

where f_1 has to satisfy certain conditions.

To get the initial condition, we need that putting $n = 0$ in (4.54) gives the case $n = 1$ in (4.55), that is

$$f_1(1) = \varphi(h).$$

For the induction step, we suppose that (4.55) is satisfied at step n and that

$$\|u_0\| < G(\delta\|q\| f_1(n+1)).$$

Then we need the following three inequalities to be satisfied

$$G(\delta\|q\|f_1(n+1) \leq G(\delta\|q\|f_1(n)), \tag{4.56a}$$

$$G(g(u_0) - \delta\|q\|f_1(n)) < G(\delta\|q\| \left[\varphi(t_{n+1}) - \varphi(t_n)\right]), \tag{4.56b}$$

$$G(g(u_n) - \delta\|q\| \left[\varphi(t_{n+1}) - \varphi(t_n)\right]) \leq G(g(u_0) - \delta\|q\|f_1(n+1)). \tag{4.56c}$$

The first inequality (4.56a) is equivalent to

$$f_1(n+1) \geq f_1(n) \qquad \forall\, n. \tag{4.57}$$

The second inequality (4.56b) becomes

$$g(u_0) - \delta\|q\|f_1(n) > \delta\|q\| \left[\varphi(t_{n+1}) - \varphi(t_n)\right),$$

and since $\|u_0\| < G(\delta\|q\|\, f_1(n+1))$, we get

$$\delta\|q\|f_1(n+1) - \delta\|q\|f_1(n) \geq \delta\|q\| \left[\varphi(t_{n+1}) - \varphi(t_n)\right],$$

that is

$$f_1(n+1) \geq f_1(n) + \varphi(t_{n+1}) - \varphi(t_n). \tag{4.58}$$

If this condition is satisfied, condition (4.57) and thus inequality (4.56a) are also satisfied. From the third inequality (4.56c), since we have $g(u_n) > g(u_0) - \delta\|q\|f_1(n)$, we obtain

$$g(u_0) - \delta\|q\|f_1(n) - \delta\|q\| \left[\varphi(t_{n+1}) - \varphi(t_n)\right] \geq g(u_0) - \delta\|q\|f_1(n+1),$$

which is condition (4.58). Thus f_1 needs to satisfy

$$f_1(1) = \varphi(h) \qquad \text{and} \qquad f_1(n+1) \geq f_1(n) + \varphi(t_{n+1}) - \varphi(t_n).$$

By induction,

$$f_1(1) = \varphi(h),$$
$$f_1(2) \geq \varphi(h) + \varphi(2h) - \varphi(h) = \varphi(2h),$$
$$f_1(3) \geq \varphi(3h),$$
$$\vdots$$
$$f_1(n) \geq \varphi(t_n).$$

Since we should choose f_1 to be as small as possible, we let $f_1(n) = \varphi(t_n)$, and we obtain by induction that

$$\|u_0\| < G(\delta\|q\|\,\varphi(t_n)) \quad \Rightarrow \quad \|u_n\| \leq G(g(\|u_0\|) - \delta\|q\|\,\varphi(t_n)).$$

Therefore the following theorem holds.

Theorem 4.22. *If the function F satisfies condition (4.7) for all $c \in (0, M)$, and either $F(0) \neq 0$ or F satisfies the hypotheses of Theorem 4.5, the scheme has a positive solution u_n for n such that $t_n = nh < T_1$, where*

$$T_1 = \varphi^{-1}\left(\frac{1}{\delta\|q\|_\infty}\, g(\|u_0\|_\infty)\right),$$

if h is small enough so that

$$\varphi(t_{n+1}) < \varphi(t_n) + \frac{M}{\delta\|q\|_\infty},$$

for all $t_n < T_1$.

4.3.2 Adjoint Splitting Method with Backward Euler

If we use the backward Euler method and the splitting method technique, we obtain the following scheme

$$u_{n+1} - h\Delta u_{n+1} = G\Big(g(u_n) - \delta q(x)\varphi(h)\Big).$$

We write it as $Au_{n+1} = f(x, u_{n+1})$ with

$$f(x,v) = \frac{1}{h} G\bigg(g(u_n) - \delta q(x)\varphi(h) \bigg) - \frac{1}{h} v.$$

A necessary condition for the scheme to be well-defined is that

$$g(u_n) - \delta q(x)\varphi(h) \in (0, M).$$

This is satisfied if

$$\varphi(h) < \frac{M}{\delta \|q\|_\infty},$$

and

$$\|u_n\|_\infty < G\bigg(\delta \|q\|_\infty \varphi(h) \bigg).$$

So Theorem 4.14 must be replaced with

Theorem 4.23. *If the function u_n is positive in Ω, continuous in $\bar{\Omega}$, and satisfies*

$$\|u_n\|_\infty < G\bigg(\delta \|q\|_\infty \varphi(h) \bigg), \tag{4.59}$$

then our scheme has a maximal nonnegative solution

$$\hat{u} \leq C_n = G\bigg(g(\|u_n\|_\infty) - \delta \|q\|_\infty \varphi(h) \bigg),$$

and a minimal solution $\bar{u} \geq 0$, and if u is a solution, then $u \in C^2(\bar{\Omega})$ and satisfies $\bar{u} \leq u \leq \hat{u}$.

Remark 4.3. *Whereas Theorems 4.3 and 4.14 were the same, condition* (4.59) *is different from condition* (4.53).

Proof. If condition (4.59) is satisfied, we have

$$f(0) = \frac{1}{h} G\bigg(g(u_n) - \delta q(x)\varphi(h) \bigg) > 0.$$

Moreover $f(C_n)$ is negative if

$$C_n \geq G\bigg(g(u_n(x)) - \delta q(x)\varphi(h) \bigg) \quad \forall x \in \Omega,$$

which is satisfied by

$$C_n = G\bigg(g(\|u_n\|_\infty) - \delta\|q\|_\infty\varphi(h)\bigg).$$

□

Since f is decreasing, Theorem 4.15 applies and the scheme has a unique solution and it is positive. To obtain a minimal bound T_1 for the blow-up time, we use the same approach as in Section 4.3.1. From Lemma 4.23, we have

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

and we look for a result of the form of (4.55). The function f_1 needs to satisfy $f_1(1) = \varphi(h)$ and

$$\begin{aligned}
&G(\delta\|q\|f_1(n+1)) \le G(\delta\|q\|f_1(n)),\\
&G(g(\|u_0\|) - \delta\|q\|f_1(n)) \le G(\delta\|q\|\varphi(h)),\\
&G(g(\|u_n\|) - \delta\|q\|\varphi(h)) \le G(g(\|u_0\|) - \delta\|q\|f_1(n+1)).
\end{aligned}$$

All three conditions are satisfied if

$$f_1(n+1) \ge f_1(n) + \varphi(h).$$

Hence we can only set $f_1(n) = \varphi(t_n)$ if φ satisfies

$$\varphi(t_{n+1}) \ge \varphi(t_n) + \varphi(h). \tag{4.60}$$

In this case we get the same result as Theorem 4.22. Condition (4.60) is satisfied in particular for all $h > 0$ and all $n \in \mathbb{N}$ if ψ is a nondecreasing function.

Theorem 4.24. *If φ satisfies condition* (4.60) *for all $t_n < T_1$, the scheme has a unique positive solution u_n for n such that $t_n = nh < T_1$, where*

$$T_1 = \varphi^{-1}\bigg(\frac{1}{\delta\|q\|_\infty}\, g(\|u_0\|_\infty)\bigg).$$

If φ does not satisfy condition (4.60), we set $f_1(n) = n\varphi(h)$, so that we obtain by induction

$$\text{if} \quad \|u_0\| < G(\delta\|q\|n\varphi(h)), \quad \text{then } \|u_n\| \leq G(g(\|u_0\|) - \delta\|q\|n\varphi(h)).$$

so that we have a solution if

$$n\varphi(h) < \frac{g(\|u_0\|)}{\delta\|q\|}.$$

So for $t_n = nh$, we get

$$t_n \frac{\varphi(h)}{h} < \frac{g(\|u_0\|)}{\delta\|q\|},$$

and for each h the bound T_1 is

$$T_1 = \frac{h}{\varphi(h)} \frac{g(\|u_0\|)}{\delta\|q\|}.$$

As the constant c defined by

$$c = \max_{0 \leq h \leq 1} \left\{ \frac{\varphi(h)}{h} \right\},$$

is finite, we can also obtain a bound independent of h

$$T_1 = \frac{g(\|u_0\|)}{c\delta\|q\|}.$$

4.4 Quasilinear Parabolic Equation with Power-Like Nonlinearities

In Section 3.2.2, we derived several schemes for the quasilinear parabolic equation with power-like nonlinearities

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

where Ω is a bounded domain in $\mathbb{R}^d$, $m > 1$ and $\alpha \geq 0$. Following the analysis for the semilinear parabolic equation, we present several results concerning the methods obtained by using the backward Euler method in the construction by variation of the constant and in the splitting method. Actually, the first B-method is deeply studied in an article by M-N Le Roux [99], so we summarize here the results she obtained. These cover existence and uniqueness of a positive solution and a minimal time of existence of the solution T_1. A specific solver based on Theorem 4.9 is presented as well as several inequalities concerning the behaviour of the numerical solution. Finally Le Roux proved that the upper bound for the numerical blow-up time is the same as the bound for the exact one. In Section 4.4.2, we prove the existence and uniqueness of the solution for the second B-method, and we obtain the same minimal time of existence T_1 as Le Roux derived for the first method.

4.4.1 Variation of the Constant and Backward Euler

The scheme derived in Section 3.2.2 by applying the variation of the constant and backward Euler is

$$u_{n+1} - u_n^{-(m-1)} u_{n+1}^m + h(m-1)[\alpha u_{n+1}^m + \Delta u_{n+1}^m] = 0.$$

If we define $v = u^m$ and $A = -\Delta$ and introduce $p = 1/m$ and $q = 1 - p$, so that $m - 1 = q/p$, the scheme becomes

$$h(Av_{n+1} - \alpha v_{n+1}) + \frac{p}{q} v_{n+1} v_n^{-q} - \frac{p}{q} v_{n+1}^p = 0, \tag{4.61}$$

which is exactly the scheme derived by Le Roux in [99]. She studied it in detail for $\alpha > \lambda_1$ as well as for $\alpha \leq \lambda_1$. We present here most results she obtained in the case where $\alpha > \lambda_1$, as this is the blow-up case. Some of these results were already shortly presented in Chapter 2.

To prove existence and uniqueness of the solution of scheme (4.61), Le Roux first uses Theorem 4.2 to prove that if

$$\|v_n\|_\infty < \left(\frac{p}{\alpha q h}\right)^{1/q}, \tag{4.62}$$

there exists a maximal solution

$$\hat{v} \leq C_n = \frac{\|v_n\|_\infty}{(1 - \frac{\alpha h p}{q}\|v_n\|_\infty^q)^{1/q}},$$

and if v is a solution of the scheme, then $0 \leq v \leq \hat{v}$.

