Finite Element Approximation of Some Degenerate/Singular Elliptic and Parabolic Equations

A thesis presented for the degree of
Doctor of Philosophy of the University of London
and the
Diploma of Imperial College
by

Mathias Schneider

Department of Mathematics
Imperial College of Science, Technology and Medicine
London SW7 2BZ

May 1997

Abstract

The main purpose of this thesis is to prove a priori error bounds for the standard continuous piecewise linear finite element approximation of a semilinear elliptic problem with a singular nonlinearity which occurs in nonlinear elasticity and a semilinear degenerate parabolic problem modelling fast diffusion.

The elliptic problem can be stated as follows: Given $\nu \in \mathbb{R}_+$, find $u > 0$ such that

$$-\Delta u = cu^{-\nu} \quad \text{in } \Omega\,, \quad u = 0 \quad \text{on } \partial\Omega\,,$$

where $\Omega \subset \mathbb{R}^d$ with sufficiently smooth boundary, $d = 1$, 2 or 3, and $c > 0$ in $\bar{\Omega}$. Although this problem has been studied in the literature from an analytical point of view, we are not aware of any finite element error analysis on it. We prove H^1 and L^∞ error bounds for the standard Galerkin finite element approximation with a (weakly) acute triangulation. Our bounds are nearly optimal. In addition, for $d = 1$ and 2 and $c \in \mathbb{R}_+$ we analyse a more practical scheme involving numerical integration on the nonlinear term. We obtain nearly optimal H^1 and L^∞ error bounds for $d = 1$. For this case we also present some numerical results.

The parabolic problem modelling fast diffusion is to find u such that

$$\partial_t \phi(u) = \Delta u + f \quad \text{in } Q_T := \Omega \times (0,T]\,, \qquad u = 0 \quad \text{on } \partial\Omega \times (0,T]\,,$$
$$\phi(u)(\cdot,0) = \phi(g(\cdot)) \quad \text{in } \Omega\,,$$

where Ω is convex polyhedral in $\mathbb{R}^d$, $d = 1$, 2 or 3, $\phi(s) := \lambda \operatorname{sign}(s)\,|s|^p$ with $\lambda > 0$ and $p > 1$ and f, g given. There is a vast literature on the singular (i.e. $p < 1$) form of the given problem, which models slow diffusion. Considerably less work has been done on the degenerate case, $p > 1$. We are aware of only one article which proves an error bound for a fully practical Galerkin scheme; and this only for $d = 1$.

We prove L^2 error bounds for the standard Galerkin finite element approximation in space and backward Euler time discretization. Our bounds hold for $d = 1, 2$ and 3. In addition, we analyse a more practical scheme involving numerical integration in space on the nonlinear terms and derive error bounds for the fully practical method. Our bounds improve on those in the literature both in terms of the order of convergence and in terms of the size of the time step required for the computation. After that we include an adaptation to our setting of a novel but abstract approach given by Rulla for the analysis of the backward Euler method.

We also deal with the finite time extinction property which distinguishes the fast from the slow diffusion equation and its appropriate numerical treatment via Le Roux's method. Numerical results for both the backward Euler and Le Roux's method are presented for $d = 1$. We conclude this thesis with a brief look at linear schemes deriving from extrapolation and relaxation.

Für meine Eltern
und
für meine Brüder

In a good wind even a turkey can fly.[1]

Acknowledgements

First and foremost I would like to thank my supervisor, Dr. John Barrett, without whose interest, help and determination this thesis could not have been written. It was he who showed me how to bang the bad guys into the other side.

Nor could I have managed without the financial assistance of the Engineering and Physical Sciences Research Council, U.K., which paid for my fees, and the support of the Studienstiftung des Deutschen Volkes, Germany, which paid for the worldly rest: bread, butter, books and the roof over my head. Many thanks to Prof. Watzlawek from the Universität Konstanz who cared at a very crucial time. I also gratefully acknowledge the receipt of the Harry Jones Scholarship through the Applied Mathematics/Numerical Analysis Section of the Maths Department at Imperial.

I am also very much indebted to Ottar, Andrew and Ross for all the hands on help with the computing system and all the philosophical discussions about football or otherwise.

Thanks very much also to my office mates of whom there have been quite a few over the last three and a half years: Heloisa, Marwan, Isabel, Ali, George, Maxine, and Peter. We did not only sit in the same office, but also in the same boat (we may not have been aware of it all the time). In this context I also have to mention "the girls from 632": April, Ioanna, Patricia, Rupa, Thuruya, and Zaineb. But I don't want to forget Mauro and Peyman either.

Well, thanks also to some members of staff: thanks to Prof. Jeff Cash for taking an interest, to Dr. Gerald Moore for our discussions about maths, to Dr. Ann Mitchell for the occasional genuine chat about art and life in general, and to Prof. Colin Atkinson for all the remarks off the record. I have had a good time at Imperial.

Finally, I have to thank my family to whom I dedicate this thesis: to my parents for always having made ends meet and to my brothers for keeping my feet on the ground.

[1] a Chinese saying, probably

Contents

1 A Semilinear Singular Elliptic Problem **11**

1.1 Introduction 11

1.2 Regularization 13

1.2.1 The Regularized Problem 13

1.2.2 The Regularization Error 20

1.3 An Alternative Regularization 25

1.3.1 The Regularized Problem 25

1.3.2 The Error for the Alternative Regularization 27

1.4 Finite Element Approximation 28

1.4.1 The Discrete Problem 28

1.4.2 The Discretization Error 30

1.4.3 Discretization and Alternative Regularization 36

1.4.4 Optimality of the Error Bounds 36

1.5 The Effect of Numerical Integration 39

1.5.1 The Discrete Problem with Numerical Integration 39

1.5.2 Error Analysis 40

1.5.3 (Nearly) Optimal Error Bounds for $d = 1$ 47

1.5.4 Numerical Integration and Alternative Regularization 51

1.6 Numerical Experiments 53

1.6.1 The Systems of Nonlinear Algebraic Equations 53

1.6.2 Numerical Results 56

2 A Semilinear Degenerate Parabolic Problem **69**

2.1 Introduction 69

2.2 Regularization 72

2.2.1 Existence of a Weak Solution to (P_ε^M) 73

2.2.2 Existence of a Weak Solution to (P_ε) 78

2.2.3 Existence of a Weak Solution to (P); Extinction in Finite Time . . . 81

2.3 A Semidiscrete Approximation 86
2.3.1 The Effect of Numerical Integration 88
2.4 A Fully Discrete Finite Element Approximation 93
2.5 An Improved Error Bound for the Time Discretization 96
2.6 Le Roux's Method for the Discretization in Time 102
2.7 Numerical Experiments 109
2.7.1 The System of Nonlinear Algebraic Equations for $(\hat{P}_N^h)$ 109
2.7.2 Numerical Results for $(\hat{P}_N^h)$ in the Separable Case 112
2.7.3 Error Analysis for the ODE 119
2.7.4 Some Numerical Results for Le Roux's Method 125
2.7.5 The Nonseparable Case and Asymptotic Profiles 131
2.8 Linear Schemes 140
2.8.1 A Scheme based on Extrapolation 140
2.8.2 A Scheme based on Relaxation 147

Bibliography **153**

List of Tables

1.1 Computational orders of convergence for $\nu = 0.5$ and $\varepsilon = h^{\frac{4}{3}}$ 63
1.2 Number of components of U_ε which are less than ε for $\nu = 0.5$ and f_ε. . . . 64
1.3 Computational orders of convergence for $\nu = 3$ and $\varepsilon = h^{\frac{1}{2}}$. 64
1.4 $p_\infty(h)$ for different ν's with $h = 2^{-15}$ and $\varepsilon = h^{\frac{2}{1+\nu}}$ 66
1.5 $p_1(h)$ for different ν's with $h = 2^{-15}$ and $\varepsilon = h^{\frac{2}{1+\nu}}$ 67
1.6 $p_1(h)$ for different ν's with $h = 2^{-15}$ and $\varepsilon = \hat{C} \times h^{\frac{4}{1+3\nu}}$ 67
1.7 $E_\infty(h)$ for different ν's with $h = 2^{-15} \approx 3.05 \times 10^{-5}$ and $\varepsilon = h^{\frac{2}{1+\nu}}$ 68

2.1 $E_u^\tau(h)$ and $q_u(h)$ for $p = 2$ with $T_{ex} = 0.25$ and $T_* = T = 0.5$ 113
2.2 $E_u^\tau(h)$ and $q_u(h)$ for $p = 5$ with $T_{ex} = 0.25$ and $T_* = T = 0.5$ 114
2.3 $E_v^\tau(h)$ and $q_v(h)$ for $p = 2$ with $T_{ex} = 0.25$ and $T_* = T = 0.5$ 114
2.4 $E_v^\tau(h)$ and $q_v(h)$ for $p = 5$ with $T_{ex} = 0.25$ and $T_* = T = 0.5$ 115
2.5 Computational and theoretical orders of convergence for different values of p 115
2.6 $E_u^\tau(h)$ and $q_u(h)$ for $p = 2$ with $T_{ex} = 0.2$ and $T = 0.4$ 116
2.7 $E_u^\tau(h)$ and $q_u(h)$ for $p = 5$ with $T_{ex} = 0.2$ and $T = 0.4$ 116
2.8 $\tilde{E}_u^\tau(h)$ and $\tilde{q}_u(h)$ for $p = 5$ with $T_{ex} = 0.2$ and $T = 0.4$ 117
2.9 Comp. orders of conv. for PDE and ODE with $T_{ex} = 0.25$ and $T = 0.5$ . . 118
2.10 $E_u(\tau)$ and $q_u(\tau)$ for ODE case with $T_{ex} = 0.2$ and $T = 0.4$ 118
2.11 Computational and theoretical orders of convergence for ODE (7.19) . . . 123
2.12 Computational orders for the first term in (7.45) 123
2.13 $q_{v,\infty}(2^{-14})$ for different values of p. 124
2.14 Comput. orders of conv. for ODE case with $p = 4$, $T_{ex} = 0.2$ and $T = 0.4$. 124
2.15 Le Roux's method: behaviour around extinction time $T_{ex} = 0.2$ 127
2.16 Test of Le Roux's method: behaviour around extinction time $T_{ex} = 0.2$ . . 128
2.17 Backward Euler method: behaviour around extinction time $T_{ex} = 0.2$ . . . 129

2.18 Fully discrete Le Roux's method: behaviour around ext. time $T_{ex} = 0.2$ . . 130
2.19 Fully discrete Le Roux's method: computational extinction times $T^{\tau}_{ex,10^{-8}}$. 132
2.20 Backward Euler method: Computational extinction time allowing 10^{-8} . . 136
2.21 Fully discrete backward Euler method: computational ext. times $T^{\tau}_{ex,10^{-8}}$. 137
2.22 $E^{\varepsilon}_{u,2}(\tau)$ and $q^{\varepsilon}_{u,2}(\tau)$ for ODE case with $T_{ex} = 0.2$ and $T = 0.4$ 145
2.23 $E^{\varepsilon}_{v,2}(\tau)$ and $q^{\varepsilon}_{v,2}(\tau)$ for ODE case with $T_{ex} = 0.2$ and $T = 0.4$ 146
2.24 A according to (8.35) and computational convergence 151
2.25 maximum error and relaxation error 151

List of Figures

1.1 Orders of convergence according to Theorem 1.1 versus optimal orders. . . 39
1.2 Orders of convergence according to Theorems 1.1 and 1.3. 51
1.3 Some piecewise lin. Galerkin approx. for $\nu = 0.5$ and comput. with f . . . 60
1.4 Some piecewise lin. Galerkin approx. for $\nu = 0.5$ and comput. with f_ε . . . 60
1.5 Some piecewise lin. Galerkin approx. for $\nu = 0.5$ and comput. with g_ε . . . 60
1.6 Some piecewise lin. Galerkin approx. for $\nu = 3$ and comput. with f 65
1.7 Some piecewise lin. Galerkin approx. for $\nu = 3$ and comput. with f_ε 65
1.8 Some piecewise lin. Galerkin approx. for $\nu = 3$ and comput. with g_ε 65

2.1 Two different initial value functions: Case (a) and Case (b). 131
2.2 Case (a): $p = 2$: Profiles of factored solution as $t \to T_{ex,10^{-8}}^{2^{-11}} = 0.130371$ . . 134
2.3 Case (a): $p = 5$: Profiles of factored solution as $t \to T_{ex,10^{-8}}^{2^{-11}} = 0.052734$ . . 134
2.4 Case (b): $p = 2$: Profiles of factored solution as $t \to T_{ex,10^{-8}}^{2^{-11}} = 0.111328$ . . 135
2.5 Case (b): $p = 5$: Profiles of factored solution as $t \to T_{ex,10^{-8}}^{2^{-11}} = 0.041504$ . . 135
2.6 Case (a): $p = 2$: Profiles of factored solution as $t \to T_{ex,10^{-8}}^{2^{-11}} = 0.133789$ . . 138
2.7 Case (a): $p = 5$: Profiles of factored solution as $t \to T_{ex,10^{-8}}^{2^{-11}} = 0.053711$ . . 138
2.8 Case (b): $p = 2$: Profiles of factored solution as $t \to T_{ex,10^{-8}}^{2^{-11}} = 0.114746$ . . 139
2.9 Case (b): $p = 5$: Profiles of factored solution as $t \to T_{ex,10^{-8}}^{2^{-11}} = 0.042967$ . . 139
2.10 *Error* for extrapolation method, $p = 2$; $C = 0.001$: ——, $C = 1$: – – – . . . 145
2.11 Relaxation method applied to ODE, $p = 2$ 149
2.12 Relaxation method applied to ODE, $p = 4$ 149

Some Preliminary Remarks

As can be seen from the Contents this thesis consists of two fairly large chapters subdivided into a number of sections and subsections. Both chapters are organised in the same way: starting from the analysis of a singular/degenerate partial differential equation via regularization we proceed to the analysis of finite element discretizations and conclude both chapters with numerical experiments. Our main results are called Theorems and auxiliary results are called Lemmas. Theorems, Lemmas and main Definitions are labelled according to chapters, e.g. **Lemma 1.5** denotes the fifth Lemma in chapter 1.

Next we point out that the two chapters were written so that they can be read independently of each other. Although this may have slightly increased the total number of pages, it should also have increased the legibility. For the same reason our formulae are labelled according to sections within the two chapters. The number of the chapter is omitted in the numbers of the formulae, e.g. (3.2) denotes formula 2 in section 3 of both chapter 1 and chapter 2. As there is no cross-referencing between the two chapters no confusion will arise.

Finally, we remark that throughout this thesis we use standard notation unless otherwise specified. In particular, the notation concerning the classical function spaces, Lebesque and Sobolev spaces and their associated (semi)norms can be found in [11], and the notation concerning the parabolic spaces in chapter 2 can be found in [48].

Chapter 1

A Semilinear Singular Elliptic Problem

1.1 Introduction

Let Ω be a bounded, connected domain in $\mathbb{R}^d$, $d = 1, 2$ or 3. The boundary $\partial\Omega$ is assumed to be C^2. We consider the following problem:

Find u such that $u > 0$ in Ω, $u = 0$ on $\partial\Omega$ and

$$(P) \qquad -\Delta u = c(x) f(u)\,,$$

where we will assume that $f(s) := s^{-\nu}$, $\nu > 0$ and c is a non-negative, bounded and measurable function.

Applications of (P) can be found for $\nu = 1$ in fluid mechanics [10] and for $\nu = 2$ in nonlinear elasticity [19]. The kind of singular nonlinearity which occurs in (P) has been considered by several authors (e.g. [12], [13], [21], [27], [42] and the references therein) from an analytical point of view. The weakest assumption on c to exclude the Laplace equation (with solution $u \equiv 0$) for (P) is

$$(A1) \qquad c \geq 0 \text{ a.e. and } meas\{x \in \Omega : c(x) > 0\} > 0\,.$$

In fact, del Pino [13] has proved that under $(A1)$ and $\partial\Omega \in C^2$ and for all $\nu > 0$ there exists a unique function $u \in C^{1,\gamma}(\Omega) \cap C(\bar{\Omega})$ for all $\gamma \in (0,1)$ which satisfies (P) in the following weak sense: $u > 0$ in Ω, $u = 0$ on $\partial\Omega$ and

$$(WP) \qquad \int_\Omega \nabla u \cdot \nabla \phi \, dx = \int_\Omega c f(u) \phi \, dx \qquad \forall \phi \in C_0^\infty(\Omega)\,.$$

In fact, $u \in C^{2,\beta}(\Omega) \cap C(\bar{\Omega})$ and thus is a classical solution of (P), if one also assumes that $c \in C^\beta(\bar{\Omega})$ for a $\beta \in (0,1)$ (cf. our remark after Lemma 1.5). This is in accordance with [42] where the existence of positive solutions of generalisations of (P) is proved.

Under the assumption

(A2) There exist constants $\alpha \geq 0$ and M, $m > 0$ such that

$$m\,[dist(x, \partial\Omega)]^{\alpha} \leq c(x) \leq M\,[dist(x, \partial\Omega)]^{\alpha} \qquad \text{a.e. in } \Omega\,.$$

Gui and Lin ([21], Theorem 2.1) have shown the following boundary behaviour of u:

(i) If $\nu < 1 + \alpha$, then there exist λ, $\Lambda > 0$ such that

$$\lambda\, d(x) \leq u(x) \leq \Lambda\, d(x) \qquad \text{for all } x \in \bar{\Omega}\,; \tag{1.1}$$

(ii) If $\nu = 1 + \alpha$, then there exist λ, Λ, $D > 0$ such that

$$\lambda\, d(x)[D - \ln d(x)]^{\frac{1}{1+\nu}} \leq u(x) \leq \Lambda\, d(x)[D - \ln d(x)]^{\frac{1}{1+\nu}} \qquad \text{for all } x \in \bar{\Omega}\,; \tag{1.2}$$

(iii) If $\nu > 1 + \alpha$, then there exist λ, $\Lambda > 0$ such that

$$\lambda\,[d(x)]^{\frac{2+\alpha}{1+\nu}} \leq u(x) \leq \Lambda\,[d(x)]^{\frac{2+\alpha}{1+\nu}} \qquad \text{for all } x \in \bar{\Omega}\,, \tag{1.3}$$

where we have used the notation $d(x) := dist(x, \partial\Omega)$. As a consequence of Remark 2.2 in [21] we have that the lower bound of (A2) is not needed to prove (1.1). On using the inequalities (1.1) – (1.3) Gui and Lin [21] were able to show that

(i) If $\nu < 1 + \alpha$, then

$$u \in C^{1,\gamma}(\bar{\Omega})\,, \tag{1.4}$$

where $\gamma = 1 + \alpha - \nu$, if $\alpha < \nu$ and γ may be any number in $(0, 1)$, if $\nu \leq \alpha$;

(ii) If $\nu = 1 + \alpha$, then

$$u \in C^{\gamma}(\bar{\Omega})\,, \quad \text{for all } \gamma \in (0, 1); \tag{1.5}$$

(iii) If $\nu > 1 + \alpha$, then

$$u \in C^{\gamma}(\bar{\Omega})\,, \quad \text{where } \gamma = \frac{2+\alpha}{1+\nu}; \tag{1.6}$$

see Theorem 1.1 in [21].

In this chapter we prove $H^1(\Omega)$ and $L^{\infty}(\Omega)$ error estimates for a finite element approximation of (P). The outline is as follows. In section 2 we consider a regularized version (P_ε) of (P) which we obtain by replacing f by a non-singular function f_ε, and bounds for the regularization error are shown. In our analysis we will use the inequalities (1.1) – (1.3) to show the following property of u:

$$\int_{0<u<\varepsilon} c f(u) u^{\mu}\, dx \leq C_* \varepsilon^{\sigma}\,, \tag{1.7}$$

where μ and σ depend on ν in (P) and α in (A2). (1.7) is the basis for our $H^1(\Omega)$ regularization bounds. In section 3 we will consider a second method for the regularization of (P) and compare its properties with the one from section 2.

In section 4 we discretize (P_ε) by applying a Galerkin approximation using continuous piecewise linear functions. For $\alpha = 0$, $H^1(\Omega)$ and $L^\infty(\Omega)$ error estimates will be proved, the latter by adapting a technique first introduced by Nochetto [32] for a semilinear elliptic problem with a free boundary. In section 5 we take into account the effect of numerical integration which is important in practice. However, we will only deal with the case $c \equiv$ constant. Finally, we present some numerical results for $d = 1$ in section 6.

In order to use the inequalities (1.1) – (1.3) and $u \in C^{2,\beta}(\Omega) \cap C(\bar{\Omega})$, we will make the following assumptions throughout this chapter: There exists a constant $\alpha \geq 0$ such that

$$A(\alpha) \begin{cases} c \in C^\beta(\bar{\Omega}) \quad \text{for some } \beta \in (0,1); \\ m\,[d(x)]^\alpha \leq c(x) \leq M\,[d(x)]^\alpha \quad \text{for some } M,\ m > 0 \text{ and all } x \in \bar{\Omega}; \\ \nu \in \mathbb{R}_+ := \{t \in \mathbb{R} : t > 0\}; \\ \partial\Omega \in C^2 \end{cases}$$

holds.

1.2 Regularization

1.2.1 The Regularized Problem

We now introduce a regularized version of (P).

Find u_ε such that $u_\varepsilon = 0$ on $\partial\Omega$ and

$$(P_\varepsilon) \qquad -\Delta u_\varepsilon = c(x) f_\varepsilon(u_\varepsilon) \quad \text{in } \Omega\,,$$

where for $\varepsilon > 0$

$$f_\varepsilon(s) := \begin{cases} f(\varepsilon) + f'(\varepsilon)(s-\varepsilon) & , \quad s \in (-\infty, \varepsilon) \\ f(s) & , \quad s \in [\varepsilon, \infty)\,. \end{cases}$$

Clearly, $f_\varepsilon \in C^1(\mathbb{R})$ with

$$f_\varepsilon(s) < f(s) \qquad \forall s \in (0,\varepsilon) \quad \text{and} \tag{2.1}$$

$$-\nu\varepsilon^{-(\nu+1)} \leq f_\varepsilon'(s) < 0 \qquad \forall s \in \mathbb{R} \tag{2.2}$$

and from (2.2) it follows that $\forall a, b \in \mathbb{R}$,

$$-\frac{\nu}{\varepsilon^{\nu+1}}\,[a-b]^2 \leq [f_\varepsilon(a) - f_\varepsilon(b)][a-b] \leq -\frac{\varepsilon^{\nu+1}}{\nu}\,[f_\varepsilon(a) - f_\varepsilon(b)]^2 \leq 0\,. \tag{2.3}$$

Definition 1.1 *u_ε is a weak solution of (P_ε), if $u_\varepsilon \in H_0^1(\Omega)$ is such that*

$$(WP_\varepsilon) \qquad \int_\Omega \nabla u_\varepsilon \cdot \nabla v\, dx = \int_\Omega c f_\varepsilon(u_\varepsilon) v\, dx \qquad \forall v \in H_0^1(\Omega)\,.$$

In order to establish existence and uniqueness of a weak solution to (P_ε), we define an operator $\mathcal{A}_\varepsilon : H_0^1(\Omega) \to [H_0^1(\Omega)]^*$ by

$$\langle \mathcal{A}_\varepsilon v, w\rangle := \int_\Omega \nabla v \cdot \nabla w \, dx - \int_\Omega c f_\varepsilon(v) w \, dx \qquad \forall v,\, w \in H_0^1(\Omega)\,.$$

It is easy to see that $f_\varepsilon(v) \in L^2(\Omega)$ for all $v \in H_0^1(\Omega)$, and that $\mathcal{A}_\varepsilon v$ is a linear functional on $H_0^1(\Omega)$. We note that for all $v_1, v_2,\, w \in H_0^1(\Omega)$,

$$\begin{aligned} |\langle \mathcal{A}_\varepsilon v_1 - \mathcal{A}_\varepsilon v_2, w\rangle| &\le \int_\Omega |\nabla(v_1 - v_2)\cdot\nabla w)|\, dx + \int_\Omega |c[f_\varepsilon(v_1) - f_\varepsilon(v_2)]w|\, dx \\ &\le \|\nabla(v_1 - v_2)\|_{0,\Omega}\, \|\nabla w\|_{0,\Omega} + C\|f_\varepsilon(v_1) - f_\varepsilon(v_2)\|_{0,\Omega}\, \|w\|_{0,\Omega} \\ &\le C(\varepsilon)\, \|v_1 - v_2\|_{1,\Omega}\, \|w\|_{1,\Omega}\,, \end{aligned} \tag{2.4}$$

where we have noted the Mean Value Theorem and (2.2).

Obviously, (WP_ε) is equivalent to the operator equation

$$(OP_\varepsilon) \qquad \mathcal{A}_\varepsilon u_\varepsilon = 0 \quad \text{in } [H_0^1(\Omega)]^* \equiv H^{-1}(\Omega)\,.$$

Lemma 1.1 *For all $\varepsilon > 0$ there exists a unique solution u_ε to (OP_ε) and hence to (WP_ε).*

Proof. $H_0^1(\Omega)$ is a separable and reflexive Banach space. We show that $\mathcal{A}_\varepsilon$ is a strictly monotone, coercive and hemicontinuous operator. The assertion then follows from the Browder-Minty-Theorem (cf. [49], p.557).

The (hemi)continuity of $\mathcal{A}_\varepsilon$ is an immediate consequence of (2.4), which implies that $\|\mathcal{A}_\varepsilon v_1 - \mathcal{A}_\varepsilon v_2\|_{H^{-1}(\Omega)} \le C(\varepsilon)\|v_1 - v_2\|_{1,\Omega}$. Furthermore, for all $v_1,\, v_2 \in H_0^1(\Omega)$ we have that

$$\langle \mathcal{A}_\varepsilon v_1 - \mathcal{A}_\varepsilon v_2, v_1 - v_2\rangle = |v_1 - v_2|_{1,\Omega}^2 - \int_\Omega c[f_\varepsilon(v_1) - f_\varepsilon(v_2)][v_1 - v_2]\, dx \ge C\|v_1 - v_2\|_{1,\Omega}^2\,,$$

where we have noted Poincaré's inequality and (2.3). Hence, $\mathcal{A}_\varepsilon$ is strongly monotone. Since strong monotonicity implies strict monotonicity and coercivity (cf. [49], p.501), the proof is complete. ∎

Since we assume that $A(\alpha)$ holds, elliptic regularity and Sobolev embedding yield that for all $p \in (1, \infty)$ and $\gamma \in (0, 1)$,

$$u_\varepsilon \in W^{2,p}(\Omega) \cap C^{2,\beta}(\Omega) \cap C^{1,\gamma}(\bar{\Omega})\,.$$

In particular, u_ε is the unique solution of (P_ε) in the classical sense.

To derive further properties of u_ε, we recall the following maximum principles:

The Strong and the Hopf Maximum Principle (cf. Theorem 3.5 and Lemma 3.4 in [17]): Suppose that $w \in C^2(\Omega) \cap C(\bar{\Omega})$ satisfies $Lw \equiv -\Delta w + \sigma(x)w \geq 0$ in Ω and

$$0 \leq \sigma \leq C \quad \text{in } \Omega. \tag{2.5}$$

(SMP) If $w \geq 0$ on $\partial\Omega$, then either $w \equiv 0$ or $w > 0$ in Ω.

(HMP) If $w = 0$ on $\partial\Omega$ and $w \not\equiv 0$ in Ω, then $\frac{\partial w}{\partial n} < 0$ on $\partial\Omega$ (provided $\frac{\partial w}{\partial n}$ exists), where n is the outward unit normal. to $\partial\Omega$.

Lemma 1.2 *Assume $A(\alpha)$ holds. Let u and u_ε be the unique solutions of (P) and (P_ε) respectively, and $\varepsilon_0 > 0$ be arbitrary. Then we have for all $\varepsilon \in (0, \varepsilon_0)$ that*

$$0 < u_{\varepsilon_0} < u_\varepsilon < u \quad \text{in } \Omega. \tag{2.6}$$

Proof. Firstly, for all $\varepsilon > 0$, u_ε satisfies

$$-\Delta u_\varepsilon(x) = c f_\varepsilon(u_\varepsilon(x)) > 0 \quad \forall x \in \Omega.$$

Hence, $u_\varepsilon \not\equiv 0$ in Ω. It follows from (SMP) (with $\sigma = 0$) that $u_\varepsilon > 0$ in Ω. To prove the last inequality in (2.6), let $w_\varepsilon \equiv u - u_\varepsilon$. On subtracting (P_ε) from (P) we obtain that

$$0 = -\Delta w_\varepsilon - c[f(u) - f_\varepsilon(u_\varepsilon)] \leq -\Delta w_\varepsilon - c[f_\varepsilon(u) - f_\varepsilon(u_\varepsilon)] \quad \text{in } \Omega,$$

where we have used that $f(s) \geq f_\varepsilon(s) \quad \forall s > 0$. The Mean Value Theorem and (2.2) yield that

$$-\Delta w_\varepsilon + \sigma_\varepsilon(x) w_\varepsilon \geq 0 \quad \text{in } \Omega.$$

Note that (2.5) holds for σ_ε. Now (SMP) yields that either $w_\varepsilon \equiv 0$ or $w_\varepsilon > 0$ in Ω. On noting that $c > 0$, $w_\varepsilon \equiv 0$ implies that

$$f(u) \equiv f_\varepsilon(u) \quad \text{in } \Omega. \tag{2.7}$$

But, from (1.1)–(1.3) we obtain that

$$u(x) < \varepsilon, \tag{2.8}$$

if $x \in \Omega$ is close to $\partial\Omega$. Since (2.8) contradicts (2.7), we conclude that $u > u_\varepsilon$. Noting this result and that $f_\varepsilon(s) \geq f_{\varepsilon_0}(s)$, $\forall s > 0$, the inequality $u_\varepsilon > u_{\varepsilon_0}$ can be proved similarly. ■

Note that the above proof did not use the upper bound in the second assumption of $A(\alpha)$. The lower bound however, was used on noting that $c > 0$ in Ω. In fact, the next Lemma shows that u_ε can be identically equal to u, if we replace the second assumption in $A(\alpha)$ by

$$c \geq 0 \quad \text{in } \bar{\Omega}.$$

With all the other assumptions of $A(\alpha)$ still in place we denote the altered set of assumptions by A_0.

Lemma 1.3 *Assume A_0 holds. Let u and u_ε be the unique solutions of (P) and (P_ε) respectively. Let $N_\varepsilon := \{x \in \Omega : u(x) < \varepsilon \text{ and } 0 < c(x)\}$. Then*

$$u_\varepsilon \equiv u \quad \text{in } \bar{\Omega} \iff N_\varepsilon = \emptyset.$$

Proof. We first prove "$\Longrightarrow$". Assume that $u_\varepsilon \equiv u$ in Ω. Then, by subtracting (P_ε) from (P) we obtain that

$$0 = c(x)[f(u(x)) - f_\varepsilon(u(x))] \qquad \forall x \in \Omega,$$

from which the result follows as $u(x) < \varepsilon$ implies that $f(u) > f_\varepsilon(u)$. Now, we prove "$\Longleftarrow$". Assume $N_\varepsilon = \emptyset$. As in the proof of Lemma 1.2 we obtain for $w_\varepsilon \equiv u - u_\varepsilon$ that either $w_\varepsilon \equiv 0$ or $w_\varepsilon > 0$ in Ω. On the other hand, by our assumption $N_\varepsilon = \emptyset$ and the definition of f_ε we obtain that

$$-\Delta w_\varepsilon = c[f(u) - f_\varepsilon(u_\varepsilon)] = \begin{cases} 0 & , \text{ if } u < \varepsilon \\ c[f_\varepsilon(u) - f_\varepsilon(u_\varepsilon)], & \text{if } \varepsilon \leq u \end{cases} \leq 0 \tag{2.9}$$

as $u \geq u_\varepsilon$, see (2.6). Now (SMP) yields that $w_\varepsilon \equiv u - u_\varepsilon \leq 0$, and hence $u \equiv u_\varepsilon$. ∎

To illustrate the last result: if $\nu < 1 + \alpha$, it follows from (1.1) that $u \geq \varepsilon_0$ in $\Omega \backslash \Omega_{\frac{\varepsilon_0}{\lambda}}$ for a given $\varepsilon_0 > 0$, where

$$\Omega_\delta := \{x \in \Omega : d(x) \leq \delta\}.$$

Hence we obtain from Lemma 1.3 that $u_\varepsilon \equiv u$ for all $\varepsilon \in (0, \varepsilon_0]$ if $c \equiv 0$ in $\Omega_{\frac{\varepsilon_0}{\lambda}}$.

To resume our main line of argument — which is to prove that u_ε converges towards u as $\varepsilon \to 0$ — we need a further technical result. From (1.1)–(1.3) it follows that u is growing away from the boundary at least linearly. The same is true for u_ε:

Lemma 1.4 *Assume $A(\alpha)$ holds. Then for all $\varepsilon > 0$ there exists a $\lambda_\varepsilon > 0$ such that*

$$\lambda_\varepsilon \, d(x) \leq u_\varepsilon(x), \quad \forall x \in \bar{\Omega}. \tag{2.10}$$

Proof. It suffices to show that there exists a $\delta = \delta(\varepsilon)$ such that

$$\lambda_\delta \, d(x) \leq u_\varepsilon(x) \,, \quad \forall x \in \Omega_\delta \,. \tag{2.11}$$

For, if (2.11) holds, (SMP) yields that

$$\lambda_\delta \, \delta \leq u_\varepsilon \qquad \text{in } \Omega \backslash \Omega_\delta \,, \tag{2.12}$$

and (2.10) follows from (2.11), (2.12) and the boundedness of Ω.

Now, the (HMP) yields that there exists a $\beta > 0$ such that

$$\frac{\partial u_\varepsilon}{\partial n} < -\beta \quad \text{on } \partial\Omega \,.$$

We choose $\bar{\delta}$ such that for all $x \in \Omega_{\bar{\delta}}$ there exists a unique $y(x) \in \partial\Omega$ such that

$$d(x) = |x - y(x)| \,,$$

where the points x and y are related by

$$x = y - d(x) n(y) \tag{2.13}$$

see [17], p.355. Let $\delta \leq \bar{\delta}$ be such that for any $x \in \Omega_\delta$,

$$\frac{\partial u_\varepsilon}{\partial n(y)}(x) \leq -\frac{\beta}{2} \,. \tag{2.14}$$

On noting (2.13) it follows from (2.14) that for any $x \in \Omega_\delta$,

$$u_\varepsilon(x) = \int_0^{d(x)} -\frac{\partial u_\varepsilon}{\partial n(y)}(y - t\, n(y)) \, dt \geq \frac{\beta}{2} d(x) \,.$$

Since x is arbitrary, we obtain (2.11) with $\lambda_\delta := \frac{\beta}{2}$. ∎

Note that analogues of the lower bounds of (1.2) and (1.3) however cannot hold for u_ε, since $u_\varepsilon \in C^{1,\gamma}(\bar{\Omega})$.

Lemma 1.5 *Assume $A(\alpha)$ holds. Let u and u_ε be the unique solutions of (P) and (P_ε) respectively. Then*

$$\lim_{\varepsilon \to 0} \|u - u_\varepsilon\|_{0,\infty,\Omega} = 0 \,.$$

Proof. For $\delta > 0$ we define $\omega_\delta := \Omega \backslash \Omega_\delta$. Clearly, ω_δ is open and for $0 < \delta' < \delta''$ we have that $\omega_{\delta''} \subset\subset \omega_{\delta'} \subset\subset \Omega$. Moreover, if δ_0 is sufficiently small, then $\partial\omega_{\delta_0}$ is C^2. We fix such a δ_0 and define a sequence $\{\Omega_j'\}_{j=1}^\infty$ of domains by $\Omega_j' := \omega_{\delta_0/j}$, $j = 1 \to \infty$. Clearly, we have that

(i) $\quad \partial\Omega_j' \in C^2$, $j = 1 \to \infty$;

(ii) $\quad \Omega_1' \subset\subset \Omega_2' \subset\subset \cdots \subset\subset \Omega$;

(iii) $\quad dist(\Omega_j', \partial\Omega) = \delta_0/j$, $j = 1 \to \infty$;

(iv) $\quad \bigcup_{j=1}^\infty \Omega_j' = \Omega$.

Now, we also fix an $\varepsilon_0 > 0$. From Lemma 1.2 we obtain that for all $\varepsilon \in (0, \varepsilon_0)$ and any subdomain $\omega \subset\subset \Omega$,

$$0 < u_{\varepsilon_0} < u_\varepsilon < u \quad \text{in } \bar{\omega}\,.$$

Hence, we have that $\{u_\varepsilon\}_{0<\varepsilon<\varepsilon_0}$ is uniformly bounded on $\bar{\omega}$. Since $f_\varepsilon(u_\varepsilon) \leq f(u_\varepsilon) \leq f(u_{\varepsilon_0})$, $\forall \varepsilon \in (0, \varepsilon_0)$, we also have that $\{f_\varepsilon(u_\varepsilon)\}_{0<\varepsilon<\varepsilon_0}$ is uniformly bounded on $\bar{\omega}$. Hence, it follows from standard interior estimates, see e.g. (4.45) in [17], that for $j = 1 \to \infty$,

$$\|u_\varepsilon\|_{C^{1,\gamma}(\bar{\Omega}'_j)} \leq C_{\gamma,j}, \quad \forall \gamma \in (0,1)\,. \tag{2.15}$$

Now, we fix γ', $\gamma \in (0,1)$ with $\gamma' < \gamma$ and define $u_k := u_{\varepsilon_0/k}$, $k = 1 \to \infty$. From (2.15) and the compactness of the embedding $C^{1,\gamma}(\bar{\Omega}'_1) \to C^{1,\gamma'}(\bar{\Omega}'_1)$, there exists a subsequence $\{u_{l_k^1}\}_{k=1}^\infty$ of the sequence $\{u_k\}_{k=1}^\infty$ which converges to a function $u^{(1)}$ in $C^{1,\gamma'}(\bar{\Omega}'_1)$ (cf. [1], p.241). Similarly, we obtain a subsequence $\{u_{l_k^2}\}_{k=1}^\infty$ of the sequence $\{u_{l_k^1}\}_{k=1}^\infty$ which converges to a function $u^{(2)}$ in $C^{1,\gamma'}(\bar{\Omega}'_2)$. By an induction procedure we obtain for each $j = 1, 2, \ldots$ a sequence $\{u_{l_k^j}\}_{k=1}^\infty$ which is convergent in $C^{1,\gamma'}(\bar{\Omega}'_j)$ to a function $u^{(j)}$ and has the property that $\{u_{l_k^{j+1}}\}_{k=1}^\infty$ is a subsequence of $\{u_{l_k^j}\}_{k=1}^\infty$ with $u^{(j+1)}|_{\bar{\Omega}'_j} = u^{(j)}|_{\bar{\Omega}'_j}$. Since $\{l_k^{j+1}\}_k$ is a subsequence of $\{l_k^j\}_k$ for all j, we have that

$$l_{j+1}^{j+1} \geq l_j^j\,, \quad l_k^j \geq k \qquad \forall j, k\,. \tag{2.16}$$

We set $u'_j = u_{l_j^j}$ for $j = 1, 2, \ldots$. Given any domain $\Omega' \subset\subset \Omega$, there exists a j_0 such that $\Omega' \subset\subset \Omega'_{j_0}$. Hence, $\{u'_j\}$ converges in $C^{1,\gamma'}(\bar{\Omega}')$ for any $\Omega' \subset\subset \Omega$, in particular there exists a $u^* \in C^{1,\gamma'}(\Omega)$ such that

$$u^*|_{\bar{\Omega}'_j} = u^{(j)}|_{\bar{\Omega}'_j}, \quad j = 1 \to \infty\,. \tag{2.17}$$

By 'construction' of u^*, Lemma 1.2 and (2.16) we have that $\forall j = 1 \to \infty$,

$$0 < u_{\varepsilon_0} \leq u'_j \leq u'_{j+1} \leq u^* \leq u \quad \text{in } \Omega\,. \tag{2.18}$$

Thus, if $x_1 \in \partial\Omega$,

$$\lim_{x \to x_1} u^*(x) = 0\,. \tag{2.19}$$

Now, we show that u^* solves (P). Firstly we show that for all $j = 2 \to \infty$,

$$-\Delta u^{(j)} = cf(u^{(j)}) \quad \text{in } \bar{\Omega}'_{j-1}\,. \tag{2.20}$$

For arbitrary but fixed $j \geq 2$ we define correspondingly to $u_{l_k^j}$ the functions $f_{l_k^j} := f_{\varepsilon_0/l_k^j}$ for $k = 1 \to \infty$. By definition we have that

$$\int_{\Omega'_j} \nabla u_{l_k^j} \cdot \nabla \phi \, dx = \int_{\Omega'_j} c f_{l_k^j}(u_{l_k^j}) \phi \, dx\,, \quad \forall \phi \in C_0^\infty(\Omega'_j)\,. \tag{2.21}$$

From $u_{l_k^j} \to u^{(j)}$ in $C^{1,\gamma'}(\bar{\Omega}'_j)$ as $k \to \infty$ we obtain that

$$\lim_{k\to\infty} \|u_{l_k^j} - u^{(j)}\|_{0,\infty,\Omega'_j} = 0\,, \tag{2.22}$$

$$\lim_{k\to\infty} \int_{\Omega'_j} \nabla u_{l_k^j} \cdot \nabla \phi \, dx = \int_{\Omega'_j} \nabla u^{(j)} \cdot \nabla \phi \, dx\,. \tag{2.23}$$

The Mean Value Theorem, Lemma 1.2, $|f'(s)| \leq |f'(s')|$, $\forall s' \leq s$, Lemma 1.4 and (iii) yield that

$$\begin{aligned} &\|f(u^{(j)}) - f_{l_k^j}(u_{l_k^j})\|_{0,\infty,\Omega'_j} \\ &\quad \leq \|f(u^{(j)}) - f(u_{l_k^j})\|_{0,\infty,\Omega'_j} + \|f(u_{l_k^j}) - f_{l_k^j}(u_{l_k^j})\|_{0,\infty,\Omega'_j} \\ &\quad \leq |f'(\lambda_{\varepsilon_0}\delta_0/j)|\,\|u^{(j)} - u_{l_k^j}\|_{0,\infty,\Omega'_j} + \|f(u_{l_k^j}) - f_{l_k^j}(u_{l_k^j})\|_{0,\infty,\Omega'_j}\,. \end{aligned} \tag{2.24}$$

Lemma 1.4, (iii) and (2.16) yield that for all $k \geq \frac{\varepsilon_0 j}{\lambda_{\varepsilon_0}\delta_0}$,

$$f(u_{l_k^j}) = f_{l_k^j}(u_{l_k^j}) \quad \text{on } \Omega'_j\,. \tag{2.25}$$

Now, we obtain from (2.24), (2.22) and (2.25) that $\lim_{k\to\infty} \|f(u^{(j)}) - f_{l_k^j}(u_{l_k^j})\|_{0,\infty,\Omega'_j} = 0$ and hence, that

$$\lim_{k\to\infty} \int_{\Omega'_j} c f_{l_k^j}(u_{l_k^j}) \phi \, dx = \int_{\Omega'_j} c f(u^{(j)}) \phi \, dx\,, \quad \forall \phi \in C_0^\infty(\Omega'_j)\,. \tag{2.26}$$

It follows from (2.21), (2.23) and (2.26) that

$$\int_{\Omega'_j} \nabla u^{(j)} \cdot \nabla \phi \, dx = \int_{\Omega'_j} c f(u^{(j)}) \phi \, dx\,, \quad \forall \phi \in C_0^\infty(\Omega'_j)\,.$$

From Theorem 8.8 and Theorem 9.19 in [17] we obtain that $u^{(j)} \in C^{2,\beta}(\bar{\Omega}'_{j-1})$ and hence that (2.20) holds.

Now, it easily follows from (ii), (iv), (2.20), (2.18), (2.19) and (2.17) that $u^* \in C^{2,\beta}(\Omega) \cap C(\bar{\Omega})$ and u^* has the properties $u^* > 0$ in Ω, $u^* = 0$ on $\partial\Omega$ and u^* satisfies (P). By uniqueness, $u^* \equiv u$.

To prove the desired convergence of u_ε towards u, we first consider the convergence of u'_j. It follows from (2.18) that for all $j \geq j_0$

$$\|u'_j - u\|_{0,\infty,\Omega} \leq \|u'_j - u\|_{0,\infty,\Omega'_{j_0}} + \|u'_{j_0} - u\|_{0,\infty,\Omega\setminus\Omega'_{j_0}}.$$

Now, let $\tau > 0$ be arbitrary but fixed. From (2.18), (1.1)–(1.3) and (iii) we have that there exists a J such that $\|u'_{j_0} - u\|_{0,\infty,\Omega\setminus\Omega'_{j_0}} \leq \|u\|_{0,\infty,\Omega\setminus\Omega'_{j_0}} \leq \tau/2$ for $j_0 \geq J$. Since u'_j converges uniformly on Ω'_{j_0}, we have that there exists a $J_1 \geq j_0$ such that $\|u'_j - u\|_{0,\infty,\Omega'_{j_0}} \leq \tau/2$ for $j \geq J_1$. Hence, we have for all $j \geq J_1$ that

$$\|u'_j - u\|_{0,\infty,\Omega} \leq \tau \tag{2.27}$$

and hence uniform convergence of u'_j towards u. The desired uniform convergence of u_ε follows easily from (2.27) and Lemma 2.3. ∎

The proof above is basically the existence proof on p.346 in [13], see also p.189 in [42]. We have filled in the details. Since we assume that $c \in C^\beta(\bar{\Omega})$, we could use a regularity argument to show that u is a classical solution of (P).

1.2.2 The Regularization Error

As was seen in Lemma 1.5, u_ε, the solution of (P_ε), converges to u, the solution of (P), in $L^\infty(\Omega)$. We will now prove bounds for $\|u - u_\varepsilon\|_{0,\infty,\Omega}$ and $|u - u_\varepsilon|_{1,\Omega}$.

Firstly we want to determine for which values of ν (depending on α) u, the solution of (WP), is in $H_0^1(\Omega)$. The following example shows that even in the case when (P) has a classical solution, this solution need not be in $H_0^1(\Omega)$. Consider the problem

$$-u'' = 4u^{-3} \qquad \text{in } \Omega \equiv (0,1) \tag{2.28}$$

with $u(0) = u(1) = 0$. On multiplying (2.28) by u' and using the method of separation of variables we obtain the classical solution of (2.28) as $u(x) = 2x^{\frac{1}{2}}(1-x)^{\frac{1}{2}}$, $x \in \Omega$. We have that $[u'(x)]^2 = -4 + x^{-1}(1-x)^{-1}$, $x \in \Omega$. Since $\int_\delta^{\frac{1}{2}} x^{-1}(1-x)^{-1}\,dx \geq \int_\delta^{\frac{1}{2}} x^{-1}\,dx$ for all $\delta \in (0, \frac{1}{2}]$, u cannot be in $H_0^1(\Omega)$.

Obviously, it follows from (1.4) that u, the solution of (WP), is in $H_0^1(\Omega)$, if $\nu < 1+\alpha$. But this result is sub-optimal as must be expected. To prove an optimal result, we need to know the boundary behaviour of $|\nabla u|$. This can be obtained from combining (1.2) and (1.3) with the next result, which is a generalization of a Harnack inequality.

Lemma 1.6 *There exists a constant $K_1 > 0$ such that if $r \in (0,1], B_{2r}(x_0) = \{x \in \mathbb{R} : |x - x_0| < 2r\} \subset \Omega$ and $v \in W^{2,q}(B_{2r}(x_0))$ where $q > d$, then*

$$(H) \qquad |\nabla v(x)| \leq K_1 \left(r\,\|\Delta v\|_{L^\infty(B_{2r}(x_0))} + \frac{1}{r}\,\|v\|_{L^\infty(B_{2r}(x_0))} \right)$$

for $x \in B_r(x_0)$ ($\|\Delta v\|_{L^\infty(B_{2r}(x_0))} = \infty$ is allowed).

Proof. See Lemma 2.19 in [12], where the result is proved for any second order uniformly elliptic differential operator with continuous coefficients. ∎

Lemma 1.7 *Assume that $A(\alpha)$ holds and let u be the unique solution of (WP). Then we have that*

$$u \in H_0^1(\Omega) \iff \nu < 3 + 2\alpha . \tag{2.29}$$

Proof. We prove "$\Longleftarrow$". We first note that by (1.4) we only have to prove the assertion for $\nu \geq 1 + \alpha$. We proceed as in the proof of Theorem 2.5 in [12]. For $x \in \Omega$, we set $r = d(x)/3$. Note that if $z \in B_{2r}(x)$, then $d(x)/3 \leq d(z) \leq 5d(x)/3$. Hence, it follows from (H) with $v = u$ (so $-\Delta u = cf(u)$) and $x_0 = x$ that

$$|\nabla u(x)| \leq K_1 \left(d(x) \, \|cf(u)\|_{L^\infty(A_x)} + \frac{1}{d(x)} \, \|u\|_{L^\infty(A_x)} \right) \tag{2.30}$$

for all $x \in \Omega$, where $A_x := \{z \in \Omega : d(x)/3 \leq d(z) \leq 5d(x)/3\}$.

If $\nu = 1 + \alpha$, we obtain from combining (1.2), $(A2)$ and (2.30) that

$$|\nabla u(x)| \leq K \, [D - \ln d(x)]^{\frac{1}{1+\nu}} , \quad \forall x \in \Omega .$$

Hence $u \in H_0^1(\Omega)$.

Similarly, if $1 + \alpha < \nu$, (1.3), $(A2)$ and (2.30) yield that

$$|\nabla u(x)| \leq K [d(x)]^{\frac{1+\alpha-\nu}{1+\nu}} , \quad \forall x \in \Omega .$$

Hence, $u \in H_0^1(\Omega)$, if $2 \left[\frac{1+\alpha-\nu}{1+\nu} \right] > -1$, i.e. if $\nu < 3 + 2\alpha$, and the proof of "$\Longleftarrow$" is complete.

For the proof of "$\Longrightarrow$" one can easily adapt the proof given on p.728 in [27]. ∎

We remark that this result is a generalization of Theorem 2 in [27] where in our notation $\alpha = 0$. However, we believe that the proof of sufficiency given in [27] is incorrect. In [12] estimates for the solution of (P) when c is constant are given in terms of a solution of the following one-dimensional problem:

Let $a > 0$ be given. Find $p \in C^2((0, a]) \cap C([0, a])$ such that

$$-p''(s) = cf(p(s)), \quad p(s) > 0 \quad \text{for } s \in (0, a], \quad p(0) = 0 . \tag{2.31}$$

When $\nu > 1$, exact solutions p for (2.31) can be found using the ansatz $p = Kx^k$, where k, K are positive constants. We find that the estimate given in Theorem 2.2 of [12] is exactly that of (1.3) when $\alpha = 0$.

The next result is crucial for our H^1 error bounds.

Lemma 1.8 *Assume $A(\alpha)$ holds. Let u be the unique solution to (WP). Then there exists $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0]$,*

$$\int_{0<u<\varepsilon} cf(u)u^\mu \, dx \leq C_* \begin{cases} \varepsilon^{1+\alpha+\mu-\nu}, & \text{if } 0 < \nu < 1+\alpha \text{ and } \nu - (1+\alpha) < \mu, \\ \varepsilon^{\mu}, & \text{if } \nu = 1+\alpha \text{ and } 0 < \mu, \\ \varepsilon^{\delta}, & \text{if } 1+\alpha < \nu, \text{ and } \mu = \frac{\nu-(1+\alpha)}{2+\alpha} + \delta \\ & \qquad \text{with } \delta > 0, \end{cases}$$

where $C_* = C_*(M, \alpha, \mu, \nu, \lambda, \Lambda, meas\, \partial\Omega)$.

Proof. Let $\bar{\delta} > 0$ be as in the proof of Lemma 1.4. If $\nu < 1+\alpha$, (1.1) holds. We choose $\varepsilon_0 < \lambda\bar{\delta}$ and obtain for all $\mu > \nu - (1+\alpha)$ that

$$\int_{0<u<\varepsilon} cf(u)u^\mu \, dx \leq \int_{\Omega_{\frac{\varepsilon}{\lambda}}} cf(u)u^\mu \, dx \leq C \int_{\partial\Omega} \int_0^{\frac{\varepsilon}{\lambda}} \eta^{\alpha+\mu-\nu} \, d\eta ds \leq C meas\, \partial\Omega \, \varepsilon^{1+\alpha+\mu-\nu}.$$

If $\nu = 1+\alpha$, (1.2) holds. Its lower bound yields that $u(x) < \varepsilon$ implies that $d(x) < \frac{\varepsilon}{\lambda}[D - \ln d(x)]^{-\frac{1}{2+\alpha}} \leq \bar{C}\frac{\varepsilon}{\lambda}$, where $\bar{C}$ is a positive constant. Hence, on choosing $\varepsilon_0 < \bar{C}^{-1}\lambda\bar{\delta}$ we obtain that

$$\int_{0<u<\varepsilon} cf(u)u^\mu \, dx \leq \int_{\Omega_{\bar{C}\frac{\varepsilon}{\lambda}}} cf(u)u^\mu \, dx \leq C \int_{\partial\Omega} \int_0^{\bar{C}\frac{\varepsilon}{\lambda}} \eta^{\mu-1} [D - \ln \eta]^{\frac{\mu-\nu}{1+\nu}} \, d\eta ds.$$

If $\mu \leq \nu$, the term $[D - \ln \eta]^{\frac{\mu-\nu}{1+\nu}}$ is bounded independently of η. Hence, we obtain the desired result for $\mu \leq \nu$. If $\mu > \nu$ then

$$\begin{aligned} \int_{0<u<\varepsilon} cf(u)u^\mu \, dx &\leq \varepsilon^{\mu-\nu} \int_{0<u<\varepsilon} cf(u)u^\nu \, dx = \varepsilon^{\mu-\nu} \int_{0<u<\varepsilon} c(x) \, dx \\ &\leq C\varepsilon^{\mu-\nu} \int_0^{\bar{C}\frac{\varepsilon}{\lambda}} \eta^\alpha \, d\eta = C_*\varepsilon^{\mu-\nu+1+\alpha} = C_*\varepsilon^\mu, \end{aligned}$$

and hence the desired result for $\mu > \nu$.

Finally, if $1+\alpha < \nu$, (1.3) holds. We choose $\varepsilon_0 < \lambda\bar{\delta}^{(2+\alpha)/(1+\nu)}$ and obtain that

$$\int_{0<u<\varepsilon} cf(u)u^\mu \, dx \leq \int_{\Omega_{(\frac{\varepsilon}{\lambda})^{\frac{1+\nu}{2+\alpha}}}} cf(u)u^\mu \, dx \leq C \, meas\, \partial\Omega \int_0^{(\frac{\varepsilon}{\lambda})^{\frac{1+\nu}{2+\alpha}}} \eta^{\alpha+\frac{2+\alpha}{1+\nu}(\mu-\nu)} \, d\eta.$$

For the existence of the last integral we require $\alpha + \frac{2+\alpha}{1+\nu}(\mu - \nu) > -1$. On choosing $\mu = \frac{\nu-(1+\alpha)}{2+\alpha} + \delta$ for a $\delta > 0$ we obtain that

$$\alpha + \frac{2+\alpha}{1+\nu}(\mu - \nu) = -1 + \frac{2+\alpha}{1+\nu}\delta$$

and hence the desired result. ∎

We can now prove $H^1(\Omega)$ and $L^\infty(\Omega)$ bounds for the regularization error.

Lemma 1.9 *Assume $A(\alpha)$ holds. Then there exists an $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0]$,*

$$(i) \qquad |u - u_\varepsilon|_{1,\Omega} \leq C \begin{cases} \varepsilon^{\frac{2+\alpha-\nu}{2}}, & \text{if} \quad 0 < \nu \leq 1+\alpha, \\ \varepsilon^{\frac{3+2\alpha-\nu}{4+2\alpha}}, & \text{if } 1+\alpha < \nu < 3+2\alpha, \end{cases}$$

$$(ii) \qquad \|u - u_\varepsilon\|_{0,\infty,\Omega} \leq C\varepsilon \quad \text{for all } \nu > 0.$$

Proof. (i) To begin with we note that by Lemma 1.7 an H^1 error bound can only be proved for $\nu \in (0, 3+2\alpha)$. Subtracting (WP_ε) from (WP) and choosing $\phi = u - u_\varepsilon$ yields $|u - u_\varepsilon|^2_{1,\Omega} = \int_\Omega c[f(u) - f_\varepsilon(u_\varepsilon)][u - u_\varepsilon]\,dx$ which we rewrite as

$$|u - u_\varepsilon|^2_{1,\Omega} - \int_\Omega c[f_\varepsilon(u) - f_\varepsilon(u_\varepsilon)][u - u_\varepsilon]\,dx = \int_\Omega c[f(u) - f_\varepsilon(u)][u - u_\varepsilon]\,dx\,.$$

Noting (2.3) we obtain that

$$|u - u_\varepsilon|^2_{1,\Omega} \leq \int_\Omega c[f(u) - f_\varepsilon(u)][u - u_\varepsilon]\,dx\,, \tag{2.32}$$

and on noting Lemma 1.2 we have for all $\mu \in [0,1]$ that

$$\begin{aligned} &\int_\Omega c[f(u) - f_\varepsilon(u)][u - u_\varepsilon]\,dx \\ &\quad = \int_{0<u_\varepsilon<u<\varepsilon} c[f(u) - f_\varepsilon(u)][u - u_\varepsilon]\,dx \leq \int_{0<u_\varepsilon<u<\varepsilon} cf(u)[u - u_\varepsilon]\,dx \\ &\quad \leq \int_{0<u_\varepsilon<u<\varepsilon} cf(u)u^\mu[u - u_\varepsilon]^{1-\mu}\,dx \leq \left[\int_{0<u<\varepsilon} cf(u)u^\mu\,dx\right]\varepsilon^{1-\mu}\,. \end{aligned} \tag{2.33}$$

(i) now follows from (2.32), (2.33) and Lemma 1.8.

(ii). Firstly we prove that for all $\delta \in (0, \varepsilon)$,

$$\|u_\delta - u_\varepsilon\|_{0,\infty,\Omega} \leq C\varepsilon\,, \tag{2.34}$$

where C can be chosen independently of δ and ε.

We follow the argument of Nochetto [32], Theorem 2.1. By subtracting (WP_ε) from (WP_δ) and choosing $\phi = e^{2p+1}$, where $e \equiv u_\delta - u_\varepsilon$ and $p \in \mathbb{R}_+$, we obtain

$$\int_\Omega \nabla e \cdot \nabla e^{2p+1}\,dx = \int_\Omega c[f_\delta(u_\delta) - f_\varepsilon(u_\varepsilon)]e^{2p+1}\,dx\,. \tag{2.35}$$

Using the chain rule and Friedrich's inequality on the left hand side of the equation yields

$$\begin{aligned} \int_\Omega \nabla e \cdot \nabla e^{2p+1}\,dx &= (2p+1)\int_\Omega (\nabla e)\cdot(e^{2p}\nabla e)\,dx = (2p+1)\int_\Omega |e^p(\nabla e)|^2\,dx \\ &= \frac{2p+1}{(p+1)^2}\int_\Omega |\nabla e^{p+1}|^2\,dx \geq C\frac{2p+1}{(p+1)^2}\|e\|^{2(p+1)}_{0,2(p+1),\Omega}\,. \end{aligned} \tag{2.36}$$

To find an upper bound for the right hand side of (2.35) we consider two different cases. If $u_\delta \geq \varepsilon$ then $f_\delta(u_\delta) = f_\varepsilon(u_\delta)$ and it follows from (2.3) that

$$[f_\delta(u_\delta) - f_\varepsilon(u_\varepsilon)]e^{2p+1} = [f_\varepsilon(u_\delta) - f_\varepsilon(u_\varepsilon)][u_\delta - u_\varepsilon]e^{2p} \leq 0\,.$$

Recalling from Lemma 1.2 that $0 < u_\varepsilon < u_\delta$ in Ω we have only to consider the case when $0 < u_\varepsilon < u_\delta < \varepsilon$. Then we have, on noting (2.1), that

$$\int_{0<u_\varepsilon<u_\delta<\varepsilon} c[f_\delta(u_\delta) - f_\varepsilon(u_\varepsilon)]e^{2p+1}\,dx \le \int_{0<u_\delta<\varepsilon} cf(u_\delta)e^{2p+1}\,dx$$
$$= \int_{0<u_\delta<\varepsilon} cf(u_\delta)|u_\delta - u_\varepsilon|^\nu e^{2p+1-\nu}\,dx \le C\int_{0<u_\delta<\varepsilon} u_\delta^{-\nu}u_\delta^\nu e^{2p+1-\nu}\,dx \le C\varepsilon^{2p+1-\nu}, \quad (2.37)$$

for all $p \ge (\nu-1)/2$. Thus, it follows from (2.35), (2.36) and (2.37) that for all $p \ge (\nu-1)/2$,

$$\|e\|_{0,2(p+1),\Omega} \le C\left[\frac{(p+1)^2}{2p+1}\right]^{\frac{1}{2(p+1)}} \varepsilon^{\frac{2p+1-\nu}{2(p+1)}}.$$

Letting $p \to \infty$ yields (2.34). Since we have that $\|u - u_\varepsilon\|_{0,\infty,\Omega} \le \|u - u_\delta\|_{0,\infty,\Omega} + \|u_\delta - u_\varepsilon\|_{0,\infty,\Omega}$, (ii) follows from (2.34) and Lemma 1.5. ∎

The following bounds for $\|cf_\varepsilon(u_\varepsilon)\|_{0,q,\Omega}$, $q = 1$, 2 or ∞ will be useful later.

Lemma 1.10 *Assume $A(\alpha)$ holds. Then there exists $\varepsilon_1 > 0$ such that for all $\varepsilon \in (0, \varepsilon_1]$,*

$$\forall q \in [1,\infty], \quad \|cf_\varepsilon(u_\varepsilon)\|_{0,q,\Omega} \le C(\alpha,\nu)\begin{cases} 1 & , \text{ if } 0 < \nu < \frac{1}{q} + \alpha, \\ (\ln\frac{1}{\varepsilon})^{\frac{1}{q}} & , \text{ if } \nu = \frac{1}{q} + \alpha, \\ \varepsilon^{\frac{1}{q}+\alpha-\nu} & , \text{ if } \frac{1}{q} + \alpha < \nu. \end{cases}$$

Proof. For ease of exposition we introduce the following constants:

$$C_1 \equiv \operatorname{meas}\partial\Omega, \quad C_2 \equiv \|c\|_{0,\infty,\Omega}, \quad C_3 \equiv \operatorname{meas}\Omega.$$

For $q \in [1,\infty)$ we have that $\int_\Omega [cf_\varepsilon(u_\varepsilon)]^q\,dx =$

$$= \int_{\Omega_{\frac{\varepsilon}{\lambda_{\varepsilon_0}}}} [cf_\varepsilon(u_\varepsilon)]^q\,dx + \int_{\Omega_{\frac{\varepsilon_1}{\lambda_{\varepsilon_0}}}\setminus\Omega_{\frac{\varepsilon}{\lambda_{\varepsilon_0}}}} [cf_\varepsilon(u_\varepsilon)]^q\,dx + \int_{\Omega\setminus\Omega_{\frac{\varepsilon_1}{\lambda_{\varepsilon_0}}}} [cf_\varepsilon(u_\varepsilon)]^q\,dx, \quad (2.38)$$

where we choose ε_1 as follows. Let $\bar\delta$ be as in the proof of Lemma 1.4. For arbitrary, but fixed $\varepsilon_0 > 0$ we can choose $\varepsilon_1 \in (0, \varepsilon_0]$ such that $\varepsilon_1 < \lambda_{\varepsilon_0}\bar\delta$ and as $f_\varepsilon(s) \le (1+\nu)\varepsilon^{-\nu}$, $\forall s \ge 0$, we obtain for any $\varepsilon \in (0, \varepsilon_1]$ that

$$\int_{\Omega_{\frac{\varepsilon}{\lambda_{\varepsilon_0}}}} [cf_\varepsilon(u_\varepsilon)]^q\,dx \le \left[M(1+\nu)\varepsilon^{-\nu}\right]^q \int_{\partial\Omega}\int_0^{\frac{\varepsilon}{\lambda_{\varepsilon_0}}} \eta^{q\alpha}\,d\eta ds$$
$$\le [M(1+\nu)]^q \frac{\lambda_{\varepsilon_0}^{-(1+q\alpha)}}{1+q\alpha} C_1\,\varepsilon^{1+q(\alpha-\nu)}. \quad (2.39)$$

As $\lambda_{\varepsilon_0} d(x) \le u_\varepsilon(x)$ for all $\varepsilon \in (0, \varepsilon_1]$ and all $x \in \bar\Omega$ we have that

$$\int_{\Omega_{\frac{\varepsilon_1}{\lambda_{\varepsilon_0}}}\setminus\Omega_{\frac{\varepsilon}{\lambda_{\varepsilon_0}}}} [cf_\varepsilon(u_\varepsilon)]^q\,dx \le \left[M\lambda_{\varepsilon_0}^{-\nu}\right]^q C_1 \int_{\frac{\varepsilon}{\lambda_{\varepsilon_0}}}^{\frac{\varepsilon_1}{\lambda_{\varepsilon_0}}} \eta^{q(\alpha-\nu)}\,d\eta$$
$$= M^q\lambda_{\varepsilon_0}^{-q\nu}C_1 \begin{cases} [1+q(\alpha-\nu)]^{-1}\left[(\frac{\varepsilon_1}{\lambda_{\varepsilon_0}})^{1+q(\alpha-\nu)} - (\frac{\varepsilon}{\lambda_{\varepsilon_0}})^{1+q(\alpha-\nu)}\right], & \text{if } \nu \ne \frac{1}{q}+\alpha, \\ \ln(\frac{\varepsilon_1}{\lambda_{\varepsilon_0}}) - \ln(\frac{\varepsilon}{\lambda_{\varepsilon_0}}) & , \text{ if } \nu = \frac{1}{q}+\alpha. \end{cases} \quad (2.40)$$

Finally, we have that

$$\int_{\Omega \setminus \Omega_{\frac{\varepsilon_1}{\lambda_{\varepsilon_0}}}} [cf_\varepsilon(u_\varepsilon)]^q \, dx \leq C_3 C_2^q \varepsilon_1^{-q\nu} . \tag{2.41}$$

Now, (2.38) – (2.41) yield that $\int_\Omega [cf_\varepsilon(u_\varepsilon)]^q \, dx \leq$

$$\leq \begin{cases} C_3 C_2^q \varepsilon_1^{-q\nu} + M^q \frac{\lambda_{\varepsilon_0}^{-(1+q\alpha)}}{1+q(\alpha-\nu)} C_1 \varepsilon_1^{1+q(\alpha-\nu)} + M^q C_1 \lambda_{\varepsilon_0}^{-(1+q\alpha)} \left(\frac{(1+\nu)^q}{1+q\alpha} - \frac{1}{1+q(\alpha-\nu)} \right) \varepsilon^{1+q(\alpha-\nu)} , \\ \quad \text{if } \nu \neq \frac{1}{q} + \alpha , \\ C_3 C_2^q \varepsilon_1^{-q\nu} + [M(1+\frac{1}{q}+\alpha)]^q \frac{\lambda_{\varepsilon_0}^{-(1+q\alpha)}}{1+q\alpha} C_1 + M^q \lambda_{\varepsilon_0}^{-q\nu} C_1 \left[\ln(\frac{\varepsilon_1}{\lambda_{\varepsilon_0}}) - \ln(\frac{\varepsilon}{\lambda_{\varepsilon_0}}) \right] , \\ \quad \text{if } \nu = \frac{1}{q} + \alpha . \end{cases}$$

On noting the inequality $a^q + b^q \leq (a+b)^q$ which is valid for all $a, b \geq 0$ and all $q \geq 1$ we obtain the desired result for $q \in [1, \infty)$. The result for $q = \infty$ follows from noting that $\|cf_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega} = \lim_{q\to\infty} \|cf_\varepsilon(u_\varepsilon)\|_{0,q,\Omega}$. ∎

1.3 An Alternative Regularization

In this section we will consider an alternative regularization of (P), i.e. f will be replaced by a non-singular function g_ε different from f_ε, see e.g. [12], [13], and [27]. In order to keep our notation simple, the notations $(P_\varepsilon), u_\varepsilon$ etc. will be kept but will correspond to the new regularization. This does not cause confusion in this section, and in later sections, we will always state which regularization we are referring to.

1.3.1 The Regularized Problem

We consider the problem:

Find u_ε such that $u_\varepsilon > -\varepsilon$ in Ω, $u_\varepsilon = 0$ on $\partial\Omega$ and

$$(P_\varepsilon) \qquad -\Delta u_\varepsilon = c(x) g_\varepsilon(u_\varepsilon) \quad \text{in } \Omega ,$$

where for $\varepsilon > 0$, $g_\varepsilon(s) := f(s+\varepsilon)$, $\forall s \in (-\varepsilon, \infty)$. Clearly, $g_\varepsilon \in C^1((-\varepsilon, \infty), \mathbb{R})$ and for all $\varepsilon \in (0, \varepsilon_0)$ we have that

$$g_{\varepsilon_0}(s) < g_\varepsilon(s) < f(s) \qquad \forall s \in \mathbb{R}_+ , \tag{3.1}$$

$$-\nu\varepsilon^{-(\nu+1)} = g'_\varepsilon(0) < g'_\varepsilon(s) < 0 \qquad \forall s \in \mathbb{R}_+ . \tag{3.2}$$

From (3.2) it follows that $\forall a, b \in \mathbb{R}_+$,

$$-\frac{\nu}{\varepsilon^{\nu+1}} [a-b]^2 \leq [g_\varepsilon(a) - g_\varepsilon(b)][a-b] \leq -\frac{\varepsilon^{\nu+1}}{\nu} [g_\varepsilon(a) - g_\varepsilon(b)]^2 \leq 0 . \tag{3.3}$$

Definition 1.2 *u_ε is a weak solution of (P_ε), if $u_\varepsilon \in H_0^1(\Omega)$, $u_\varepsilon > -\varepsilon$ in Ω and*

$$(WP_\varepsilon) \qquad \int_\Omega \nabla u_\varepsilon \cdot \nabla \phi \, dx = \int_\Omega c g_\varepsilon(u_\varepsilon)\phi \, dx \qquad \forall \phi \in H_0^1(\Omega) .$$

Lemma 1.11 *For all $\varepsilon > 0$ there exists a unique solution u_ε to (WP_ε).*

Proof. Since g_ε is monotone, there can be at most one solution to (WP_ε). To prove existence, we set

$$\tilde{g}_\varepsilon(s) := \begin{cases} g_\varepsilon(0) + g_\varepsilon'(0)s , & s \in (-\infty, 0) , \\ g_\varepsilon(s) \qquad\quad , & s \in [0, \infty) . \end{cases}$$

Since $\tilde{g}_\varepsilon$ is monotone and has a bounded derivative, one can show as in section 2 that there exists a (unique) solution $\tilde{u}_\varepsilon \in H_0^1(\Omega)$ to $(W\tilde{P}_\varepsilon)$, where $(W\tilde{P}_\varepsilon)$ denotes (WP_ε) with g_ε replaced by $\tilde{g}_\varepsilon$. Moreover, we obtain that for all $p \in (1, \infty)$ and $\gamma \in (0, 1)$,

$$\tilde{u}_\varepsilon \in W^{2,p}(\Omega) \cap C^{2,\beta}(\Omega) \cap C^{1,\gamma}(\bar{\Omega}) . \tag{3.4}$$

As in the proof of Lemma 1.2 (SMP) yields that $\tilde{u}_\varepsilon > 0$ in Ω. Since $\tilde{g}_\varepsilon(s) = g_\varepsilon(s)$, $\forall s > 0$, $\tilde{u}_\varepsilon$ is the unique solution of (WP_ε), and by (3.4) also the unique solution of (P_ε). ∎

Lemma 1.12 *Assume $A(\alpha)$ holds. Let u and u_ε be the unique solutions of (P) and (P_ε) respectively, and $\varepsilon_0 > 0$ be arbitrary. Then we have for all $\varepsilon \in (0, \varepsilon_0)$ that*

$$0 < u_{\varepsilon_0} < u_\varepsilon < u \leq u_\varepsilon + \varepsilon \quad \text{in } \Omega .$$

Proof. The first three inequalities can be proved as in Lemma 1.2 with f_ε replaced by g_ε. To prove the last inequality, we set $w_\varepsilon = u_\varepsilon + \varepsilon - u$. On subtracting (P) from (P_ε) and on noting the definition of g_ε and the Mean Value Theorem we obtain that

$$0 = -\Delta(u_\varepsilon - u) - c[g_\varepsilon(u_\varepsilon) - f(u)] = -\Delta(u_\varepsilon + \varepsilon - u) - c[f(u_\varepsilon + \varepsilon) - f(u)] \leq -\Delta w_\varepsilon + \sigma w_\varepsilon .$$

Since $\sigma \geq 0$ in Ω and $w_\varepsilon \geq 0$ on $\partial\Omega$, it follows from the weak maximum principle (cf. Corollary 3.2 in [17]) that $w_\varepsilon \geq 0$ and the proof is complete. ∎

Lemma 1.12 is different from Lemma 1.2 in that now u is "sandwiched" between u_ε and $u_\varepsilon + \varepsilon$. As an immediate consequence we obtain that

$$\|u - u_\varepsilon\|_{0,\infty,\Omega} \leq \varepsilon \tag{3.5}$$

and hence that

$$\lim_{\varepsilon \to 0} \|u - u_\varepsilon\|_{0,\infty,\Omega} = 0 . \tag{3.6}$$

We remind the reader that in section 2 a lot of work had to be done to prove (3.6) for the regularization f_ε in Lemma 1.5. (3.6) was then used in the proof of Lemma 1.9(ii). Also we note that (3.5) is better than the bound of Lemma 1.9(ii) since $C = 1$ as opposed to an unknown C before.

We remark however that the new regularization which basically consists of shifting the whole graph of f to the left effects the solution globally. There are no longer non-trivial cases in which $u_\varepsilon \equiv u$:

Lemma 1.13 *Assume $A_0(\alpha)$ holds. Let u and u_ε be the unique solutions of (P) and (P_ε) respectively. Then* $u_\varepsilon \equiv u$ *in* $\bar{\Omega} \iff c \equiv 0$.