A direct argument ensures uniqueness of a positive solution, however since $u \equiv 0$ is a solution of the scheme, extra work is required to ensure the existence of a positive solution. This solution is obtained by minimizing the functional

$$J_n(v) = \int_\Omega |\nabla v|^2 \, dx + \int_\Omega \left(\frac{p}{qh} v_n^{-q} - \alpha\right) v^2 \, dx,$$

on $K = \{v \in H_0^1(\Omega) | \int_\Omega v^{p+1} \, dx = 1\}$. Denoting by ψ_n the non-negative solution of this optimization problem, the unique positive solution is given by

$$v_{n+1} = \left(\frac{p}{qh J_n(\psi_n)}\right)^{1/q} \psi_n.$$

Then Le Roux obtained the lower bound $T_1 = \frac{p}{\alpha q}\|v_0\|_\infty^{-q}$ on the numerical blow-up time (see Theorem 4.27), and since scheme (4.61) is nonlinear, she used Theorem 4.9 with $\varphi(x) = \alpha - \frac{p}{qh} v_n^{-q}$, to derive a specific solver for the problem.

Le Roux also derived several inequalities concerning the numerical solution obtained by scheme (4.61): the existence of a subsolution

$$v_{n+1} \geq \left(\frac{t_n}{t_{n+1}}\right)^{1/q} v_n,$$

and two bounds on the rate of growth. The first one

$$v_{n+1} \leq \left(\frac{T_2 - t_n}{T_2 - t_{n+1}}\right)^{1/q} v_n,$$

where $T_2 = p/(qC_0)$ and C_0 is a constant satisfying

$$Av_0 - \alpha v_0 + C_0 v_0^p \geq 0,$$

can be related to the solution

$$v(t) = \left(\frac{p}{\alpha q(T-t)}\right)^{1/q}$$

of $pv^{-q}v_t = \alpha v$. The second bound, given in the following theorem, is linked to the numerical blow-up time, for which we also get an upper bound.

Theorem 4.25 (Le Roux)**.** *If $\alpha > \lambda_1$, there exists T^* depending on h and v_0 such that the numerical solution v_n exists for $nh < T^*$ and becomes infinite at T^*. In addition, we have*

$$\|v_n\|_{p+1} \leq \left(\frac{T^*}{T^* - t_n}\right)^{1/q} \|v_0\|_{p+1}.$$

If $\int_\Omega (|\nabla v_0|^2 - \alpha v_0^2)dx$ is negative, we have the estimate

$$T^* \leq \frac{p}{q} \frac{-\|v_0\|_{p+1}^{2-q}}{\int_\Omega (|\nabla v_0|^2 - \alpha v_0^2)dx}.$$

This theorem is obtained by introducing the functional

$$F(v) = \frac{\int_\Omega (|\nabla v|^2 - \alpha v^2)dx}{\|v\|_{p+1}^2}.$$

Le Roux proved that the sequence $\{F(v_n)\}$ is decreasing with

$$\frac{q}{p} hF(v_{n+1}) \leq \|v_{n+1}\|_{p+1}^{-q} - \|v_n\|_{p+1}^{-q} \leq \frac{q}{p} hF(v_n).$$

This inequality is used to derive the upper bound for the blow-up time given in the theorem. Indeed, if $\alpha > \lambda_1$ and $F(v_0) < 0$, we obtain

$$\|v_{n+1}\|_{p+1}^{-q} \leq \|v_0\|_{p+1}^{-q} + \frac{q}{p} t_n F(v_0).$$

As the right-hand side of this inequality is negative if

$$t_n > T_{\max} := \frac{p}{q} \frac{\|v_0\|_{p+1}^{-q}}{-F(v_0)},$$

there must be $t_{\tilde{n}} < T_{\max}$ for which (4.62) is not satisfied anymore and thus

$$\|v_{\tilde{n}}\|_\infty \geq \left(\frac{p}{\alpha q h}\right)^{1/q}.$$

As the constant on the right-hand side can be made as large as desired by decreasing h, Le Roux refers to that time $t_{\tilde{n}}$ as the numerical blow-up time T^* and says that the numerical solution becomes infinite at T^*. As we already pointed out in Chapter 2, the numerical solution should not be computed further as the numerical result may become irrelevant.

4.4.2 Adjoint Splitting Method with Backward Euler Method

In this section we prove existence and uniqueness of the solution of the scheme obtained using the backward Euler method in the adjoint splitting method. This scheme is given by

$$u_{n+1} = \left(u_n^{-(m-1)} - \alpha(m-1)h\right)^{-1/(m-1)} + h\Delta u_{n+1}^m.$$

Introducing $Av = -\Delta v$, $v = u^m$, $p = 1/m$ and $q = 1 - p$, it becomes

$$Av_{n+1} = f(x, v_{n+1}) = \frac{-1}{h} v_{n+1}^p + \frac{1}{h}\left(v_n^{-q}(x) - \alpha\frac{qh}{p}\right)^{-p/q}. \tag{4.63}$$

For this scheme to be well-defined, we need condition (4.62) to be satisfied. Moreover, under this condition, we have

$$f(x,0) = \frac{1}{h}\left(v_n^{-q} - \alpha\frac{qh}{p}\right)^{-p/q} > 0,$$

for all x. The constant C_n is a supersolution if $f(C_n) \leq 0$, that is

$$C_n^p \geq (v_n^{-q} - \alpha\frac{qh}{p})^{-p/q},$$

and since

$$\left[\frac{1}{v_n^{-q} - \alpha\frac{qh}{p}}\right]^{1/q} \leq \left[\frac{1}{\|v_n\|_\infty^{-q} - \alpha\frac{qh}{p}}\right]^{1/q},$$

we can take

$$C_n = \left[\frac{1}{\|v_n\|_\infty^{-q} - \alpha \frac{qh}{p}} \right]^{1/q}.$$

Hence we can apply Theorem 4.2 to obtain the same result as Le Roux:

Theorem 4.26. *If the function v_n is positive in Ω, continuous in $\bar{\Omega}$, and satisfies*

$$\|v_n\|_\infty < \left(\frac{p}{\alpha q h} \right)^{\frac{1}{q}}, \tag{4.64}$$

then scheme (4.63) *has a maximal nonnegative solution*

$$\hat{v} \leq C_n = \left[\frac{1}{\|v_n\|_\infty^{-q} - \alpha \frac{qh}{p}} \right]^{1/q},$$

and a minimal solution $\bar{v} \geq 0$ and if v is a solution, $v \in C^2(\bar{\Omega})$ and satisfies $\bar{v} \leq v \leq \hat{v}$.

Since $v \equiv 0$ is not a solution of the scheme (4.63), this theorem implies the existence of a non-identically zero nonnegative solution, which by the Maximum Principle, must be positive. Moreover, since f is decreasing in v, Theorem 4.15 applies and the solution is unique.

It remains to prove that the condition (4.64) is satisfied for a positive number of steps. Since Theorem 4.26 is the same result as the one given by Le Roux [99], we obtain the same lower bound T_1:

Theorem 4.27. *Scheme* (4.63) *has a unique positive solution v_n for n such that $t_n = n\Delta t < T_1$, where*

$$T_1 = \frac{p}{\alpha q} \|v_0\|_\infty^{-q}.$$

Proof. We have seen that if $\|v_n\|_\infty^q < \frac{p}{q\Delta t \alpha}$, then we have a supersolution

$$C_n = \frac{\|v_n\|_\infty}{(1 - \alpha \frac{qh}{p} \|v_n\|_\infty^q)^{1/q}},$$

and $\|v_{n+1}\|_\infty \leq C_n$. We can rewrite this as

$$\|v_{n+1}\|_\infty \leq \frac{\|v_n\|_\infty}{(1-\alpha\Delta t\frac{q}{p}\|v_n\|_\infty^q)^{1/q}} \qquad \text{if } \|v_n\|_\infty^q < \frac{p}{q\alpha\Delta t}. \tag{4.65}$$

We want to prove by induction that

$$\|v_n\|_\infty \leq \frac{\|v_0\|_\infty}{(1-\alpha t_n\frac{q}{p}\|v_0\|_\infty^q)^{1/q}} \qquad \text{if } \|v_0\|_\infty^q < \frac{p}{q\alpha t_n}.$$

If $n=1$, we have $\|v_0\|_\infty^q < \frac{p}{q\alpha\Delta t}$ implies

$$\|v_1\|_\infty \leq \frac{\|v_0\|_\infty}{(1-\alpha\Delta t\frac{qh}{p}\|v_0\|_\infty^q)^{1/q}}.$$

Now suppose that

$$\|v_0\|_\infty^q < \frac{p}{q\alpha t_n} \quad \Rightarrow \quad \|v_n\|_\infty \leq \frac{\|v_0\|_\infty}{(1-\alpha t_n\frac{q}{p}\|v_0\|_\infty^q)^{1/q}}$$

and assume that $\|v_0\|_\infty^q < \frac{p}{q\alpha t_{n+1}} < \frac{p}{q\alpha t_n}$, so that, by induction hypothesis,

$$\|v_n\|_\infty^q \leq \frac{\|v_0\|_\infty^q}{1-\alpha t_n\frac{q}{p}\|v_0\|_\infty^q} = \frac{1}{\|v_0\|_\infty^{-q}-\alpha t_n\frac{q}{p}},$$

but since $\|v_0\|_\infty^q < \frac{p}{q\alpha t_{n+1}}$, that is $\Delta t\|v_0\|_\infty^q + t_n\|v_0\|_\infty^q < \frac{p}{q\alpha}$ and $\|v_0\|_\infty^{-q} - \alpha t_n\frac{q}{p} > \Delta t\frac{q\alpha}{p}$, we get that

$$\|v_n\|_\infty^q < \frac{p}{q\alpha\Delta t},$$

and we can apply (4.65) to get

$$\|v_{n+1}\|_\infty \leq \frac{\|v_n\|_\infty}{(1-\alpha\Delta t\frac{q}{p}\|v_n\|_\infty^q)^{1/q}},$$

that is

$$\|v_{n+1}\|_\infty^q \leq \frac{1}{\|v_n\|_\infty^{-q}-\alpha\Delta t\frac{q}{p}},$$

and we use

$$\|v_n\|_\infty^q \leq \frac{\|v_0\|_\infty^q}{1-\alpha t_n\frac{q}{p}\|v_0\|_\infty^q},$$

to obtain

$$\|v_{n+1}\|_\infty^q \leq \frac{1}{\frac{(1-\alpha t_n\|v_0\|_\infty^q q/p)}{\|v_0\|_\infty^q}-\alpha\Delta t q/p} = \frac{\|v_0\|_\infty^q}{1-\alpha t_{n+1}\frac{q}{p}\|v_0\|_\infty^q},$$

which concludes the proof of the induction. □

Conclusion

We addressed in this thesis the delicate problem of approximating a blow-up solution using numerical methods. We introduced new fixed-step methods, called B-methods, that are designed to properly reproduce a blow-up solution. These methods can be constructed in different ways but the constructions are all based on the same idea: since the diffusion part of the equation becomes less essential as we get closer to the blow-up, it is relevant to consider the simplified equation obtained by removing the diffusion part. If the exact solution of this simplified equation can be explicitly derived, it is wise to exploit this information, in order to construct more efficient numerical methods. We presented two types of construction that follow different approaches.

The first simply consists in a splitting method. The right-hand side of the original equation is decomposed into two parts, and generally only the exact flow of the simplified equation (obtained by removing the diffusion part) can be derived, so that a first-order numerical method is used to solve the sub-equation with the diffusion part. The exact flow and the numerical method are then composed in order to obtain a consistent method for the original problem.