Proof. "$\Longleftarrow$" obvious. "$\Longrightarrow$" similar to the proof in Lemma 1.3. ∎

We also note that Lemma 1.4 holds for the new regularization.

1.3.2 The Error for the Alternative Regularization

Since the Lemmas 1.7 and 1.8 only involve f and u, they remain valid. We now prove an analogue for Lemma 1.9.

Lemma 1.14 *Assume $A(\alpha)$ holds. Then there exists an $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0]$,*

$$(i) \qquad |u - u_\varepsilon|_{1,\Omega} \leq C \begin{cases} \varepsilon & , \text{ if } \quad 0 < \nu < \alpha, \\ \varepsilon \left(\log \frac{1}{\varepsilon}\right)^{\frac{1}{2}} & , \text{ if } \quad \nu = \alpha, \\ \varepsilon^{\frac{2+\alpha-\nu}{2}} & , \text{ if } \quad \alpha < \nu \leq 1+\alpha, \\ \varepsilon^{\frac{3+2\alpha-\nu}{4+2\alpha}} & , \text{ if } 1+\alpha < \nu < 3+2\alpha, \end{cases}$$

$$(ii) \qquad \|u - u_\varepsilon\|_{0,\infty,\Omega} \leq \varepsilon \quad \text{for all } \nu > 0.$$

Proof. (ii) was noted above in (3.5). Proof of (i): Similarly to (2.32) we obtain that

$$|u - u_\varepsilon|^2_{1,\Omega} \leq (c[f(u) - g_\varepsilon(u)], u - u_\varepsilon), \tag{3.7}$$

Let $w := c[f(u) - g_\varepsilon(u)][u - u_\varepsilon]$. Then we have that

$$|u - u_\varepsilon|^2_{1,\Omega} \leq \int_{0<u<\varepsilon} w\,dx + \int_{\varepsilon \leq u < \varepsilon_0} w\,dx + \int_{\varepsilon_0 \leq u} w\,dx = I_1 + I_2 + I_3. \tag{3.8}$$

Similarly to (2.33) and on noting Lemma 1.8 we obtain for all $\mu \in [0,1]$ that

$$I_1 \leq \varepsilon^{1-\mu} \int_{0<u<\varepsilon} c f(u) u^\mu\,dx \leq C \begin{cases} \varepsilon^{2+\alpha-\nu} & , \text{ if } \quad 0 < \nu \leq 1+\alpha, \\ \varepsilon^{\frac{3+2\alpha-\nu}{2+\alpha}} & , \text{ if } 1+\alpha < \nu < 3+2\alpha. \end{cases} \tag{3.9}$$

The Mean Value Theorem and (ii) yield that

$$I_3 = \int_{\varepsilon_0 \le u} c[f(u) - f(u+\varepsilon)][u - u_\varepsilon]\, dx \le C\varepsilon^2\,, \tag{3.10}$$

$$I_2 \le \varepsilon^2 \int_{\varepsilon \le u < \varepsilon_0} c|f'(u)|\, dx\,. \tag{3.11}$$

To bound the right hand side of (3.11), we proceed as in the proof of Lemma 1.8 to which we refer to for the choice of ε_0. If $\nu < 1+\alpha$, we obtain that

$$I_2 \le C\varepsilon^2 \int_{\frac{\varepsilon}{\Lambda} \le d(x) \le \frac{\varepsilon_0}{\lambda}} c|f'(u)|\, dx \le C\varepsilon^2 \int_\varepsilon^{\varepsilon_0} \eta^\alpha \eta^{-(1+\nu)}\, d\eta \le C \begin{cases} \varepsilon^{2+[\alpha-\nu]_-}\,, & \text{if } \nu \ne \alpha\,, \\ \varepsilon^2 \log \frac{1}{\varepsilon}\,, & \text{if } \nu = \alpha\,, \end{cases} \tag{3.12}$$

and if $1+\alpha \le \nu < 3+2\alpha$, that

$$I_2 \le C\varepsilon^2 \int_{[\frac{\varepsilon}{\Lambda}]^{\frac{1+\nu}{2+\alpha}} \le d(x) \le [\frac{\varepsilon_0}{\lambda}]^{\frac{1+\nu}{2+\alpha}}} c|f'(u)|\, dx \le C\varepsilon^2 \int_{\varepsilon^{\frac{1+\nu}{2+\alpha}}}^{\varepsilon_0^{\frac{1+\nu}{2+\alpha}}} \eta^\alpha \eta^{-\frac{2+\alpha}{1+\nu}(1+\nu)}\, d\eta \le C\varepsilon^{\frac{3+2\alpha-\nu}{2+\alpha}}\,. \tag{3.13}$$

The desired result follows from (3.8), (3.9), (3.10), (3.12), and (3.13). ∎

We note that the bounds of Lemma 1.14(i) are only different from those of Lemma 1.9(i) when $0 < \nu \le \alpha$. Finally, since Lemma 1.4 holds for the new regularization, it is easy to see that Lemma 1.10 holds with f_ε replaced by g_ε.

1.4 Finite Element Approximation

We will now introduce and analyse a finite element approximation of (WP_ε). As in the two previous sections when we considered the solution to two regularized versions of our problem, existence and uniqueness of a discrete solution do not require any boundary assumptions.

In subsections 4.1 and 4.2 we consider the regularization of section 2, whereas in subsection 4.3 we show that the results obtained in subsections 4.1 and 4.2 remain valid for the alternative regularization of section 3. Finally, we discuss the optimality of our error bounds in subsection 4.4.

1.4.1 The Discrete Problem

Let Ω^h be a polyhedral approximation of Ω defined by $\bar{\Omega}^h \equiv \bigcup_{\tau \in T^h} \bar{\tau}$, where T^h is a quasi-uniform partitioning and h is the maximum diameter (i.e. side-length) over all simplices τ of T^h. Note that $\Omega^h \equiv \Omega$ if Ω is convex polyhedral. Let I be the set of nodes and $\{x_i\}_{i \in I}$ the coordinates of the vertices of this partitioning. It is assumed that $x_i \in \partial\Omega$ for all $x_i \in \partial\Omega^h$ and all h, so that $\mathrm{dist}(\partial\Omega, \partial\Omega^h) \le Ch^2$ if $\partial\Omega$ is C^2. For ease of exposition,

it is also assumed that $\Omega^h \subseteq \Omega$. (Note that this last stipulation is made purely for the purposes of simplifying the following analysis; it is not essential.) We also assume that for each simplex τ not all vertices lie on $\partial\Omega$. The following function spaces are introduced:

$$\begin{aligned} S^h &\equiv \{\chi \in C(\bar{\Omega}^h) : \chi|_\tau \text{ is linear } \forall \tau \in T^h\}, \\ S_0^h &\equiv \{\chi \in S^h : \chi|_{\partial\Omega^h} = 0\}. \end{aligned}$$

On defining the functions $\chi_i \in S^h$ by $\chi_i(x_j) := \delta_{i,j}$ for all $i, j \in I$, and the set of internal nodes by $I_{int} := \{i \in I : x_i \notin \partial\Omega^h\}$, the set $\{\chi_i : i \in I\}$ (resp. $\{\chi_i : i \in I_{int}\}$) is a basis for S^h (resp. S_0^h).

It will be convenient to extend the elements of S_0^h to the whole of $\bar{\Omega}$ by defining

$$\chi|_{\bar{\Omega}\setminus\bar{\Omega}^h} \equiv 0. \tag{4.1}$$

We also introduce the interpolation operator $\pi^h : C(\bar{\Omega}) \to S^h$ which is defined to be such that for any $\psi \in C(\bar{\Omega})$, $\pi^h\psi \in S^h$ satisfies $\pi^h\psi(x_i) = \psi(x_i)\ \forall i \in I$. The standard interpolation result, see [11], p.123, states that for $m = 0$ or 1, $s \in [1, \infty]$ and $\sigma \in [1, \infty]$, provided that $W^{2,\sigma}(\tau) \subset W^{m,s}(\tau) \cap C(\bar{\tau})$,

$$|\psi - \pi^h\psi|_{m,s,\tau} \leq Ch^{d(\frac{1}{s}-\frac{1}{\sigma})+2-m}|\psi|_{2,\sigma,\tau} \qquad \forall \psi \in W^{2,\sigma}(\tau) \text{ and } \forall \tau \in T^h. \tag{4.2}$$

We want to consider the following discretization of (P_ε):

Find $u_\varepsilon^h \in S_0^h$ such that

$$(P_\varepsilon^h) \qquad \int_\Omega \nabla u_\varepsilon^h \cdot \nabla\chi \, dx = \int_\Omega c f_\varepsilon(u_\varepsilon^h)\chi \, dx \qquad \forall \chi \in S_0^h.$$

Existence and uniqueness can be shown along the lines of the proof of Lemma 1.1.

In order to show that u_ε^h is non-negative in Ω, we need the Discrete Maximum Principle which in turn requires a further assumption on the partitioning T^h of Ω^h. Throughout the following we will assume

(WA) The triangulation is such that (i) for $d = 2$ it is **w**eakly **a**cute; that is, for any pair of adjacent triangles the sum of opposite angles relative to a common side does not exceed π; and (ii) for $d = 3$ the angle between any two faces of the same tetrahedron must not exceed $\frac{\pi}{2}$.

We note the following consequences of (WA).

$$\int_\Omega |\nabla\pi^h[\chi]_\pm|^2 \, dx \leq \int_\Omega \nabla\chi \cdot \nabla\pi^h[\chi]_\pm \, dx \qquad \forall \chi \in S_0^h, \tag{4.3}$$

e.g. see [22], Lemma 1.

$$\int_{\Omega^h} \nabla\chi_i \cdot \nabla\chi_j \, dx \leq 0 \qquad \forall i, j \in I,\ i \neq j, \tag{4.4}$$

and the stiffness matrix K defined by

$$K := \left(\int_\Omega \nabla\chi_i \cdot \nabla\chi_j \, dx \right)_{i,j \in I_{int}},$$

is an M matrix, e.g. see [34].

Discrete Maximum Principle (DMP):

Let T^h be a weakly acute triangulation of Ω^h and $v^h \in S_0^h$ satisfy

$$\int_\Omega \nabla v^h \cdot \nabla\chi \, dx + \int_\Omega \pi^h(\sigma v^h \chi) \, dx \geq 0\,, \qquad \forall \chi \in S_0^h,\ \chi \geq 0, \tag{4.5}$$

where $\sigma \in C(\bar{\Omega})$ and $\sigma \geq 0$. Then $0 \leq v^h$ in $\bar{\Omega}$.

Proof. (cf. proof of Theorem 8.1 in [17]). On letting $\chi := -\pi^h[v^h]_-$ (4.5) yields that

$$\int_\Omega \nabla v^h \cdot \nabla \pi^h[v^h]_- \, dx \leq -\int_\Omega \pi^h(\sigma v^h \pi^h[v^h]_-) \, dx \leq 0\,. \tag{4.6}$$

On noting (4.3) and Poincaré's inequality we obtain from (4.6) that $\pi^h[v^h]_- = 0$ and hence, $v^h \geq 0$. ∎

Since c, $f_\varepsilon \geq 0$, we obtain from (P_ε^h) that

$$\int_\Omega \nabla u_\varepsilon^h \cdot \nabla\chi \geq 0, \qquad \forall \chi \in S_0^h,\ \chi \geq 0\,.$$

Hence, (DMP) (with $\sigma \equiv 0$) yields that

$$0 \leq u_\varepsilon^h \quad \text{in } \bar{\Omega}\,. \tag{4.7}$$

Later we will also need the following consequence of (WA):

$$\begin{aligned} &\alpha \in W^{1,\infty}(\mathbb{R}),\ \alpha(0) = 0 \quad \text{and } 0 \leq \alpha' \leq L_\alpha < \infty \text{ a.e. in } \mathbb{R} \quad \Longrightarrow \\ &|\nabla \pi^h[\alpha(\chi)]|_{0,\Omega}^2 \leq L_\alpha \langle \nabla\chi, \nabla \pi^h[\alpha(\chi)] \rangle \qquad \forall \chi \in S_0^h\,, \end{aligned} \tag{4.8}$$

e.g. see §2.4.2 in [34].

1.4.2 The Discretization Error

We now want to derive bounds for $|u_\varepsilon - u_\varepsilon^h|_{1,\Omega}$ and $\|u_\varepsilon - u_\varepsilon^h\|_{0,\infty,\Omega}$. The following inverse inequality and standard discrete Sobolev embedding result will be required later: for $d = 1, 2, 3$ and for any $r \in [1, \infty]$,

$$|\chi|_{1,r,\Omega^h} \leq Ch^{-1} |\chi|_{0,r,\Omega^h} \quad \forall \chi \in S^h\,; \tag{4.9}$$

and for $d = 1, 2$,

$$\|\chi\|_{0,\infty,\Omega^h} \leq C \left(\ln \frac{1}{h} \right)^{\frac{d-1}{2}} |\chi|_{1,\Omega^h} \quad \forall \chi \in S_0^h\,. \tag{4.10}$$

We also recall the definition of the elliptic projection operator Q^h.
$Q^h \in \mathcal{L}(H_0^1(\Omega), S_0^h)$ is defined to be such that for any $\psi \in H_0^1(\Omega)$

$$\int_{\Omega^h} \nabla(\psi - Q^h\psi) \cdot \nabla\chi = 0 \quad \forall \chi \in S_0^h \,. \tag{4.11}$$

The following estimate is standard. For $\psi \in H_0^1(\Omega) \cap H^2(\Omega)$ and for $m = 1$ and 2 we have that

$$|\psi - Q^h\psi|_{0,\Omega} + h|\psi - Q^h\psi|_{1,\Omega} \leq Ch^m \|\psi\|_{m,\Omega} \,. \tag{4.12}$$

Later we will also need the following result. For all $\psi \in H_0^1(\Omega) \cap H^2(\Omega)$ we have for h sufficiently small that

$$\|(I - Q^h)\psi\|_{0,\infty,\Omega^h} \leq C \left(\ln \frac{1}{h}\right)^r \|(I - \pi^h)\psi\|_{0,\infty,\Omega^h} \,, \tag{4.13}$$

where $r = 0$, if $d = 1$, and $r = 1$, if $d = 2$ or 3. For the proof of this result for $d = 2$ or 3 see [41]. If $d = 1$, it is well known that

$$Q^h \equiv \pi^h \quad \text{on} \quad H_0^1(\Omega) \,. \tag{4.14}$$

For notational convenience we define $h_d := h \left(\ln \frac{1}{h}\right)^r$, where r is as in (4.13).

Lemma 1.15 *Assume $A(\alpha)$ holds. Then there exist $\varepsilon_1 > 0$ and $H_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_1]$, $h \in (0, H_0]$ the following bounds hold:*

$$\begin{aligned}
&(i) \quad \textit{for } m = 0 \textit{ or } 1, \\
&\qquad |(I - Q^h)u_\varepsilon|_{m,\Omega} \leq Ch^{2-m}\|u_\varepsilon\|_{2,\Omega} \leq Ch^{2-m}\|cf_\varepsilon(u_\varepsilon)\|_{0,\Omega} \\
&\qquad\qquad \leq Ch^{2-m} \begin{cases} 1 \,, & \textit{if } \nu < \frac{1}{2} + \alpha \,, \\ \left(\ln \frac{1}{\varepsilon}\right)^{\frac{1}{2}} \,, & \textit{if } \nu = \frac{1}{2} + \alpha \,, \\ \varepsilon^{\frac{1}{2}+\alpha-\nu} \,, & \textit{if } \frac{1}{2} + \alpha < \nu \,, \end{cases} \\
&(ii) \quad \|(I - \pi^h)u_\varepsilon\|_{0,\infty,\Omega^h} \leq Ch_d h\|cf_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega} \leq C\, h_d h\, \varepsilon^{[\alpha-\nu]-} \,, \\
&(iii) \quad \|(I - Q^h)u_\varepsilon\|_{0,\infty,\Omega^h} \leq C\, h_d^2\, \varepsilon^{[\alpha-\nu]-} \,.
\end{aligned}$$

Proof. (i) follows from (4.12), elliptic regularity and Lemma 1.10 with $q = 2$, and holds for all $h > 0$. To prove (ii) we note that (4.2) yields for $d = 1, 2, 3$ and $\sigma \geq 1$ that

$$\|(I - \pi^h)u_\varepsilon\|_{0,\infty,\Omega^h} \leq Ch^{2-\frac{d}{\sigma}}|u_\varepsilon|_{2,\sigma,\Omega^h} \,. \tag{4.15}$$

For $d = 1$ we can choose $\sigma = \infty$, and the desired result follows from (4.15), the equation $-u_\varepsilon'' = cf_\varepsilon(u_\varepsilon)$ and Lemma 1.10 with $q = \infty$. For $d = 2$ or 3 we use the Calderón–Zygmund inequality, see [17], to bound $|u_\varepsilon|_{2,\sigma,\Omega^h}$: For all $\sigma \in (1, \infty)$,

$$|u_\varepsilon|_{2,\sigma,\Omega} \leq C(\sigma)\|cf_\varepsilon(u_\varepsilon)\|_{0,\sigma,\Omega} \,, \tag{4.16}$$

where

$$C(\sigma) \equiv \begin{cases} C(\sigma-1)^{-1}, & 1 < \sigma \leq 2 \\ C\sigma, & 2 \leq \sigma < \infty \end{cases} \tag{4.17}$$

from tracing the constants in the singular integrals involved, see e.g. [43], p.48. Combining (4.15) with $\sigma = \ln \frac{1}{h}$, (4.16) and noting (4.17) we obtain that for sufficiently small h,

$$\|(I-\pi^h)u_\varepsilon\|_{0,\infty,\Omega^h} \leq C\left(\ln\frac{1}{h}\right) h^{2+\frac{d}{\ln h}} \|cf_\varepsilon(u_\varepsilon)\|_{0,\ln\frac{1}{h},\Omega}\,.$$

Noting that $h^{\frac{d}{\ln h}} = e^d$ and Lemma 1.10 yields

$$\|(I-\pi^h)u_\varepsilon\|_{0,\infty,\Omega^h} \leq C\left(\ln\frac{1}{h}\right) h^2 \|cf_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega} \leq C\, h_d h\, \varepsilon^{[\alpha-\nu]-}\,,$$

and hence (ii). (iii) follows from (ii) and (4.13). ∎

In the following $(\cdot,\cdot)$ denotes the inner product in $L^2(\Omega)$.

Lemma 1.16 *Assume $A(\alpha)$ holds. Then there exists $\varepsilon_1 > 0$ such that for all $\varepsilon \in (0,\varepsilon_1]$ and $h > 0$,*

$$|Q^h u_\varepsilon - u_\varepsilon^h|_{1,\Omega} \leq Ch\left[\|cf_\varepsilon(u_\varepsilon)\|_{0,\Omega} + \|cf_\varepsilon(u_\varepsilon^h)\|_{0,\Omega}\right].$$

Proof. Define $e^h := Q^h u_\varepsilon - u_\varepsilon^h \in S_0^h$. By subtracting (P_ε^h) from (WP_ε), choosing $\chi \equiv e^h$ and noting (4.11) we obtain

$$(\nabla e^h, \nabla e^h) = (c[f_\varepsilon(u_\varepsilon) - f_\varepsilon(u_\varepsilon^h)], e^h)\,,$$

which we rewrite as

$$|e^h|_{1,\Omega}^2 - (c[f_\varepsilon(u_\varepsilon) - f_\varepsilon(u_\varepsilon^h)], u_\varepsilon - u_\varepsilon^h) = (c[f_\varepsilon(u_\varepsilon) - f_\varepsilon(u_\varepsilon^h)], Q^h u_\varepsilon - u_\varepsilon)\,.$$

Noting (2.3) and using Cauchy's inequality and the first two inequalities in Lemma 1.15(i) yields the desired result. ∎

To bound $\|Q^h u_\varepsilon - u_\varepsilon^h\|_{0,\infty,\Omega^h}$ we adapt the technique introduced by Nochetto [32]. This involves introducing the following dual problem:

Find $D \in H_0^1(\Omega) \cap H^2(\Omega)$ s.t.

$$-\Delta D + \Theta D = \delta\,, \tag{4.18}$$

where

$$\Theta(x) \equiv \begin{cases} -c\left(\frac{f_\varepsilon(u_\varepsilon)-f_\varepsilon(u_\varepsilon^h)}{u_\varepsilon - u_\varepsilon^h}\right)(x), & \text{if } x \in \Omega^h \text{ and } u_\varepsilon(x) \neq u_\varepsilon^h(x) \\ 0, & \text{otherwise,} \end{cases} \tag{4.19}$$

and δ is defined to be a regularisation of the Dirac delta function satisfying

$$0 \le \delta \le C(\gamma h)^{-d}, \int_\Omega \delta = 1 \tag{4.20}$$

and zero outside some open ball of radius γh about some point x_0, where γ and x_0 are to be chosen. Note that by (2.3) and the boundedness of c

$$0 \le \Theta(x) \le C\varepsilon^{-(\nu+1)}, \quad \forall x \in \bar{\Omega}, \tag{4.21}$$

and hence there exists a unique solution to (4.18), see [17], Theorem 8.3.

Lemma 1.17 *The unique solution D of (4.18) satisfies*

$$\|\Delta D\|_{0,1,\Omega} + \|\Theta D\|_{0,1,\Omega} \le C.$$

Proof. See Lemma 2.1 in [32]. ∎

Lemma 1.18 *For h sufficiently small we have that*

$$\|(I - Q^h)D\|_{0,1,\Omega} \le Ch_d^2.$$

Proof. For $d = 1$ the bound follows from (4.14), (4.2) and Lemma 1.17. For $d = 2$ or 3 see Lemma 2.2 in [32]. ∎

Lemma 1.19 *Assume that $A(\alpha)$ holds. Then there exist $\varepsilon_1 > 0$ and $H_1 > 0$ such that for all $\varepsilon \in (0, \varepsilon_1]$, $h \in (0, H_1]$,*

$$\begin{aligned} &\|Q^h u_\varepsilon - u_\varepsilon^h\|_{0,\infty,\Omega} \\ &\quad \le \|c[f_\varepsilon(u_\varepsilon) - f_\varepsilon(u_\varepsilon^h)]\|_{0,\infty,\Omega^h} \|(I - Q^h)D\|_{0,1,\Omega^h} + C\|(I - Q^h)u_\varepsilon\|_{0,\infty,\Omega^h} \\ &\quad \le Ch_d^2 \left[\|cf_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega} + \|cf_\varepsilon(u_\varepsilon^h)\|_{0,\infty,\Omega} + \varepsilon^{[\alpha-\nu]-} \right]. \end{aligned} \tag{4.22}$$

Proof. Define $e^h \equiv Q^h u_\varepsilon - u_\varepsilon^h \in S_0^h$. Let $x_0 \in \Omega^h$ be such that $\|e^h\|_{0,\infty,\Omega} = |e^h(x_0)|$ and, noting (4.1) and (4.20), choose $x_1 \in B_{\gamma h}(x_0)$ satisfying

$$e^h(x_1) = \int_\Omega e^h \delta = \int_{\Omega^h} e^h \delta.$$

By the Mean Value theorem and (4.9),

$$|e^h(x_0)| \le |e^h(x_1)| + \gamma h \|e^h\|_{1,\infty,\Omega} \le |e^h(x_1)| + C\gamma |e^h(x_0)|,$$

and so for γ sufficiently small

$$\|e^h\|_{0,\infty,\Omega} = |e^h(x_0)| \le C|e^h(x_1)| = C \left| \int_{\Omega^h} e^h \delta \right|. \tag{4.23}$$

By subtracting (P_ε^h) from (WP_ε) and noting (4.1) and (4.11) we obtain the following error equation

$$\int_{\Omega^h} [\nabla e^h \cdot \nabla \chi + \Theta e \chi] = 0 \quad \forall \chi \in S_0^h , \tag{4.24}$$

where $e \equiv u_\varepsilon - u_\varepsilon^h$. Now, by (4.18), (4.24), (4.11), (4.19) and Lemma 1.17 it follows that

$$\begin{aligned} \int_{\Omega^h} e^h \delta &= \int_{\Omega^h} [\nabla e^h \cdot \nabla D + \Theta e^h D] \\ &= \int_{\Omega^h} [\nabla e^h \cdot \nabla (I - Q^h) D + \Theta (e^h D - e Q^h D)] \\ &= \int_{\Omega^h} [\Theta (u_\varepsilon - u_\varepsilon^h)(I - Q^h) D - \Theta D (I - Q^h) u_\varepsilon] \\ &\leq \int_{\Omega^h} |c[f_\varepsilon(u_\varepsilon) - f_\varepsilon(u_\varepsilon^h)](I - Q^h) D| + C \| (I - Q^h) u_\varepsilon \|_{0,\infty,\Omega^h} . \end{aligned} \tag{4.25}$$

The first inequality in (4.22) now follows from (4.25) and Hölder's inequality, and the second from noting Lemmas 1.15(iii) and 1.18. ∎

In order to use the bounds of Lemma 1.15 and Lemma 1.16 [resp. Lemma 1.19] to bound $\|u_\varepsilon - u_\varepsilon^h\|_{0,\Omega}$ [resp. $\|u_\varepsilon - u_\varepsilon^h\|_{0,\infty,\Omega}$], we need bounds for $\|c f_\varepsilon(u_\varepsilon^h)\|_{0,\Omega}$ and $\|c f_\varepsilon(u_\varepsilon^h)\|_{0,\infty,\Omega}$. Clearly, the best thing would be to have an analogue of Lemma 1.10 for u_ε^h. Unfortunately, we were only able to prove such an analogue for $\alpha = 0$, see Lemma 1.20 below. The essential ingredient of the proof is that u_ε^h is growing at least linearly away from the boundary, independently of ε. We will establish this result first.

It follows from Lemma 1.10, the boundedness of c and (4.7) that

$$\|c f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega},\ \|c f_\varepsilon(u_\varepsilon^h)\|_{0,\infty,\Omega} \leq C \varepsilon^{-\nu} . \tag{4.26}$$

Letting $H_2 := \min\{H_0, H_1\}$ we obtain from the Lemmas 1.4, 1.15(iii), 1.19 and (4.26) that there exist C^*, $\varepsilon_1 > 0$ such that for all $\varepsilon \in (0, \varepsilon_1]$ and $h \in (0, H_2]$,

$$\begin{aligned} & u_\varepsilon^h(x) - \frac{1}{2} \lambda_{\varepsilon_1} d(x) \\ &= [u_\varepsilon(x) - \lambda_{\varepsilon_1} d(x)] + \frac{1}{2} \lambda_{\varepsilon_1} d(x) + [u_\varepsilon^h(x) - u_\varepsilon(x)] \\ &\geq \frac{1}{2} \lambda_{\varepsilon_1} d(x) - C^* \varepsilon^{-\nu} h_d^2 , \qquad \forall x \in \bar{\Omega} . \end{aligned} \tag{4.27}$$

Hence, we obtain from (4.27) that for all $\varepsilon \in (0, \varepsilon_1]$ and $h \in (0, H_2]$,

$$u_\varepsilon^h(x) \geq \frac{1}{2} \lambda_{\varepsilon_1} d(x) , \qquad \forall x \in \Omega \setminus \Omega_{\frac{2\varepsilon}{\lambda_{\varepsilon_1}}} . \tag{4.28}$$

if

$$\varepsilon \geq C^* h_d^{\frac{2}{1+\nu}} . \tag{4.29}$$

Lemma 1.20 *Assume $A(0)$ holds. Then there exist ε_2, $H_2 > 0$ and $C^* > 0$ such that for all $h \in (0, H_2]$ and $\varepsilon \in [C^* h_d^{\frac{2}{1+\nu}}, \varepsilon_2]$ we have that*

$$\forall q \in [1,\infty], \quad \|cf_\varepsilon(u_\varepsilon^h)\|_{0,q,\Omega} \leq C(\nu) \begin{cases} 1 & , \text{ if } 0 < \nu < \frac{1}{q}, \\ \left(\ln \frac{1}{\varepsilon}\right)^{\frac{1}{q}} & , \text{ if } \nu = \frac{1}{q}, \\ \varepsilon^{\frac{1}{q}-\nu} & , \text{ if } \frac{1}{q} < \nu. \end{cases}$$

Proof. Let $\bar{\delta}$ be as in the proof of Lemma 1.4 and ε_1, λ_{ε_1}, H_2 and C^* be as in (4.27). We choose $\varepsilon_2 := \min\{\frac{1}{2}\lambda_{\varepsilon_1}\bar{\delta}, \varepsilon_1\}$.

For $q \in [1,\infty)$ we have that

$$\begin{aligned} &\int_\Omega [cf_\varepsilon(u_\varepsilon^h)]^q \, dx \\ &= \int_{\Omega_{\frac{2\varepsilon}{\lambda_{\varepsilon_1}}}} [cf_\varepsilon(u_\varepsilon^h)]^q \, dx + \int_{\Omega_{\frac{2\varepsilon_2}{\lambda_{\varepsilon_1}}} \setminus \Omega_{\frac{2\varepsilon}{\lambda_{\varepsilon_1}}}} [cf_\varepsilon(u_\varepsilon^h)]^q \, dx + \int_{\Omega \setminus \Omega_{\frac{2\varepsilon_2}{\lambda_{\varepsilon_1}}}} [cf_\varepsilon(u_\varepsilon^h)]^q \, dx. \end{aligned} \tag{4.30}$$

Note that as $\varepsilon_2 \leq \frac{1}{2}\lambda_{\varepsilon_1}\bar{\delta}$ we have that

$$\Omega_{\frac{2\varepsilon_2}{\lambda_{\varepsilon_1}}} \subset \Omega_{\bar{\delta}}. \tag{4.31}$$

As we only consider those ε for which (4.29) holds, we have also (4.28). Using (4.31) and (4.28) one can now proceed from (4.30) as in the proof of Lemma 1.10 to obtain the desired result. ∎

We can now summarize this section under the assumption $A(0)$. Firstly we remark that it follows from comparing the bounds of Lemma 1.15 with those of Lemmas 1.16 and 1.19 on noting Lemma 1.20 that the bounds for the error $u_\varepsilon - u_\varepsilon^h$ are of the same order as those for $u_\varepsilon - Q^h u_\varepsilon$.

Theorem 1.1 *Assume $A(0)$ holds. We then have for h sufficiently small that*

(*i*) $$|u - u_\varepsilon^h|_{1,\Omega}$$
$$\leq C \begin{cases} h & , \text{ if } C^* h_d^{\frac{2}{1+\nu}} \leq \varepsilon \leq \tilde{C} h^{\frac{2}{2-\nu}} \text{ and } 0 < \nu < \frac{1}{2}, \\ h\left(\ln \frac{1}{h}\right)^{\frac{1}{2}} & , \text{ if } \varepsilon = \tilde{C} h^{\frac{4}{3}} \text{ and } \nu = \frac{1}{2}, \\ h^{\frac{2-\nu}{1+\nu}} & , \text{ if } \varepsilon = \tilde{C} h^{\frac{2}{1+\nu}} \text{ and } \frac{1}{2} < \nu \leq 1, \\ h^{\frac{3-\nu}{1+3\nu}} & , \text{ if } \varepsilon = \tilde{C} h^{\frac{4}{1+3\nu}} \text{ and } 1 < \nu < 3, \end{cases}$$
where $\tilde{C}$ has to be chosen sufficiently large,

(*ii*) $$\|u - u_\varepsilon^h\|_{0,\infty,\Omega^h} \leq C h_d^{\frac{2}{1+\nu}}, \text{ if } \varepsilon = \tilde{C} h_d^{\frac{2}{1+\nu}} \text{ and } 0 < \nu.$$

Proof. From equating the bounds of the Lemmas 1.9(i), 1.15(i), 1.16 and noting Lemma 1.20 we obtain that

$$|u-u_\varepsilon^h|_{1,\Omega} \le C \begin{cases} \varepsilon^{1-\frac{\nu}{2}}, & \text{if } 0<\nu\le 1 \\ \varepsilon^{\frac{3-\nu}{4}}, & \text{if } 1<\nu<3 \end{cases} + Ch \begin{cases} 1, & \text{if } \nu<\frac{1}{2}, \\ (\ln\frac{1}{\varepsilon})^{\frac{1}{2}}, & \text{if } \nu=\frac{1}{2}, \\ \varepsilon^{\frac{1}{2}-\nu}, & \text{if } \frac{1}{2}<\nu, \end{cases} \tag{4.32}$$

from which (i) follows on equating ε and h, where we note that $\tilde{C}$ has to be chosen sufficiently large in order to use Lemma 1.20. Similarly, we obtain from the Lemmas 1.9(ii), 1.15(iii), 1.19 and (4.26) that

$$\|u-u_\varepsilon^h\|_{0,\infty,\Omega^h} \le C\varepsilon + Ch_d^2\varepsilon^{-\nu} \le Ch_d^{\frac{2}{1+\nu}},$$

if $\varepsilon = \tilde{C}h^{\frac{2}{1+\nu}}$ and $0<\nu$. ∎

We remark that for $0<\nu\le 1$ the same choice of the regularization parameter ε can be used for both the $L^\infty(\Omega)$ and the $H^1(\Omega)$ error bound without much altering the bounds in (4.32); more precisely, for $\varepsilon = \tilde{C}h_d^{\frac{2}{1+\nu}}$ we have that

$$|u-u_\varepsilon^h|_{1,\Omega} \le C \begin{cases} h, & \text{if } 0<\nu<\frac{1}{2}, \\ h_d^{\frac{2-\nu}{1+\nu}}, & \text{if } \frac{1}{2}\le\nu\le 1. \end{cases}$$

1.4.3 Discretization and Alternative Regularization

We now show that the above results apply to the discretization of the alternative regularization which we introduced in section 3.

We define (P_ε^h) as in 4.1 with f_ε replaced by g_ε. Existence and uniqueness of a solution u_ε^h to (P_ε^h) can be proved similarly as in the proof of Lemma 1.11. We just note that instead of (SMP) we have to use (DMP) which yields that

$$0 \le u_\varepsilon^h \quad \text{in } \bar{\Omega}.$$

Since Lemma 1.10 holds for f_ε replaced by g_ε, the same is true for the Lemmas 1.15 and 1.16. On replacing f_ε by g_ε in (4.19) and on noting (3.3) (4.21) remains valid, and so do the Lemmas 1.17–1.19. Also the analogue of (4.26) holds, and hence we obtain analogues for Lemma 1.20 and Theorem 1.1. Note that in the proof of the analogue of Theorem 1.1 one now uses Lemma 1.14 instead of Lemma 1.9: the fact that Lemma 1.14 differs from Lemma 1.9 when $0<\nu\le\alpha$ has no effect, since we assume $A(0)$ in Theorem 1.1.

1.4.4 Optimality of the Error Bounds

To obtain an idea how good the bounds of Theorem 1.1 are, we compare them with standard bounds for the interpolation error: $u-\pi^h u$. We adopt the convention that, if these interpolation error bounds are the same as our bounds on $u-u_\varepsilon^h$, we say that our bounds are optimal.

Firstly, we consider the L^∞ bounds. For $m = 0$ or 1 and $\gamma \in [0,1]$ we have the following standard estimate for the interpolation error:

$$\|u - \pi^h u\|_{0,\infty,\Omega} \leq Ch^{m+\gamma} \|u\|_{C^{m,\gamma}(\bar{\Omega})}, \tag{4.33}$$

if $u \in C^{m,\gamma}(\bar{\Omega})$. Combining (4.33) and (1.4)–(1.6) we obtain that

$$\|u - \pi^h u\|_{0,\infty,\Omega} \leq C \begin{cases} h^{2-\nu}, & \text{if } \nu < 1, \\ h^{1-\delta}, & \text{if } \nu = 1, \\ h^{\frac{2}{1+\nu}}, & \text{if } 1 < \nu, \end{cases} \tag{4.34}$$

where $\delta \in (0,1)$ is arbitrary. Comparing (4.34) with the bound of Theorem 1.1(ii), we see that our bound is quasi-optimal for $\nu \geq 1$. For $\nu \in (0,1)$, we have to compare $h^{2-\nu}$ and $h^{\frac{2}{1+\nu}}$. An easy calculation shows that the maximum difference between $2-\nu$ and $\frac{2}{1+\nu}$ for $\nu \in (0,1)$ occurs at $\nu_{\max} = \sqrt{2} - 1$ and is approximately equal to 0.1716, which is not too bad considering that $2 - \nu_{\max} \approx 1.5858$ and $\frac{2}{1+\nu_{\max}} \approx 1.4142$.