The idea of the second approach is quite innovative. We look for a solution u in a specific form: by plugging in the original equation the solution of the simplified equation in which the constant of integration K is considered as a function, we obtain a differential equation for this function K. Any numerical method can be applied to

this differential equation, and the corresponding B-method is obtained by rewriting the resulting scheme with the original unknown function u.

These two types of B-methods have been presented in detail. However one can think of other ways of constructing such methods. For example, even if the scheme derived by Le Roux in [99] can be obtained using the construction by variation of the constant, Le Roux's approach was different. We briefly explain a generalization of her approach on the semilinear problem $u_t = \Delta u + \delta F(u)$. Using the notation introduced in Section 3.1.1, the exact scheme of the simplified equation $u_t = \delta F(u)$ can be written as

$$u_{n+1} = G(g(u_n) - \delta h). \tag{4.66}$$

To construct a B-method, the idea is to first isolate δ in (4.66) to get

$$\frac{g(u_n) - g(u_{n+1})}{h} = \delta,$$

and then to multiply each side of the resulting scheme by an approximation of $F(u)$. For example, this could be $F(u_n)$ if we use forward Euler, $F(u_{n+1})$ if we use backward Euler, $F(\frac{u_n+u_{n+1}}{2})$ if we use the midpoint rule or $\frac{F(u_n)+F(u_{n+1})}{2}$ if we use the trapezoidal rule. Finally, the terms left aside (diffusion part) are added accordingly. The schemes obtained using forward Euler and backward Euler, respectively

$$-\frac{g(u_{n+1}) - g(u_n)}{h} F(u_n) = \delta F(u_n) + \Delta u_n,$$

and

$$-\frac{g(u_{n+1}) - g(u_n)}{h} F(u_{n+1}) = \delta F(u_{n+1}) + \Delta u_{n+1},$$

are exactly the same as the ones obtained using the construction by variation of the constant. Indeed, for these methods the differential equation for K leads to

$$\frac{K_{n+1} - K_n}{h} = \frac{g(u_{n+1}) - g(u_n) + \delta h}{h} = \frac{-1}{F(u)} \Delta u,$$

where u in the right-hand side is u_n if we use forward Euler and u_{n+1} if we use backward Euler. However, for more complex methods, like the midpoint or trapezoidal rules, the two constructions lead to different schemes. Moreover, both schemes

$$-\frac{g(u_{n+1})-g(u_n)}{h}F\left(\frac{u_n+u_{n+1}}{2}\right)=\delta F\left(\frac{u_n+u_{n+1}}{2}\right)+\Delta\left(\frac{u_n+u_{n+1}}{2}\right),$$

and

$$-\frac{g(u_{n+1})-g(u_n)}{h}\left(\frac{F(u_n)+F(u_{n+1})}{2}\right)=\delta\left(\frac{F(u_n)+F(u_{n+1})}{2}\right)+\frac{\Delta u_n+\Delta u_{n+1}}{2},$$

are second-order methods. This third way to construct B-methods is clearly of interest and would have a place in a larger study of B-methods. It illustrates the fact that the theory of B-methods is only at its beginnings and that other types of methods can be developed.

As shown by the numerous numerical examples we presented in Chapter 3, the B-methods we constructed bring a clear improvement compared to the original standard methods. The theoretical results proved in Chapter 4 also reinforce these observations as the behavior of the numerical solutions obtained using the selected B-methods was proven to be very similar to the behavior of the exact solution. Unfortunately, the theoretical results we presented are somehow unsatisfying as they only represent the beginning of a more thorough study of B-methods. In particular, we chose to concentrate on two B-methods applied to two different problems. In future work, similar results should be proved for a larger variety of schemes and problems.

The methods of construction we presented are aimed at relatively simple problems since they require that the solution of the simplified equation can be explicitly written. And even if this condition is satisfied, some difficulties can weaken the interest of the methods: an example of the limits of the construction by variation of the constant was given in Section 3.2.1. As a next step, one should try to find a way to overcome these difficulties in order to widen the application area.

Finally, one can see from the numerical experiments presented in this thesis that the type of B-methods leading to the best results depends on the problem. While on some problems splitting methods give better results than methods obtained by variation of the constant, on other problems it is the opposite. At the same time as new types of B-methods are developed, more extended numerical experiments can lead to the creation of a large database on which a deeper study of the comparative performances of the different B-methods can be performed. This study could lead to the development of guidelines aimed at advising users in their selection of numerical methods. Moreover one can expect that new avenues of research to improve B-methods can be drawn from this reflection.

Bibliography

[1] L. M. Abia, J. C. López-Marcos, and J. Martínez. Blow-up for semidiscretizations of reaction-diffusion equations. *Appl. Numer. Math.*, 20(1-2):145–156, 1996. Workshop on the method of lines for time-dependent problems (Lexington, KY, 1995).

[2] L. M. Abia, J. C. López-Marcos, and J. Martínez. On the blow-up time convergence of semidiscretizations of reaction-diffusion equations. *Appl. Numer. Math.*, 26(4):399–414, 1998.

[3] L. M. Abia, J. C. López-Marcos, and J. Martínez. The Euler method in the numerical integration of reaction-diffusion problems with blow-up. *Appl. Numer. Math.*, 38(3):287–313, 2001.

[4] G. Acosta, R. G. Durán, and J. D. Rossi. An adaptive time step procedure for a parabolic problem with blow-up. *Computing*, 68(4):343–373, 2002.

[5] G. Acosta, J. Fernández Bonder, P. Groisman, and J. D. Rossi. Simultaneous vs. non-simultaneous blow-up in numerical approximations of a parabolic system with non-linear boundary conditions. *M2AN Math. Model. Numer. Anal.*, 36(1):55–68, 2002.

[6] R. P. Agarwal. *Difference equations and inequalities*, volume 155 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker Inc., New York, 1992. Theory, methods, and applications.

[7] S. Agmon, A. Douglis, and L. Nirenberg. Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions. I. *Comm. Pure Appl. Math.*, 12:623–727, 1959.

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

[9] S. B. Angenent and J. J. L. Velázquez. Asymptotic shape of cusp singularities in curve shortening. *Duke Math. J.*, 77(1):71–110, 1995.

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

[11] J. M. Ball. Remarks on blow-up and nonexistence theorems for nonlinear evolution equations. *Quart. J. Math. Oxford Ser. (2)*, 28(112):473–486, 1977.

[12] C. Bandle and H. Brunner. Numerical analysis of semilinear parabolic problems with blow-up solutions. *Rev. Real Acad. Cienc. Exact. Fís. Natur. Madrid*, 88(2-3):203–222, 1994.

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

[14] P. Baras and L. Cohen. Sur l'explosion totale après $T_{\max}$ de la solution d'une équation de la chaleur semi-linéaire. *C. R. Acad. Sci. Paris Sér. I Math.*, 300(10):295–298, 1985.

[15] P. Baras and L. Cohen. Complete blow-up after $T_{\max}$ for the solution of a semilinear heat equation. *J. Funct. Anal.*, 71(1):142–174, 1987.

[16] G. Barro, B. Mampassi, L. Some, J. M. Ntaganda, and O. So. Full discretization of some reaction diffusion equation with blow up. *Cent. Eur. J. Math.*, 4(2):260–269 (electronic), 2006.

[17] J. Bebernes, A. Bressan, and D. Eberly. A description of blowup for the solid fuel ignition model. *Indiana Univ. Math. J.*, 36(2):295–305, 1987.

[18] J. Bebernes and D. Eberly. *Mathematical problems from combustion theory*, volume 83 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1989.

[19] J. Bebernes and A. Lacey. Finite-time blowup for a particular parabolic system. *SIAM J. Math. Anal.*, 21(6):1415–1425, 1990.

[20] J. W. Bebernes and D. R. Kassoy. A mathematical analysis of blowup for thermal reactions—the spatially nonhomogeneous case. *SIAM J. Appl. Math.*, 40(3):476–484, 1981.

[21] H. Bellout. Blow-up of solutions of parabolic equations with nonlinear memory. *J. Differential Equations*, 70(1):42–68, 1987.

[22] H. Bellout. A criterion for blow-up of solutions to semilinear heat equations. *SIAM J. Math. Anal.*, 18(3):722–727, 1987.

[23] M. Berger and R. V. Kohn. A rescaling algorithm for the numerical calculation of blowing-up solutions. *Comm. Pure Appl. Math.*, 41(6):841–863, 1988.

[24] L. Bieberbach. $\Delta u = e^u$ und die automorphen Funktionen. *Math. Ann.*, 77(2):173–212, 1916.

[25] J. L. Bona, V. A. Dougalis, O. A. Karakashian, and W. R. McKinney. Computations of blow-up and decay for periodic solutions of the generalized Korteweg-de Vries-Burgers equation. *Appl. Numer. Math.*, 10(3-4):335–355, 1992. A Festschrift to honor Professor Garrett Birkhoff on his eightieth birthday.

[26] H. Brezis and L. Oswald. Remarks on sublinear elliptic equations. *Nonlinear Anal.*, 10(1):55–64, 1986.

[27] C. J. Budd, J. Chen, W. Huang, and R. D. Russell. Moving mesh methods with applications to blow-up problems for PDEs. In *Numerical analysis 1995 (Dundee, 1995)*, volume 344 of *Pitman Res. Notes Math. Ser.*, pages 1–18. Longman, Harlow, 1996.

[28] C. J. Budd, S. Chen, and R. D. Russell. New self-similar solutions of the nonlinear Schrödinger equation with moving mesh computations. *J. Comput. Phys.*, 152(2):756–789, 1999.

[29] C. J. Budd, G. J. Collins, and V. A. Galaktionov. An asymptotic and numerical description of self-similar blow-up in quasilinear parabolic equations. *J. Comput. Appl. Math.*, 97(1-2):51–80, 1998.

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

[31] S. Chandrasekar. *An Introduction to the Theory of Stellar Structures*. Dover, N.Y., 1957.

[32] S. Chen. A sufficient condition for blowup solutions of nonlinear heat equations. *J. Math. Anal. Appl.*, 293(1):227–236, 2004.

[33] X.-Y. Chen and H. Matano. Convergence, asymptotic periodicity, and finite-point blow-up in one-dimensional semilinear heat equations. *J. Differential Equations*, 78(1):160–190, 1989.

[34] Y. G. Chen. Asymptotic behaviours of blowing-up solutions for finite difference analogue of $u_t = u_{xx} + u^{1+\alpha}$. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.*, 33(3):541–574, 1986.

[35] Y. G. Chen. Blow-up solutions to a finite difference analogue of $u_t = \Delta u + u^{1+\alpha}$ in N-dimensional balls. *Hokkaido Math. J.*, 21(3):447–474, 1992.

[36] R. Chiao, E. Garmire, and C. Townes. Self-trapping of optical beams. *Physical review letters*, 13(15):479–482, 1964.

[37] M. Chipot and F. B. Weissler. Some blowup results for a nonlinear parabolic equation with a gradient term. *SIAM J. Math. Anal.*, 20(4):886–907, 1989.

[38] A. J. Chorin. Estimates of intermittency, spectra, and blow-up in developed turbulence. *Comm. Pure Appl. Math.*, 34(6):853–866, 1981.

[39] J. Coleman and C. Sulem. Numerical simulation of blow-up solutions of the vector nonlinear Schrödinger equation. *Phys. Rev. E (3)*, 66(3):036701, 14, 2002.

[40] R. Courant and D. Hilbert. *Methods of mathematical physics. Vol. II: Partial differential equations.* (Vol. II by R. Courant.). Interscience Publishers (a division of John Wiley & Sons), New York-Lon don, 1962.

[41] J. Dold. Analysis of the early stage of thermal runaway. *Quarterly journal of mechanics and applied mathematics*, 38:361–387, 1985.

[42] R. G. Duran, J. I. Etcheverry, and J. D. Rossi. Numerical approximation of a parabolic problem with a nonlinear boundary condition. *Discrete Contin. Dynam. Systems*, 4(3):497–506, 1998.

[43] A. Duvnjak and H. J. Eberl. Time-discretization of a degenerate reaction-diffusion equation arising in biofilm modeling. *Electron. Trans. Numer. Anal.*, 23:15–37, 2006.

[44] C. Elliott and B. Suydam. Self-focusing phenomena in air-glass laser structures. *IEEE Journal of quantum electronics*, 11(11):863–867, 1975.

[45] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.

[46] J. Fernández Bonder, P. Groisman, and J. D. Rossi. On numerical blow-up sets. *Proc. Amer. Math. Soc.*, 130(7):2049–2055 (electronic), 2002.

[47] R. Ferreira, P. Groisman, and J. D. Rossi. Numerical blow-up for a nonlinear problem with a nonlinear boundary condition. *Math. Models Methods Appl. Sci.*, 12(4):461–483, 2002.

[48] R. Ferreira, P. Groisman, and J. D. Rossi. Adaptive numerical schemes for a parabolic problem with blow-up. *IMA J. Numer. Anal.*, 23(3):439–463, 2003.

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

[50] D. Frank-Kamenetskii. *Diffusion and Heat Transfer in Chemical Kinetics*. New York: Plenum Press, 1969.

[51] 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.

[52] 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.

[53] A. Friedman and B. McLeod. Blow-up of positive solutions of semilinear heat equations. *Indiana Univ. Math. J.*, 34(2):425–447, 1985.

[54] H. Fujita. On the blowing up of solutions of the Cauchy problem for $u_t = \Delta u + u^{1+\alpha}$. *J. Fac. Sci. Univ. Tokyo Sect. I*, 13:109–124 (1966), 1966.

[55] H. Fujita. On the nonlinear equations $\Delta u + e^u = 0$ and $\partial v/\partial t = \Delta v + e^v$. *Bull. Amer. Math. Soc.*, 75:132–135, 1969.

[56] H. Fujita. On some nonexistence and nonuniqueness theorems for nonlinear parabolic equations. In *Nonlinear Functional Analysis (Proc. Sympos. Pure Math., Vol. XVIII, Part 1, Chicago, Ill., 1968)*, pages 105–113. Amer. Math. Soc., Providence, R.I., 1970.

[57] H. Fujita. On the asymptotic stability of solutions of the equation $v_t = \Delta v + e^v$. In *Proc. Internat. Conf. on Functional Analysis and Related Topics (Tokyo, 1969)*, pages 252–259. Univ. of Tokyo Press, Tokyo, 1970.

[58] 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, 551556.}. *Differentsial'nye Uravneniya*, 17(5):836–842, 956, 1981.

[59] 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.

[60] L. Gang and B. D. Sleeman. Nonexistence of global solutions to systems of semilinear parabolic equations. *J. Differential Equations*, 104(1):147–168, 1993.

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

[62] Y. Giga and R. V. Kohn. Asymptotically self-similar blow-up of semilinear heat equations. *Comm. Pure Appl. Math.*, 38(3):297–319, 1985.

[63] Y. Giga and R. V. Kohn. Characterizing blowup using similarity variables. *Indiana Univ. Math. J.*, 36(1):1–40, 1987.

[64] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*. Springer-Verlag, Berlin, 1977. Grundlehren der Mathematischen Wissenschaften, Vol. 224.

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

[66] 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.

[67] M. Goldman. Strong turbulence of plasma-waves. *Reviews of modern physics*, 56(4):709–735, 1984.

[68] S. Goldstein. On laminar boundary-layer flow near a position of separation. *Quart. J. Mech. Appl. Math.*, 1:43–69, 1948.

[69] P. Groisman, F. Quirós, and J. D. Rossi. Non-simultaneous blow-up in a numerical approximation of a parabolic system. *Comput. Appl. Math.*, 21(3):813–831, 2002.

[70] P. Groisman and J. D. Rossi. Asymptotic behaviour for a numerical approximation of a parabolic problem with blowing up solutions. *J. Comput. Appl. Math.*, 135(1):135–155, 2001.

[71] M. E. Gurtin and R. C. MacCamy. On the diffusion of biological populations. *Math. Biosci.*, 33(1-2):35–49, 1977.

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

[73] K. Hayakawa. On nonexistence of global solutions of some semilinear parabolic differential equations. *Proc. Japan Acad.*, 49:503–505, 1973.

[74] W. Huang, J. Ma, and R. D. Russell. A study of moving mesh PDE methods for numerical simulation of blowup in reaction diffusion equations. *J. Comput. Phys.*, 227(13):6532–6552, 2008.

[75] W. Huang, Y. Ren, and R. D. Russell. Moving mesh methods based on moving mesh partial differential equations. *J. Comput. Phys.*, 113(2):279–290, 1994.

[76] W. Huang, Y. Ren, and R. D. Russell. Moving mesh partial differential equations (MMPDES) based on the equidistribution principle. *SIAM J. Numer. Anal.*, 31(3):709–730, 1994.

[77] T. Ishiwata and M. Tsutsumi. Semidiscretization in space of nonlinear degenerate parabolic equations with blow-up of the solutions. *J. Comput. Math.*, 18(6):571–586, 2000.

[78] D. D. Joseph and T. S. Lundgren. Quasilinear Dirichlet problems driven by positive sources. *Arch. Rational Mech. Anal.*, 49:241–269, 1972/73.

[79] D. D. Joseph and E. M. Sparrow. Nonlinear diffusion induced by nonlinear sources. *Quart. Appl. Math.*, 28:327–342, 1970.

[80] A. K. Kapila. Reactive-diffusive system with Arrhenius kinetics: dynamics of ignition. *SIAM J. Appl. Math.*, 39(1):21–36, 1980.

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

[82] D. R. Kassoy and J. Poland. The thermal explosion confined by a constant temperature boundary. I. The induction period solution. *SIAM J. Appl. Math.*, 39(3):412–430, 1980.

[83] O. Kavian. Remarks on the large time behaviour of a nonlinear diffusion equation. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 4(5):423–452, 1987.

[84] J. L. Kazdan and F. W. Warner. Curvature functions for compact 2-manifolds. *Ann. of Math. (2)*, 99:14–47, 1974.

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

[86] J. B. Keller. On solutions of $\Delta u = f(u)$. *Comm. Pure Appl. Math.*, 10:503–510, 1957.

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

[88] P. Kelley. Self-focusing of optimal beams. *Physical Review Letters*, 15(26):1005–&, 1965.

[89] R. J. Knops, H. A. Levine, and L. E. Payne. Non-existence, instability, and growth theorems for solutions of a class of abstract nonlinear equations with applications to nonlinear elastodynamics. *Arch. Rational Mech. Anal.*, 55:52–72, 1974.

[90] K. Kobayashi, T. Sirao, and H. Tanaka. On the growing up problem for semi-linear heat equations. *J. Math. Soc. Japan*, 29(3):407–424, 1977.

[91] A. Krzywicki and T. Nadzieja. Some results concerning the Poisson-Boltzmann equation. *Zastos. Mat.*, 21(2):265–272, 1991.

[92] A. A. Lacey. Mathematical analysis of thermal runaway for spatially inhomogeneous reactions. *SIAM J. Appl. Math.*, 43(6):1350–1366, 1983.

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

[94] A. A. Lacey. Thermal runaway in a non-local problem modelling Ohmic heating. I. Model derivation and some special cases. *European J. Appl. Math.*, 6(2):127–144, 1995.

[95] A. A. Lacey. Thermal runaway in a non-local problem modelling Ohmic heating. II. General proof of blow-up and asymptotics of runaway. *European J. Appl. Math.*, 6(3):201–224, 1995.

[96] A. A. Lacey and D. Tzanetis. Complete blow-up for a semilinear diffusion equation with a sufficiently large initial condition. *IMA J. Appl. Math.*, 41(3):207–215, 1988.

[97] A.-Y. Le Roux and M.-N. Le Roux. Numerical solution of a nonlinear reaction diffusion equation. *J. Comput. Appl. Math.*, 173(2):211–237, 2005.

[98] A.-Y. Le Roux and M.-N. Le Roux. Numerical solution of a Cauchy problem for nonlinear reaction diffusion processes. *J. Comput. Appl. Math.*, 214(1):90–110, 2008.

[99] M.-N. Le Roux. Semidiscretization in time of nonlinear parabolic equations with blowup of the solution. *SIAM J. Numer. Anal.*, 31(1):170–195, 1994.

[100] M.-N. Le Roux. Numerical solution of fast diffusion or slow diffusion equations. *J. Comput. Appl. Math.*, 97(1-2):121–136, 1998.

[101] M.-N. Le Roux. Numerical solution of nonlinear reaction diffusion processes. *SIAM J. Numer. Anal.*, 37(5):1644–1656, 2000.

[102] M.-N. Le Roux and P.-E. Mainge. Numerical solution of a fast diffusion equation. *Math. Comp.*, 68(226):461–485, 1999.

[103] 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.

[104] H. A. Levine. Some nonexistence and instability theorems for solutions of formally parabolic equations of the form $Pu_t = -Au+\mathcal{F}(u)$. *Arch. Rational Mech. Anal.*, 51:371–386, 1973.

[105] H. A. Levine. Instability and nonexistence of global solutions to nonlinear wave equations of the form $Pu_{tt} = -Au + \mathcal{F}(u)$. *Trans. Amer. Math. Soc.*, 192:1–21, 1974.

[106] H. A. Levine. On the nonexistence of global solutions to a nonlinear Euler-Poisson-Darboux equation. *J. Math. Anal. Appl.*, 48:646–651, 1974.

[107] H. A. Levine. Some additional remarks on the nonexistence of global solutions to nonlinear wave equations. *SIAM J. Math. Anal.*, 5:138–146, 1974.

[108] H. A. Levine. Nonexistence of global weak solutions to some properly and improperly posed problems of mathematical physics: the method of unbounded Fourier coefficients. *Math. Ann.*, 214:205–220, 1975.

[109] H. A. Levine and L. E. Payne. Nonexistence theorems for the heat equation with nonlinear boundary conditions and for the porous medium equation backward in time. *J. Differential Equations*, 16:319–334, 1974.

[110] H. A. Levine and L. E. Payne. Some nonexistence theorems for initial-boundary value problems with nonlinear boundary constraints. *Proc. Amer. Math. Soc.*, 46:277–284, 1974.

[111] H. A. Levine and L. E. Payne. On the nonexistence of global solutions to some abstract Cauchy problems of standard and non standard types. *Rend. Mat. (6)*, 8(2):413–428, 1975.

[112] H. A. Levine and P. E. Sacks. Some existence and nonexistence theorems for solutions of degenerate parabolic equations. *J. Differential Equations*, 52(2):135–161, 1984.

[113] W. Liniger and R. A. Willoughby. Efficient integration methods for stiff systems of ordinary differential equations. *SIAM J. Numer. Anal.*, 7:47–66, 1970.