Now we turn to the H^1 bounds. Clearly, for $\nu < \frac{1}{2}$ our bound is optimal. For $\nu \in [\frac{1}{2}, 3)$ we only consider the case $d = 1$. From (4.2) we obtain for $\sigma \in [1, \infty]$ that

$$|u - \pi^h u|_{1,\Omega} \leq Ch^{\frac{3}{2}-\frac{1}{\sigma}} |u|_{2,\sigma,\Omega} \leq Ch^{\frac{3}{2}-\frac{1}{\sigma}} |f(u)|_{0,\sigma,\Omega}. \tag{4.35}$$

If $\nu < 1$, we obtain from (1.1) that $\lambda d(x) \leq u(x)$ for all $x \in \bar{\Omega}$ and hence $|f(u)|_{0,\sigma,\Omega} \leq C$, if $\sigma < \frac{1}{\nu}$. Choosing $\sigma = \frac{1}{\nu(1+\delta)}$ we obtain from (4.35) that

$$|u - \pi^h u|_{1,\Omega} \leq Ch^{\frac{3}{2}-\nu(1+\delta)} = Ch^{\frac{3}{2}-\nu-\delta},$$

i.e. we have to compare $h^{\frac{3}{2}-\nu}$ and $h^{\frac{2-\nu}{1+\nu}}$. Again, an easy calculation shows that the maximum difference between $\frac{3}{2} - \nu$ and $\frac{2-\nu}{1+\nu}$ for $\nu \in [\frac{1}{2}, 1)$ occurs at $\nu_{\max} = \sqrt{3} - 1$. We have that $\frac{3}{2} - \nu_{\max} \approx 0.7679$ and $\frac{2-\nu_{\max}}{1+\nu_{\max}} = \sqrt{3} - 1 \approx 0.7321$. Hence, the maximal difference is approximately equal to 0.0358. Thus we consider our error bound of Theorem 1.1(i) to be "nearly optimal" for $\nu \in [\frac{1}{2}, 1)$.

Now we consider the case when $\nu \in (1,3)$. We cannot apply (4.35), since by (1.3) $f(u) \notin L^1(\Omega)$. We will use the following generalization of (4.35) for "fractional derivatives":

$$|u - \pi^h u|_{1,\Omega} \leq Ch^\beta |u|_{1+\beta,2,\Omega}, \tag{4.36}$$

where for $\beta \in (0,1)$,

$$|u|^2_{1+\beta,2,\Omega} := \int_\Omega \int_\Omega \frac{|u'(x) - u'(y)|^2}{|x-y|^{1+2\beta}} \, dxdy, \tag{4.37}$$

cf. [16]. We may assume w.l.o.g. that $\Omega \equiv (0,1)$. On noting (1.3) we obtain that

$$|u'(x) - u'(y)| = \left| \int_y^x u''(s) \, ds \right| = \left| \int_y^x c[u(s)]^{-\nu} \, ds \right| \leq C \left| \int_y^x s^{-\frac{2\nu}{1+\nu}} \, ds \right| \leq C \left| x^{\frac{1-\nu}{1+\nu}} - y^{\frac{1-\nu}{1+\nu}} \right|. \tag{4.38}$$

As for the integration in (4.37) we note that by the symmetry of the integrand,

$$\int_0^1 \int_0^1 \ldots dxdy = \int_0^1 \left[\int_0^x \ldots dy\right] dx + \int_0^1 \left[\int_x^1 \ldots dy\right] dx$$
$$= \int_0^1 \left[\int_0^x \ldots dy\right] dx + \int_0^1 \left[\int_0^y \ldots dx\right] dy = 2\int_0^1 \left[\int_0^x \ldots dy\right] dx\,. \qquad (4.39)$$

From (4.38) and (4.39) we obtain that

$$\begin{aligned} |u|^2_{1+\beta,2,\Omega} &\le C\int_0^1 \left[\int_0^x \frac{(x^{\frac{1-\nu}{1+\nu}} - y^{\frac{1-\nu}{1+\nu}})^2}{(x-y)^{1+2\beta}}\, dy\right] dx \\ &\le C\int_0^1 x^{2\frac{1-\nu}{1+\nu}-(1+2\beta)} \left[\int_0^x \frac{(1-[\frac{y}{x}]^{\frac{1-\nu}{1+\nu}})^2}{(1-[\frac{y}{x}])^{1+2\beta}}\, dy\right] dx \\ &\le C\left[\int_0^1 x^{2(\frac{1-\nu}{1+\nu}-\beta)}\, dx\right] \left[\int_0^1 \frac{(1-z^{\frac{1-\nu}{1+\nu}})^2}{(1-z)^{1+2\beta}}\, dz\right], \end{aligned} \qquad (4.40)$$

where we have used the substitution $z \equiv y/x$. The first integral on the right hand side of (4.40) exists, if $(1-\nu)/(1+\nu) > \beta - \frac{1}{2}$, i.e. if

$$\beta < \frac{3-\nu}{2(1+\nu)}\,. \qquad (4.41)$$

We note that

$$\frac{(1-z^{\frac{1-\nu}{1+\nu}})^2}{(1-z)^{1+2\beta}} = z^{2\frac{1-\nu}{1+\nu}} \frac{(1-z^{\frac{\nu-1}{1+\nu}})^2}{(1-z)^{1+2\beta}} \le z^{2\frac{1-\nu}{1+\nu}} (1-z)^{1-2\beta},$$

as $\frac{\nu-1}{1+\nu} \in (0, \frac{1}{2})$ and $1 - z^{\frac{\nu-1}{1+\nu}} \le 1 - z$, and hence

$$\int_0^1 \frac{(1-z^{\frac{1-\nu}{1+\nu}})^2}{(1-z)^{1+2\beta}}\, dz < \infty,$$

if (i) $2(1-\nu)/(1+\nu) > -1$ and (ii) $1 - 2\beta > -1$. But (i) holds as $\nu \in (1,3)$, and (ii) holds as $\beta \in (0,1)$.

On combining (4.36), (4.37), (4.40), and (4.41) we obtain that for all $\delta > 0$,

$$|u - \pi^h u|_{1,\Omega} \le C h^{\frac{3-\nu}{2(1+\nu)}-\delta}, \qquad (4.42)$$

i.e. we have to compare $h^{\frac{3-\nu}{2(1+\nu)}}$ and $h^{\frac{3-\nu}{1+3\nu}}$. A straightforward calculation shows that the maximum difference between $\frac{3-\nu}{2(1+\nu)}$ and $\frac{3-\nu}{1+3\nu}$ for $\nu \in (1,3)$ occurs at $\nu_{\max} = (1+\sqrt{5})/2$ and is approximately equal to 0.0278, which is again not too bad considering that $\frac{3-\nu_{\max}}{2(1+\nu_{\max})} \approx 0.2639$ and $\frac{3-\nu_{\max}}{1+3\nu_{\max}} \approx 0.2361$.

Finally, we note that the analysis for $\nu = 1$ is the same as for $\nu \in (1,3)$. This is because from (1.2) and (SMP) it follows that $\tilde{\lambda} d(x) \le u(x)$ for all $x \in \bar{\Omega}$, which is the lower bound of (1.3) for $\alpha = 0$ and $\nu = 1$. In fact, (4.42) shows that our estimate in Theorem 1.1 is optimal for $\nu = 1$.

Figure 1.1 shows the orders of convergence for $\nu \in (0,3)$ according to Theorem 1.1 compared to the optimal orders of convergence obtained from the interpolation error. In the upper part of the figure the graphs refer to the L^∞ error, in the lower part to the H^1 error.

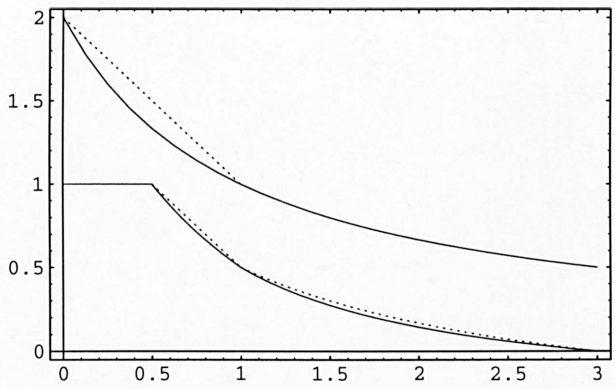


Figure 1.1: Orders of convergence according to Theorem 1.1 (—) versus optimal orders $(\cdots)$.

1.5 The Effect of Numerical Integration

(P_ε^h) involves exact integration of the nonlinear term, and as it is virtually impossible to implement such a scheme in practice, it is of interest to introduce and analyse a method involving numerical integration. As in section 4 we will first consider the regularization of section 2 in subsections 5.1–5.3.

1.5.1 The Discrete Problem with Numerical Integration

A quadrature rule is introduced as follows: given a simplex $\tau \in T^h$ having vertices $\{a_i\}_{i=1}^{d+1}$, for $\psi \in C(\bar{\tau})$, define

$$Q_\tau(\psi) := \int_\tau \pi^h \psi \equiv \frac{1}{d+1} meas(\tau) \sum_{i=1}^{d+1} \psi(a_i)\,.$$

Then

$$\int_{\Omega^h} \psi_1 \psi_2 \approx (\psi_1, \psi_2)^h := \sum_{\tau \in T^h} Q_\tau(\psi_1 \psi_2) \equiv \sum_{\tau \in T^h} \int_\tau \pi^h(\psi_1 \psi_2)\,, \tag{5.1}$$

for all $\psi_1, \psi_2 \in C(\bar{\Omega}^h)$. Also required is the fact that for all $\chi \in S^h$,

$$\|\chi\|_{0,\Omega^h}^2 \le |\chi|_h^2 \le C \|\chi\|_{0,\Omega^h}^2\,, \tag{5.2}$$

where $|\chi|_h := [(\chi,\chi)^h]^{\frac{1}{2}}$.

The standard Galerkin finite element approximation in the presence of numerical integration is now defined as follows:

Find $\hat{u}_\varepsilon^h \in S_0^h$ such that

$$(\hat{P}_\varepsilon^h) \qquad \int_\Omega \nabla \hat{u}_\varepsilon^h \cdot \nabla \chi \, dx = (c f_\varepsilon(\hat{u}_\varepsilon^h), \chi)^h \qquad \forall \chi \in S_0^h \,.$$

The proof of existence and uniqueness of a solution $\hat{u}_\varepsilon^h$ to $(\hat{P}_\varepsilon^h)$ is similar to the proof of Lemma 1.1: We again consider the corresponding operator equation. We define an operator $\hat{\mathcal{A}}_\varepsilon^h : S_0^h \to [S_0^h]^*$ by

$$\langle \hat{\mathcal{A}}_\varepsilon^h v^h, \chi \rangle := \int_\Omega \nabla v^h \cdot \nabla \chi \, dx - \int_\Omega \pi^h [c f_\varepsilon(v^h) \chi] \, dx \qquad \forall v^h, \, \chi \in S_0^h \,.$$

We note that since $c f_\varepsilon(v^h) \in C(\bar{\Omega}^h)$ we have that $\|c f_\varepsilon(v^h)\|_{0,\infty,\Omega^h} \leq C(\varepsilon, v^h)$ for all $v^h \in S_0^h$. Moreover, we have for all $v_1^h, v_2^h, \chi \in S_0^h$,

$$\begin{aligned} |\langle \hat{\mathcal{A}}_\varepsilon^h v_1^h - \hat{\mathcal{A}}_\varepsilon^h v_2^h, \chi \rangle| &\leq \int_\Omega |\nabla (v_1^h - v_2^h) \cdot \nabla \chi)| \, dx + \int_\Omega \pi^h |c[f_\varepsilon(v_1^h) - f_\varepsilon(v_2^h)] \chi| \, dx \\ &\leq \|\nabla (v_1^h - v_2^h)\|_{0,\Omega} \, \|\nabla \chi\|_{0,\Omega} + C(\varepsilon) |v_1^h - v_2^h|_h |\chi|_h \\ &\leq C(\varepsilon) \, \|v_1^h - v_2^h\|_{1,\Omega} \, \|\chi\|_{1,\Omega} \,, \end{aligned} \tag{5.3}$$

where we have noted the Mean Value Theorem and (5.2). Note that (5.3) implies the hemicontinuity of $\hat{\mathcal{A}}_\varepsilon^h$. Moreover, similarly to $\mathcal{A}_\varepsilon$, $\hat{\mathcal{A}}_\varepsilon^h$ is strongly monotone.

On noting the (DMP) we obtain similarly as in the proof of Lemma 1.2 that for all $\varepsilon \in (0, \varepsilon_0]$ and $h > 0$,

$$0 \leq \hat{u}_{\varepsilon_0}^h \leq \hat{u}_\varepsilon^h \qquad \text{in } \bar{\Omega} \,. \tag{5.4}$$

1.5.2 Error Analysis

Our next goal is to find bounds for $|u_\varepsilon - \hat{u}_\varepsilon^h|_{1,\Omega}$ and $\|u_\varepsilon - \hat{u}_\varepsilon^h\|_{0,\infty,\Omega}$. Ideally one would want to show that the (nearly) optimal bounds of Theorem 1.1 still hold when numerical integration is involved.

In this subsection we prove error bounds for $d = 1$ and 2. Although these bounds will be seen to be suboptimal, our approach is fairly general and aims at pointing out the difficulties in the analysis. In the next subsection we will refine our analysis for $d = 1$. To simplify our analysis, we will from on assume that c is a positive constant.

We begin with some more notation. Let $\mathcal{G} \in \mathcal{L}(L^2(\Omega), H_0^1(\Omega) \cap H^2(\Omega))$ be the "inverse Laplacian" defined by

$$\int_\Omega \nabla \mathcal{G} \eta \cdot \nabla v \, dx = \int_\Omega \eta \, v \, dx \qquad \forall v \in H_0^1(\Omega) \,. \tag{5.5}$$

Let $\mathcal{G}^h \in \mathcal{L}(L^2(\Omega^h), S_0^h)$ be the "discrete inverse Laplacian" defined by

$$\int_\Omega \nabla \mathcal{G}^h \eta \cdot \nabla \chi \, dx = \int_\Omega \eta \chi \, dx \qquad \forall \chi \in S_0^h \tag{5.6}$$

and $\hat{\mathcal{G}}^h \in \mathcal{L}(C(\bar{\Omega}^h), S_0^h)$ be the "discrete inverse Laplacian in the presence of numerical integration" defined by

$$\int_\Omega \nabla \hat{\mathcal{G}}^h \eta \cdot \nabla \chi \, dx = (\eta, \chi)^h \qquad \forall \chi \in S_0^h \,. \tag{5.7}$$

It is easy to see that $u_\varepsilon = c\mathcal{G} f_\varepsilon(u_\varepsilon)$, $u_\varepsilon^h = c\mathcal{G}^h f_\varepsilon(u_\varepsilon^h)$, $\hat{u}_\varepsilon^h = c\hat{\mathcal{G}}^h f_\varepsilon(\hat{u}_\varepsilon^h)$ and

$$Q^h \mathcal{G} \equiv \mathcal{G}^h \tag{5.8}$$

and hence it follows from (4.12) and (4.13) that for all $\eta \in L^2(\Omega)$,

$$|(\mathcal{G} - \mathcal{G}^h)\eta|_{0,\Omega} + h|(\mathcal{G} - \mathcal{G}^h)\eta|_{1,\Omega} \leq Ch^2 \|\mathcal{G}\eta\|_{2,\Omega} \leq Ch^2 \|\eta\|_{0,\Omega} \,, \tag{5.9}$$

$$\|(\mathcal{G} - \mathcal{G}^h)\eta\|_{0,\infty,\Omega^h} \leq C \left(\ln \frac{1}{h}\right)^r \|(I - \pi^h)\mathcal{G}\eta\|_{0,\infty,\Omega^h} \,, \tag{5.10}$$

where $r = 0$, if $d = 1$, and $r = 1$, if $d = 2$ or 3.

We also note two technical facts. For $\chi_1, \chi_2 \in S^h$ we have that

$$\left| \int_\tau \chi_1 \chi_2 - Q_\tau(\chi_1 \chi_2) \right| \leq Ch^2 |\chi_1|_{1,\tau} |\chi_2|_{1,\tau} \,, \tag{5.11}$$

for all $\tau \in T^h$. Furthermore, under the assumption $A(0)$ one can proceed as in the proof of Lemma 1.10 to show that for all $\varepsilon \in (0, \varepsilon_1]$ and $q \in [1, \infty]$,

$$\|f_\varepsilon'(u_\varepsilon)\|_{0,q,\Omega} \leq C\varepsilon^{-(\nu + [1 - \frac{1}{q}])} \,. \tag{5.12}$$

Lemma 1.21 *Assume that $A(0)$ holds and that $c \in \mathbb{R}_+$. Then we have for sufficiently small ε and h that*

$$\begin{aligned} &|c\mathcal{G}^h f_\varepsilon(u_\varepsilon) - \hat{u}_\varepsilon^h|_{1,\Omega} \\ &\quad \leq c[|\mathcal{G}^h(I - \pi^h) f_\varepsilon(u_\varepsilon)|_{1,\Omega} + |(\mathcal{G}^h - \hat{\mathcal{G}}^h)\pi^h f_\varepsilon(u_\varepsilon)|_{1,\Omega}] + C\varepsilon^{-\frac{\nu+1}{2}} h^2 \|f_\varepsilon(u_\varepsilon)\|_{0,\Omega} \,. \end{aligned}$$

Proof. We define $\hat{e}^h := c\mathcal{G}^h f_\varepsilon(u_\varepsilon) - \hat{u}_\varepsilon^h$. On noting (5.6) and (5.7) it follows from $(\hat{P}_\varepsilon^h)$ and (WP_ε) that for all $\chi \in S_0^h$,

$$\begin{aligned} &(\nabla \hat{e}^h, \nabla \chi) - c(f_\varepsilon(u_\varepsilon) - f_\varepsilon(\hat{u}_\varepsilon^h), \chi)^h \\ &\quad = c(f_\varepsilon(u_\varepsilon), \chi) - c(f_\varepsilon(u_\varepsilon), \chi)^h = c(\nabla(\mathcal{G}^h - \hat{\mathcal{G}}^h) f_\varepsilon(u_\varepsilon), \nabla \chi) \,. \end{aligned}$$

Hence, on letting $\chi \equiv \hat{e}^h$ and noting (5.8) we have that

$$\begin{aligned} &|\hat{e}^h|_{1,\Omega}^2 - c(f_\varepsilon(u_\varepsilon) - f_\varepsilon(\hat{u}_\varepsilon^h), \pi^h u_\varepsilon - \hat{u}_\varepsilon^h)^h \\ &\quad = c(\nabla(\mathcal{G}^h - \hat{\mathcal{G}}^h) f_\varepsilon(u_\varepsilon), \nabla \hat{e}^h) + c(f_\varepsilon(u_\varepsilon) - f_\varepsilon(\hat{u}_\varepsilon^h), (Q^h - \pi^h) u_\varepsilon)^h. \end{aligned}$$

On noting (2.3), (5.2) and Young's inequality we obtain that

$$\begin{aligned} &|\hat{e}^h|^2_{1,\Omega} + c\nu^{-1}\varepsilon^{\nu+1}|f_\varepsilon(u_\varepsilon) - f_\varepsilon(\hat{u}^h_\varepsilon)|^2_h \\ &\quad \le \; c^2|(\mathcal{G}^h - \hat{\mathcal{G}}^h)f_\varepsilon(u_\varepsilon)|^2_{1,\Omega} + c\nu\varepsilon^{-(\nu+1)}|(\pi^h - Q^h)u_\varepsilon|^2_h \,. \end{aligned} \tag{5.13}$$

From (5.2), (4.2) and Lemma 1.15(i) we have that

$$\begin{aligned} &|(\pi^h - Q^h)u_\varepsilon|_h \le C\|(\pi^h - Q^h)u_\varepsilon\|_{0,\Omega^h} \\ &\quad \le \; C\,[\|(I - \pi^h)u_\varepsilon\|_{0,\Omega^h} + \|(I - Q^h)u_\varepsilon\|_{0,\Omega^h}] \le C\,h^2\|f_\varepsilon(u_\varepsilon)\|_{0,\Omega}\,. \end{aligned} \tag{5.14}$$

The desired result follows from (5.13), (5.14) and noting that $\hat{\mathcal{G}}^h = \hat{\mathcal{G}}^h\pi^h$. ∎

Lemma 1.22 *Let $d = 1$ or 2. Assume that $A(0)$ holds and that $c \in \mathbb{R}_+$. Then we have for sufficiently small ε and h that*

$$|\mathcal{G}^h(I-\pi^h)f_\varepsilon(u_\varepsilon)|_{1,\Omega} \le Ch^2\left(\ln\frac{1}{h}\right)^{\frac{d-1}{2}}|f_\varepsilon(u_\varepsilon)|_{2,1,\Omega} \le Ch^2\left(\ln\frac{1}{h}\right)^{\frac{d-1}{2}}\begin{cases} \varepsilon^{-(1+\nu)}\,, & \text{if } \nu < 1, \\ \varepsilon^{-2}\ln\frac{1}{\varepsilon}, & \text{if } \nu = 1, \\ \varepsilon^{-2\nu}\,, & \text{if } 1 < \nu. \end{cases}$$

Proof. Let $E^h = \mathcal{G}^h(I - \pi^h)f_\varepsilon(u_\varepsilon)$. From (5.6), (4.10) and (4.2) we obtain that

$$\begin{aligned} |E^h|^2_{1,\Omega} &= \int_\Omega [(I - \pi^h)f_\varepsilon(u_\varepsilon)]\,E^h\,dx \le |(I - \pi^h)f_\varepsilon(u_\varepsilon)|_{0,1,\Omega}\,\|E^h\|_{0,\infty,\Omega} \\ &\le \; C\,h^2\left(\ln\frac{1}{h}\right)^{\frac{d-1}{2}}|f_\varepsilon(u_\varepsilon)|_{2,1,\Omega}\,|E^h|_{1,\Omega}\,. \end{aligned} \tag{5.15}$$

On expanding $|f_\varepsilon(u_\varepsilon)|_{2,1,\Omega}$ (see Theorem 7.8 and 7.5 in [17]) and on noting Cauchy's inequalities in $\mathbb{R}^d$ and $L^2(\Omega)$ we obtain that

$$|f_\varepsilon(u_\varepsilon)|_{2,1,\Omega} \le C[\|f'_\varepsilon(u_\varepsilon)\|_{0,\Omega}|u_\varepsilon|_{2,\Omega} + \|f''_\varepsilon(u_\varepsilon)\,|\nabla u_\varepsilon|^2\,\|_{0,1,\Omega}]\,. \tag{5.16}$$

Since $f''_\varepsilon(t) \ge 0$ for almost all $t \in \mathbb{R}$, it follows from Green's formula that

$$\begin{aligned} &\|f''_\varepsilon(u_\varepsilon)\,|\nabla u_\varepsilon|^2\,\|_{0,1,\Omega} = \int_\Omega f''_\varepsilon(u_\varepsilon)\,|\nabla u_\varepsilon|^2\,dx \\ &= \int_\Omega \nabla f'_\varepsilon(u_\varepsilon)\cdot\nabla u_\varepsilon\,dx = \int_{\partial\Omega} f'_\varepsilon(u_\varepsilon)\frac{\partial u_\varepsilon}{\partial n}\,ds - \int_\Omega f'_\varepsilon(u_\varepsilon)\Delta u_\varepsilon\,dx \\ &\le \int_{\partial\Omega} f'_\varepsilon(u_\varepsilon)\frac{\partial u_\varepsilon}{\partial n}\,ds \le \|f'_\varepsilon(u_\varepsilon)\|_{0,\infty,\partial\Omega}\int_{\partial\Omega}|\frac{\partial u_\varepsilon}{\partial n}|\,ds\,, \end{aligned}$$

where we have noted that $f'_\varepsilon(u_\varepsilon)$, $\Delta u_\varepsilon \le 0$. The (HMP) yields that $\frac{\partial u_\varepsilon}{\partial n}(x) < 0$ for all $x \in \partial\Omega$. Hence, we can again use Green's formula and obtain that

$$\begin{aligned} \|f''_\varepsilon(u_\varepsilon)\,|\nabla u_\varepsilon|^2\,\|_{0,1,\Omega} &\le \|f'_\varepsilon(u_\varepsilon)\|_{0,\infty,\partial\Omega}\int_\Omega -\Delta u_\varepsilon\,dx \\ &\le c\,\|f'_\varepsilon(u_\varepsilon)\|_{0,\infty,\partial\Omega}\|f_\varepsilon(u_\varepsilon)\|_{0,1,\Omega}\,. \end{aligned} \tag{5.17}$$

On noting elliptic regularity we obtain from (5.16) and (5.17) that

$$|f_\varepsilon(u_\varepsilon)|_{2,1,\Omega} \leq C\,[\|f'_\varepsilon(u_\varepsilon)\|_{0,\Omega}\|f_\varepsilon(u_\varepsilon)\|_{0,\Omega} + \|f'_\varepsilon(u_\varepsilon)\|_{0,\infty,\partial\Omega}\|f_\varepsilon(u_\varepsilon)\|_{0,1,\Omega}]\,. \tag{5.18}$$

The desired result follows from (5.15), (5.18) and the bounds of Lemma 1.10 and (5.12). ■

Since we believe that the bounds of Lemma 1.10 are optimal, we also believe that the bound on $|f_\varepsilon(u_\varepsilon)|_{2,1,\Omega}$ in Lemma 1.22 is optimal. To see this, consider the following simple identities for $d = 1$, where $\Omega = (0,1)$.

$$\begin{aligned} |f_\varepsilon(u_\varepsilon)|_{2,1,\Omega} &= \int_0^1 [f_\varepsilon(u_\varepsilon)]''\,dx = [f_\varepsilon(u_\varepsilon)]'(1) - [f_\varepsilon(u_\varepsilon)]'(0) = f'_\varepsilon(0)\,[u'_\varepsilon(1) - u'_\varepsilon(0)] \\ &= f'_\varepsilon(0)\int_0^1 u''_\varepsilon(x)\,dx = -f'_\varepsilon(0)\int_0^1 c f_\varepsilon(u_\varepsilon)(x)\,dx = C\varepsilon^{-\nu+1}|f_\varepsilon(u_\varepsilon)|_{0,1,\Omega}\,. \end{aligned}$$

On noting Lemma 1.10 we obtain the given bound.

The proof of Lemma 1.22 also shows why we have only dealt with the simplest case $c \equiv$ constant. The transfer of the proof above to the more general case $A(0)$ is complicated: To obtain bounds for $|cf_\varepsilon(u_\varepsilon)|_{2,1,\Omega}$ one has to bound $\|\,|\nabla c|\,f_\varepsilon(u_\varepsilon)\|_{0,q,\Omega}$, $\|cf'_\varepsilon(u_\varepsilon)\|_{0,q,\Omega}$ and $\|\,|\nabla c|\,f'_\varepsilon(u_\varepsilon)\|_{0,q,\Omega}$ for suitable q and $\|\frac{\partial c}{\partial x_i \partial x_j} f_\varepsilon(u_\varepsilon)\|_{0,1,\Omega}$. Although one can do this similarly to Lemma 1.10, this would not only require additional assumptions on c but also lead to a very complicated error bound, since many different cases would have to be distinguished.

Lemma 1.23 *Assume that $A(0)$ holds and that $c \in \mathbb{R}_+$. Then we have for sufficiently small ε and h, satisfying $\varepsilon \geq \hat{C}h^{\frac{2}{1+\nu}}$ for arbitrarily chosen $\hat{C}$, that*

$$\begin{aligned} &|(\mathcal{G}^h - \hat{\mathcal{G}}^h)\pi^h f_\varepsilon(u_\varepsilon)|_{1,\Omega} \\ &\leq\ C\varepsilon^{-\frac{\nu+1}{2}}h^2[\|f_\varepsilon(u_\varepsilon)\|_{0,\Omega} + \varepsilon^{-\frac{\nu}{2}}\|f_\varepsilon(u_\varepsilon)\|^{\frac{1}{2}}_{0,1,\Omega}] \leq C\,h^2 \begin{cases} \varepsilon^{-\frac{1}{2}(1+2\nu)}\,, & \text{if } \nu < 1\,, \\ \varepsilon^{-\frac{3}{2}}\left(\ln\frac{1}{\varepsilon}\right)^{\frac{1}{2}}, & \text{if } \nu = 1\,, \\ \varepsilon^{-\frac{3\nu}{2}}\,, & \text{if } 1 < \nu\,. \end{cases} \end{aligned}$$

Proof. From (5.6), (5.7) and (5.11) we obtain that

$$|(\mathcal{G}^h - \hat{\mathcal{G}}^h)\pi^h f_\varepsilon(u_\varepsilon)|_{1,\Omega} \leq C\,h^2\,|\pi^h f_\varepsilon(u_\varepsilon)|_{1,\Omega}\,. \tag{5.19}$$

To bound $|\pi^h f_\varepsilon(u_\varepsilon)|_{1,\Omega}$, we set $\tilde{f}_\varepsilon(s) := \varepsilon^{-\nu} - f_\varepsilon(s)$, $\forall s \in \mathbb{R}$. We have that $\tilde{f}_\varepsilon(0) = 0$ and $0 \leq \tilde{f}'_\varepsilon(s) \leq \nu\varepsilon^{-(\nu+1)}$, $\forall s \in \mathbb{R}$. It now follows from (4.8) that

$$\begin{aligned} |\pi^h f_\varepsilon(u_\varepsilon)|^2_{1,\Omega} = |\pi^h \tilde{f}_\varepsilon(u_\varepsilon)|^2_{1,\Omega} &\leq \nu\varepsilon^{-(\nu+1)}\,(\nabla\pi^h u_\varepsilon, \nabla\pi^h \tilde{f}_\varepsilon(u_\varepsilon)) \\ &= -\nu\varepsilon^{-(\nu+1)}\,(\nabla\pi^h u_\varepsilon, \nabla\pi^h f_\varepsilon(u_\varepsilon))\,. \end{aligned} \tag{5.20}$$

Next we have that

$$-(\nabla\pi^h u_\varepsilon, \nabla\pi^h f_\varepsilon(u_\varepsilon)) = (\nabla(u_\varepsilon - \pi^h u_\varepsilon), \nabla\pi^h f_\varepsilon(u_\varepsilon)) - (\nabla u_\varepsilon, \nabla\pi^h f_\varepsilon(u_\varepsilon))$$
$$\leq Ch\,\|f_\varepsilon(u_\varepsilon)\|_{0,\Omega}\,|\pi^h f_\varepsilon(u_\varepsilon)|_{1,\Omega} + (\Delta u_\varepsilon, \pi^h f_\varepsilon(u_\varepsilon)) - \int_{\partial\Omega} \frac{\partial u_\varepsilon}{\partial n}\pi^h f_\varepsilon(u_\varepsilon)\,ds\,, \quad (5.21)$$

where we have noted (4.2), Lemma 1.15(i) and have used Cauchy's inequality and Green's formula. On noting that $\Delta u_\varepsilon \leq 0$, $\pi^h f_\varepsilon(u_\varepsilon) \geq 0$ and noting Young's inequality and again Green's formula, it follows from (5.20) and (5.21) that

$$\varepsilon^{\nu+1}|\pi^h f_\varepsilon(u_\varepsilon)|^2_{1,\Omega} \leq Ch^2\varepsilon^{-(\nu+1)}\|f_\varepsilon(u_\varepsilon)\|^2_{0,\Omega} + C\varepsilon^{-\nu}\int_\Omega -\Delta u_\varepsilon\,dx\,. \quad (5.22)$$

Hence, if $\varepsilon \geq \hat{C}h^{\frac{2}{1+\nu}}$, the assertion follows from (5.19) and (5.22) on noting the bounds of Lemma 1.10. ∎

On comparing the bounds of the Lemmas 1.21–1.23 we obtain that the bounds in Lemma 1.21 (for $d = 1$ or 2) are dominated by the bound on $|\mathcal{G}^h(I - \pi^h)f_\varepsilon(u_\varepsilon)|_{1,\Omega}$, i.e.

$$|c\mathcal{G}^h f_\varepsilon(u_\varepsilon) - \hat{u}^h_\varepsilon|_{1,\Omega} \leq C\,h^2\left(\ln\frac{1}{h}\right)^{\frac{d-1}{2}}\begin{cases}\varepsilon^{-(1+\nu)}\,, & \text{if } \nu < 1\,,\\ \varepsilon^{-2}\ln\frac{1}{\varepsilon}, & \text{if } \nu = 1\,,\\ \varepsilon^{-2\nu}\,, & \text{if } 1 < \nu\,,\end{cases} \quad \text{if } \varepsilon \geq \hat{C}h^{\frac{2}{1+\nu}}\,. \quad (5.23)$$

Theorem 1.2 *Let $d = 1$ or 2. Assume that $A(0)$ holds and that $c \in \mathbb{R}_+$. We then have for arbitrarily chosen $\hat{C}$ and h sufficiently small that*

$$(i) \quad |u - \hat{u}^h_\varepsilon|_{1,\Omega} \leq C\left(\ln\frac{1}{h}\right)^{\frac{d-1}{2}}\begin{cases} h^{\frac{4-2\nu}{4+\nu}}\,, & \text{if } \varepsilon = \hat{C}h^{\frac{4}{4+\nu}} \text{ and } 0 < \nu < 1,\\ h^{\frac{2}{5}}\ln\frac{1}{h}\,, & \text{if } \varepsilon = \hat{C}h^{\frac{4}{5}} \text{ and } \nu = 1,\\ h^{\frac{6-2\nu}{3+7\nu}}\,, & \text{if } \varepsilon = \hat{C}h^{\frac{8}{3+7\nu}} \text{ and } 1 < \nu < 3,\end{cases}$$

$$(ii) \quad \|u - \hat{u}^h_\varepsilon\|_{0,\infty,\Omega^h} \leq C\begin{cases} h_d^{\frac{2}{2+\nu}}\,, & \text{if } \varepsilon = \hat{C}h_d^{\frac{2}{2+\nu}} \text{ and } 0 < \nu < 1,\\ h_d^{\frac{2}{3}}\ln\frac{1}{h_d}\,, & \text{if } \varepsilon = \hat{C}h_d^{\frac{2}{3}} \text{ and } \nu = 1,\\ h_d^{\frac{2}{1+2\nu}}\,, & \text{if } \varepsilon = \hat{C}h_d^{\frac{2}{1+2\nu}} \text{ and } 1 < \nu.\end{cases}$$

Proof. From equating the bounds of Lemma 1.9(i), (5.9), Lemma 1.10 and (5.23) we obtain that

$$|u - \hat{u}^h_\varepsilon|_{1,\Omega} \leq |u - u_\varepsilon|_{1,\Omega} + c|(\mathcal{G} - \mathcal{G}^h)f_\varepsilon(u_\varepsilon)|_{1,\Omega} + |c\mathcal{G}^h f_\varepsilon(u_\varepsilon) - \hat{u}^h_\varepsilon|_{1,\Omega}$$
$$\leq C\begin{cases}\varepsilon^{1-\frac{\nu}{2}}, & \text{if } 0 < \nu \leq 1\\ \varepsilon^{\frac{3-\nu}{4}}\,, & \text{if } 1 < \nu < 3\end{cases} + Ch\begin{cases}1\,, & \text{if } 0 < \nu < \frac{1}{2}\\ \left(\ln\frac{1}{\varepsilon}\right)^{\frac{1}{2}}, & \text{if } \nu = \frac{1}{2}\\ \varepsilon^{\frac{1}{2}-\nu}\,, & \text{if } \frac{1}{2} < \nu\end{cases}$$
$$+ Ch^2\left(\ln\frac{1}{h}\right)^{\frac{d-1}{2}}\begin{cases}\varepsilon^{-(1+\nu)}\,, & \text{if } 0 < \nu < 1\\ \varepsilon^{-2}\ln\frac{1}{\varepsilon}\,, & \text{if } \nu = 1\\ \varepsilon^{-2\nu}\,, & \text{if } 1 < \nu\end{cases} \quad (5.24)$$

from which (i) follows on equating ε and h.

It follows from (5.8) and Lemma 1.15(iii) that

$$\|(\mathcal{G}-\mathcal{G}^h)f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega^h} \le C\, h_d^2 \varepsilon^{-\nu}\,. \tag{5.25}$$

Hence, we obtain from Lemma 1.9(ii), (5.25) and (5.23) on noting (4.10) that

$$\begin{aligned}\|u-\hat{u}_\varepsilon^h\|_{0,\infty,\Omega^h} &\le \|u-u_\varepsilon\|_{0,\infty,\Omega} + c\|(\mathcal{G}-\mathcal{G}^h)f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega^h} + \|c\mathcal{G}^h f_\varepsilon(u_\varepsilon)-\hat{u}_\varepsilon^h\|_{0,\infty,\Omega^h}\\ &\le\ C\varepsilon + Ch_d^2\varepsilon^{-\nu} + Ch^2\left(\ln\frac{1}{h}\right)^{d-1}\begin{cases}\varepsilon^{-(1+\nu)}\,, & \text{if } 0<\nu<1\,,\\ \varepsilon^{-2}\ln\frac{1}{\varepsilon}\,, & \text{if } \nu=1\,,\\ \varepsilon^{-2\nu}\,, & \text{if } 1<\nu\,.\end{cases}\end{aligned}$$

(ii) follows on equating ε and h_d. ∎

Clearly, the bounds of Theorem 1.2 are suboptimal compared to those of Theorem 1.1. We point out again that in our analysis above the discretization error is dominated by the bounds of Lemma 1.22. Hence, by our remark after Lemma 1.22 we need to improve (5.15) to improve on our error bounds. In the following subsection we will thus try to bound $\|\mathcal{G}^h(I-\pi^h)f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega^h}$ without using the discrete Sobolev embedding (4.10).

Before we do this, we note that

$$\begin{aligned}\|u_\varepsilon-\hat{u}_\varepsilon^h\|_{0,\infty,\Omega^h} &\le \|(I-\pi^h)u_\varepsilon\|_{0,\infty,\Omega^h}\\ &\quad + \|\pi^h u_\varepsilon - c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega^h} + \|c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon)-\hat{u}_\varepsilon^h\|_{0,\infty,\Omega^h}.\end{aligned} \tag{5.26}$$

On noting Lemma 1.15(ii) we have only to work on the last two terms of (5.26). In fact, the next Lemma shows that it suffices to bound $\|\pi^h u_\varepsilon - c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega^h}$.