[114] P.-E. Maingé. Discretization of the Cauchy problem for a fast diffusion equation. *J. Comput. Appl. Math.*, 213(1):95–110, 2008.

[115] D. McLaughlin, G. Papanicolaou, C. Sulem, and P.-L. Sulem. The focusing singularity of the cubic Schrödinger equation. *Physical Review A*, 34(2):1200–1210, 1986.

[116] P. Meier. *Explosion con Lösungen semilinearer parabolischer Differenitalgleichungen*. PhD thesis, Universität Basel.

[117] P. Meier. Existence et non-existence de solutions globales d'une équation de la chaleur semi-linéaire: extension d'un théorème de Fujita. *C. R. Acad. Sci. Paris Sér. I Math.*, 303(13):635–637, 1986.

[118] P. Meier. Blow-up of solutions of semilinear parabolic differential equations. *Z. Angew. Math. Phys.*, 39(2):135–149, 1988.

[119] P. Meier. On the critical exponent for reaction-diffusion equations. *Arch. Rational Mech. Anal.*, 109(1):63–71, 1990.

[120] F. Merle and Y. Tsutsumi. L^2 concentration of blow-up solutions for the nonlinear Schrödinger equation with critical power nonlinearity. *J. Differential Equations*, 84(2):205–214, 1990.

[121] R. Meyer-Spasche. Difference schemes of optimum degree of implicitness for a family of simple ODEs with blow-up solutions. *J. Comput. Appl. Math.*, 97(1-2):137–152, 1998.

[122] R. Meyer-Spasche. Variable-coefficient difference schemes for quasilinear evolution problems. In *Numerical methods and applications*, volume 2542 of *Lecture Notes in Comput. Sci.*, pages 36–47. Springer, Berlin, 2003.

[123] R. Meyer-Spasche. On difference schemes for quasilinear evolution problems. *Electron. Trans. Numer. Anal.*, 27:78–93 (electronic), 2007.

[124] R. Meyer-Spasche and D. Düchs. A general method for obtaining unconventional and nonstandard difference schemes. *Dynam. Contin. Discrete Impuls. Systems*, 3(4):453–467, 1997.

[125] R. E. Mickens. *Nonstandard finite difference models of differential equations*. World Scientific Publishing Co. Inc., River Edge, NJ, 1994.

[126] C. E. Mueller and F. B. Weissler. Single point blow-up for a general semilinear heat equation. *Indiana Univ. Math. J.*, 34(4):881–913, 1985.

[127] T. Nakagawa. Blowing up of a finite difference solution to $u_t = u_{xx} + u_2$.. *Appl. Math. Optim.*, 2(4):337–350, 1975/76.

[128] T. Nakagawa, T. Ikeda, and T. Ushijima. Numerical analysis of some blow-up problems of the semilinear evolution equation. In H. Fujita, editor, *Functional Analysis and Numerical Analysis*, pages 337–359. Japan Society for the Promotion of Science, 1978.

[129] T. Nakagawa and T. Ushijima. Finite element analysis of semilinear heat equation of blow-up type. *Topics in numerical analysis, III*, pages 275–291, 1977. Academic Press, London.

[130] H. Nawa and M. Tsutsumi. On blow-up for the pseudo-conformally invariant nonlinear Schrödinger equation. *Funkcial. Ekvac.*, 32(3):417–428, 1989.

[131] M. A. Rammaha. On the blowing up of solutions to nonlinear wave equations in two space dimensions. *J. Reine Angew. Math.*, 391:55–64, 1988.

[132] P. Rosenbloom. Singularities of solutions of nonlinear hyperbolic equations. *Preliminary report, Bull. Amer. Math. Soc.*, 60:343, 1954.

[133] M.-N. L. Roux and H. .Wilhelmsson. External boundary effects on simultaneous diffusion and reaction processes. *Phys. Scripta*, 40:674–681, 1989.

[134] P. Sacks. Continuity of solutions of a singular parabolic equation. *Nonlinear Anal.*, 7(4):387–409, 1983.

[135] A. A. Samarskii, V. A. Galaktionov, S. P. Kurdyumov, and A. P. Mikhailov. *Blow-up in quasilinear parabolic equations*, volume 19 of *de Gruyter Expositions in Mathematics*. Walter de Gruyter & Co., Berlin, 1995.

[136] J. M. Sanz-Serna and J. G. Verwer. A study of the recursion $y_{n+1} = y_n + \tau y_n^m$. *J. Math. Anal. Appl.*, 116(2):456–464, 1986.

[137] D. H. Sattinger. *Topics in stability and bifurcation theory*. Springer-Verlag, Berlin, 1973.

[138] J. Schaeffer. Finite-time blow-up for $u_{tt} - \Delta u = H(u_r, u_t)$. *Comm. Partial Differential Equations*, 11(5):513–543, 1986.

[139] M. Schechter. On L^p estimates and regularity. I. *Amer. J. Math.*, 85:1–13, 1963.

[140] A. R. Soheili and S. Salahshour. Moving mesh method with local time step refinement for blow-up problems. *Appl. Math. Comput.*, 195(1):76–85, 2008.

[141] 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.

[142] P. Souplet and F. B. Weissler. Self-similar subsolutions and blowup for nonlinear parabolic equations. *J. Math. Anal. Appl.*, 212(1):60–74, 1997.

[143] K. Stewart and T. Geveci. Numerical experiments with a nonlinear evolution equation which exhibits blow-up. *Appl. Numer. Math.*, 10(2):139–147, 1992.

[144] B. Straughan. *Instability, nonexistence and weighted energy methods in fluid dynamics and related theories*, volume 74 of *Research Notes in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1982.

[145] B. Straughan. *Explosive instabilities in mechanics*. Springer-Verlag, Berlin, 1998.

[146] A. M. Stuart and M. S. Floater. On the computation of blow-up. *European J. Appl. Math.*, 1(1):47–71, 1990.

[147] S. Sugitani. On nonexistence of global solutions for some nonlinear integral equations. *Osaka J. Math.*, 12:45–51, 1975.

[148] C. Sulem and P.-L. Sulem. *The nonlinear Schrödinger equation*, volume 139 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1999. Self-focusing and wave collapse.

[149] P.-L. Sulem, C. Sulem, and A. Patera. Numerical simulation of singular solutions to the two-dimensional cubic Schrödinger equation. *Comm. Pure Appl. Math.*, 37(6):755–778, 1984.

[150] B. Suydam. Effect of refractive-index nonlinearity on optical quality of high-power laser-beams. *IEEE Journal of quantum electronics*, 11(6):225–230, 1975.

[151] V. Talanov. Self-focusing of wave beams in nonlinear media. *JETP Letters-USSR*, 2(5):138–141, 1965.

[152] Y. Tourigny and M. Grinfeld. Deciphering singularities by discrete methods. *Math. Comp.*, 62(205):155–169, 1994.

[153] Y. Tourigny and J. M. Sanz-Serna. The numerical study of blowup with application to a nonlinear Schrödinger equation. *J. Comput. Phys.*, 102(2):407–416, 1992.

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

[155] M. Tsutsumi. On solutions of semilinear differential equations in a Hilbert space. *Math. Japon.*, 17:173–193, 1972.

[156] 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.

[157] T. K. Ushijima. On the approximation of blow-up time for solutions of nonlinear parabolic equations. *Publ. Res. Inst. Math. Sci.*, 36(5):613–640, 2000.

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

[159] V. M. Vold, R.D. *Colloid and Interface Chemistry.* Addision-Wesley, Reading-Mass, 1983.

[160] F. B. Weissler. Single point blow-up for a semilinear initial value problem. *J. Differential Equations*, 55(2):204–224, 1984.

[161] F. B. Weissler. An L^∞ blow-up estimate for a nonlinear heat equation. *Comm. Pure Appl. Math.*, 38(3):291–295, 1985.

[162] V. Zakharov. Collapse of langmuir waves. *Soviet Physics JETP*, 35:908–922, 1972.

[163] V. Zakharov. *Handbook of plasma physics (vol.2)*, volume 2, chapter Collapse of self-focusing of Langmuir waves. Elsevier, New-York, 1984.

[164] V. Zakharov, S. Musher, and A. Rubenchik. Hamiltonian approach to the description of non-linear plasma phenomena. *Phys. Rep.*, 129:286–366, 1985.

Appendix A: Additional Numerical Experiments

A.1 A Semilinear Parabolic Equation with Different Initial Conditions

Most of the experiments we present are done with the initial conditions $u_0(x) = \cos(\pi x/2)$. This symmetric bell-shaped function is concave on the whole interval so that the second derivative is negative everywhere. As mentioned in Chapter 4, the negativity of the Laplacian plays a part in the results concerning the rate of growth of the numerical solution. Actually, whatever the shape of the initial condition is, the Laplacian quickly becomes negative on the whole interval. To show that B-methods are efficient no matter what the initial conditions are, we applied them to the semilinear equation (3.1) with $F(u) = e^u$ and $\delta = 3$ with different initial conditions.

As a first example, we used $u_0(x) = (x^2 - 1)^2$ on $[-1, 1]$, whose particularity is that the second derivative is positive close to the boundary. The blow-up time can be approximated by $T_b \approx 0.1830$ and for Figures A.1 and A.2, we used $T_f = 0.1829$.

For the second example, we used the initial condition: $u_0(x) = (1-x^2)(8x^2+1)$ on $[-1, 1]$, whose particularity is that the second derivative is positive in the middle of

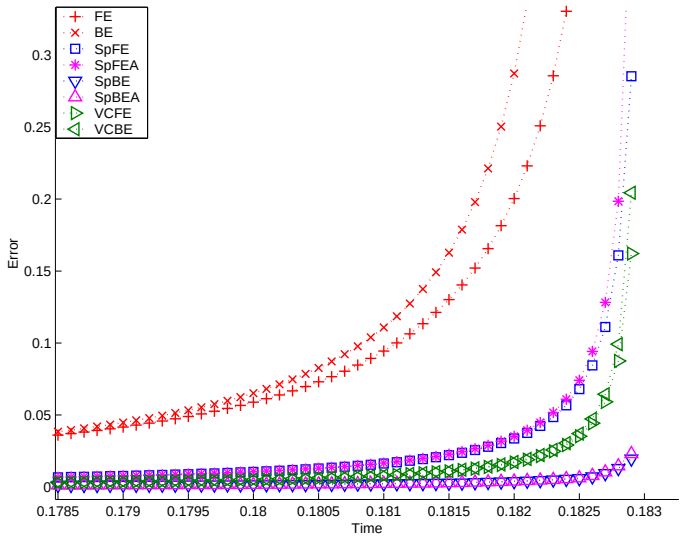

Figure A.1: Error for first-order methods applied to the semilinear parabolic equation with $u_0(x) = (x^2 - 1)^2$, for timesteps close to $T_f = 0.1829$.