Lemma 1.24 *Let u_ε and $\hat{u}_\varepsilon^h$ be the unique solutions of (P_ε) and $(\hat{P}_\varepsilon^h)$. Then*

$$\|c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon)-\hat{u}_\varepsilon^h\|_{0,\infty,\Omega^h} \le \|\pi^h u_\varepsilon - c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega^h}$$

Proof. We set $\xi_\varepsilon^h := \|\pi^h u_\varepsilon - c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega^h}$. Then we have that

$$c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon) - \xi_\varepsilon^h \le \pi^h u_\varepsilon \le c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon) + \xi_\varepsilon^h \tag{5.27}$$

and hence by the monotonicity of f_ε,

$$f_\varepsilon(c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon)+\xi_\varepsilon^h) \le f_\varepsilon(\pi^h u_\varepsilon) \le f_\varepsilon(c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon)-\xi_\varepsilon^h)\,. \tag{5.28}$$

Note that for fixed ε and h, ξ_ε^h is a constant and hence we obtain from (5.28) that

$$\begin{aligned}\int_\Omega \nabla(c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon)-\xi_\varepsilon^h)\cdot\nabla\chi\,dx &= c\,(f_\varepsilon(u_\varepsilon),\chi)^h = c\,(f_\varepsilon(\pi^h u_\varepsilon),\chi)^h\\ &\le c\,(f_\varepsilon(c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon)-\xi_\varepsilon^h),\chi)^h \qquad \forall\chi\in S_0^h,\ \chi\ge 0\end{aligned} \tag{5.29}$$

and similarly that

$$\int_\Omega \nabla(c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon) + \xi_\varepsilon^h) \cdot \nabla\chi \, dx \geq c\,(f_\varepsilon(c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon) + \xi_\varepsilon^h), \chi)^h \qquad \forall \chi \in S_0^h,\ \chi \geq 0\,. \tag{5.30}$$

Now we replace ξ_ε^h by $v_\varepsilon^h \in S_0^h$ defined by

$$v_\varepsilon^h(x_i) = \xi_\varepsilon^h \qquad \text{for all interior nodes } x_i\,.$$

As $\pi^h u_\varepsilon,\ \hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon) \in S_0^h$, it follows that (5.27), and hence (5.28) remain true for v_ε^h in place of ξ_ε^h. On noting (4.4) we have for all $i \in I_{int}$ that

$$\begin{aligned}\int_\Omega \nabla v_\varepsilon^h \cdot \nabla\chi_i \, dx &= \xi_\varepsilon^h \sum_{j\in I_{int}} (\nabla\chi_j, \nabla\chi_i) \\ &\geq \xi_\varepsilon^h \left[\sum_{j\in I_{int}} (\nabla\chi_j, \nabla\chi_i) + \sum_{j\in I\setminus I_{int}} (\nabla\chi_j, \nabla\chi_i) \right] \\ &= \xi_\varepsilon^h\,(\nabla 1, \nabla\chi_i) = 0\,.\end{aligned}$$

Hence we obtain similarly to (5.29) and (5.30) that

$$\int_\Omega \nabla(c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon) - v_\varepsilon^h) \cdot \nabla\chi \, dx \leq c\,(f_\varepsilon(c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon) - v_\varepsilon^h), \chi)^h \quad \forall \chi \in S_0^h,\ \chi \geq 0, \tag{5.31}$$

$$\int_\Omega \nabla(c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon) + v_\varepsilon^h) \cdot \nabla\chi \, dx \geq c(f_\varepsilon(c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon) + v_\varepsilon^h), \chi)^h \quad \forall \chi \in S_0^h,\ \chi \geq 0. \tag{5.32}$$

On noting (DMP) it follows from (5.31) and (5.32) that

$$c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon) - v_\varepsilon^h \leq \hat{u}_\varepsilon^h \leq c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon) + v_\varepsilon^h\,,$$

which implies the desired result, since $v_\varepsilon^h \leq \xi_\varepsilon^h$. ∎

Next we write

$$\|\pi^h u_\varepsilon - c\hat{\mathcal{G}}^h f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega^h} \leq \|u_\varepsilon - \pi^h u_\varepsilon\|_{0,\infty,\Omega^h} + c\,\|(\mathcal{G} - \hat{\mathcal{G}}^h) f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega^h} \tag{5.33}$$

and split the second term on the right hand side again

$$(\mathcal{G} - \hat{\mathcal{G}}^h) f_\varepsilon(u_\varepsilon) \equiv (\mathcal{G} - \mathcal{G}^h) f_\varepsilon(u_\varepsilon) + \mathcal{G}^h(I - \pi^h) f_\varepsilon(u_\varepsilon) + (\mathcal{G}^h - \hat{\mathcal{G}}^h)\pi^h f_\varepsilon(u_\varepsilon)\,. \tag{5.34}$$

Lemma 1.25 *Assume that $A(0)$ holds and that $c \in \mathbb{R}_+$. Then we have for ε and h sufficiently small that*

$$\|u_\varepsilon - \hat{u}_\varepsilon^h\|_{0,\infty,\Omega^h} \leq C\varepsilon^{-\nu} h_d^2 + 2c\,[\|\mathcal{G}^h(I-\pi^h) f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega^h} + \|(\mathcal{G}^h - \hat{\mathcal{G}}^h)\pi^h f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega^h}]\,.$$

Proof. The assertion follows from (5.26), Lemma 1.15(ii), Lemma 1.24, (5.33), (5.34), and (5.25). ∎

1.5.3 (Nearly) Optimal Error Bounds for $d = 1$

In this subsection we consider the special case $d = 1$. Our aim is to improve on the bounds of Theorem 1.2 and to recover the bounds of Theorem 1.1 as far as possible. First we will deal with the L^∞ error bounds, i.e. with the terms $\|\mathcal{G}^h(I - \pi^h)f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega^h}$ and $\|(\mathcal{G}^h - \hat{\mathcal{G}}^h)\pi^h f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega^h}$ of Lemma 1.25. In the light of our remark after Theorem 1.2 we will then also be able to improve on the H^1 error bound. As before we assume w.l.o.g. that in one dimension $\Omega \equiv (0,1)$.

The crucial ingredient of our approach is the known simple behaviour of the Green's function of $-\Delta$ for $d = 1$. Moreover, our analysis will implicitly make use of the fact that the main contributions to the error occur in the vicinity of the boundary $\partial\Omega$, e.g. see p.655 in [4].

Given a quasi-uniform mesh on Ω, i.e. a set $\{x_j : j = 0 \to N+1$ such that $0 = x_0 < x_1 < \ldots < x_{N+1} = 1\}$, we use the following notation

$$I_{j-\frac{1}{2}} := (x_{j-1}, x_j),\ h_{j-\frac{1}{2}} := x_j - x_{j-1} \text{ for } j = 1 \to N+1, \quad \text{and } h := \max_{j=1\to N+1} h_{j-\frac{1}{2}}\,.$$

We take h to be the parameter for our meshes and assume that $h \to 0$. Note that $\Omega \equiv \Omega^h$.

It is easily deduced that for y a mesh point, $g_y^h \in S_0^h$ defined by

$$g_y^h(x) := \begin{cases} x(1-y)\,, & \text{if } 0 \le x \le y\,, \\ (1-x)y\,, & \text{if } y \le x \le 1 \end{cases} \tag{5.35}$$

satisfies

$$\int_\Omega g_y^{h\prime}\,\chi'\,dx = \chi(y) \qquad \forall \chi \in S_0^h\,. \tag{5.36}$$

Moreover, for all mesh points $y \in \Omega$ we have that

$$|g_y^h|_{1,\infty,\Omega} \le 1\,. \tag{5.37}$$

Also note for future reference that

$$[f_\varepsilon(u_\varepsilon)]'' = f_\varepsilon'(u_\varepsilon)u_\varepsilon'' + f_\varepsilon''(u_\varepsilon)(u_\varepsilon')^2 = -cf_\varepsilon'(u_\varepsilon)f_\varepsilon(u_\varepsilon) + f_\varepsilon''(u_\varepsilon)(u_\varepsilon')^2 \ge 0\,. \tag{5.38}$$

Lemma 1.26 *Let $d = 1$. For ε and h sufficiently small we have that*

$$\|\mathcal{G}^h(I - \pi^h)f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega} \le C\varepsilon^{-\nu}h^2\,.$$

Proof. We have from (5.6) and (5.36) that there exists a mesh point $y \in \Omega$ such that

$$\begin{aligned}\|\mathcal{G}^h(I - \pi^h)f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega} &= |[\mathcal{G}^h(I - \pi^h)f_\varepsilon(u_\varepsilon)](y)| = |\int_\Omega g_y^h\,(I - \pi^h)f_\varepsilon(u_\varepsilon)\,dx| \\ &\le \sum_{j=1}^{N+1} \|(I - \pi^h)f_\varepsilon(u_\varepsilon)\|_{0,1,I_{j-\frac{1}{2}}} \|g_y^h\|_{0,\infty,I_{j-\frac{1}{2}}} \end{aligned} \tag{5.39}$$

From (5.35) we have that $\|g_y^h\|_{0,\infty,I_{j-\frac{1}{2}}} \le Ch$ for $j = 1$ and $N+1$ and hence, that

$$\begin{aligned} &\|(I-\pi^h)f_\varepsilon(u_\varepsilon)\|_{0,1,I_{j-\frac{1}{2}}}\,\|g_y^h\|_{0,\infty,I_{j-\frac{1}{2}}} \\ &\quad\le\ Ch^2\|(I-\pi^h)f_\varepsilon(u_\varepsilon)\|_{0,\infty,I_{j-\frac{1}{2}}} \le C\varepsilon^{-\nu}h^2 \qquad \text{for } j=1 \text{ and } N+1\,. \end{aligned} \tag{5.40}$$

Since $\mathcal{G}^h(I-\pi^h)f_\varepsilon(u_\varepsilon) \in S_0^h$ and y is a mesh point, there exists $J_0 \in \{1,\dots,N\}$ such that $y = x_{J_0} \in (0,1)$. From (5.35) we have that

$$\|g_y^h\|_{0,\infty,I_{j-\frac{1}{2}}} \le \begin{cases} (1-y)(x+h) & \text{for all } x \in I_{j-\frac{1}{2}}\,, \text{ if } j \le J_0\,, \\ y(1-x+h) & \text{for all } x \in I_{j-\frac{1}{2}}\,, \text{ if } J_0 < j\,. \end{cases} \tag{5.41}$$

(5.41) and the quasi-uniformity of the mesh yield that

$$\|g_y^h\|_{0,\infty,I_{j-\frac{1}{2}}} \le C \begin{cases} (1-y)x & \text{for all } x \in I_{j-\frac{1}{2}}\,, \text{ if } 2 \le j \le J_0\,, \\ y(1-x) & \text{for all } x \in I_{j-\frac{1}{2}}\,, \text{ if } J_0 < j \le N\,. \end{cases} \tag{5.42}$$

(5.42), (4.2) and (5.38) yield that

$$\begin{aligned} \sum_{j=2}^{J_0} \|(I-\pi^h)f_\varepsilon(u_\varepsilon)\|_{0,1,I_{j-\frac{1}{2}}}\,\|g_y^h\|_{0,\infty,I_{j-\frac{1}{2}}} &\le Ch^2(1-y)\int_{x_1}^{y} x\,[f_\varepsilon(u_\varepsilon)]''(x)\,dx \\ &\le\ Ch^2(1-y)\int_0^y x\,[f_\varepsilon(u_\varepsilon)]''(x)\,dx\,. \end{aligned} \tag{5.43}$$

Similarly, we have that

$$\sum_{j=J_0+1}^{N} \|(I-\pi^h)f_\varepsilon(u_\varepsilon)\|_{0,1,I_{j-\frac{1}{2}}}\,\|g_y^h\|_{0,\infty,I_{j-\frac{1}{2}}} \le Ch^2 y\int_y^1 (1-x)\,[f_\varepsilon(u_\varepsilon)]''(x)\,dx\,. \tag{5.44}$$

Hence, on noting (5.43) and (5.44) we have that

$$\sum_{j=2}^{N} \|(I-\pi^h)f_\varepsilon(u_\varepsilon)\|_{0,1,I_{j-\frac{1}{2}}}\,\|g_y^h\|_{0,\infty,I_{j-\frac{1}{2}}} \le Ch^2\,\Lambda\,, \tag{5.45}$$

where

$$\Lambda := \left[(1-y)\int_0^y x\,[f_\varepsilon(u_\varepsilon)]''\,dx + y\int_y^1 (1-x)\,[f_\varepsilon(u_\varepsilon)]''\,dx\right].$$

Integration by parts yields that

$$\Lambda = f_\varepsilon(0) - f_\varepsilon(u_\varepsilon(y)) \le C\varepsilon^{-\nu}\,. \tag{5.46}$$

The desired result follows from (5.39), (5.40) and (5.45) on noting (5.46). ∎

(HMP) yields that $u_\varepsilon'(0) > 0$ and $u_\varepsilon'(1) < 0$. From (P_ε) we obtain that $u_\varepsilon'' < 0$ in $[0,1]$ and that u_ε is symmetric about $\xi = 1/2$. Hence, for all $\varepsilon > 0$ we have that

$$u_\varepsilon' > 0 \quad \text{in } [0,\tfrac{1}{2})\,, \qquad u_\varepsilon'(\tfrac{1}{2}) = 0\,, \qquad u_\varepsilon' < 0 \quad \text{in } (\tfrac{1}{2},1]\,. \tag{5.47}$$

From this and the Lemmas 1.2 and 1.4 we obtain for all $\varepsilon \in (0, \varepsilon_0]$ and $x \in [\frac{1}{4}, \frac{3}{4}]$ that

$$|u_\varepsilon'(x)| = \left| u_\varepsilon'(\tfrac{1}{2}) - c \int_{\frac{1}{2}}^{x} f_\varepsilon(u_\varepsilon)\, dx \right| = c \operatorname{sign}(\tfrac{1}{2} - x) \int_x^{\frac{1}{2}} f_\varepsilon(u_\varepsilon)\, dx \le C\,. \tag{5.48}$$

Lemma 1.27 *Let $d = 1$. For ε and h sufficiently small we have that*

$$\|(\mathcal{G}^h - \hat{\mathcal{G}}^h)\pi^h f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega} \le C\varepsilon^{-\nu}h^2\,.$$

Proof. From (5.6), (5.7), (5.36), (5.1), (4.2), and (5.37) we have that there exists a mesh point $y \in \Omega$ such that

$$\begin{aligned}
&\|(\mathcal{G}^h - \hat{\mathcal{G}}^h)\pi^h f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega} \\
&\quad = |[(\mathcal{G}^h - \hat{\mathcal{G}}^h)\pi^h f_\varepsilon(u_\varepsilon)](y)| = \Big| \int_\Omega [\pi^h f_\varepsilon(u_\varepsilon)] g_y^h\, dx - (\pi^h f_\varepsilon(u_\varepsilon), g_y^h)^h \Big| \\
&\quad \le \sum_{j=1}^{N+1} \|(I - \pi^h)[g_y^h \pi^h f_\varepsilon(u_\varepsilon)]\|_{0,1,I_{j-\frac{1}{2}}} \le C \sum_{j=1}^{N+1} h_{j-\frac{1}{2}}^2 |\pi^h f_\varepsilon(u_\varepsilon)|_{1,I_{j-\frac{1}{2}}} |g_y^h|_{1,I_{j-\frac{1}{2}}} \\
&\quad \le C \sum_{j=1}^{N+1} h_{j-\frac{1}{2}}^3 |\pi^h f_\varepsilon(u_\varepsilon)|_{1,\infty,I_{j-\frac{1}{2}}} = C \sum_{j=1}^{N+1} h_{j-\frac{1}{2}}^2 |f_\varepsilon(u_\varepsilon(x_j)) - f_\varepsilon(u_\varepsilon(x_{j-1}))|.
\end{aligned} \tag{5.49}$$

For h sufficiently small there exist $J_1, J_2 \in \{1, \ldots, N\}$ such that $x_{J_1} \in (\frac{1}{4}, \frac{1}{2})$ and $x_{J_2} \in (\frac{1}{2}, \frac{3}{4})$. From (5.47) we obtain that

$$\sum_{j=1}^{J_1} h_{j-\frac{1}{2}}^2 |f_\varepsilon(u_\varepsilon(x_j)) - f_\varepsilon(u_\varepsilon(x_{j-1}))| \le C\, h^2 [f_\varepsilon(u_\varepsilon(x_0)) - f_\varepsilon(u_\varepsilon(x_{J_1}))] \le C\varepsilon^{-\nu}h^2 \tag{5.50}$$

and analogously that

$$\sum_{j=J_2+1}^{N} h_{j-\frac{1}{2}}^2 |f_\varepsilon(u_\varepsilon(x_j)) - f_\varepsilon(u_\varepsilon(x_{j-1}))| \le C\varepsilon^{-\nu}h^2\,. \tag{5.51}$$

Finally, by the Mean Value Theorem, the Lemmas 1.2 and 1.4 and (5.48) we obtain that there exist $\zeta_{j-\frac{1}{2}} \in I_{j-\frac{1}{2}}$, $j = J_1 + 1 \to J_2$, such that

$$\sum_{j=J_1+1}^{J_2} h_{j-\frac{1}{2}}^2 |f_\varepsilon(u_\varepsilon(x_j)) - f_\varepsilon(u_\varepsilon(x_{j-1}))| \le \sum_{j=J_1+1}^{J_2} h_{j-\frac{1}{2}}^3 |f_\varepsilon'(u_\varepsilon(\zeta_{j-\frac{1}{2}}))|\, |u_\varepsilon'(\zeta_{j-\frac{1}{2}})| \le Ch^2. \tag{5.52}$$

The desired result follows from (5.49)–(5.52). ∎

We can now summarize our results for $d = 1$.

Theorem 1.3 *Let $d = 1$. Assume that $A(0)$ holds and that $c \in \mathbb{R}_+$. We then have for arbitrarily chosen $\hat{C}$ and h sufficiently small that*

$$(i)\quad |u - \hat{u}_\varepsilon^h|_{1,\Omega} \leq C \begin{cases} h & , \text{ if } \hat{C}h^{\frac{4}{3+\nu}} \leq \varepsilon \leq \hat{C}h^{\frac{2}{2-\nu}} \text{ and } 0 < \nu \leq \frac{1}{3}, \\ h^{\frac{2(2-\nu)}{3+\nu}} & , \text{ if } \varepsilon = \hat{C}h^{\frac{4}{3+\nu}} \text{ and } \frac{1}{3} < \nu < \frac{1}{2}, \\ h^{\frac{6}{7}} \left(\ln \frac{1}{h}\right)^{\frac{1}{2}} & , \text{ if } \varepsilon = \hat{C}h^{\frac{8}{7}} \text{ and } \nu = \frac{1}{2}, \\ h^{\frac{2(2-\nu)}{3+\nu}} & , \text{ if } \varepsilon = \hat{C}h^{\frac{4}{3+\nu}} \text{ and } \frac{1}{2} < \nu < 1, \\ h^{\frac{1}{2}} \left(\ln \frac{1}{h}\right)^{\frac{1}{2}} & , \text{ if } \varepsilon = \hat{C}h \text{ and } \nu = 1, \\ h^{\frac{3-\nu}{1+3\nu}} & , \text{ if } \varepsilon = \hat{C}h^{\frac{4}{1+3\nu}} \text{ and } 1 < \nu < 3, \end{cases}$$

$$(ii)\quad \|u - \hat{u}_\varepsilon^h\|_{0,\infty,\Omega} \leq C\, h^{\frac{2}{1+\nu}}, \text{ if } \varepsilon = \hat{C}h^{\frac{2}{1+\nu}} \text{ and } 0 < \nu.$$

Proof. (i) On noting (5.15), Lemma 1.22 and Lemma 1.26 we obtain that

$$\begin{aligned} |\mathcal{G}^h(I - \pi^h) f_\varepsilon(u_\varepsilon)|_{1,\Omega}^2 &\leq \|(I - \pi^h) f_\varepsilon(u_\varepsilon)\|_{0,1,\Omega} \|\mathcal{G}^h(I - \pi^h) f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega} \\ &\leq C\, h^4 \begin{cases} \varepsilon^{-(1+2\nu)}, & \text{if } \nu < 1, \\ \varepsilon^{-3} \ln \frac{1}{\varepsilon}, & \text{if } \nu = 1, \\ \varepsilon^{-3\nu}, & \text{if } 1 < \nu. \end{cases} \end{aligned} \tag{5.53}$$

Hence, we have from (5.24), Lemma 1.9(i), Lemma 1.21, Lemma 1.10, (5.53) and Lemma 1.23 that

$$\begin{aligned} |u - \hat{u}_\varepsilon^h|_{1,\Omega} &\leq |u - u_\varepsilon|_{1,\Omega} + C\varepsilon^{-\frac{\nu+1}{2}} h^2 \|f_\varepsilon(u_\varepsilon)\|_{0,\Omega} \\ &\quad + c[|(\mathcal{G} - \mathcal{G}^h) f_\varepsilon(u_\varepsilon)|_{1,\Omega} + |\mathcal{G}^h(I - \pi^h) f_\varepsilon(u_\varepsilon)|_{1,\Omega} + |(\mathcal{G}^h - \hat{\mathcal{G}}^h)\pi^h f_\varepsilon(u_\varepsilon)|_{1,\Omega}] \\ &\leq C \begin{cases} \varepsilon^{1-\frac{\nu}{2}}, & \text{if } 0 < \nu \leq 1 \\ \varepsilon^{\frac{3-\nu}{4}}, & \text{if } 1 < \nu < 3 \end{cases} + Ch \begin{cases} 1, & \text{if } \nu < \frac{1}{2} \\ \left(\ln \frac{1}{\varepsilon}\right)^{\frac{1}{2}}, & \text{if } \nu = \frac{1}{2} \\ \varepsilon^{\frac{1}{2}-\nu}, & \text{if } \frac{1}{2} < \nu \end{cases} \\ &\quad + C\, h^2 \begin{cases} \varepsilon^{-\frac{1}{2}(1+2\nu)}, & \text{if } \nu < 1, \\ \varepsilon^{-\frac{3}{2}} \left(\ln \frac{1}{\varepsilon}\right)^{\frac{1}{2}}, & \text{if } \nu = 1, \\ \varepsilon^{-\frac{3\nu}{2}}, & \text{if } 1 < \nu. \end{cases} \end{aligned} \tag{5.54}$$

The desired bounds of (i) follow from (5.54), if we take ε as prescribed above.

(ii) From the Lemmas 1.9(ii) and 1.25–1.27 we obtain that

$$\begin{aligned} \|u - \hat{u}_\varepsilon^h\|_{0,\infty,\Omega} &\leq \|u - u_\varepsilon\|_{0,\infty,\Omega} + C\varepsilon^{-\nu} h^2 \\ &\quad + c[\|\mathcal{G}^h(I - \pi^h) f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega} + \|(\mathcal{G}^h - \hat{\mathcal{G}}^h)\pi^h f_\varepsilon(u_\varepsilon)\|_{0,\infty,\Omega}] \\ &\leq C\varepsilon + C\varepsilon^{-\nu} h^2, \end{aligned}$$

and the desired bound for $\varepsilon = \hat{C}h^{\frac{2}{1+\nu}}$. ∎

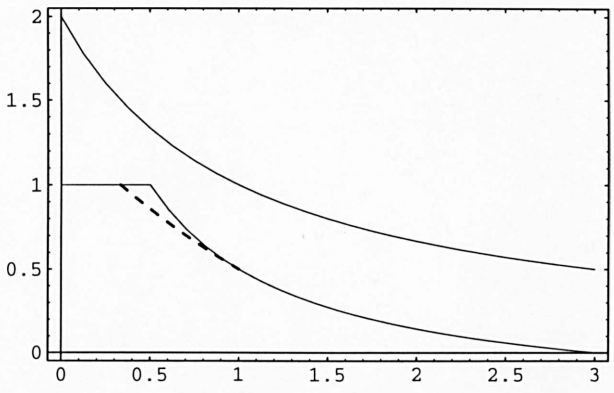


Figure 1.2: Orders of convergence according to Theorems 1.1 (——) and 1.3 (– – –).

Note that the L^∞ error bounds of Theorems 1.1 and 1.3 are the same, and that the H^1 error bounds differ only when $\frac{1}{3} < \nu \leq 1$. Figure 1.2 is an analogue to Figure 1.1. This time we compare the orders of convergence according to Theorems 1.1 and 1.3 for $\nu \in (0, 3)$. As before the upper part of the figure refers to the L^∞ error, and the lower part to the H^1 error.

1.5.4 Numerical Integration and Alternative Regularization

In this subsection we again replace f_ε by g_ε, i.e. we consider the problem:

Find $\hat{u}_\varepsilon^h \in S_0^h$ such that

$$(\hat{P}_\varepsilon^h) \qquad \int_\Omega \nabla \hat{u}_\varepsilon^h \cdot \nabla \chi \, dx = (c g_\varepsilon(\hat{u}_\varepsilon^h), \chi)^h \qquad \forall \chi \in S_0^h \,.$$

Existence and uniqueness of a solution $\hat{u}_\varepsilon^h$ to $(\hat{P}_\varepsilon^h)$ follow similarly as in the proof of Lemma 1.11. Next we note the following inequalities. For fixed $h > 0$ and $\varepsilon_0 > 0$ we have that

$$0 < \hat{u}_{\varepsilon_0}^h \quad \text{in } \Omega^h \,, \tag{5.55}$$

$$\hat{u}_{\varepsilon'}^h \leq \hat{u}_\varepsilon^h \leq \hat{u}_{\varepsilon'}^h + \varepsilon' - \varepsilon \quad \text{in } \bar{\Omega} \quad \text{for } 0 < \varepsilon < \varepsilon' \leq \varepsilon_0 \,. \tag{5.56}$$

Since the stiffness matrix K of the problem $(\hat{P}_\varepsilon^h)$ is an M matrix, we have that $K^{-1} \geq 0$ and each row of K^{-1} has at least one element strictly greater than zero. If we also note that no interior node lies on $\partial\Omega$ we obtain (5.55). (5.56) can be proved similarly as in Lemma 1.12 on noting (DMP) instead of (WMP). The non-discrete analogues of (5.55) and (5.56) were used in [12] to prove existence of a solution u to (P). We will use the above inequalities to prove existence of a solution to the following discrete unregularized problem with numerical integration.

Find $\hat{u}^h \in S_0^h$, $\hat{u}^h > 0$ in Ω^h, such that

$$(\hat{P}^h) \qquad \int_\Omega \nabla \hat{u}^h \cdot \nabla \chi \, dx = (c f(\hat{u}^h), \chi)^h \qquad \forall \chi \in S_0^h \,.$$

Lemma 1.28 *For all $h > 0$ there exists a unique solution $\hat{u}^h$ to $(\hat{P}^h)$. Moreover, for all $\varepsilon \in (0, \varepsilon_0)$ we have that*

$$0 < \hat{u}^h_{\varepsilon_0} \le \hat{u}^h_{\varepsilon} \le \hat{u}^h \le \hat{u}^h_{\varepsilon} + \varepsilon \quad \text{in } \Omega^h , \tag{5.57}$$

where $\hat{u}^h_{\varepsilon}$ denotes the unique solution of $(\hat{P}^h_{\varepsilon})$.

Proof. We fix $h > 0$ and $\varepsilon_0 > 0$, and consider the sequence $\{\hat{u}^h_{\varepsilon_0/j}\}_j$, $j = 1, 2, \ldots$, in S^h_0. By (5.56) we have that $\{\hat{u}^h_{\varepsilon_0/j}\}_j$ is a Cauchy sequence in S^h_0 with respect to the L^∞ norm. Hence, there exists a $\hat{u}^h \in S^h_0$ such that

$$\lim_{j\to\infty} \|\hat{u}^h_{\varepsilon_0/j} - \hat{u}^h\|_{0,\infty,\Omega} = 0 . \tag{5.58}$$

Obviously, (5.57) follows from (5.55), (5.56) and (5.58). Since S^h_0 is finite dimensional, we also have convergence with respect to the H^1 norm, which implies that

$$\lim_{j\to\infty} \int_\Omega \nabla \hat{u}^h_{\varepsilon_0/j} \cdot \nabla \chi \, dx = \int_\Omega \nabla \hat{u}^h \cdot \nabla \chi \, dx \qquad \forall \chi \in S^h_0 . \tag{5.59}$$

On noting $(\hat{P}^h_{\varepsilon})$ and (5.59) it is clear that $\hat{u}^h$ will be a solution of $(\hat{P}^h)$, if one can show that

$$\lim_{j\to\infty} \int_\Omega \pi^h [c g_{\varepsilon_0/j}(\hat{u}^h_{\varepsilon_0/j})\chi] \, dx = \int_\Omega \pi^h [c f(\hat{u}^h)\chi] \, dx \qquad \forall \chi \in S^h_0 . \tag{5.60}$$

Note that it suffices to prove (5.60) for the standard basis functions χ_i of S^h_0, i.e. for each fixed $i \in I_{int}$, we need to show that

$$\lim_{j\to\infty} g_{\varepsilon_0/j}(\hat{u}^h_{\varepsilon_0/j}(x_i)) = f(\hat{u}^h(x_i)) . \tag{5.61}$$

The proof of (5.61) is easy. On noting the Mean Value Theorem, (5.57) and (5.58) we have that

$$\begin{aligned} |g_{\varepsilon_0/j}(\hat{u}^h_{\varepsilon_0/j}(x_i)) - f(\hat{u}^h(x_i))| &= |f(\hat{u}^h_{\varepsilon_0/j}(x_i) + \varepsilon_0/j) - f(\hat{u}^h(x_i))| \\ &\le C\, |f'(\hat{u}^h_{\varepsilon_0}(x_i))|\, [|\hat{u}^h_{\varepsilon_0/j}(x_i) - \hat{u}^h(x_i)| + \varepsilon_0/j] \to 0 \quad \text{as } j \to \infty , \end{aligned}$$

and hence the proof is complete. ∎

It is straightforward to check that all results of subsections 5.2 and 5.3 remain valid when f_ε is replaced by g_ε. In particular we have the error bounds of the Theorems 1.2 and 1.3. Moreover, it follows from (5.57) that

$$\|\hat{u}^h - \hat{u}^h_{\varepsilon}\|_{0,\infty,\Omega} \le \varepsilon .$$

Combining this result with Theorems 1.2(ii) and 1.3(ii) we obtain that these results remain valid with $\hat{u}^h_{\varepsilon}$ replaced by $\hat{u}^h$, i.e. for $d = 1$ we have that

$$\|u - \hat{u}^h\|_{0,\infty,\Omega} \le C\, h^{\frac{2}{1+\nu}} . \tag{5.62}$$

Hence, we have obtained an a priori L^∞ error bound for a practical discretization of (P) which does not involve regularization.

1.6 Numerical Experiments

In this section we deal with the practical solution of $(\hat{P}_\varepsilon^h)$ and $(\hat{P}^h)$. We choose the Newton method to solve the nonlinear systems of equations which correspond to $(\hat{P}_\varepsilon^h)$ and $(\hat{P}^h)$. We describe this in subsection 6.1. We then report on some numerical results in the case $d = 1$ in subsection 6.2. There we will compare the results of computations done with f, f_ε and g_ε.

1.6.1 The Systems of Nonlinear Algebraic Equations

In the following we will always assume $A(0)$ and that we are given a partitioning T^h of Ω^h which satisfies the assumption (WA) of section 4.1. In practice, one considers a family $\{T^h\}_h$, and $(\hat{P}_\varepsilon^h)$ (resp. $(\hat{P}^h)$) is solved for some selected h. In this subsection we let $h, \varepsilon > 0$ be given and fixed.

We first deal with $(\hat{P}_\varepsilon^h)$ for the regularization f_ε of section 2. Choosing the standard basis functions $\{\chi_i\}_{i=1}^N$ for S_0^h as in subsection 4.1 leads to the following nonlinear system for $(\hat{P}_\varepsilon^h)$:

Find $\hat{u}_\varepsilon^h = \sum_{j=1}^N U_{\varepsilon,j}\chi_j \in S_0^h$, i.e. find $U_\varepsilon = (U_{\varepsilon,1}, \ldots, U_{\varepsilon,N})^t \in \mathbb{R}^N$, such that

$$KU_\varepsilon = \Phi_\varepsilon(U_\varepsilon)\,, \tag{6.1}$$

where

$$K := \left(\int_\Omega \nabla\chi_i \cdot \nabla\chi_j \, dx\right)_{i,j=1}^N,$$

is the stiffness matrix and the map $\Phi_\varepsilon : \mathbb{R}^N \to \mathbb{R}^N$ is given by its component functions

$$\Phi_\varepsilon^i(V) = c(x_i)\, f_\varepsilon(V_i) \int_\Omega \chi_i \, dx\,, \qquad V \in \mathbb{R}^N,\ i = 1 \to N\,. \tag{6.2}$$

(6.2) shows that Φ_ε is a continuously differentiable (with respect to V) diagonal map.

We also have that

$$Q_\varepsilon \le D\Phi_\varepsilon(V) \le \mathbf{0}\,, \qquad \forall V \in \mathbb{R}^N, \tag{6.3}$$

where $D\Phi_\varepsilon$ is the Jacobian of Φ_ε, $\mathbf{0}$ is the $N \times N$ zero matrix and Q_ε is a diagonal $N \times N$ matrix with

$$-C(\varepsilon) \le (Q_\varepsilon)_{ii}, \qquad i = 1 \to N\,, \tag{6.4}$$

where $C(\varepsilon) := [\max_{x\in\bar{\Omega}} c(x)]\nu\varepsilon^{-(\nu+1)} > 0$. As mentioned in section 4.1, K is an M matrix under assumption (WA) and so must be $K - D\Phi_\varepsilon(V)$ for all $V \in \mathbb{R}^N$ and $K - Q_\varepsilon$ (see 2.4.11 in [36]). Moreover, Φ_ε is a convex map, i.e.

$$\Phi_\varepsilon(V) - \Phi_\varepsilon(V') \ge D\Phi_\varepsilon(V')(V - V')\,, \qquad \forall V, V' \in \mathbb{R}^N. \tag{6.5}$$

Hence the Newton iteration

$$\left[K - D\Phi_\varepsilon(U_\varepsilon^k)\right] \delta U_\varepsilon^k = [\Phi_\varepsilon(U_\varepsilon^k) - KU_\varepsilon^k], \tag{6.6}$$

$$U_\varepsilon^{k+1} = U_\varepsilon^k + \delta U_\varepsilon^k \tag{6.7}$$

converges globally, i.e. for any $U_\varepsilon^0 \in \mathbb{R}^N$, towards the solution U_ε of (6.1), and the sequence of iterates has the following monotonicity property

$$U_\varepsilon^1 \leq U_\varepsilon^2 \leq \cdots \leq U_\varepsilon^k \leq U_\varepsilon . \tag{6.8}$$

(see 13.3.7 in [36], where a concave nonlinearity, i.e. $-\Phi_\varepsilon$ is convex, is assumed to prove global, monotonically decreasing convergence for (6.6), (6.7). However, the proof can easily be adapted to a convex nonlinearity.) Moreover, since f_ε, and hence Φ_ε, is in $C^{1,1}(\mathbb{R})$, the iteration converges quadratically in a neighbourhood of U_ε (see the proof of 11.2.2 in [36]).

Next, we deal with the unregularized problem $(\hat{P}^h)$. With K as above, we now have to solve the problem:

Find $\hat{u}^h = \sum_{j=1}^N U_j \chi_j \in S_0^h$, i.e. find $U = (U_1, \ldots, U_N)^t \in \mathbb{R}^N$, such that

$$KU = \Phi(U) , \tag{6.9}$$

where the map $\Phi : \mathbb{R}_+^N \to \mathbb{R}^N$ is given by the component functions

$$\Phi^i(V) = c(x_i)\, f(V_i) \int_\Omega \chi_i \, dx , \qquad V \in \mathbb{R}_+^N, \; i = 1 \to N . \tag{6.10}$$

Existence and uniqueness of U satisfying (6.9) follow from Lemma 1.28. Again we want to solve (6.9) by the Newton method:

$$[K - D\Phi(U^k)]\delta U^k = [\Phi(U^k) - KU^k], \tag{6.11}$$

$$U^{k+1} = U^k + \delta U^k . \tag{6.12}$$

Since $\Phi(V)$ is only defined for $V \in \mathbb{R}_+^N$, it is not clear that for any given $U^0 \in \mathbb{R}_+^N$ the Newton iteration (6.11), (6.12) is well defined let alone converges. Hence, we need a sufficient condition for the choice of $U^0 \in \mathbb{R}_+^N$.

Lemma 1.29 *Suppose $U^0 \in \mathbb{R}_+^N$ satisfies*

$$KU^0 \leq \Phi(U^0) . \tag{6.13}$$

Then the iterates $\{U^k\}$ given by (6.11), (6.12) satisfy for $k \geq 0$,

$$U^k \leq U^{k+1} , \tag{6.14}$$

in particular $U^k \in \mathbb{R}_+^N$, and they converge to the solution U of (6.9).

Proof. Since $U^0 \in \mathbb{R}^N_+$, $[K - D\Phi(U^0)]$ and $[\Phi(U^0) - KU^0)]$ are defined. Moreover, the former is an M matrix. Hence, it follows from (6.13), (6.11) and (6.12) that

$$U^0 \leq U^1 . \tag{6.15}$$

On noting that (6.5) remains true for $V, V' \in \mathbb{R}^N_+$, if Φ_ε is replaced by Φ, and that $[K - D\Phi(V)]$ is an M matrix for all $V \in \mathbb{R}^N_+$, one can now prove by induction that for $k \geq 0$,

$$\begin{aligned} U^k &\in \mathbb{R}^N_+ , \\ KU^k &\leq \Phi(U^k) , \\ U^k &\leq U^{k+1} , \end{aligned} \tag{6.16}$$

which proves (6.14). Next we show that for $k \geq 0$,

$$U^k \leq U . \tag{6.17}$$

We set $\delta^k := U - U^k \in \mathbb{R}^N$. On noting (6.9), (6.16) and (6.5) for Φ we obtain that

$$K\delta^k \geq \Phi(U) - \Phi(U^k) = \Phi(U^k + \delta^k) - \Phi(U^k) \geq D\Phi(U^k)\delta^k ,$$

and hence,

$$[K - D\Phi(U^k)]\delta^k \geq 0 ,$$

from which (6.17) follows. Since the sequence of iterates $\{U^k\}$ is increasing and bounded above, it converges to a $\bar{U} \in \mathbb{R}^N_+$. On taking the limit as $k \to \infty$ in (6.11) we obtain that

$$K\bar{U} = \Phi(\bar{U}) .$$

Hence, by uniqueness, $\bar{U} = U$, and the proof is complete. ∎

We now give a practical method of finding a $U^0 \in \mathbb{R}^N_+$ which satisfies (6.13). First, for a given mesh, we solve the linear problem

$$K\tilde{U}^0 = \tilde{1}^0 , \tag{6.18}$$

where $\tilde{1}^0 \in \mathbb{R}^N_+$ is given by

$$\tilde{1}^0_i = \int_\Omega \chi_i \, dx , \qquad i = 1 \to N . \tag{6.19}$$

Since K is an M matrix, we have that

$$\tilde{U}^0 \in \mathbb{R}^N_+ . \tag{6.20}$$

For $\mu > 0$ to be determined we set

$$U^0 := \mu \tilde{U}^0 . \tag{6.21}$$

By (6.20) we have that $U^0 \in \mathbb{R}^N_+$, and it follows from (6.21), (6.18), (6.19) and (6.10) that U^0 satisfies (6.13), if

$$\mu \int_\Omega \chi_i \, dx \le c(x_i) \, \mu^{-\nu} [\tilde{U}^0_i]^{-\nu} \int_\Omega \chi_i \, dx , \qquad i = 1 \to N .$$

Hence, for any

$$\mu \le \left([\min_{i=1\to N} c(x_i)] [\max_{i=1\to N} \tilde{U}^0_i]^{-\nu} \right)^{\frac{1}{1+\nu}} \tag{6.22}$$

the vector U^0 given by (6.21) satisfies (6.13).

Finally, we consider $(\hat{P}^h_\varepsilon)$ with the regularization g_ε of section 3. Analogously to Lemma 1.29 we obtain that the iterates of (6.6), (6.7) with f_ε in (6.2) replaced by g_ε are well defined and converge to the solution U_ε of (6.1), if we start the iteration with $U^0_\varepsilon \in \mathbb{R}^N$ satisfying

$$U^0_{\varepsilon,i} > -\varepsilon , \qquad i = 1 \to N \tag{6.23}$$

and

$$K U^0_\varepsilon \le \Phi_\varepsilon(U^0_\varepsilon) . \tag{6.24}$$

In practice, we set

$$U^0_\varepsilon := \mu_\varepsilon \tilde{U}^0 , \tag{6.25}$$

where $\tilde{U}^0$ satisfies (6.18), (6.19), and $\mu_\varepsilon > 0$. Again, $U^0_\varepsilon \in \mathbb{R}^N_+$, in particular (6.23) holds. Similarly to above, we obtain that U^0_ε satisfies (6.24), if

$$\mu_\varepsilon \int_\Omega \chi_i \, dx \le c(x_i) \, [\mu_\varepsilon \tilde{U}^0_i + \varepsilon]^{-\nu} \int_\Omega \chi_i \, dx , \qquad i = 1 \to N .$$

If $\mu_\varepsilon \le \varepsilon$, we have that $\varepsilon^{-\nu}[\tilde{U}^0_i + 1]^{-\nu} \le [\mu_\varepsilon \tilde{U}^0_i + \varepsilon]^{-\nu}$. Hence, if for any

$$\mu_\varepsilon \le \min\{[\min_{i=1\to N} c(x_i)] [1 + \max_{i=1\to N} \tilde{U}^0_i]^{-\nu} \varepsilon^{-\nu}, \varepsilon\} \tag{6.26}$$

the vector U^0_ε given by (6.25) satisfies (6.24).

1.6.2 Numerical Results

In this subsection we present some numerical results of the following problem in $\mathbb{R}$:

$$-u'' = cu^{-\nu} \quad \text{on } (0,1), \; u(0) = 0 \text{ and } u(1) = a , \tag{6.27}$$

where $c \in \mathbb{R}_+$ and $a \ge 0$. (6.27) is of type (P) when $a = 0$, and we have discussed its fully practical finite element approximation in section 5. If $a > 0$, i.e for inhomogeneous

Dirichlet boundary conditions, our analysis can be adapted without effecting the error bounds of Theorems 1.1, 1.2 and 1.3. The case $a > 0$ is included here, since for $\nu > 1$ it is easy to write down the exact solutions for (6.27). For example, for

$$\nu > 1\,, \qquad a = 1\,, \qquad c = 2\,\frac{\nu - 1}{(1+\nu)^2} \tag{6.28}$$

the exact solution of (6.27) is given by

$$u(x) := x^{\frac{2}{1+\nu}}\,, \qquad x \in [0,1]\,. \tag{6.29}$$

We choose equidistant meshes T^h on $[0,1]$, i.e. for fixed h there exists an $N \in \mathbb{N}$ such that $h = \frac{1}{N+1}$, and the mesh points are given by $x_i = ih$, $i = 0 \to N+1$.

Straightforward calculations show that

$$K \;=\; \frac{1}{h}\begin{pmatrix} 2 & -1 & 0 & \cdots & \cdots & 0 \\ -1 & 2 & -1 & \ddots & & \vdots \\ 0 & -1 & 2 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & -1 & 2 & -1 \\ 0 & \cdots & \cdots & 0 & -1 & 2 \end{pmatrix}$$

and that

$$\Phi^i_\varepsilon(V) \;=\; h\,c\,\phi_\varepsilon(V_i) + \delta_{iN}\,\frac{a}{h}\,, \qquad i = 1 \to N\,, \tag{6.30}$$

where $\phi_\varepsilon \equiv f_\varepsilon$ or g_ε, and

$$\Phi^i(V) \;=\; h\,c\,f(V_i) + \delta_{iN}\,\frac{a}{h}\,, \qquad i = 1 \to N\,. \tag{6.31}$$

Since we have included the case $a > 0$, we have to remark on the convergence of the Newton iterations (6.6), (6.7) and (6.11), (6.12).

First we consider (6.30) for $\phi_\varepsilon \equiv f_\varepsilon$. For any $a > 0$, (6.3)–(6.5) remain true. Hence the Newton iteration (6.6), (6.7) converges for arbitrary $U^0_\varepsilon \in \mathbb{R}^N$.

As for the convergence of (6.11), (6.12) for Φ in (6.31) with $a > 0$, we note that Lemma 1.29 remains true for this case. Hence, we only have to give an adapted practical method of finding a $U^0 \in \mathbb{R}^N_+$ which satisfies (6.13). Instead of (6.18), (6.19) we now solve

$$K\tilde{U}^a = \tilde{1}^a\,, \tag{6.32}$$

where $\tilde{1}^a \in \mathbb{R}^N_+$ is given by

$$\tilde{1}^a_i = \int_\Omega \chi_i\,dx + \delta_{iN}\frac{a}{h} = h + \delta_{iN}\frac{a}{h}\,, \qquad i = 1 \to N\,. \tag{6.33}$$

Next, we set

$$U^0 = \mu^a \tilde{U}^a . \tag{6.34}$$

It is easy to see that U^0 satisfies (6.13), if

$$\mu^a := \min\{\mu, 1\}, \tag{6.35}$$

where μ is as in (6.22) with $\tilde{U}^0$ replaced by $\tilde{U}^a$.

Similarly, one see that for (6.30) with $\phi_\varepsilon \equiv g_\varepsilon$ and $a > 0$, the Newton iteration (6.6), (6.7) converges, if we start with

$$U_\varepsilon^0 = \mu_\varepsilon^a \tilde{U}^a , \tag{6.36}$$

where $\tilde{U}_\varepsilon^a$ satisfies (6.32), (6.33) and

$$\mu_\varepsilon^a = \min\{\mu_\varepsilon, 1\} , \tag{6.37}$$

and μ_ε as in (6.26) with $\tilde{U}^0$ replaced by $\tilde{U}^a$.

We also note that for $d = 1$ we do not have to solve the systems (6.18), (6.19) and (6.32), (6.33) computationally. It is well-known, see (4.14), that

$$\tilde{U}^a = \pi^h u^a , \tag{6.38}$$

where we have used the standard identification $S_0^h \equiv \mathbb{R}^N$, and

$$u^a(x) \equiv \frac{1}{2} x(2a + 1 - x) \tag{6.39}$$

is the solution of

$$-u'' = 1 \quad \text{on } (0,1),\ u(0) = 0 \text{ and } u(1) = a .$$

We use a direct method to solve (6.6) and (6.11). As $K - D\Phi_\varepsilon(V)$ and $K - D\Phi(V)$ are M matrices, we can use Gaussian elimination without pivoting. As they are also tridiagonal, one can store them as $N \times 3$ or $3 \times N$ matrices. Algorithms which are adapted to this storage are available and solve (6.6) with $O(N)$ multiplications (as compared to $O(N^3)$ multiplications which are needed to solve a linear system directly where all the elements of the corresponding matrix are nonzero). We use the FORTRAN 77 code for such an algorithm given in [31], p.424. Alternatively, on noting that $K - D\Phi_\varepsilon(V)$ and $K - D\Phi(V)$ are also symmetric (Stieltjes) and positive definite matrices, one could use Cholesky factorization to solve the linear systems (6.6) and (6.11).

For our computations we choose $h = 2^{-l}$, i.e. $N(l) = 2^l - 1$, $l = 1 \to L$. For fixed h we stop the Newton iteration, if

$$\max_{i=1 \to N(l)} |(\delta U^k)_i| \leq Tol := 10^{-8} ,$$

where as an initial guess U^0 for the iteration we use the piecewise linear interpolant of the result for the case $N(l-1)$. If $l = 1$, i.e. $N = 1$, we start the iteration with U^0 as prescribed above. Although we did not prove convergence of the Newton iterations for this choice of the initial guesses when $N > 1$, they perform very well in practice, since the initial guesses are by 'construction' close to the sought solution.

All our computations were done on an IBM RS6000 workstation using FORTRAN 77 with *real*8* precision. We used MATHEMATICA to generate the graphs in this section.

We begin with showing some qualitative behaviour of the computational solutions for zero boundary conditions, i.e. $a = 0$. We choose the following parameters:

$$\nu \equiv \tfrac{1}{2}, \qquad c \equiv 1, \qquad \hat{C} \equiv 1, \qquad \varepsilon \equiv h^{\frac{2}{1+\nu}}, \tag{6.40}$$

where c is from (6.27) and $\hat{C}$ from Theorem 1.3.

We start the computations for $h = \frac{1}{2}$ and use as an initial guess for the Newton iterations U^0 given by (6.21). On noting (6.22), (6.26) and (6.38), (6.39) we have to choose

$$U^0 \leq 0.25 \tag{6.41}$$

for the computation with f and

$$U^0 \leq 2^{-\frac{13}{3}}. \tag{6.42}$$

for the computation with g_ε. In order to have comparable computations, we start in every case, i.e. for f, f_ε and g_ε, with $U^0 \equiv 2^{-\frac{13}{3}} \approx 0.0496$.

The Figures 1.3 – 1.5 show some piecewise linear Galerkin approximations of u for the parameters as in (6.40) and $h = 2^{-1}, 2^{-2}, 2^{-3}$ and 2^{-10}, where for f_ε and g_ε, $\varepsilon = h^{\frac{4}{3}}$ is chosen according to Theorem 5.2. The graphs for $h = 2^{-10}$ appear to be smooth in the scale of the Figures below, and graphs for smaller values of h cannot be distinguished in the same scale.

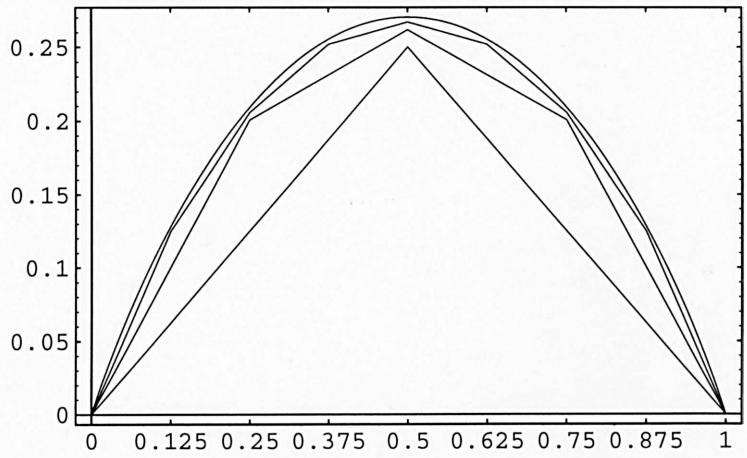


Figure 1.3: Some piecewise lin. Galerkin approx. for $\nu = 0.5$ and comput. with f

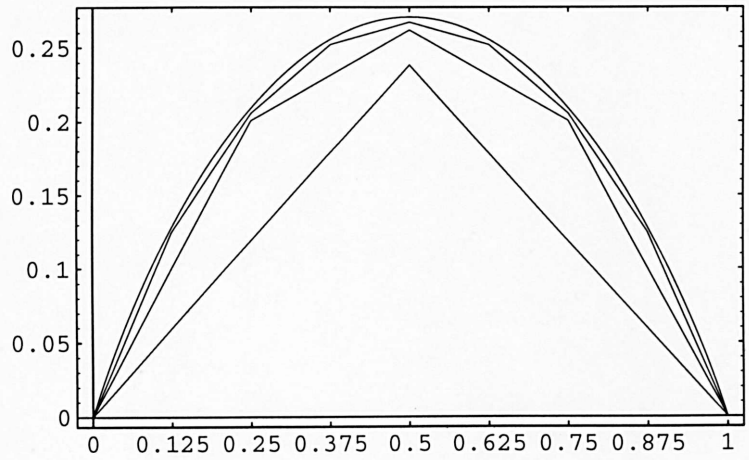


Figure 1.4: Some piecewise lin. Galerkin approx. for $\nu = 0.5$ and comput. with f_ε

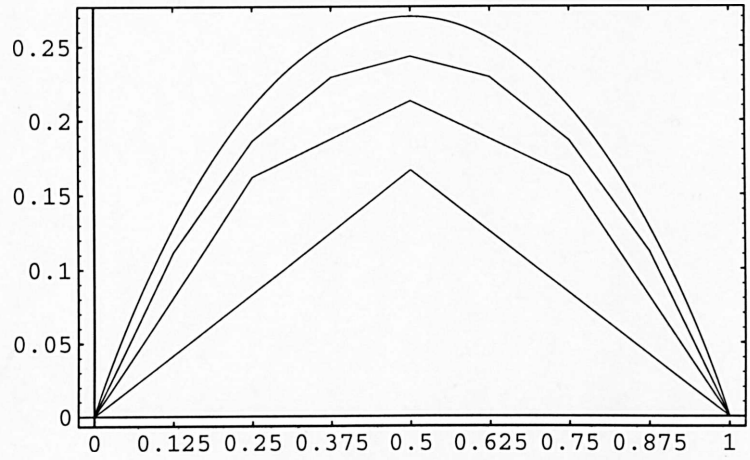


Figure 1.5: Some piecewise lin. Galerkin approx. for $\nu = 0.5$ and comput. with g_ε

Next, we compare the error estimates of Theorem 1.3 with some computational results. On using the notation $E_\infty(h) := \|\pi^h u - v\|_{0,\infty,\Omega}$, where $v \equiv \hat{u}^h$ or $\hat{u}^h_\varepsilon$, we have that

$$\|u - v\|_{0,\infty,\Omega} \leq \|u - \pi^h u\|_{0,\infty,\Omega} + E_\infty(h) \tag{6.43}$$

Clearly, exact values for $E_\infty(h)$ can be obtained computationally, if the exact solution u is known. For the interpolation error we have the bounds of (4.34) which are the same as those of Theorem 1.3(ii), resp. (5.62), when $\nu \geq 1$ and slightly better, when $0 < \nu < 1$.

On assuming that

$$E_\infty(h) \approx Ch^{p_\infty}, \quad p_\infty > 0,$$

we obtain that $2^{p_\infty} E_\infty(h) - E_\infty(2h) \approx 0$ and hence

$$p_\infty \approx p_\infty(h) := \log_2 \frac{E_\infty(2h)}{E_\infty(h)}. \tag{6.44}$$

Similarly, with $E_1(h) := |\pi^h u - v|_{1,\Omega}$ and $E_1(h) \approx Ch^{p_1}$, $p_1 > 0$, we have that

$$|u - v|_{1,\Omega} \leq |u - \pi^h u|_{1,\Omega} + E_1(h), \tag{6.45}$$

$$p_1 \approx p_1(h) := \log_2 \frac{E_1(2h)}{E_1(h)}. \tag{6.46}$$

From our discussion in section 4.4 we have that

$$|u - \pi^h u|_{1,\Omega} \leq C \begin{cases} h & , \text{ if } 0 < \nu < \frac{1}{2}, \\ h^{\frac{3}{2}-\nu-\delta} & , \text{ if } \frac{1}{2} \leq \nu < 1, \\ h^{\frac{3-\nu}{2(1+\nu)}-\delta} & , \text{ if } 1 \leq \nu < 3, \end{cases} \tag{6.47}$$

where $\delta > 0$ is arbitrary.

We need the exact solution u of (6.27) to compute $E_1(h)$ and $E_\infty(h)$. In fact, for ν and c as in (6.40) we can obtain u:

Noting the uniqueness of u it must also be symmetric about $\frac{1}{2}$, since $\bar{u}$ defined by $\bar{u}(x) := u(1-x)$, $x \in [0,1]$, solves (6.27). Hence, it suffices to solve the problem

$$-w'' = w^{-\frac{1}{2}} \quad \text{on } (0,1),\ w(0) = w'(\tfrac{1}{2}) = 0. \tag{6.48}$$

Then, u defined by

$$u(x) := \begin{cases} w(x) & , \text{ if } 0 \leq x \leq \frac{1}{2}, \\ w(1-x), & \text{if } \frac{1}{2} \leq x \leq 1 \end{cases} \tag{6.49}$$

is the solution of (6.27).

To solve (6.48) we multiply the differential equation by w' and integrate to obtain for $x \in [0, \frac{1}{2}]$ that

$$w'(x) = 2\left[w^{\frac{1}{2}}(\tfrac{1}{2}) - w^{\frac{1}{2}}(x)\right]^{\frac{1}{2}}.$$

The method of separation of variables yields that

$$\int_{w(x)}^{M} \frac{ds}{(M^{\frac{1}{2}} - s^{\frac{1}{2}})^{\frac{1}{2}}} = 1 - 2x\,, \tag{6.50}$$

where $M := w(\frac{1}{2})$. Using the substitution $t := M^{-1}s$ we obtain from (6.50) that

$$M^{\frac{3}{4}} \int_{\frac{w(x)}{M}}^{1} \frac{dt}{(1 - t^{\frac{1}{2}})^{\frac{1}{2}}} = 1 - 2x\,.$$

Hence, on noting that

$$\int \frac{dt}{(1 - t^{\frac{1}{2}})^{\frac{1}{2}}} = 4\left[\tfrac{1}{3}(1 - t^{\frac{1}{2}})^{\frac{3}{2}} - (1 - t^{\frac{1}{2}})^{\frac{1}{2}}\right]$$

we have for $x \in [0, \frac{1}{2}]$ that

$$\left[1 - \left(\frac{w(x)}{M}\right)^{\frac{1}{2}}\right]^{\frac{1}{2}} - \tfrac{1}{3}\left[1 - \left(\frac{w(x)}{M}\right)^{\frac{1}{2}}\right]^{\frac{3}{2}} = \tfrac{1}{4}M^{-\frac{3}{4}}(1 - 2x)\,. \tag{6.51}$$

On choosing $x := 0$ (6.51) yields $M = (\frac{3}{8})^{\frac{4}{3}}$. With

$$y(x) := \left[1 - \left(\frac{w(x)}{M}\right)^{\frac{1}{2}}\right]^{\frac{1}{2}} \tag{6.52}$$

(6.51) becomes a cubic equation for $y(x)$:

$$y(x) - \tfrac{1}{3}[y(x)]^3 = \tfrac{2}{3}(1 - 2x)\,. \tag{6.53}$$

(6.53) has three real roots:

$$y_k(x) := -2\cos(\tfrac{1}{3}[\cos^{-1}(1 - 2x) + 2k\pi])\,, \quad k = 0, 1, 2\,, \tag{6.54}$$

in particular,

$$y_0(\tfrac{1}{2}) = -\sqrt{3}\,, \quad y_1(\tfrac{1}{2}) = \sqrt{3}\,, \quad y_2(\tfrac{1}{2}) = 0\,. \tag{6.55}$$

(6.52) yields that $y(\frac{1}{2}) = 0$. Therefore we must choose for $k = 2$ in (6.54). Hence, we obtain from combining (6.49), (6.52) and (6.54) with $k = 2$ the exact solution u of (6.27) as

$$u(x) := (\tfrac{3}{8})^{\frac{4}{3}} \begin{cases} \{1 - 4\cos^2(\frac{1}{3}[\cos^{-1}(1 - 2x) + 4\pi])\}\,, & \text{if } 0 \le x \le \frac{1}{2} \\ \{1 - 4\cos^2(\frac{1}{3}[\cos^{-1}(2x - 1) + 4\pi])\}\,, & \text{if } \frac{1}{2} \le x \le 1\,. \end{cases}$$

Table 1.1 shows the computational orders of convergence, $p_1(h)$, $p_\infty(h)$, for the computations with f, f_ε and g_ε. The parameters were chosen as in (6.40), in particular $\varepsilon = h^{\frac{4}{3}}$, by Theorem 1.3. We also include the number of Newton iterations, n_{it}, that were needed for each h to solve (6.6) or (6.11).

	f			f_ε			g_ε		
h	n_{it}	$p_1(h)$	$p_\infty(h)$	n_{it}	$p_1(h)$	$p_\infty(h)$	n_{it}	$p_1(h)$	$p_\infty(h)$
2^{-1}	6	—	—	2	—	—	4	—	—
2^{-2}	5	.7642	1.2289	4	1.4262	1.8908	4	.6064	.8613
2^{-3}	4	.8326	1.3182	4	.8326	1.3182	4	.9225	1.0802
2^{-4}	4	.8656	1.3297	4	.8656	1.3297	4	1.0512	1.2025
2^{-5}	3	.8979	1.3493	3	.8979	1.3493	3	1.1087	1.2647
2^{-6}	3	.9252	1.3712	3	.9252	1.3712	3	1.1366	1.2971
2^{-7}	3	.9460	1.3905	3	.9460	1.3905	3	1.1505	1.3147
2^{-8}	3	.9612	1.4087	3	.9612	1.4087	3	1.1562	1.3246
2^{-9}	2	.9723	1.4233	2	.9723	1.4233	3	1.1564	1.3303
2^{-10}	2	.9802	1.4358	2	.9802	1.4358	2	1.1523	1.3337
2^{-11}	2	.9859	1.4461	2	.9859	1.4461	2	1.1445	1.3356
2^{-12}	2	.9899	1.4548	2	.9899	1.4548	2	1.1339	1.3366
2^{-13}	2	.9928	1.4620	2	.9928	1.4620	2	1.1211	1.3371
2^{-14}	2	.9949	1.4681	2	.9949	1.4681	2	1.1070	1.3373
2^{-15}	2	.9964	1.4731	2	.9964	1.4731	2	1.0924	1.3372

Table 1.1: Computational orders of convergence for $\nu = 0.5$ and $\varepsilon = h^{\frac{4}{3}}$

As for the numbers of Table 1.1, we first note that the values of $p_\infty(h)$, h small, for f and f_ε are better than the value predicted by (5.62) and Theorem 1.3(ii): $p_\infty = \frac{4}{3} \approx 1.3333$. In fact, these values suggest optimal convergence: (4.34) for $\nu = \frac{1}{2}$ yields $p_\infty^{opt} = 1.5000$. However, the values of $p_\infty(h)$ for g_ε are in good agreement with our theory.

The values of $p_1(h)$ are too good compared to our prediction from (5.54), $p_1 = \frac{2}{3} \approx 0.6667$, for f_ε and g_ε. Although we do not have a theory for the value of p_1 for f, the values of $p_1(h)$ for f (as well as those for f_ε and g_ε) in the above Table suggest again optimal convergence: (6.47) for $\nu = \frac{1}{2}$ yields $p_1^{opt} = 1$.

To conclude our discussion of Table 1.1, we now explain why the values of $p_1(h)$, $p_\infty(h)$ for f and f_ε are the same when $h \geq 2^{-3}$. The Newton iteration (6.6), (6.7) yields the computational solution U_ε of (6.1). However, from the definition of Φ_ε and f_ε we have that (6.6), (6.7) are only different from (6.11), (6.12), if at least one component of U_ε^k is less than ε. Table 1.2 shows the number n_ε of components of the computed solution U_ε which are less than ε. Since for each h the final iterates U_ε^k are close to U_ε, Table 1.2 implies that these iterates are in fact computed with f instead of f_ε for $h \geq 2^{-2}$. Hence, they converge to U, the solution of (6.9), and we obtain the same computational orders for f and f_ε in Table 1.1 for $h \geq 2^{-3}$. Obviously, the same effect cannot happen for g_ε as it is a global regularization.

h	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}	2^{-6}	2^{-7}	2^{-8}
n_ε	1	0	0	0	0	0	0	0
h	2^{-9}	2^{-10}	2^{-11}	2^{-12}	2^{-13}	2^{-14}	2^{-15}	—
n_ε	0	0	0	0	0	0	0	—

Table 1.2: Number of components of U_ε which are less than ε for $\nu = 0.5$ and f_ε.

Next we present some numerical results when the parameters ν, a and c are as in (6.28). The parameter $\hat{C}$ of Theorem 1.3 is again chosen to be equal to 1 and $\varepsilon = h^{\frac{2}{1+\nu}}$. Again we start the computation for $h = \frac{1}{2}$ but use as an initial guess U^0 given by (6.34). In order to use U^0 for the computations with f, f_ε and g_ε, we choose $\mu^a_{(\varepsilon)}$ to satisfy both (6.35) and (6.37). Table 1.3 is an analogue of Table 1.1 for $\nu = 3$.

	f			f_ε			g_ε		
h	n_{it}	$p_1(h)$	$p_\infty(h)$	n_{it}	$p_1(h)$	$p_\infty(h)$	n_{it}	$p_1(h)$	$p_\infty(h)$
2^{-1}	10	—	—	2	—	—	5	—	—
2^{-2}	5	-.0077	.2023	4	-.0709	.1394	3	-.4120	-.1889
2^{-3}	6	.0391	.4402	4	.0218	.4242	4	-.1733	.1904
2^{-4}	6	.0200	.4863	4	.0155	.4826	4	-.0701	.2476
2^{-5}	6	.0075	.4967	4	.0063	.4958	4	-.0223	.3266
2^{-6}	5	.0025	.4991	4	.0022	.4990	4	-.0035	.3770
2^{-7}	5	.0008	.4998	4	.0007	.4997	4	.0021	.4114
2^{-8}	5	.0002	.5000	4	.0002	.4999	4	.0027	.4357
2^{-9}	5	.0001	.5000	4	.0001	.5000	4	.0021	.4526
2^{-10}	5	.0000	.5000	4	.0000	.5000	4	.0013	.4646
2^{-11}	5	.0000	.5000	4	.0000	.5000	4	.0008	.4734
2^{-12}	5	.0000	.5000	4	.0000	.5000	4	.0004	.4799
2^{-13}	5	.0000	.5000	4	.0000	.5000	4	.0002	.4847
2^{-14}	5	.0000	.5000	4	.0000	.5000	4	.0001	.4883
2^{-15}	5	.0000	.5000	4	.0000	.5000	4	.0001	.4910

Table 1.3: Computational orders of convergence for $\nu = 3$ and $\varepsilon = h^{\frac{1}{2}}$.

For small h, the computational orders $p_\infty(h)$ for f and f_ε are in very good agreement with (5.62) Theorem 1.3. As for g_ε the figures are slightly worse but still in good agreement. Once again, the values of $p_1(h)$ are better than compared with the prediction from (5.54), $p_1 = -\frac{1}{4}$, for f_ε and g_ε.

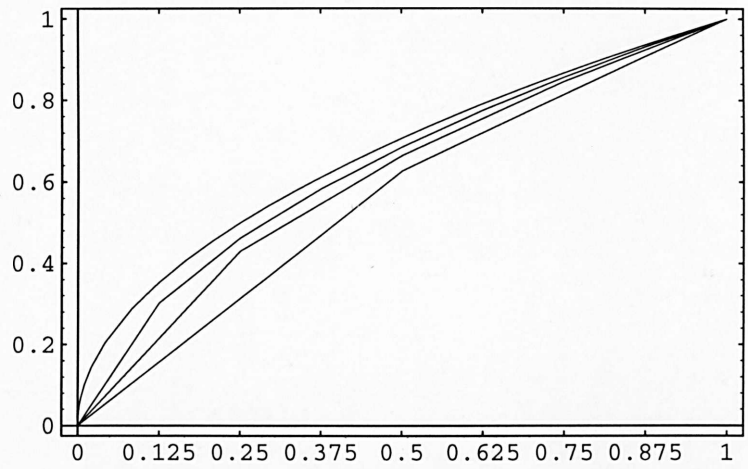


Figure 1.6: Some piecewise lin. Galerkin approx. for $\nu = 3$ and comput. with f

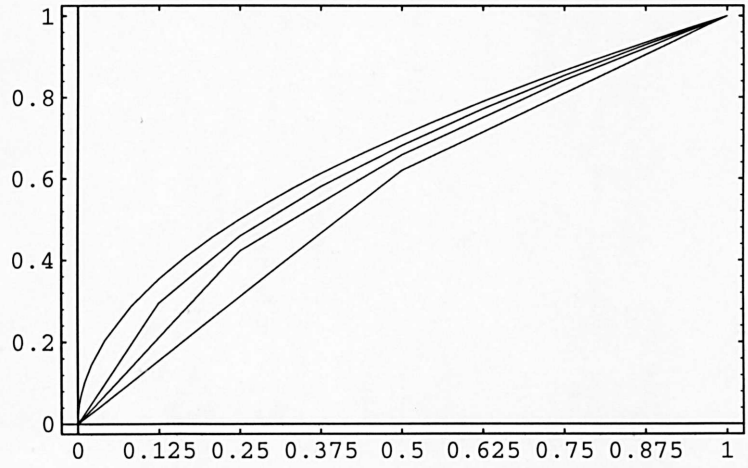


Figure 1.7: Some piecewise lin. Galerkin approx. for $\nu = 3$ and comput. with f_ε

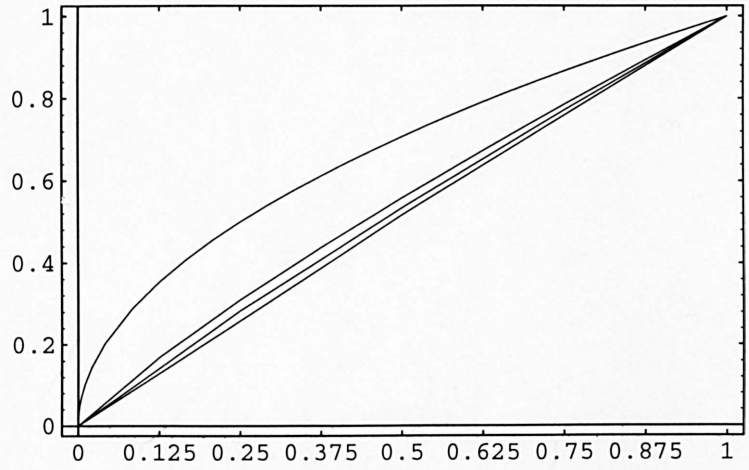


Figure 1.8: Some piecewise lin. Galerkin approx. for $\nu = 3$ and comput. with g_ε

Figures 1.6 – 1.8 are analogues to Figures 1.3 – 1.5 for $\nu = 3$. The smooth graph in the Figures above is the graph of the exact solution $u(x) = x^{\frac{1}{2}}$.

In Table 1.4 we show numerical results for different ν's with $h = 2^{-15}$ and $\varepsilon = h^{\frac{2}{1+\nu}}$. As for the values of $p_\infty(2^{-15})$ we have again very good agreement with Theorem 1.3 for f and f_ε, and good agreement for g_ε. It is however remarkable how well $p_1(2^{-15})$ agrees with the optimal rate of convergence given by (6.47), see Table 1.5.

	f		f_ε		g_ε		Theorem 1.3
ν	n_{it}	$p_\infty(h)$	n_{it}	$p_\infty(h)$	n_{it}	$p_\infty(h)$	$\frac{2}{1+\nu}$
1.0001	1	.9973	1	.9973	2	.8820	1.0000
1.1	2	.9523	2	.9523	2	.8565	.9524
1.2	3	.9091	2	.9091	3	.8304	.9091
1.3	3	.8696	3	.8696	3	.8043	.8696
1.4	3	.8333	3	.8333	3	.7787	.8333
1.5	3	.8000	3	.8000	3	.7534	.8000
2	4	.6667	3	.6667	4	.6437	.6667
2.5	5	.5714	4	.5714	4	.5579	.5714
3	5	.5000	3	.5000	4	.4910	.5000
4	6	.4000	4	.4000	4	.3951	.4000
5	7	.3333	5	.3333	5	.3301	.3333
6	8	.2857	5	.2857	5	.2334	.2857
7	8	.2500	5	.2500	5	.2481	.2500
8	9	.2222	5	.2222	5	.2207	.2222
9	10	.2000	5	.2000	5	.1987	.2000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
19	17	.1000	5	.1000	5	.0994	.1000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
99	72	.0200	6	.0200	6	.0198	.0200

Table 1.4: $p_\infty(h)$ for different ν's with $h = 2^{-15}$ and $\varepsilon = h^{\frac{2}{1+\nu}}$

	f	f_ε	g_ε	(6.47)	(5.54)
ν	$p_1(h)$	$p_1(h)$	$p_1(h)$	$\frac{3-\nu}{2(1+\nu)}$	$\frac{2-\nu}{1+\nu}$
1.0001	.4999	.4999	.4991	.5000	.4999
1.1	.4524	.4524	.4520	.4523	.4286
1.2	.4091	.4091	.4089	.4091	.3636
1.3	.3696	.3696	.3695	.3696	.3043
1.4	.3333	.3333	.3334	.3333	.2500
1.5	.3000	.3000	.3000	.3000	.2000
2	.1667	.1667	.1667	.1667	(.0000)
2.5	.0714	.0714	.0715	.0714	(-.1429)
3.0	.0000	.0000	.0001	(.0000)	(-.2500)

Table 1.5: $p_1(h)$ for different ν's with $h = 2^{-15}$ and $\varepsilon = h^{\frac{2}{1+\nu}}$

We note that for our computations we have so far chosen $\varepsilon = h^{\frac{2}{1+\nu}}$. However, by Theorem 1.3(i) we should choose $\varepsilon = \hat{C}\, h^{\frac{4}{1+3\nu}}$ for the computations of $p_1(h)$, when $1 < \nu < 3$. It turns out that now the choice of $\hat{C}$ determines which kind of convergence will be seen computationally. In Table 1.6 we show $p_1(h)$ for $\hat{C} = 0.001$ and $\hat{C} = 100$. The numbers for $\hat{C} = 0.001$ are in very good agreement with the optimal orders of convergence from (6.47), whereas the numbers for $\hat{C} = 100$ are in good agreement with Theorem 1.3.