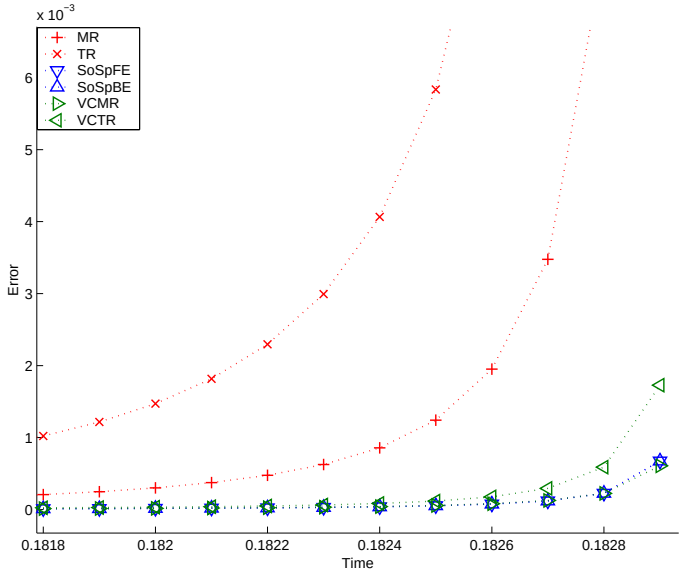

Figure A.2: Error for second-order methods applied to the semilinear parabolic equation with $u_0(x) = (x^2 - 1)^2$, for timesteps close to $T_f = 0.1829$.

the interval. The blow-up time can be approximated by $T_b \approx 0.0587$, so we computed the solutions using $h = 0.0001$ up to $T_f = 0.0586$. The errors are plotted in Figures A.3 and A.4.

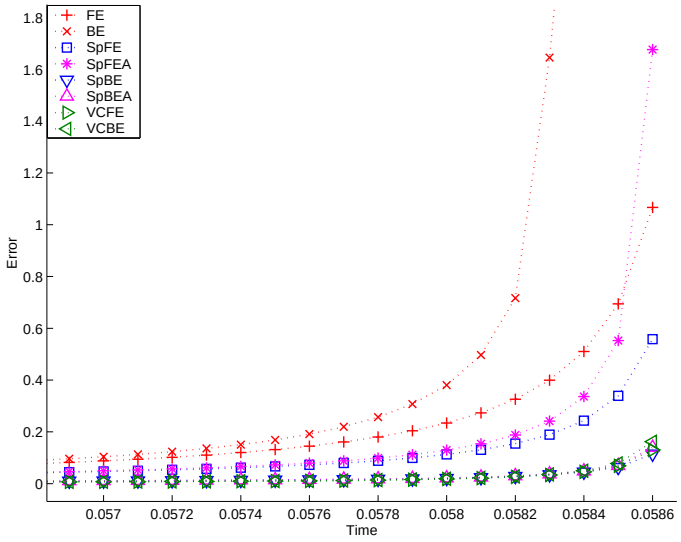

Figure A.3: Error for second-order methods applied to the semilinear parabolic equation with $u_0(x) = (1 - x^2)(8x^2 + 1)$, for timesteps close to $T_f = 0.0586$.

A.2 Semilinear Parabolic Equations with Different Functions F

In this section we present the results of numerical experiments for the semilinear parabolic equation (3.1) with $F(u) = (u + \alpha)^{p+1}$ and $F(u) = (u + 1)(\ln(u + 1))^{p+1}$.

For the first example, $F(u) = (u + \alpha)^{p+1}$, we used $\delta = 3$, $\alpha = 2$ and $p = 1$. The initial condition is $u_0(x) = \cos(\pi x/2)$ on $\Omega = [-1, 1]$. The blow-up time is approximately $T_b \approx 0.1209$.

For Figures A.5 and A.6 we computed the solution up to $T_f = 0.1150$, using

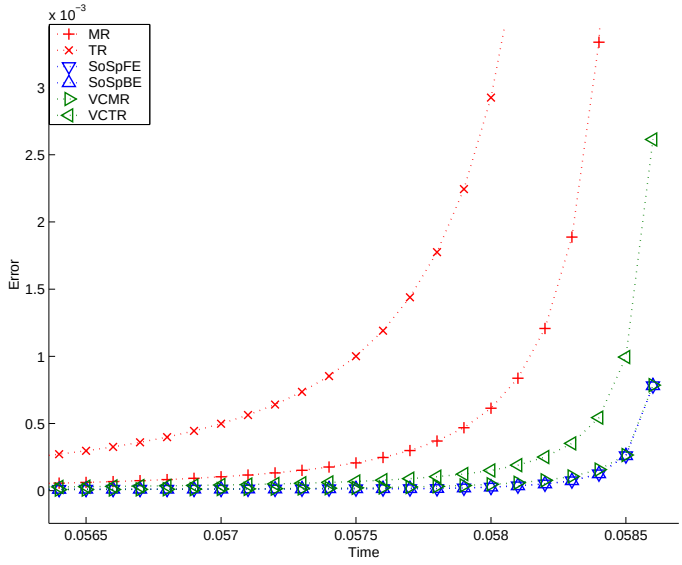

Figure A.4: Error for second-order methods applied to the semilinear parabolic equation with $u_0(x) = (1 - x^2)(8x^2 + 1)$, for timesteps close to $T_f = 0.0586$.

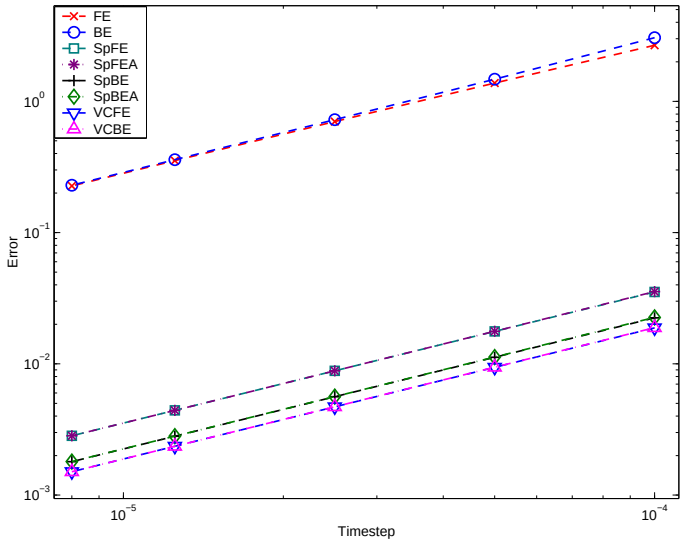

Figure A.5: Error at $T_f = 0.1150$ for first-order methods applied to the semilinear equation with $F(u) = (u + 2)^2$, with different values of h.

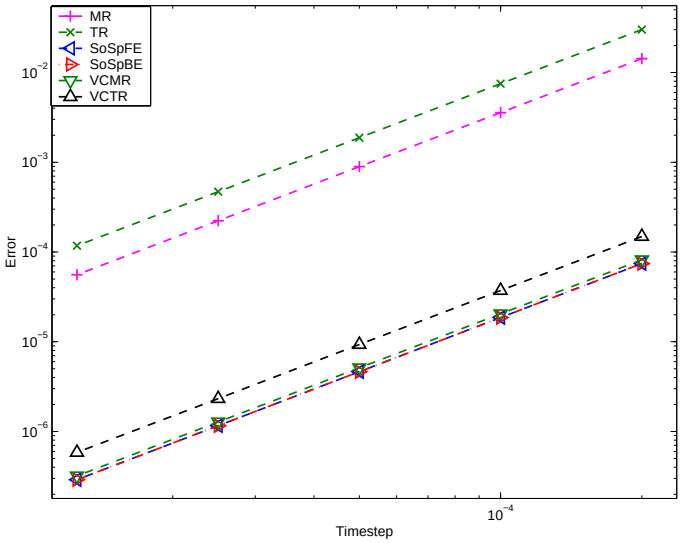

Figure A.6: Error at $T_f = 0.1150$ for second-order methods applied to the semilinear equation with $F(u) = (u+2)^2$, with different values of h.

Timestep	0.0001	5e-005	2.5e-005	1.25e-005	8e-006
FE	2.67	1.38	0.701	0.353	0.227
BE	3.06	1.47	0.725	0.359	0.229
SpFE	0.0353	0.0177	0.00884	0.00442	0.00283
SpFEA	0.0355	0.0177	0.00885	0.00442	0.00283
SpBE	0.0224	0.0112	0.00562	0.00281	0.0018
SpBEA	0.0226	0.0113	0.00563	0.00281	0.0018
VCFE	0.0188	0.0094	0.0047	0.00235	0.00151
VCBE	0.0189	0.00942	0.00471	0.00235	0.00151

Table A.1: Error at $T_f = 0.1150$ for first-order methods applied to the semilinear equation with $F(u) = (u+2)^2$.

Timestep	0.0002	0.0001	5e-005	2.5e-005	1.25e-005
MR	0.0143	0.00357	0.000893	0.000223	5.58e-005
TR	0.0301	0.00752	0.00188	0.00047	0.000117
SoSpFE	7.43e-005	1.86e-005	4.65e-006	1.16e-006	2.9e-007
SoSpBE	7.43e-005	1.86e-005	4.65e-006	1.16e-006	2.9e-007
VCMR	8.18e-005	2.05e-005	5.12e-006	1.28e-006	3.2e-007
VCTR	0.000149	3.72e-005	9.31e-006	2.33e-006	5.84e-007

Table A.2: Error at $T_f = 0.1150$ for second-order methods applied to the semilinear equation with $F(u) = (u+2)^2$.

the stepsizes $h = 0.0001$, 0.00005, 0.000025, 0.0000125 and 0.000008 for first-order methods and $h = 0.0002$, 0.0001, 0.00005, 0.000025 and 0.0000125 for second-order methods. The errors are listed in Tables A.1 and A.2.

For Figures A.7 and A.8, we used $h = 0.0001$ and computed the solutions up to $T_f = 0.1200$.

For the second example, $F(u) = (u+1)(\ln(u+1))^{p+1}$, we used $\delta = 6$ and $p = 1$. The initial condition is $u_0(x) = \cos(\pi x/2)$ on $\Omega = [-1, 1]$. The blow-up time is approximately $T_b \approx 0.3426$.

For Figures A.9 and A.10 we computed the solution up to $T_f = 0.3000$, using the stepsizes $h = 0.0001$, 0.00005, 0.000025, 0.0000125 and 0.000008 for first-order methods and $h = 0.0008$, 0.0004,0.0002, 0.0001 and 0.00005 for second-order methods. The errors are listed in Tables A.3 and A.4.

For Figures A.11 and A.12, we used $h = 0.0001$ and computed the solutions up to $T_f = 0.3280$.

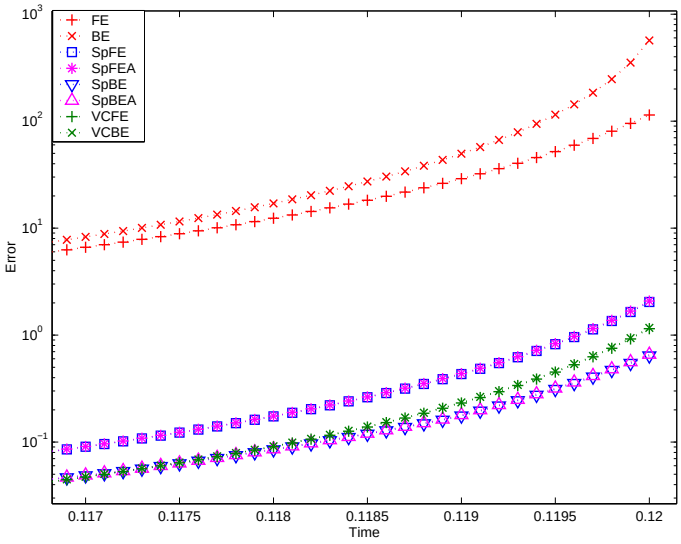

Figure A.7: Error for first-order methods applied to the semilinear equation with $F(u) = (u+2)^2$, for timesteps close to $T_f = 0.1200$.