	$\hat{C} = 0.001$			$\hat{C} = 100$		
	f_ε	g_ε	(6.47)	f_ε	g_ε	Theorem 1.3
ν	$p_1(h)$	$p_1(h)$	$\frac{3-\nu}{2(1+\nu)}$	$p_1(h)$	$p_1(h)$	$\frac{3-\nu}{1+3\nu}$
1.0001	.4999	.4999	.5000	.4995	.4791	.4999
1.1	.4524	.4523	.4523	.4414	.4214	.4419
1.2	.4091	.4089	.4091	.3908	.3703	.3913
1.3	.3696	.3692	.3696	.3464	.3243	.3469
1.4	.3333	.3329	.3333	.3071	.2826	.3077
1.5	.3000	.2994	.3000	.2721	.2440	.2727
2	.1667	.1658	.1667	.1408	.0856	.1429
2.5	.0714	.0704	.0714	.0416	-.0059	.0588
3.0	.0000	-.0010	(.0000)	-.0406	-.0483	(.0000)

Table 1.6: $p_1(h)$ for different ν's with $h = 2^{-15}$ and $\varepsilon = \hat{C} \times h^{\frac{4}{1+3\nu}}$

Tables 1.5 and 1.6 show a fundamental problem when one uses regularization in practical computations. If the regularization parameter ε is small (as in Table 1.5 and in Table 1.6 for $\hat{C} = 0.001$), it can happen that the computation does not see the effect of the regularization, and thus the error of the computation is just the discretization error. It is however likely that our theoretical bound for the discretization error is suboptimal. This is why we see a better order of convergence than predicted by Theorem 1.3. On the other hand, when ε is large, regularization does play a rôle in the practical computations and the computational error consists of both the regularization and discretization error. The figures in Table 1.6 for $\hat{C} = 100$ agree with the prediction of Theorem 1.3(i), where ε was chosen in terms of h in order to balance these two errors.

We conclude this section with the remark that the computations presented above may not be practical after all. Especially if one is interested in high accuracy, i.e. in small $E_\infty(h)$, for large values of ν, then the above computations yield very poor results. This is because in our computations we have chosen an equidistant mesh which is not appropriate for the kind of singular behaviour displayed by the exact solution near the boundary, see Figures 1.6 – 1.8 and Table 1.7 where we have listed $E_\infty(2^{-15})$ for the computations corresponding to Table 1.4. Note however that $\varepsilon = h^{\frac{2}{1+\nu}} = 2^{-\frac{30}{1+\nu}}$.

In practice, one would of course use adaptive meshes and be interested in a posteriori error bounds. But that is a completely different story....

ν	f	f_ε	g_ε
1.0001	.00000000	.00000000	.00000001
2	.00011126	.00011727	.00090577
3	.00081333	.00092742	.00541156
4	.00245714	.00302192	.01547681
5	.00495774	.00653462	.03107407
6	.00802051	.01126045	.05107385
7	.01134680	.01687784	.07411220
8	.01470950	.02307486	.09898692
9	.01795932	.02959537	.12476470
⋮	⋮	⋮	⋮
19	.03772635	.08954411	.35328263
⋮	⋮	⋮	⋮
99	.03101551	.21542397	.81202102

Table 1.7: $E_\infty(h)$ for different ν's with $h = 2^{-15} \approx 3.05 \times 10^{-5}$ and $\varepsilon = h^{\frac{2}{1+\nu}}$

Chapter 2

A Semilinear Degenerate Parabolic Problem

2.1 Introduction

Let Ω be a convex polyhedral domain in $\mathbb{R}^d$, $1 \leq d \leq 3$, with boundary $\partial\Omega$. We now consider the following problem:

(P) Find $u(x,t)$ such that

$$\partial_t v = \Delta u + f \quad \text{in } Q_T := \Omega \times (0,T]\,, \quad v \equiv \phi(u)\,; \quad u = 0 \quad \text{on } \partial\Omega \times (0,T]\,,$$
$$v(\cdot,0) = \phi(g(\cdot)) \quad \text{in } \Omega\,;$$

where, for given constants $\lambda > 0$ and $p > 1$,

$$\phi(s) := \lambda \operatorname{sign}(s)\,|s|^p \quad \forall s \in \mathbb{R}\,. \tag{1.1}$$

We will begin our analysis with the following assumptions on the given data:

$$f \in L^2(Q_T)\,, \quad g \in L^2(\Omega)\,,$$

but these assumptions will be sharpened as we proceed.

The model (P) is widely used for physical processes. Obviously, when $p = 1$ then (P) is the standard mass diffusion or heat transfer equation with u representing mass density or temperature. For $p \neq 1$ we list some examples which we have mainly taken from [35]. We also refer the reader to the references therein.

In the case when $0 < p < 1$, (P) can be used as a model for the diffusion of gas in porous media; the gas density being represented by v. A similar model is the Stefan problem for heat transfer involving two phases (liquid-solid): $\phi(s)$ in (1.1) is replaced by $\phi(s) := c\,s + H(s)$, where c is a positive constant and H the Heaviside graph. In the Stefan problem u represents temperature and v enthalpy. There is an abundance of literature on both the porous media equation and the Stefan problem.

We consider the case $p > 1$. For $d = 1$, $p = 2$ and $f = 0$, (P) with v representing particle density is mentioned in the literature as a model for plasma diffusion across magnetic fields, see [7] and the references therein. For general $p > 1$ and $d = 1, 2$ and 3, (P) can also be considered a crude approximation of models of nonstationary filtration, see again [35]. Here, we point out that although there is some physical justification for considering (P) we are only aware of one article [29] dealing with a finite element error analysis on it. This is all the more surprising since it has been known for quite some time that the qualitative behaviour of the solutions for $p > 1$ is in stark contrast to the qualitative behaviour for $0 < p < 1$, see below.

Depending on the range of the parameter p in the constitutive relation (1.1) (P) is called a singular parabolic problem when $0 < p < 1$ (since in this case the graph of ϕ has a singularity at the origin), and it is called a degenerate parabolic (or elliptic-parabolic) problem when $p > 1$ (since then the parabolic equation above may degenerate into an elliptic equation). Unfortunately, we have to remark that there is no agreement in the literature as to which case should be called singular and which degenerate. Some authors call the case $0 < p < 1$ degenerate and $p > 1$ singular (since they replace $v \equiv \phi(u)$ by its equivalent $u \equiv \phi^{-1}(v)$). Other authors, like [35] mentioned above, use "degenerate" for the whole range of $p > 0$, $p \neq 1$ in order to contrast it with "regular" nonlinear parabolic problems. We stick to the first definition in this thesis.

The solutions for the singular and the degenerate parabolic equations differ in their qualitative behaviour in time. We first mention the question of compact support arising for the Cauchy problem, i.e. (P) with $f \equiv 0$, $\Omega = \mathbb{R}^d$ and no boundary conditions. It is well-known (cf. [47]) that when $0 < p < 1$ a solution for a compactly supported initial function stays compactly supported for all $t > 0$ and that the interfaces (i.e. the boundary of supp u) move with finite speed. Therefore the singular case is sometimes referred to as "slow diffusion". This is in contrast to the standard heat transfer equation ($p = 1$) where the solution has an infinite speed of propagation in $\mathbb{R}^d$, $d = 1, 2, 3$, for any $t > 0$. As for the degenerate case $p > 1$, Sabinina [40] showed that the solution also has an infinite speed of propagation in $\mathbb{R}^1$ (we do not know whether there is an extension of her proof to $d = 2$ and 3). In addition and more importantly to our considerations, it is also shown in [40] that for the initial boundary value problem (P) with $f \equiv 0$ and $d = 1$, the solution for a non-identically zero initial function has to vanish identically in Ω at a certain time $T_{ex} > 0$. This important result can be extended to $d = 2$ and, under a restriction on the range of p to $d = 3$; we refer to subsection 2.2.3 where we report further on this *finite time extinction property*. Because of this property the degenerate problem (P) is sometimes also called the "fast diffusion problem".

The layout of this chapter is as follows. In section 2.2 we begin with the analysis of (P) by introducing the notion of a weak solution to (P) in an appropriate function space. Uniqueness of a weak solution will be easily seen. Although existence proofs based on different approaches exist in the literature (e.g. see [37], [2], [14] and the references therein) we will give a full existence proof based on the standard techniques of introducing a regularized problem (P_ε), deriving a priori estimates for a semidiscretization of (P_ε) in time and then using compactness arguments (our proof is closest to that given in [37]). The obtained a priori estimates for (P) and (P_ε) play a crucial rôle in later sections. We conclude section 2.2 with reporting on the finite time extinction property.

In section 2.3 we introduce semidiscrete approximations in space, (P^h) and (P_ε^h). The standard continuous piecewise linear Galerkin approximation will be used for the discretization in space, but (P^h) and (P_ε^h) will still be continuous in time. A priori error bounds in $L^2(Q_T)$ will be derived first on assuming exact integration and then on taking into account the effect of numerical integration in space.

Fully discrete problems $(\hat{P}_N^h)$ and $(\hat{P}_{\varepsilon,N}^h)$ will then be introduced in section 2.4 by using the backward Euler method for the time discretization. A priori error bounds are obtained from combining the bounds of section 2.3 with a priori bounds for the time discretization similar to the ones derived in section 2.2. The work done in sections 2.2 and 2.3 allows section 2.4 to be rather short. Nonetheless, the obtained a priori error bound is both sharper and more general than the existing one in the literature (cf. [29]).

In section 2.5 Rulla's novel approach [39] for the analysis of the backward Euler method in the setting of an abstract Cauchy problem will be adapted to our concrete situation. This way we are able to improve on the a priori error bounds of section 2.4.

In section 2.6 we report on Le Roux's scheme for the semidiscretization of (P) in time. This scheme has the interesting feature of capturing the extinction in finite time property of the continuous problem. We introduce a fully discrete version of the method for later reference, but unfortunately it is not clear how to prove a priori error bounds.

Numerical results are included in section 2.7 for homogeneous (P) and $d = 1$. First we compare our a priori error bounds of section 2.5 with computational results. Then in analogy to the extinction time of the continuous solution we observe a computational extinction time for the discrete solutions. The asymptotic profiles in space of the discrete solutions approaching extinction time will be compared with the asymptotic profile of the continuous solution.

Finally, since linear numerical schemes based on extrapolation and relaxation have proved to be useful methods for the numerical solution of slow diffusion problems (cf. [34] and the references therein) we will have a brief look at these methods transferred to an ODE model of the fast diffusion problem. This is done in the concluding section 2.8.

2.2 Regularization

We now introduce a regularized version of (P).

(P_ε) Find $u_\varepsilon(x,t)$ such that

$$\partial_t v_\varepsilon = \Delta u_\varepsilon + f \quad \text{in } Q_T\,, \quad v_\varepsilon \equiv \phi_\varepsilon(u_\varepsilon)\,; \quad u_\varepsilon = 0 \quad \text{on } \partial\Omega \times (0,T]\,,$$
$$v_\varepsilon(\cdot,0) = \phi_\varepsilon(g(\cdot)) \quad \text{in } \Omega\,;$$

where we define for all $s \in \mathbb{R}$

$$\phi_\varepsilon(s) := \begin{cases} \lambda\varepsilon\, s\,, & |s| \le \varepsilon^{\frac{1}{p-1}}\,, \\ \phi(s)\,, & \varepsilon^{\frac{1}{p-1}} \le |s|\,. \end{cases} \tag{2.1}$$

Since $\phi_\varepsilon'(s) \ge \lambda\varepsilon > 0\ \forall s \in \mathbb{R}$, (P_ε) is not degenerate. Furthermore the following inequalities are easily verified for all $\varepsilon \in (0,\varepsilon_0]$:

$$\lambda\varepsilon[a-b]^2 \le [\phi_\varepsilon(a) - \phi_\varepsilon(b)][a-b] \le (\lambda\varepsilon)^{-1}[\phi_\varepsilon(a) - \phi_\varepsilon(b)]^2 \quad \forall a,\, b \in \mathbb{R}\,, \tag{2.2}$$

and $\forall M \in \mathbb{R}^+$,

$$L_M^{-1}[\phi_\varepsilon(a) - \phi_\varepsilon(b)]^2 \le [\phi_\varepsilon(a) - \phi_\varepsilon(b)][a-b] \le L_M[a-b]^2 \quad \forall |a|,\, |b| \le M\,, \tag{2.3}$$

where $L_M := \lambda \max\{pM^{p-1}, \varepsilon_0\}$. To define the notion of a solution of (P_ε), we make the following assumption on f and g:

$$(D0) \qquad f \in L^2(Q_T), \quad g \in L^2(\Omega)\,.$$

Definition 2.1 *u_ε is a weak solution of (P_ε) if $u_\varepsilon \in L^2(0,T;H_0^1(\Omega))$, $v_\varepsilon \equiv \phi_\varepsilon(u_\varepsilon) \in H^1(0,T;L^2(\Omega))$ and for all $\eta \in L^2(0,T;H_0^1(\Omega))$ we have that*

$$\int_{Q_T} \{[\partial_t v_\varepsilon]\eta + \nabla u_\varepsilon \cdot \nabla\eta - f\eta\}\, dx\, dt = 0\,, \tag{2.4}$$

$$v_\varepsilon(\cdot,0) = \phi_\varepsilon(g(\cdot)) \quad \text{in } \Omega\,. \tag{2.5}$$

We remark that since $H^1(0,T;L^2(\Omega)) \subset C([0,T];L^2(\Omega))$ the initial condition makes sense. We define a weak solution u to (P) analogously by replacing ϕ_ε, u_ε and v_ε by ϕ, u and v.

Firstly, we note that there can be at most one solution to (P_ε). Suppose there are two solutions $u_\varepsilon^{(1)}$ and $u_\varepsilon^{(2)}$. Then we have that for all $T' \in (0,T]$ and $\eta \in L^2(0,T';H_0^1(\Omega))$,

$$\int_{Q_{T'}} \{\partial_t[\phi_\varepsilon(u_\varepsilon^{(1)}) - \phi_\varepsilon(u_\varepsilon^{(2)})]\eta + \nabla(u_\varepsilon^{(1)} - u_\varepsilon^{(2)}) \cdot \nabla\eta\}\, dx\, dt = 0\,.$$

Choosing $\eta(x,t) = \int_t^{T'}(u_\varepsilon^{(1)} - u_\varepsilon^{(2)})(x,s)\, ds$ and integrating by parts yields

$$\int_{Q_{T'}} [\phi_\varepsilon(u_\varepsilon^{(1)}) - \phi_\varepsilon(u_\varepsilon^{(2)})][u_\varepsilon^{(1)} - u_\varepsilon^{(2)}]\, dx\, dt + \tfrac{1}{2}\left| \int_0^{T'} (u_\varepsilon^{(1)} - u_\varepsilon^{(2)})\, dt \right|_{1,\Omega}^2 = 0\,. \tag{2.6}$$

On noting that the first term in (2.6) is nonnegative by the monotonicity of ϕ_ε we obtain that

$$\left| \int_0^{T'} (u_\varepsilon^{(1)} - u_\varepsilon^{(2)})\, dt \right|_{1,\Omega} = 0\,, \quad \forall T' \in (0,T]\,,$$

and hence by Poincaré's inequality,

$$\int_0^{T'} (u_\varepsilon^{(1)} - u_\varepsilon^{(2)})\, dt = 0\,, \quad \forall T' \in (0,T]\,,$$

which implies that $u_\varepsilon^{(1)} = u_\varepsilon^{(2)}$ a.e. in $[0,T]$. Uniqueness of a weak solution to (P) can be proved similarly.

To prove existence of a solution to (P_ε) we introduce a problem (P_ε^M) with a Lipschitz continuous nonlinearity ϕ_ε^M. It is convenient to fix an $\varepsilon_0 > 0$ and to choose M such that

$$\varepsilon_0^{\frac{1}{p-1}} \leq M\,.$$

For all $\varepsilon \in (0, \varepsilon_0]$ and $s \in \mathbb{R}$ we define

$$\phi_\varepsilon^M(s) := \begin{cases} \phi_\varepsilon(M) + \phi_\varepsilon'(M)(s-M) & , \quad M \leq s\,, \\ \phi_\varepsilon(s) & , \quad -M \leq s \leq M\,, \\ \phi_\varepsilon(-M) + \phi_\varepsilon'(-M)(s+M) & , \quad s \leq -M\,. \end{cases} \tag{2.7}$$

Obviously, ϕ_ε^M satisfies (2.2) and (2.3) for all a, $b \in \mathbb{R}$ with $L_M = \lambda p M^{p-1}$.

We define the problem (P_ε^M) as (P_ε) with ϕ_ε, u_ε and v_ε replaced by ϕ_ε^M, u_ε^M and v_ε^M. Uniqueness of u_ε^M is shown as for u_ε.

Our program now is the following. Firstly we will prove existence of a weak solution u_ε^M to (P_ε^M) by considering a semidiscrete approximation in time. Then we will show that for M sufficiently large u_ε^M is a weak solution of (P_ε). After that we prove the existence of a weak solution of (P) by letting $\varepsilon \to 0$.

2.2.1 Existence of a Weak Solution to (P_ε^M)

For $N \in \mathbb{N}$ we define the time step $\tau := \frac{T}{N}$. We strengthen our assumption $(D0)$ to

$$(D1) \qquad f \in H^1(0,T;L^2(\Omega))\,, \quad g \in H_0^1(\Omega)$$

and consider the following problem:

$(P_{\varepsilon,N}^M)$ For $n = 1 \to N$ find $u_{\varepsilon,N}^{M,n} \in H_0^1(\Omega)$ such that for all $w \in H_0^1(\Omega)$,

$$\begin{aligned} \int_\Omega \frac{\phi_\varepsilon^M(u_{\varepsilon,N}^{M,n}) - \phi_\varepsilon^M(u_{\varepsilon,N}^{M,n-1})}{\tau}\, w\, dx + \int_\Omega \nabla u_{\varepsilon,N}^{M,n} \cdot \nabla w\, dx &= \int_\Omega f_N^n w\, dx\,, \\ \phi_\varepsilon^M(u_{\varepsilon,N}^{M,0}) &= \phi_\varepsilon^M(g)\,, \end{aligned}$$

where $f_N^n(\cdot) := f(\cdot, n\tau)$.

The first equation in $(P^M_{\varepsilon,N})$ can be rewritten as

$$\int_\Omega \phi^M_\varepsilon(u^{M,n}_{\varepsilon,N})w\,dx + \tau\int_\Omega \nabla u^{M,n}_{\varepsilon,N}\cdot\nabla w\,dx = \tau\int_\Omega f^n_N w\,dx + \int_\Omega \phi^M_\varepsilon(u^{M,n-1}_{\varepsilon,N})w\,dx \qquad (2.8)$$

For given $u^{M,n-1}_{\varepsilon,N}$ (2.8) can be considered as an operator equation

$$\mathcal{A}u^{M,n}_{\varepsilon,N} = b \in [H^1_0(\Omega)]^* \qquad (2.9)$$

on defining $\mathcal{A} : H^1_0(\Omega) \to H^{-1}(\Omega) \equiv [H^1_0(\Omega)]^*$ ($(\,\cdot,\cdot\,)$ denoting the $H^{-1}(\Omega) - H^1_0(\Omega)$ duality pairing) by

$$(\mathcal{A}\bar w, w) := (\mathcal{A}_1\bar w, w) + (\mathcal{A}_2\bar w, w) := \int_\Omega \phi^M_\varepsilon(\bar w)w\,dx + \tau\int_\Omega \nabla\bar w\cdot\nabla w\,dx\,, \quad \forall \bar w, w \in H^1_0(\Omega)\,,$$

which is well defined since $\phi^M_\varepsilon(\bar w) \in L^2(\Omega)$ for all $\bar w \in L^2(\Omega) \supset H^1_0(\Omega)$.

$\mathcal{A}_2$ is strongly monotone and continuous. Since by (2.3) $\mathcal{A}_1$ is monotone, $\mathcal{A}$ will be strongly monotone and continuous, if $\mathcal{A}_1$ is continuous. On noting Hölder's inequality and (2.3) we obtain for all $\bar w_1$, $\bar w_2$, $w \in H^1_0(\Omega)$ that

$$\begin{aligned} |(\mathcal{A}_1\bar w_1 - \mathcal{A}_1\bar w_2, w)| &= \left|\int_\Omega [\phi^M_\varepsilon(\bar w_1) - \phi^M_\varepsilon(\bar w_2)]w\,dx\right| \le L_M|\bar w_1 - \bar w_2|_{0,\Omega}|w|_{0,\Omega} \\ &\le L_M\|\bar w_1 - \bar w_2\|_{1,\Omega}\|w\|_{1,\Omega}\,. \end{aligned}$$

Hence, $\mathcal{A}_1$ is (Lipschitz) continuous.

Existence and uniqueness of a solution to (2.9) and hence to $(P^M_{\varepsilon,N})$ now follow from the Brouwder-Minty-Theorem on Monotone Operators (cf. [49], p.557).

Our main aim in this subsection is to show that a solution to (P^M_ε) can be obtained from $(P^M_{\varepsilon,N})$ by letting $N \to \infty$. To this end we need a priori estimates. To prove these estimates we recall a 'summation by parts' formula (where we use superscripts):

$$\sum_{n=1}^N a^n(b^n - b^{n-1}) = a^N b^N - a^0 b^0 - \sum_{n=1}^N (a^n - a^{n-1})b^{n-1}\,, \qquad (2.10)$$

and the discrete Gronwall's inequality (cf. [18], p.176): Let $\{a^n\}$, $\{b^n\}$, $\{c^n\}$ be three sequences of non-negative real numbers such that $\{c^n\}$ is non-decreasing and

$$a^m + b^m \le c^m + \zeta\sum_{n=0}^{m-1} a^n \quad \text{for } m \ge 1 \text{ and } \zeta > 0, \quad \text{with } a^0 + b^0 \le c^0\,.$$

Then these three sequences satisfy

$$a^m + b^m \le c^m \exp(m\zeta) \quad \text{for } m \ge 0\,. \qquad (2.11)$$

Lemma 2.1 *Under assumption (D1) we have that*

$$\lambda\varepsilon\tau\sum_{n=1}^N \left|\frac{u^{M,n}_{\varepsilon,N} - u^{M,n-1}_{\varepsilon,N}}{\tau}\right|^2_{0,\Omega} + \frac{\tau}{L_M}\sum_{n=1}^N \left|\frac{\phi^M_\varepsilon(u^{M,n}_{\varepsilon,N}) - \phi^M_\varepsilon(u^{M,n-1}_{\varepsilon,N})}{\tau}\right|^2_{0,\Omega} + \max_{n=1\to N}|u^{M,n}_{\varepsilon,N}|^2_{1,\Omega} \le C, \qquad (2.12)$$

where C is independent of M and N.

Proof. On testing the first equation of $(P^M_{\varepsilon,N})$ with $u^{M,n}_{\varepsilon,N} - u^{M,n-1}_{\varepsilon,N}$ and noting (2.2) and (2.3) we obtain that

$$\tfrac{1}{2}\lambda\varepsilon\tau\left|\frac{u^{M,n}_{\varepsilon,N} - u^{M,n-1}_{\varepsilon,N}}{\tau}\right|^2_{0,\Omega} + \tfrac{1}{2}L_M^{-1}\tau\left|\frac{\phi^M_\varepsilon(u^{M,n}_{\varepsilon,N}) - \phi^M_\varepsilon(u^{M,n-1}_{\varepsilon,N})}{\tau}\right|^2_{0,\Omega} + \int_\Omega \nabla u^{M,n}_{\varepsilon,N}\cdot\nabla(u^{M,n}_{\varepsilon,N} - u^{M,n-1}_{\varepsilon,N})\,dx \leq \int_\Omega f^n_N(u^{M,n}_{\varepsilon,N} - u^{M,n-1}_{\varepsilon,N})\,dx\,. \quad (2.13)$$

On noting the inequality $\xi(\xi-\eta) \geq \frac{1}{2}(|\xi|^2 - |\eta|^2)$, $\forall \xi,\eta \in \mathbb{R}$, and summing (2.13) from $n=1$ to m $(1 \leq m \leq N)$ we obtain that

$$\tfrac{1}{2}\lambda\varepsilon\tau\sum_{n=1}^m\left|\frac{u^{M,n}_{\varepsilon,N} - u^{M,n-1}_{\varepsilon,N}}{\tau}\right|^2_{0,\Omega} + \tfrac{1}{2}L_M^{-1}\tau\sum_{n=1}^m\left|\frac{\phi^M_\varepsilon(u^{M,n}_{\varepsilon,N}) - \phi^M_\varepsilon(u^{M,n-1}_{\varepsilon,N})}{\tau}\right|^2_{0,\Omega} + \tfrac{1}{2}|u^{M,m}_{\varepsilon,N}|^2_{1,\Omega} \leq \tfrac{1}{2}|g|^2_{1,\Omega} + \sum_{n=1}^m\int_\Omega f^n_N(u^{M,n}_{\varepsilon,N} - u^{M,n-1}_{\varepsilon,N})\,dx\,. \quad (2.14)$$

From (2.10) we have that

$$\begin{aligned}
&\sum_{n=1}^m\int_\Omega f^n_N(u^{M,n}_{\varepsilon,N} - u^{M,n-1}_{\varepsilon,N})\,dx\\
&= \int_\Omega f^m_N u^{M,m}_{\varepsilon,N}\,dx - \int_\Omega f(\cdot,0)g\,dx - \sum_{n=1}^m\int_\Omega\left[\int_{(n-1)\tau}^{n\tau}\partial_t f\,dt\right]u^{M,n-1}_{\varepsilon,N}\,dx\\
&\leq C\|f\|^2_{L^\infty(0,T;L^2(\Omega))} + \tfrac{1}{4}|u^{M,m}_{\varepsilon,N}|^2_{1,\Omega} + \tfrac{1}{2}|g|^2_{1,\Omega}\\
&\qquad + \left|\sum_{n=1}^m\int_{(n-1)\tau}^{n\tau}\left[\int_\Omega \partial_t f\,u^{M,n-1}_{\varepsilon,N}\,dx\right]dt\right|\\
&\leq C\|f\|^2_{L^\infty(0,T;L^2(\Omega))} + \tfrac{1}{4}|u^{M,m}_{\varepsilon,N}|^2_{1,\Omega} + \tfrac{1}{2}|g|^2_{1,\Omega}\\
&\qquad + \left[\int_0^{m\tau}|\partial_t f|^2_{0,\Omega}\,dt\right]^{\frac{1}{2}}\left[\tau\sum_{n=1}^m|u^{M,n-1}_{\varepsilon,N}|^2_{0,\Omega}\right]^{\frac{1}{2}}, \quad (2.15)
\end{aligned}$$

where we have noted Young's, Poincaré's and Hölder's inequalities. On noting Young's and Poincaré's inequalities again we obtain that

$$\left[\int_0^{m\tau}|\partial_t f|^2_{0,\Omega}\,dt\right]^{\frac{1}{2}}\left[\tau\sum_{n=1}^m|u^{M,n-1}_{\varepsilon,N}|^2_{0,\Omega}\right]^{\frac{1}{2}} \leq C|\partial_t f|^2_{0,Q_T} + C\tau\sum_{n=0}^{m-1}|u^{M,n}_{\varepsilon,N}|^2_{1,\Omega}\,. \quad (2.16)$$

Now, it follows from (2.14), (2.15) and (2.16) that

$$\tfrac{1}{2}\lambda\varepsilon\tau\sum_{n=1}^m\left|\frac{u^{M,n}_{\varepsilon,N} - u^{M,n-1}_{\varepsilon,N}}{\tau}\right|^2_{0,\Omega} + \tfrac{1}{2}L_M^{-1}\tau\sum_{n=1}^m\left|\frac{\phi^M_\varepsilon(u^{M,n}_{\varepsilon,N}) - \phi^M_\varepsilon(u^{M,n-1}_{\varepsilon,N})}{\tau}\right|^2_{0,\Omega} + \tfrac{1}{4}|u^{M,m}_{\varepsilon,N}|^2_{1,\Omega} \leq |g|^2_{1,\Omega} + C\|f\|^2_{L^\infty(0,T;L^2(\Omega))} + C|\partial_t f|^2_{0,Q_T} + C\tau\sum_{n=0}^{m-1}|u^{M,n}_{\varepsilon,N}|^2_{1,\Omega}\,, \quad (2.17)$$

and hence by (2.11),

$$|u_{\varepsilon,N}^{M,m}|_{1,\Omega}^2 \leq C\left[|g|_{1,\Omega}^2 + \|f\|_{L^\infty(0,T;L^2(\Omega))}^2 + |\partial_t f|_{0,Q_T}^2\right]\exp(Cm\tau) \leq C, \quad m = 1 \to N, \tag{2.18}$$

since $m \leq N = \frac{T}{\tau}$. The desired bound (2.12) now follows from (2.17) and (2.18). ∎

Let $u_{\varepsilon,N}^M \in L^\infty(0,T;H_0^1(\Omega))$, $f_N \in L^\infty(0,T;L^2(\Omega))$ and $u_{\varepsilon,N}^{M,\mathcal{L}} \in H^1(0,T;H_0^1(\Omega))$ be such that

$u_{\varepsilon,N}^M(t) := u_{\varepsilon,N}^{M,n}$, $f_N(t) := f_N^n$ and

$u_{\varepsilon,N}^{M,\mathcal{L}}(t) := \tau^{-1}\left[(t-(n-1)\tau)u_{\varepsilon,N}^{M,n} + (n\tau - t)u_{\varepsilon,N}^{M,n-1}\right]$,

if $t \in ((n-1)\tau, n\tau]$; $u_{\varepsilon,N}^M(0) \equiv u_{\varepsilon,N}^{M,\mathcal{L}}(0) := g$.

Similarly, let $v_{\varepsilon,N}^M \in L^\infty(0,T;L^2(\Omega))$ and $v_{\varepsilon,N}^{M,\mathcal{L}} \in H^1(0,T;L^2(\Omega))$ be such that

$v_{\varepsilon,N}^M(t) := \phi_\varepsilon^M(u_{\varepsilon,N}^{M,n})$ and

$v_{\varepsilon,N}^{M,\mathcal{L}}(t) := \tau^{-1}\left[(t-(n-1)\tau)\phi_\varepsilon^M(u_{\varepsilon,N}^{M,n}) + (n\tau - t)\phi_\varepsilon^M(u_{\varepsilon,N}^{M,n-1})\right]$,

if $t \in ((n-1)\tau, n\tau]$; $v_{\varepsilon,N}^M(0) \equiv v_{\varepsilon,N}^{M,\mathcal{L}}(0) := \phi_\varepsilon^M(g)$.

On noting the definitions above and $(P_{\varepsilon,N}^M)$ it follows that

$$\int_{Q_T}\{[\partial_t v_{\varepsilon,N}^{M,\mathcal{L}}]\eta + \nabla u_{\varepsilon,N}^M \cdot \nabla\eta - f_N\eta\}\,dx\,dt = 0, \quad \forall \eta \in L^2(0,T;H_0^1(\Omega))\,. \tag{2.19}$$

As a consequence of Lemma 2.1 we have the following a priori estimates:

$$\varepsilon^{\frac{1}{2}}|\partial_t u_{\varepsilon,N}^{M,\mathcal{L}}|_{0,Q_T} + L_M^{-\frac{1}{2}}|\partial_t v_{\varepsilon,N}^{M,\mathcal{L}}|_{0,Q_T} + \|u_{\varepsilon,N}^{M,\mathcal{L}}\|_{L^\infty(0,T;H^1(\Omega))} + \|u_{\varepsilon,N}^M\|_{L^\infty(0,T;H^1(\Omega))} \leq C\,, \tag{2.20}$$

where C is independent of M and N.

These a priori estimates will enable us to use two standard compactness results, the Banach-Alaoglu and Rellich's Theorems, in the proof of the main result of this subsection, the existence of a solution u_ε^M to (P_ε^M). This will be done by letting $N \to \infty$ in (2.19). As a byproduct we will obtain that the bounds of (2.20) also hold for u_ε^M. For a further bound on u_ε^M we will use elliptic regularity (cf. [20]) which requires a further assumption:

$$(B1) \quad \partial\Omega \in C^{1,1} \quad \text{or} \quad \Omega \text{ is convex polyhedral}\,.$$

We also recall the following well-known embedding and identification results:

$$H_0^1(\Omega) \subset L^2(\Omega) \subset [H_0^1(\Omega)]^* \equiv H^{-1}(\Omega)\,, \tag{2.21}$$

$$L^\infty(0,T;H_0^1(\Omega)) = L^1(0,T;H^{-1}(\Omega))^*\,. \tag{2.22}$$

For further reference we note the following results which are easy to prove

$$|u_{\varepsilon,N}^{M,\mathcal{L}} - u_{\varepsilon,N}^M|_{0,Q_T} \leq \tau\,|\partial_t u_{\varepsilon,N}^{M,\mathcal{L}}|_{0,Q_T}\,, \tag{2.23}$$

$$|v_{\varepsilon,N}^{M,\mathcal{L}} - v_{\varepsilon,N}^M|_{0,Q_T} \leq \tau\,|\partial_t v_{\varepsilon,N}^{M,\mathcal{L}}|_{0,Q_T}\,, \tag{2.24}$$

$$|f - f_N|_{0,Q_T} \leq \tau\,|\partial_t f|_{0,Q_T}\,. \tag{2.25}$$

Lemma 2.2 *Suppose that assumption* $(D1)$ *holds. Then for all* $\varepsilon \in (0, \varepsilon_0]$ *there exists a unique weak solution* u_ε^M *of* (P_ε^M). *Furthermore, we have that*

$$\varepsilon^{\frac{1}{2}} |\partial_t u_\varepsilon^M|_{0,Q_T} + L_M^{-\frac{1}{2}} |\partial_t v_\varepsilon^M|_{0,Q_T} + \|u_\varepsilon^M\|_{L^\infty(0,T;H^1(\Omega))} \leq C\,. \tag{2.26}$$

If in addition $(B1)$ *holds, we have that*

$$L_M^{-\frac{1}{2}} \|u_\varepsilon^M\|_{L^2(0,T;H^2(\Omega))} \leq C\,. \tag{2.27}$$

Proof. In the following we adopt the standard practice of not distinguishing notationally a subsequence from the sequence it is derived from.

It follows from the Banach-Alaoglu Theorem on noting the bounds on the first and third term in (2.20) that we can extract a subsequence $\{u_{\varepsilon,N}^{M,\mathcal{L}}\}$ such that

$$u_{\varepsilon,N}^{M,\mathcal{L}} \to u_\varepsilon^M \qquad \text{weak}^* \text{ in } L^\infty(0,T;H_0^1(\Omega))\,, \tag{2.28}$$

$$u_{\varepsilon,N}^{M,\mathcal{L}} \text{ [resp. } \partial_t u_{\varepsilon,N}^{M,\mathcal{L}}] \to u_\varepsilon^M \text{ [resp. } \partial_t u_\varepsilon^M] \qquad \text{weakly in } L^2(Q_T)\,, \tag{2.29}$$

for the compatibility of the generalized derivative and the weak limit in (2.29) see e.g. [48], Proposition 23.19. Also note that the limits in (2.28) and (2.29) are the same. This is because (2.28) on noting (2.21) and (2.22) implies weak convergence in $L^2(Q_T)$.

By Rellich's Theorem, the bounds on the first and third term in (2.20) also imply that we can choose the above subsequence such that

$$u_{\varepsilon,N}^{M,\mathcal{L}} \to u_\varepsilon^M \quad \text{strongly in } L^2(Q_T)\,, \tag{2.30}$$

see e.g. [30], p.12. From noting (2.23) and the bounds on the first and last term in (2.20) we obtain that (2.30) and (2.28) remain true for a suitably chosen subsequence of $\{u_{\varepsilon,N}^M\}$.

Our first aim is to show convergence of the first term in (2.19). On noting (2.3) it follows from (2.30) for $\{u_{\varepsilon,N}^M\}$ that

$$v_{\varepsilon,N}^M \equiv \phi_\varepsilon^M(u_{\varepsilon,N}^M) \to \phi_\varepsilon^M(u_\varepsilon^M) \equiv v_\varepsilon^M \quad \text{strongly in } L^2(Q_T)\,.$$

Hence, on noting (2.24) and the second bound in (2.20) we have that

$$v_{\varepsilon,N}^{M,\mathcal{L}} \to v_\varepsilon^M \quad \text{strongly in } L^2(Q_T)\,. \tag{2.31}$$

Since by the second bound in (2.20) $\{\partial_t v_{\varepsilon,N}^{M,\mathcal{L}}\}$ must converge weakly in $L^2(Q_T)$, we have similarly to (2.29) that $v_\varepsilon^M \in H^1(0,T;L^2(\Omega))$ and that

$$\int_{Q_T} [\partial_t v_{\varepsilon,N}^{M,\mathcal{L}}]\eta \, dx\, dt \longrightarrow \int_{Q_T} [\partial_t v_\varepsilon^M]\eta \, dx\, dt \quad \text{as } N \to \infty, \quad \forall \eta \in L^2(Q_T)\,. \tag{2.32}$$

The convergence of the second and the third term in (2.19) is easy to show. It follows from (2.28) for $\{u_{\varepsilon,N}^M\}$ on noting (2.22) and from (2.25) that

$$\int_{Q_T} \nabla u_{\varepsilon,N}^M \cdot \nabla \eta \, dx \, dt \longrightarrow \int_{Q_T} \nabla u_\varepsilon^M \cdot \nabla \eta \, dx \, dt \quad \text{as } N \to \infty, \quad \forall \eta \in L^2(0,T;H_0^1(\Omega)), \tag{2.33}$$

$$\int_{Q_T} f_N \eta \, dx \, dt \longrightarrow \int_{Q_T} f \eta \, dx \, dt \quad \text{as } N \to \infty, \quad \forall \eta \in L^2(Q_T). \tag{2.34}$$

We have from (2.19), (2.32), (2.33) and (2.34) that u_ε^M satisfies (2.4). To prove that (2.5) holds, let $\varphi \in C^1([0,T];\mathbb{R})$ with $\varphi(0) = 1$ and $\varphi(T) = 0$, $w \in L^2(\Omega)$ be arbitrary and $\eta \equiv \varphi w \in H^1(0,T;L^2(\Omega))$. On integrating both sides of (2.32) by parts and noting (2.31) we obtain that

$$\int_\Omega [v_\varepsilon^M(x,0) - \phi_\varepsilon^M(g(x))] w(x) \, dx = 0,$$

and hence that $v_\varepsilon^M(\cdot,0) = \phi_\varepsilon^M(g(\cdot))$, since $w \in L^2(\Omega)$ was arbitrary. This completes the proof of existence. As remarked earlier, uniqueness is a consequence of the monotonicity of ϕ_ε^M.