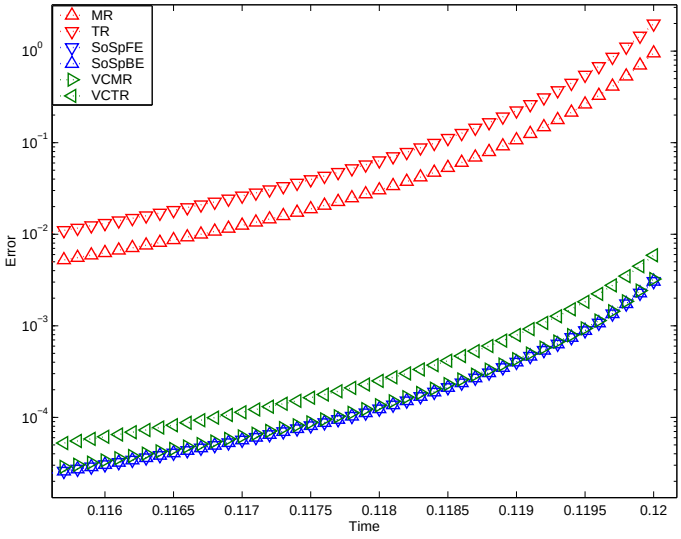

Figure A.8: Error for second-order methods applied to the semilinear equation with $F(u) = (u+2)^2$, for timesteps close to $T_f = 0.1200$.

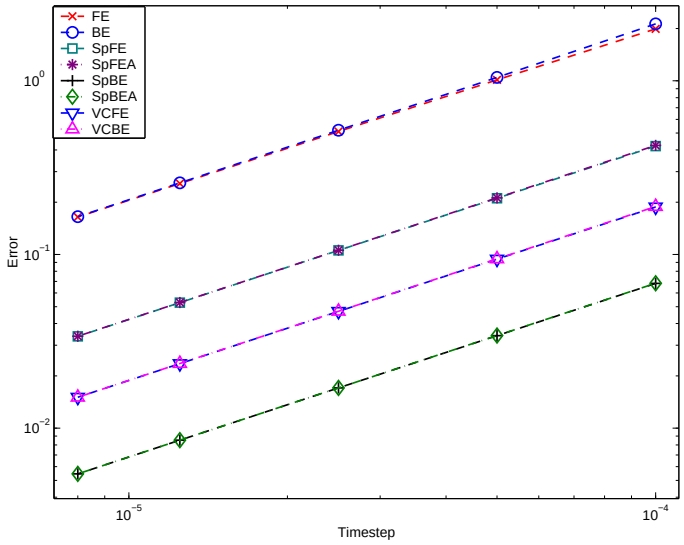

Figure A.9: Error at $T_f = 0.3000$ for first-order methods applied to the semilinear equation with $F(u) = (u+1)(\ln(u+1))^2$, with different values of h.

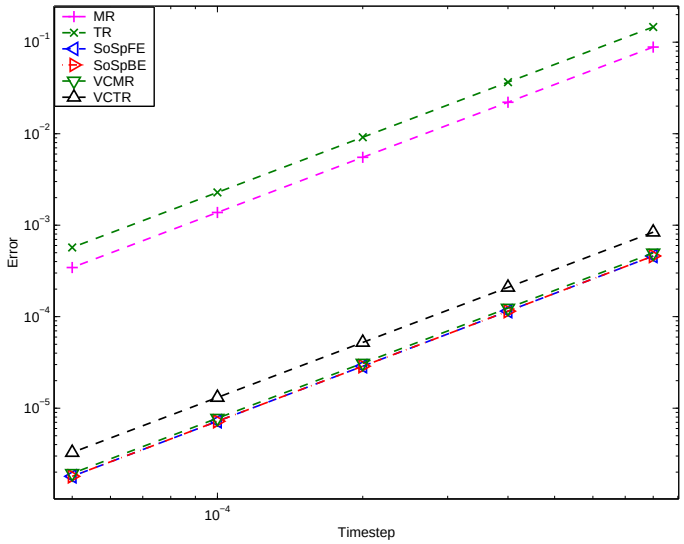

Figure A.10: Error at $T_f = 0.3000$ for second-order methods applied to the semilinear equation with $F(u) = (u+1)(\ln(u+1))^2$, with different values of h.

Timestep	0.0001	5e-005	2.5e-005	1.25e-005	8e-006
FE	1.99	1.01	0.51	0.256	0.164
BE	2.13	1.05	0.519	0.258	0.165
SpFE	0.42	0.211	0.106	0.0528	0.0338
SpFEA	0.425	0.212	0.106	0.0529	0.0338
SpBE	0.0681	0.0341	0.017	0.00852	0.00545
SpBEA	0.0683	0.0341	0.017	0.00852	0.00545
VCFE	0.188	0.094	0.047	0.0235	0.0151
VCBE	0.189	0.0942	0.0471	0.0235	0.0151

Table A.3: Error at $T_f = 0.3000$ for first-order methods applied to the semilinear equation with $F(u) = (u+1)(\ln(u+1))^2$.

Timestep	0.0008	0.0004	0.0002	0.0001	5e-005
MR	0.0884	0.0221	0.00552	0.00138	0.000345
TR	0.147	0.0366	0.00913	0.00228	0.000571
SoSpFE	0.00046	0.000115	2.88e-005	7.19e-006	1.8e-006
SoSpBE	0.00046	0.000115	2.88e-005	7.19e-006	1.8e-006
VCMR	0.000497	0.000124	3.1e-005	7.76e-006	1.94e-006
VCTR	0.000838	0.000209	5.24e-005	1.31e-005	3.27e-006

Table A.4: Error at $T_f = 0.3000$ for second-order methods applied to the semilinear equation with $F(u) = (u+1)(\ln(u+1))^2$.

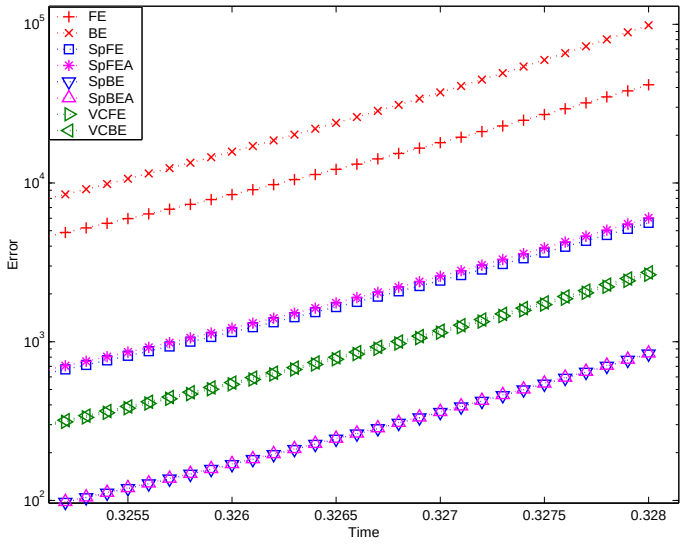

Figure A.11: Error for first-order methods applied to the semilinear equation with $F(u) = (u+1)(\ln(u+1))^2$, for timesteps close to $T_f = 0.3280$.

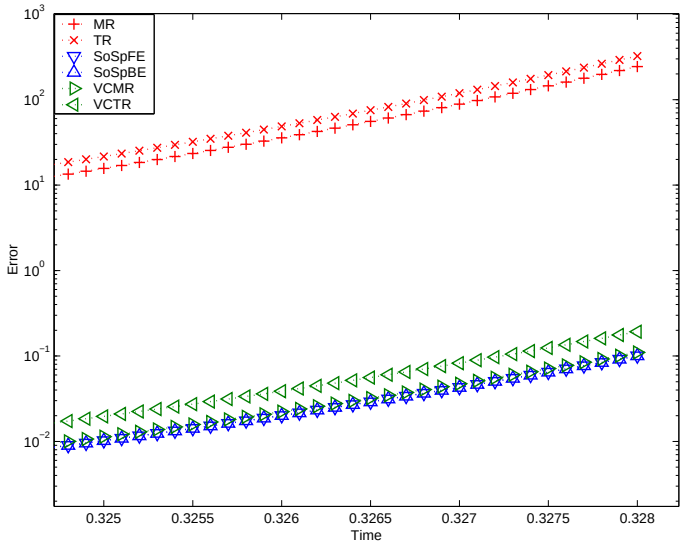

Figure A.12: Error for second-order methods applied to the semilinear equation with $F(u) = (u+1)(\ln(u+1))^2$, for timesteps close to $T_f = 0.3280$.

A.3 Semilinear System

In this section we present the results of numerical experiments for the system of semilinear parabolic equations (3.24) 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 $T_b \approx 0.1181$.

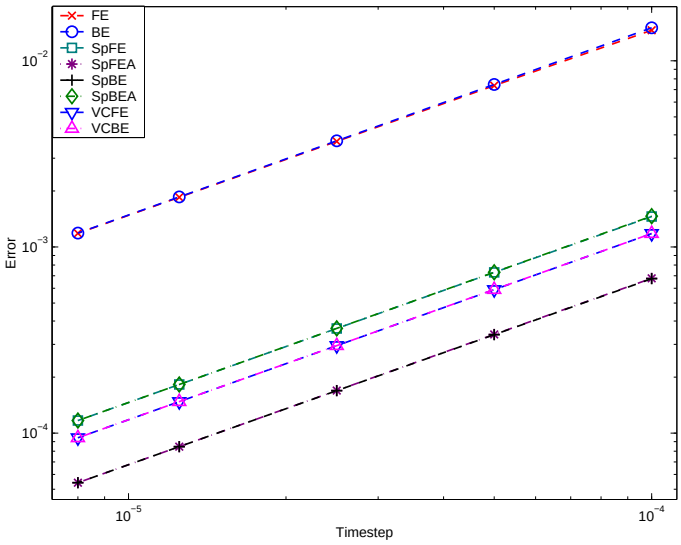

Figure A.13: Error at $T_f = 0.1100$ for first-order methods applied to the system of semilinear equations with different values of h.

For Figures A.13 and A.14 we computed the solution up to $T_f = 0.1100$, using the stepsizes $h = 0.0001$, 0.00005, 0.000025, 0.0000125 and 0.000008 for first-order methods and $h = 0.0004$, 0.0002, 0.0001, 0.00005 and 0.000025 for second-order methods. The errors are listed in Tables A.1 and A.2.

For Figures A.15 and A.16, we used $h = 0.0001$ and computed the solutions up to $T_f = 0.1170$.

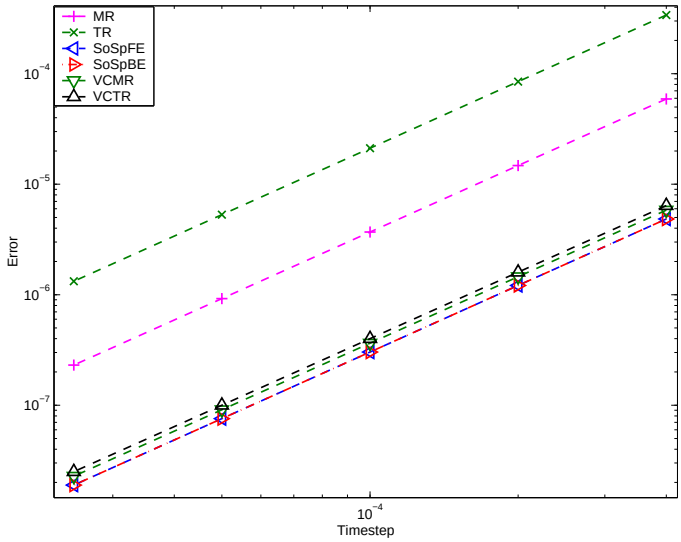

Figure A.14: Error at $T_f = 0.1100$ for second-order methods applied to the system of semilinear equations with different values of h.

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

Table A.5: Error at $T_f = 0.1100$ for first-order methods applied to the system of semilinear equations.

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 A.6: Error at $T_f = 0.1100$ for second-order methods applied to the system of semilinear equations.

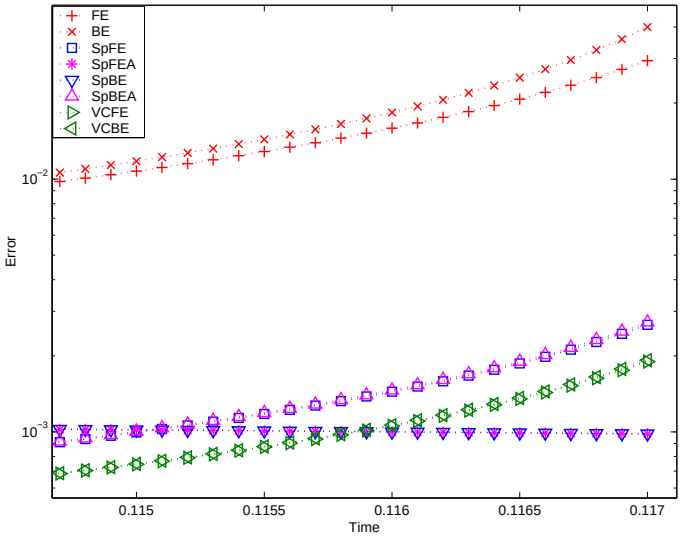

Figure A.15: Error for first-order methods applied to the system of semilinear equations, for timesteps close to $T_f = 0.1170$.

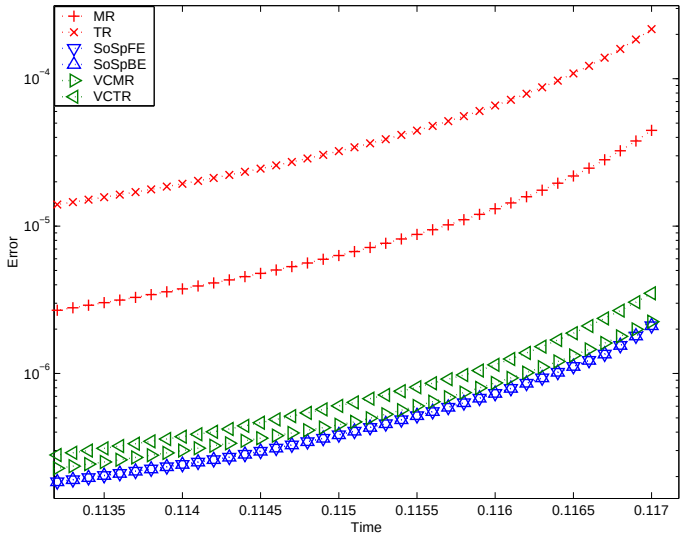

Figure A.16: Error for second-order methods applied to the system of semilinear equations, for timesteps close to $T_f = 0.1170$.

A.4 "Accretive" Equation (3.38)

In this section we present the results of numerical experiments for the "accretive" equation (3.38) with $\delta = 3$. The initial conditions are $u_0(x) = \cos(\pi x/2)$ and $u_{t0}(x) = \cos(\pi x/2)$ on $\Omega = [-1, 1]$. The blow-up time is approximately $T_b \approx 0.1483$.

For Figures A.17 and A.18 we computed the solution up to $T_f = 0.1450$, using the stepsizes $h = 0.00025$, 0.000125, 0.00005, 0.000025 and 0.0000125 for first-order methods and $h = 0.005$, 0.001, 0.0005, 0.00025 and 0.000125 for second-order methods. The errors are listed in Tables A.7 and A.8.

For Figures A.19 and A.20, we used $h = 0.0001$ and computed the solutions up to $T_f = 0.1482$.

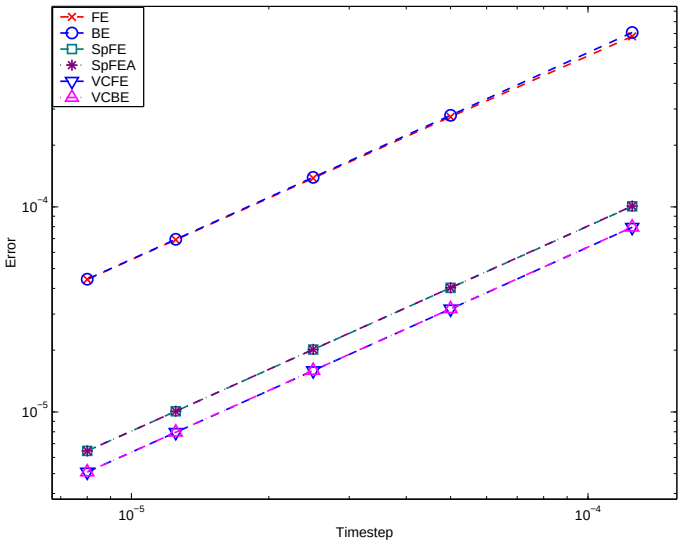

Figure A.17: Error at $T_f = 0.1450$ for first-order methods applied to the accretive equation, with different values of h.

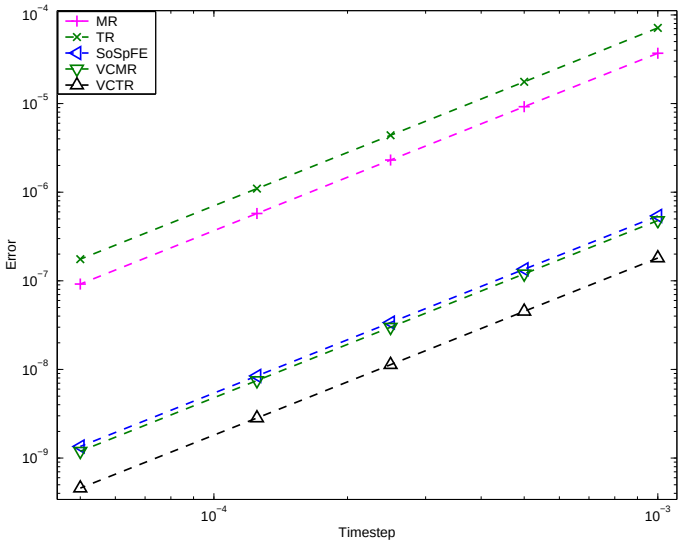

Figure A.18: Error at $T_f = 0.1450$ for second-order methods applied to the accretive equation, with different values of h.

Timestep	0.000125	5e-005	2.5e-005	1.25e-005	8e-006
FE	0.000679	0.000275	0.000138	6.91e-005	4.43e-005
BE	0.000709	0.00028	0.000139	6.94e-005	4.44e-005
SpFE	0.000101	4.03e-005	2.02e-005	1.01e-005	6.45e-006
SpFEA	0.000101	4.03e-005	2.02e-005	1.01e-005	6.45e-006
VCFE	7.95e-005	3.18e-005	1.59e-005	7.95e-006	5.09e-006
VCBE	7.96e-005	3.18e-005	1.59e-005	7.95e-006	5.09e-006

Table A.7: Error at $T_f = 0.1450$ for first-order methods applied to the accretive equation.

Timestep	0.001	0.0005	0.00025	0.000125	5e-005
MR	3.7e-005	9.22e-006	2.3e-006	5.76e-007	9.21e-008
TR	7.13e-005	1.76e-005	4.39e-006	1.1e-006	1.75e-007
SoSpFE	5.41e-007	1.35e-007	3.38e-008	8.45e-009	1.35e-009
VCMR	4.79e-007	1.2e-007	3e-008	7.49e-009	1.19e-009
VCTR	1.81e-007	4.53e-008	1.13e-008	2.83e-009	4.58e-010

Table A.8: Error at $T_f = 0.1450$ for second-order methods applied to the accretive equation.

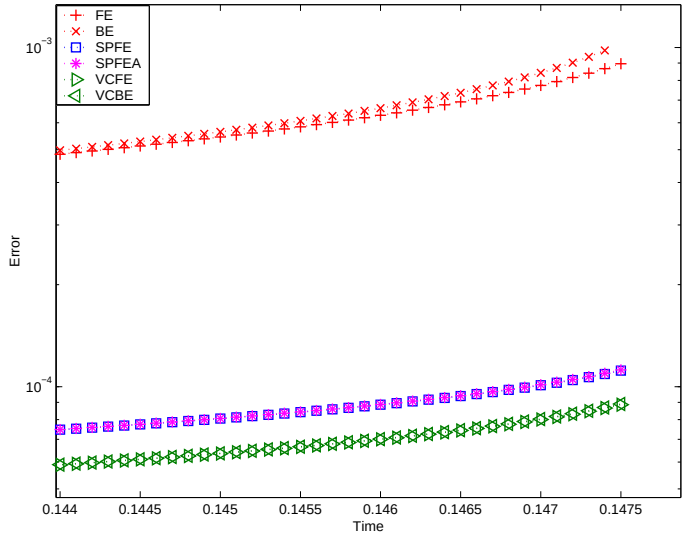

Figure A.19: Error for first-order methods applied to the accretive equation, for timesteps close to $T_f = 0.1482$.

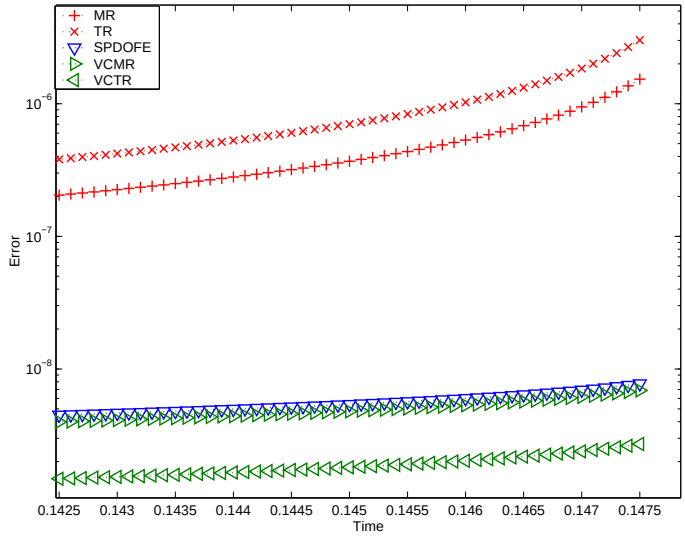

Figure A.20: Error for second-order methods applied to the accretive equation, for timesteps close to $T_f = 0.1482$.