Finally, we want to show the a priori bounds (2.26) and (2.27). The bound on the first term in (2.26) follows from (2.29) on noting the bound on the first term in (2.20), cf. [48], Proposition 21.23c. Similarly, we obtain the bound on the second term from the weak convergence of $\partial_t v_{\varepsilon,N}^{M,\mathcal{L}}$; whereas the bound on the third term in (2.26) follows from (2.28) on noting the bound on the third term in (2.20), cf. [48], Proposition 21.26b. To prove (2.27) note that since u_ε^M satisfies (2.4) we have for almost all $t \in [0,T]$ that

$$\Delta u_\varepsilon^M(\cdot,t) = -f(\cdot,t) + \partial_t v_\varepsilon^M(\cdot,t) \quad \text{in } \Omega \tag{2.35}$$

in a weak sense. (2.27) now follows from (2.35) and elliptic regularity on noting the second bound in (2.26). ∎

2.2.2 Existence of a Weak Solution to (P_ε)

To prove existence of a weak solution u_ε to (P_ε), we first prove a uniform a priori bound for u_ε^M on Q_T which only depends on the data f and g. From this the existence result will easily follow.

We need a stronger assumption on the data. We will assume that

$$(D2) \qquad f \in H^1(0,T;L^2(\Omega)) \cap L^\infty(Q_T), \quad g \in H_0^1(\Omega) \cap L^\infty(\Omega).$$

It follows from $(D2)$ that there exist constants $\underline{f}$, $\underline{g}$, $\overline{f}, \overline{g}$ such that

$$\underline{f} \le f \le \overline{f} \quad \text{a.e. in } Q_T, \quad \underline{g} \le g \le \overline{g} \quad \text{a.e. in } \Omega,$$

where without loss of generality we may assume that

$$\underline{f}, \underline{g} \le 0 \le \overline{f}, \overline{g}. \tag{2.36}$$

We also need the following well-known results:

$$\beta \in C(\mathbb{R}) \text{ [and } \beta(0) = 0], \beta' \in L^\infty(\mathbb{R}) \quad \text{and} \quad \eta \in W^{1,q}(\Omega) \text{ [resp. } W_0^{1,q}(\Omega)] \Longrightarrow$$
$$\beta(\eta) \in W^{1,q}(\Omega) \text{ [resp. } W_0^{1,q}(\Omega)] \quad \text{and} \quad \nabla(\beta(\eta)) = \beta'(\eta)\nabla\eta \quad \text{a.e. in } \Omega, \tag{2.37}$$

e.g. see [38], Theorem 3.3.

Lemma 2.3 *Assume that (D2) holds. If M satisfies*

$$M^p \geq \lambda^{-1} \max\{|\underline{f}|, \overline{f}\}T + \left[\max\{|\underline{g}|, \overline{g}, \varepsilon_0^{\frac{1}{p-1}}\}\right]^p, \tag{2.38}$$

then we have for all $\varepsilon \in (0, \varepsilon_0]$ that

$$-M \leq u_\varepsilon^M \leq M \quad \text{a.e. in } Q_T. \tag{2.39}$$

In addition, if $f, g \geq 0$ we have that $u_\varepsilon^M \geq 0$.

Proof. The idea of the proof is to exhibit a super- and a subsolution of (P_ε^M) which can be bounded in terms of the data.

Let $\overline{u}_\varepsilon^M$ be defined by

$$\overline{u}_\varepsilon^M(\cdot, t) := (\phi_\varepsilon^M)^{-1}(\overline{f}t + \phi_\varepsilon^M(\overline{g})), \quad \forall t \in [0, T].$$

It is easy to see that $\overline{u}_\varepsilon^M \in L^2(0,T; H^1(\Omega))$, $\overline{v}_\varepsilon^M \equiv \phi_\varepsilon^M(\overline{u}_\varepsilon^M) \in H^1(0,T; L^2(\Omega))$ and on noting (2.36) that for all $T' \in [0,T]$ and $\eta \in L^2(0,T'; H_0^1(\Omega))$ with $\eta \geq 0$ in $Q_{T'}$,

$$\int_{Q_{T'}} \{\partial_t[\overline{v}_\varepsilon^M - v_\varepsilon^M]\eta + \nabla[\overline{u}_\varepsilon^M - u_\varepsilon^M] \cdot \nabla\eta\} \, dx \, dt \geq 0. \tag{2.40}$$

Moreover, we have that $\overline{v}_\varepsilon^M(\cdot, 0) \geq \phi_\varepsilon^M(g)$, and $\overline{u}_\varepsilon^M \geq 0$ on $\partial\Omega \times [0,T]$. Hence, $\overline{u}_\varepsilon^M$ is a supersolution of (P_ε^M).

On noting the first part of (2.37) it follows from the Lipschitz continuity of ϕ_ε^M and (2.36) that we may choose $\eta \equiv -[\phi_\varepsilon^M(\overline{u}_\varepsilon^M) - \phi_\varepsilon^M(u_\varepsilon^M)]_- \equiv -[\overline{v}_\varepsilon^M - v_\varepsilon^M]_-$ in (2.40) to obtain that

$$\int_{Q_{T'}} \{\partial_t[\overline{v}_\varepsilon^M - v_\varepsilon^M][\overline{v}_\varepsilon^M - v_\varepsilon^M]_- + \nabla[\overline{u}_\varepsilon^M - u_\varepsilon^M] \cdot \nabla[\overline{v}_\varepsilon^M - v_\varepsilon^M]_-\} \, dx \, dt \leq 0. \tag{2.41}$$

By the second part of (2.37) and the strict monotonicity of ϕ_ε^M we have that

$$\begin{aligned} \nabla[\overline{u}_\varepsilon^M - u_\varepsilon^M]_- &= \nabla[\overline{u}_\varepsilon^M - u_\varepsilon^M] \, \chi\{\overline{v}_\varepsilon^M < v_\varepsilon^M\}, \\ \partial_t[\overline{v}_\varepsilon^M - v_\varepsilon^M]_- &= \partial_t[\overline{v}_\varepsilon^M - v_\varepsilon^M] \, \chi\{\overline{v}_\varepsilon^M < v_\varepsilon^M\}, \end{aligned}$$

where χ is the characteristic function. Hence, we obtain from (2.41) that

$$\int_{Q_{T'}} \{\partial_t[\overline{v}_\varepsilon^M - v_\varepsilon^M]_-[\overline{v}_\varepsilon^M - v_\varepsilon^M]_- + \nabla[\overline{u}_\varepsilon^M - u_\varepsilon^M]_- \cdot \nabla[\overline{v}_\varepsilon^M - v_\varepsilon^M]_-\} \, dx \, dt \leq 0. \tag{2.42}$$

Noting that by (2.37),

$$\int_{Q_{T'}} \nabla[\overline{u}_\varepsilon^M - u_\varepsilon^M]_- \cdot \nabla[\overline{v}_\varepsilon^M - v_\varepsilon^M]_- \, dx\, dt = \int_0^{T'} \int_{\Omega_t} (\phi_\varepsilon^M)'(u_\varepsilon^M) |\nabla u_\varepsilon^M|^2 \, dx\, dt \geq 0 \,,$$

where $\Omega_t := \{x \in \Omega : \overline{u}_\varepsilon^M(x,t) < u_\varepsilon^M(x,t)\}$, it follows from (2.42) that for all $t \in [0,T]$

$$|[\phi_\varepsilon^M(\overline{u}_\varepsilon^M) - \phi_\varepsilon^M(u_\varepsilon^M)]_-(\cdot,t)|_{0,\Omega}^2 \leq |[\phi_\varepsilon^M(\overline{u}_\varepsilon^M) - \phi_\varepsilon^M(u_\varepsilon^M)]_-(\cdot,0)|_{0,\Omega}^2 = 0 \,.$$

Hence, on noting the monotonicity of ϕ_ε^M we have that

$$u_\varepsilon^M \leq \overline{u}_\varepsilon^M \quad \text{a.e. in } Q_T \,. \tag{2.43}$$

Similarly, one can show that for $\underline{u}_\varepsilon^M$ defined by

$$\underline{u}_\varepsilon^M(\cdot,t) = (\phi_\varepsilon^M)^{-1}(\underline{f}t + \phi_\varepsilon^M(\underline{g})) \,, \quad \forall t \in [0,T] \,,$$

one has that $u_\varepsilon^M \geq \underline{u}_\varepsilon^M$. If f, $g \geq 0$, one can take $\underline{f}$, $\underline{g} \equiv 0$ and it follows that $u_\varepsilon^M \geq 0$.

We obtain from (2.36) that $\phi_\varepsilon^M(\overline{g}) \leq \lambda[\max\{\overline{g}, \varepsilon_0^{\frac{1}{p-1}}\}]^p$ and hence that

$$0 \leq \overline{f}t + \phi_\varepsilon^M(\overline{g}) \leq \overline{f}T + \phi_\varepsilon^M(\overline{g}) \leq \phi_\varepsilon^M(M) \equiv \lambda M^p \,, \quad \forall t \in [0,T] \,, \tag{2.44}$$

if

$$M^p \geq \lambda^{-1}\overline{f}T + \left[\max\{\overline{g}, \varepsilon_0^{\frac{1}{p-1}}\}\right]^p \,. \tag{2.45}$$

It then follows from (2.43), the definition of $\overline{u}_\varepsilon^M$ and (2.44) that

$$u_\varepsilon^M \leq \overline{u}_\varepsilon^M \leq M \quad \text{a.e. in } Q_T \,.$$

Similarly, one can show that

$$-M \leq \underline{u}_\varepsilon^M \leq u_\varepsilon^M \quad \text{a.e. in } Q_T \,,$$

if

$$M^p \geq \lambda^{-1}|\underline{f}|T + \left[\max\{|\underline{g}|, \varepsilon_0^{\frac{1}{p-1}}\}\right]^p \,. \tag{2.46}$$

Since (2.45) and (2.46) are satisfied by assumption (2.38), the proof is complete. ∎

Lemma 2.4 *Assume that (D2) holds. Then for all $\varepsilon \in (0, \varepsilon_0]$ there exists a unique weak solution u_ε to (P_ε). Moreover, there exists an $M > 0$ such that for all $\varepsilon \in (0, \varepsilon_0]$*

$$-M \leq u_\varepsilon \leq M \quad \textit{a.e. in } Q_T \,, \tag{2.47}$$

and if f, $g \geq 0$ we have that $u_\varepsilon \geq 0$.

Furthermore, we have for all $\varepsilon \in (0, \varepsilon_0]$

$$\varepsilon^{\frac{1}{2}}|\partial_t u_\varepsilon|_{0,Q_T} + L_M^{-\frac{1}{2}}|\partial_t v_\varepsilon|_{0,Q_T} + \|u_\varepsilon\|_{L^\infty(0,T;H^1(\Omega))} \leq C \,. \tag{2.48}$$

If in addition (B1) holds, we have that

$$L_M^{-\frac{1}{2}}\|u_\varepsilon\|_{L^2(0,T;H^2(\Omega))} \leq C \,. \tag{2.49}$$

Proof. On choosing M such that (2.38) holds we obtain from Lemma 2.3 that for all $\varepsilon \in (0, \varepsilon_0]$ (2.39) holds and hence that $\phi_\varepsilon^M(u_\varepsilon^M) \equiv \phi_\varepsilon(u_\varepsilon^M)$, from which it follows that u_ε^M, the solution of (P_ε^M), is also a solution of (P_ε). Uniqueness was proved earlier, see (2.6) and below. Obviously, (2.47), (2.48) and (2.49) follow from (2.39), (2.26) and (2.27). ∎

2.2.3 Existence of a Weak Solution to (P); Extinction in Finite Time

Theorem 2.1 *Under assumption (D2) there exists a unique weak solution u of (P). In addition the bound (2.47) holds true for u, and if f, $g \geq 0$ then $u \geq 0$.*

Furthermore, we have that $u \in L^\infty(0,T;H_0^1(\Omega))$ and $v \in H^1(0,T;L^2(\Omega))$; and $u \in L^2(0,T;H^2(\Omega))$ if in addition (B1) holds.

Finally, for all $\varepsilon \in (0, \varepsilon_0]$ we have that

$$\varepsilon |u - u_\varepsilon|^2_{0,Q_T} + |v - v_\varepsilon|^2_{0,Q_T} \leq C \{A_\varepsilon^u(T)\, \varepsilon^{\frac{p+1}{p-1}} + B_\varepsilon\, \varepsilon^{\frac{2p}{p-1}}\} \leq C \varepsilon^{\frac{p+1}{p-1}}, \tag{2.50}$$

where $A_\varepsilon^u(T) := \int_0^T meas(\Omega_\varepsilon^u(t))\, dt$, $\Omega_\varepsilon^u(t) := \{x \in \Omega : |u(x,t)| < \varepsilon^{\frac{1}{p-1}}\}$ and $B_\varepsilon := meas\{x \in \Omega : |g(x)| < \varepsilon^{\frac{1}{p-1}}\}$.

Proof. Uniqueness was remarked upon earlier, see (2.6) and below. To prove existence of a weak solution u to (P) we let $\varepsilon \to 0$ in the weak formulation of (P_ε). Let $\varepsilon_0 \geq \varepsilon_j \geq \varepsilon_i > 0$ and $e \equiv u_{\varepsilon_i} - u_{\varepsilon_j}$. On subtracting the weak formulation of (P_{ε_j}) from that of (P_{ε_i}), choosing the test function $\eta(x,t) \equiv \int_t^T e(x,s)\, ds$ and integrating by parts we obtain that

$$\int_{Q_T} [\phi_{\varepsilon_i}(u_{\varepsilon_i}) - \phi_{\varepsilon_j}(u_{\varepsilon_j})]\, e\, dx\, dt + \tfrac{1}{2} \left| \int_0^T e\, dt \right|^2_{1,\Omega} = \int_\Omega [\phi_{\varepsilon_i}(g) - \phi_{\varepsilon_j}(g)][\int_0^T e\, dt]\, dx\,,$$

which we rewrite as

$$\begin{aligned} &\int_{Q_T} [\phi_{\varepsilon_j}(u_{\varepsilon_i}) - \phi_{\varepsilon_j}(u_{\varepsilon_j})]\, e\, dx\, dt + \tfrac{1}{2} \left| \int_0^T e\, dt \right|^2_{1,\Omega} \\ &\quad = \int_{Q_T} [\phi_{\varepsilon_j}(u_{\varepsilon_i}) - \phi_{\varepsilon_i}(u_{\varepsilon_i})]\, e\, dx\, dt + \int_\Omega [\phi_{\varepsilon_i}(g) - \phi_{\varepsilon_j}(g)][\int_0^T e\, dt]\, dx\,. \end{aligned}$$

Using (2.2) and (2.3) on noting (2.47) and Young's and Poincaré's inequalities yields

$$\begin{aligned} &\varepsilon_j\, |u_{\varepsilon_i} - u_{\varepsilon_j}|^2_{0,Q_T} + |\phi_{\varepsilon_j}(u_{\varepsilon_i}) - \phi_{\varepsilon_j}(u_{\varepsilon_j})|^2_{0,Q_T} + \left| \int_0^T (u_{\varepsilon_i} - u_{\varepsilon_j})\, dt \right|^2_{1,\Omega} \leq \\ &C \left[\varepsilon_j^{-1} |\phi_{\varepsilon_j}(u_{\varepsilon_i}) - \phi_{\varepsilon_i}(u_{\varepsilon_i})|^2_{0,Q_T} + |\phi_{\varepsilon_i}(g) - \phi_{\varepsilon_j}(g)|^2_{0,\Omega} \right] \leq C \left[\varepsilon_j^{\frac{p+1}{p-1}} A_{\varepsilon_j}^{u_{\varepsilon_i}}(T) + \varepsilon_j^{\frac{2p}{p-1}} B_{\varepsilon_j} \right] \end{aligned} \tag{2.51}$$

Thus $\{u_\varepsilon\}$ is a Cauchy sequence in $L^2(Q_T)$, i.e. $u_\varepsilon \to u$ in $L^2(Q_T)$ as $\varepsilon \to 0$, in particular we have pointwise convergence almost everywhere and hence (2.47) holds true for u in place of u_ε.

Due to the last bound in (2.48) there exists a subsequence such that

$$u_\varepsilon \to u \quad \text{weak}^* \text{ in } L^\infty(0,T;H_0^1(\Omega))\,,$$

in particular $u \in L^\infty(0,T;H_0^1(\Omega))$, cf. (2.28) and below. Therefore, we can pass to the limit in the weak formulation of (P_ε), if we can show that

$$\partial_t v_\varepsilon \to \partial_t v \quad \text{weakly in } L^2(Q_T)\,. \tag{2.52}$$

We have from (2.47) for u_ε and u, the Lipschitz continuity of ϕ on $[-M,M]$ and (2.1) that

$$\begin{aligned} |v - v_\varepsilon|_{0,Q_T} &= |\phi(u) - \phi_\varepsilon(u_\varepsilon)|_{0,Q_T} \le |\phi(u) - \phi(u_\varepsilon)|_{0,Q_T} + |\phi(u_\varepsilon) - \phi_\varepsilon(u_\varepsilon)|_{0,Q_T} \\ &\le L_M |u - u_\varepsilon|_{0,Q_T} + C\varepsilon^{\frac{p}{p-1}}\,. \end{aligned} \tag{2.53}$$

Now (2.52) and $v \in H^1(0,T;L^2(\Omega))$ follow from (2.53) and the second bound in (2.48), cf. (2.29) and below. Moreover v satisfies the initial condition, since $\phi_\varepsilon(g) \to \phi(g)$ strongly in $L^2(\Omega)$.

If $(B1)$ holds, it can be shown similarly to (2.27) on noting $\partial_t v \in L^2(Q_T)$ that $u \in L^2(0,T;H^2(\Omega))$, cf. (2.35).

It is left to show (2.50). On replacing ϕ_{ε_i} and u_{ε_i} by ϕ and u and ε_j by ε we obtain analogously to (2.51) that

$$\begin{aligned} &\varepsilon\,|u - u_\varepsilon|^2_{0,Q_T} + |\phi_\varepsilon(u) - v_\varepsilon|^2_{0,Q_T} + \left|\int_0^T (u - u_\varepsilon)\,dt\right|^2_{1,\Omega} \le \\ &C\left[\varepsilon^{-1}|\phi_\varepsilon(u) - v|^2_{0,Q_T} + |\phi(g) - \phi_\varepsilon(g)|^2_{0,\Omega}\right] \le C\left[\varepsilon^{\frac{p+1}{p-1}} A_\varepsilon^u(T) + \varepsilon^{\frac{2p}{p-1}} B_\varepsilon\right] \end{aligned}$$

and hence the desired result (2.50). ∎

Now we turn to the important case $f \equiv 0$, i.e. (P) homogeneous. By the above there exists a unique weak solution u to (P) for *all* $T > 0$. The interesting feature of u however is that it becomes extinct after a *finite* time T_{ex} which was first proved by Sabinina [40] for the case $d = 1$. It is claimed in [28] that Sabinina's proof can be adapted for $d = 2$ and 3. Below we will prove extinction in finite time by adapting the proof of [28] for semidiscrete solutions, cf. section 2.6. To this end we recall a further technical result.

Note that since

$$\phi(u) \in H^1(0,T;L^2(\Omega)) \tag{2.54}$$

we have by Sobolev's Embedding Theorem that $\phi(u) \in C([0,T];L^2(\Omega))$ and hence, by the definition of ϕ, that $u \in C([0,T];L^{2p}(\Omega))$. Thus, since $2p > p+1$ for $p > 1$, we have that

$$u \in L^2(0,T;H_0^1(\Omega)) \cap C([0,T];L^{p+1}(\Omega))\,, \tag{2.55}$$

in particular, we obtain that $u(\cdot,0) = g(\cdot)$. On noting (2.54) and (2.55) it can be shown analogously to the proof of Lemma 1.2 in [37] that

$$\int_{Q_T} [\partial_t \phi(u)]\, u \, dx dt = \lambda \tfrac{p}{p+1} \left(|u(\cdot,T)|^{p+1}_{0,p+1,\Omega} - |u(\cdot,0)|^{p+1}_{0,p+1,\Omega} \right) ; \tag{2.56}$$

which is not too surprising since we formally have that

$$(\partial_t u^p) u \equiv (\partial_t u^p)(u^p)^{\frac{1}{p}} \equiv \tfrac{p}{p+1} \partial_t \left([u^p]^{\frac{p+1}{p}} \right) \equiv \tfrac{p}{p+1} \partial_t (u^{p+1}) .$$

We also recall the Sobolev embedding $H^1_0(\Omega) \subset L^{p+1}(\Omega)$ with

$$|w|_{0,p+1,\Omega} \leq C_p |w|_{1,\Omega} \tag{2.57}$$

for all $w \in H^1_0(\Omega)$, provided $p \geq 0$ is restricted by

$$p < \infty \quad \text{if } d = 2 \qquad \text{and} \qquad p < 5 \quad \text{if } d = 3 . \tag{2.58}$$

Since we consider $p \in (1, \infty)$ only, it is no restriction to rewrite (2.58) for clarity as

$$(S) \qquad 1 < p < \infty \quad \text{if } d \leq 2 \qquad \text{and} \qquad 1 < p < 5 \quad \text{if } d = 3 .$$

Lemma 2.5 *Assume that* (S) *and* (D_2) *hold and that* $f \equiv 0$. *Then there exists a* $T_{ex} \in \mathbb{R}_+$ *such that*

$$u(\cdot,t) \equiv 0 , \qquad \forall t \geq T_{ex} . \tag{2.59}$$

Moreover, with C_p *as in (2.57) we have the following bound on the extinction time* T_{ex}*:*

$$T_{ex} \leq \mathcal{T}_{ex} \equiv \lambda \tfrac{p}{p-1} C_p^2 |u(\cdot,0)|^{p-1}_{0,p+1,\Omega} , \tag{2.60}$$

Proof. As mentioned above we adapt the proof given in [28]. On testing (P) with u and noting (2.56) we obtain that

$$\lambda \tfrac{p}{p+1} \left(|u(\cdot,T)|^{p+1}_{0,p+1,\Omega} - |u(\cdot,0)|^{p+1}_{0,p+1,\Omega} \right) = - \int_0^T |u(\cdot,t)|^2_{1,\Omega} \, dt . \tag{2.61}$$

This equation holds for all $T > 0$. We also recall that $u \in L^\infty(0,T; H^1_0(\Omega))$. Hence, on introducing the function $a(t) \equiv |u(\cdot,t)|_{0,p+1,\Omega}$, $t \in \mathbb{R}_+$, we obtain from (2.61) that a^{p+1} is Lipschitz continuous and that

$$\lambda \tfrac{p}{p+1} \tfrac{d}{dt} a^{p+1}(t) = -|u(\cdot,t)|^2_{1,\Omega} \leq -C_p^{-2} a^2(t) \quad \text{a.e. in } \mathbb{R}_+ , \tag{2.62}$$

where we have also noted (2.57).

Our aim is to show that there exists a $T_{ex} \in \mathbb{R}_+$ such that

$$a(T_{ex}) \equiv 0\,. \tag{2.63}$$

For (2.63) is equivalent to $u(\cdot, T_{ex}) \equiv 0$ and the desired result (2.59) then follows from the uniqueness of the weak solution.

For the purpose of contradiction we assume that

$$a(t) > 0 \quad \forall t \geq 0\,. \tag{2.64}$$

Let $T > 0$ be arbitrary but fixed. Since a^{p+1} is (Lipschitz) continuous, it attains its minimum b, say, on $[0,T]$. By (2.64), we have that $b > 0$. Hence, on noting the Mean Value Theorem we have for any t, $t_0 \in [0,T]$ that

$$|a^{p-1}(t) - a^{p-1}(t_0)| = |\,[a^{p+1}(t)]^{\frac{p-1}{p+1}} - [a^{p+1}(t_0)]^{\frac{p-1}{p+1}}\,| \leq \tfrac{p-1}{p+1} b^{-\frac{2}{p+1}}\,|a^{p+1}(t) - a^{p+1}(t_0)|\,.$$

Thus, it follows from the Lipschitz continuity of a^{p+1} that a^{p-1} is Lipschitz continuous on any interval $[0,T]$, $T > 0$, in particular $\frac{d}{dt}a^{p-1}(t)$ exists a.e. in $\mathbb{R}_+$. Hence, on noting the chain rule we have that

$$\tfrac{d}{dt}a^{p+1}(t) = \tfrac{d}{dt}(a^{p-1}(t))^{\frac{p+1}{p-1}} = \tfrac{p+1}{p-1}\,a^2(t)\,\tfrac{d}{dt}a^{p-1}(t) \quad \text{a.e. in } \mathbb{R}_+\,.$$

Consequently, we obtain from (2.62) that

$$\lambda \tfrac{p}{p-1}\,\tfrac{d}{dt}a^{p-1}(t) \leq -C_p^{-2} \quad \text{a.e. in } \mathbb{R}_+\,.$$

By the aforementioned properties of a^{p-1} we can integrate the last inequality on $[0,T]$, see e.g. [48], Corollary 23.22, and obtain that

$$a^{p-1}(T) \leq a^{p-1}(0) - \lambda^{-1}\tfrac{p-1}{p}C_p^{-2}\,T\,.$$

It follows that $a^{p-1}(T_{ex}) \equiv 0$ which contradicts (2.64). Thus we have shown that u becomes extinct at a certain time T_{ex}. The argument above also shows that (2.60) holds. ∎

We note that Lemma 2.5 contains the trivial case $u \equiv 0$, if $u(\cdot,0) \equiv 0$. We also mention that Lemma 2.5 can be sharpened under additional assumptions: For instance, Sabinina [40] showed that if $d = 1$ and g non-negative and not identically 0 then there exists a $T_{ex} > 0$ such that (2.59) holds and that u, the weak solution of (P), is a positive classical solution on $\Omega \times (0, T_{ex})$. To illustrate this result we consider the following example which is based on [8].

For $d = 1$, $\Omega \equiv (0,1)$ and $p > 1$ there exists a unique $w \in H_0^1(\Omega)$ which satisfies

$$-w'' = pw^p \text{ and } w > 0 \quad \text{in } \Omega,$$

see e.g. [8]. Now, for fixed $T_{ex} \in (0,T)$, let $k \equiv [(p-1)T_{ex}]^{\frac{1}{p-1}}$, $\lambda \equiv 1$, $g \equiv kw$ and $f \equiv 0$.

Given these data, we have from Theorem 2.1 that there exists a unique weak solution u of (P), and u is non-negative on Q_T, since $f, g \geq 0$. In fact, the data have been chosen so that u is the following separable function, which is easy to check:

$$u(x,t) = w(x)T^*(t), \ \forall (x,t) \in \Omega \times [0,T], \tag{2.65}$$

where

$$T^*(t) := \begin{cases} [(p-1)(T_{ex}-t)]^{\frac{1}{p-1}}, & t \in [0, T_{ex}), \\ 0 \qquad\qquad\qquad\quad , & t \in [T_{ex}, T]. \end{cases} \tag{2.66}$$

From elliptic regularity we have that $w \in C^2(\Omega) \cap C(\bar{\Omega})$ and hence it is easy to check from (2.65) and (2.66) that Sabinina's result holds in this special case.

Assuming extinction in a finite time T_{ex} Sabinina's result about the existence of positive classical solutions on $\Omega \times (0, T_{ex})$ was extended to $d > 1$ in [26] where it is also assumed that $\partial\Omega$ is smooth, $g \in L^\infty(\Omega)$ and $p \geq 2$.

The example above raises the question whether the positive separable solutions of (P) are not in fact typical rather than special solutions. Indeed there is another feature of separable solutions which shows that they are quite typical. This feature is stability which was first studied by Berryman and Holland [8]. Below we state their main result.

For $d = 1, 2, 3$ and $\lambda = 1$, $u_{sep}(x,t) \equiv w(x)T^*(t)$ is a positive separable solution of (P) if

$$-\Delta w = pw^p, \quad w > 0 \ \text{ in } \Omega, \quad w = 0 \ \text{ on } \partial\Omega \tag{2.67}$$

and $T^*(t)$ is as in (2.66). To state the main result of [8] the following regularity requirement is needed for the solutions u of (P): There exists a $\gamma \in (0, T_{ex})$ such that

$$(R) \qquad u_t,\ u_{x_i},\ u_{x_i t},\ u_{x_i x_i} \in C(\bar{\Omega} \times (\gamma, T_{ex})) \ \text{ and } \ \partial\Omega \text{ is smooth}.$$

Lemma 2.6 *(main result of [8]) Assume that (S) and (R) hold.*
Let $W(x,t) := u(x,t)/T^*(t)$ for all $(x,t) \in \Omega \times (0, T_{ex})$.

(i) Then there exists an increasing sequence of times $t_n \nearrow T_{ex}$ and a classical solution $w \in H_0^1(\Omega)$ of (2.67) such that $W(\cdot, t_n) \to w(\cdot)$ in $H_0^1(\Omega)$.

(ii) If there is a unique solution to (2.67) (which e.g. is the case when $d = 1$), then $W(\cdot, t) \to w(\cdot)$ in $H_0^1(\Omega)$ as $t \nearrow T_{ex}$.

It is claimed in [26] that (i) and (ii) can be shown without using the strong regularity requirement (R).

Finally, to conclude this section on the regularization we remark that it is also easy to derive from (2.65) and (2.66) that $u \in H^1(0,T; L^2(\Omega))$ if and only if $p \in (1,3)$.

2.3 A Semidiscrete Approximation

We now consider a continuous piecewise linear finite element approximation to (P_ε), which is continuous in time.

To keep this exposition simple, we will assume that

$$(B2) \qquad \Omega \text{ is convex polyhedral.}$$

Let $\mathcal{T}^h$ be a regular partitioning of Ω, i.e. $\bar{\Omega} \equiv \bigcup_{\kappa\in\mathcal{T}^h} \bar{\kappa}$, where h is the maximum diameter (i.e. side-length) over all simplices κ of $\mathcal{T}^h$. We denote the set of vertices of $\mathcal{T}^h$ by $\{x_i\}_{i\in I}$ and introduce the following function spaces:

$$\begin{aligned} S^h &\equiv \{\chi \in C(\bar{\Omega}^h) : \chi|_\kappa \text{ is linear } \forall \kappa \in \mathcal{T}^h\} \\ S_0^h &\equiv \{\chi \in S^h : \chi|_{\partial\Omega^h} = 0\} \end{aligned}$$

We also introduce the interpolation operator $\pi^h : C(\bar{\Omega}) \to S^h$ which is defined to be such that for any $\psi \in C(\bar{\Omega})$, $\pi^h\psi \in S^h$ satisfies $\pi^h\psi(x_i) = \psi(x_i)\ \forall i \in I$; and the elliptic projection operator $Q^h : H_0^1(\Omega) \to S_0^h$ which is such that for any $\psi \in H_0^1(\Omega)$, $Q^h\psi$ satisfies $\int_\Omega \nabla(\psi - Q^h\psi)\cdot\nabla\chi\,dx = 0\ \forall\chi \in S_0^h$. We note the standard approximation results:

For $\psi \in W^{2,q}(\Omega)$, $m = 0$ and 1 and $q \in [1,\infty]$ if $d \le 2$ and $q \in (\frac{3}{2},\infty]$ if $d = 3$, we have that

$$|(I - \pi^h)\psi|_{m,q,\kappa} \le Ch^{2-m}|\psi|_{2,q,\kappa} \quad \forall\kappa \in \mathcal{T}^h\,; \tag{3.1}$$

and for $\psi \in H_0^1(\Omega)\cap H^2(\Omega)$ and $m = 1$ and 2,

$$|(I - Q^h)\psi|_{0,\Omega} + h|(I - Q^h)\psi|_{1,\Omega} \le Ch^m|\psi|_{m,\Omega}\,. \tag{3.2}$$

We also note that

$$|(I - \pi^h)\phi_\varepsilon^M(\chi)|_{0,\Omega} \le Ch|\pi^h(\phi_\varepsilon^M(\chi))|_{1,\Omega} \qquad \forall\chi \in S_0^h\,, \tag{3.3}$$

see [15], p.68.

We have seen in the previous section that under assumption $(D2)$, (2.47) holds true for both u_ε and u, if M is chosen sufficiently large, i.e. such that (2.38) holds. As a consequence we have that

$$\phi_\varepsilon^M(u_\varepsilon) = \phi_\varepsilon(u_\varepsilon),\ \phi_\varepsilon^M(u) = \phi_\varepsilon(u),\ \phi^M(u) = \phi(u),\ \phi^M(u_\varepsilon) = \phi(u_\varepsilon)\,. \tag{3.4}$$

We will from now on always assume that $(D2)$ holds and that M is fixed but satisfies (2.38). By (3.4) it is justified to use ϕ_ε^M [resp. ϕ^M] instead of ϕ_ε [resp. ϕ] in our (semi)discrete models. Moreover, we will strengthen $(D2)$ to

$$(D3) \qquad f \in H^1(0,T;L^2(\Omega))\cap L^\infty(Q_T)\,,\quad g \in H^2(\Omega)\cap H_0^1(\Omega)\,.$$

Analogously to (3.4) we have that

$$\phi_\varepsilon^M(g) = \phi_\varepsilon(g),\ \phi_\varepsilon^M(\pi^h g) = \phi_\varepsilon(\pi^h g),\ \phi^M(g) = \phi(g),\ \phi^M(\pi^h g) = \phi(\pi^h g)\,. \tag{3.5}$$

We consider the following continuous in time finite element approximation to (P_ε):

(P_ε^h) Find $u_\varepsilon^h \in L^2(0,T;S_0^h)$ with $v_\varepsilon^h \equiv \phi_\varepsilon^M(u_\varepsilon^h) \in H^1(0,T;L^2(\Omega))$ such that for all $\chi \in L^2(0,T;S_0^h)$,

$$\begin{aligned}\int_{Q_T} \{[\partial_t v_\varepsilon^h]\chi + \nabla u_\varepsilon^h \cdot \nabla \chi - f\chi\}\,dx\,dt &= 0\,,\\ v_\varepsilon^h(\cdot,0) &= \phi_\varepsilon^M(\pi^h g(\cdot)) \equiv \phi_\varepsilon(\pi^h g(\cdot))\,.\end{aligned}$$

Note the slight abuse of notation, since $v_\varepsilon^h \notin S^h$. However, this notation is very concise. It also suppresses the dependence on M which is justified since we consider a fixed M. (P^h), u^h and v^h are defined analogously.

Theorem 2.2 *Assume that (D3) and (B2) hold. Then for all $\varepsilon \in (0,\varepsilon_0]$ and $h > 0$ there exists a unique weak solution u_ε^h to (P_ε^h). Furthermore, we have that*

$$\varepsilon^{\frac{1}{2}}|\partial_t u_\varepsilon^h|_{0,Q_T} + L_M^{-\frac{1}{2}}|\partial_t v_\varepsilon^h|_{0,Q_T} + \|u_\varepsilon^h\|_{L^\infty(0,T;H^1(\Omega))} \le C\,, \tag{3.6}$$

$$\varepsilon|u_\varepsilon - u_\varepsilon^h|^2_{0,Q_T} + |v_\varepsilon - v_\varepsilon^h|^2_{0,Q_T} \le Ch^4\,. \tag{3.7}$$

Proof. Existence and uniqueness can be proved as in Lemmas 2.1 and 2.2; and (3.6) as (2.26). We now give the proof of (3.7), which is similar to that of (2.50). From (3.4), (P_ε) and (P_ε^h) we have for all $\chi \in L^2(0,T;S_0^h)$ that

$$\int_{Q_T} \{\partial_t[\phi_\varepsilon^M(u_\varepsilon) - \phi_\varepsilon^M(u_\varepsilon^h)]\chi + \nabla[Q^h u_\varepsilon - u_\varepsilon^h]\cdot\nabla\chi\}\,dx\,dt = 0\,.$$

Setting $e_\varepsilon^h \equiv Q^h u_\varepsilon - u_\varepsilon^h$ and choosing $\chi(x,t) \equiv \int_t^T e_\varepsilon^h(x,s)\,ds$, we obtain that

$$\int_{Q_T} [\phi_\varepsilon^M(u_\varepsilon) - \phi_\varepsilon^M(u_\varepsilon^h)]e_\varepsilon^h\,dx\,dt + \tfrac{1}{2}\left|\int_0^T e_\varepsilon^h\,dt\right|^2_{1,\Omega} = \int_\Omega [\phi_\varepsilon^M(g) - \phi_\varepsilon^M(\pi^h g)][\int_0^T e_\varepsilon^h\,dt]\,dx\,.$$

Hence, on noting Young's and Poincaré's inequalities,

$$\begin{aligned}&\int_{Q_T} [\phi_\varepsilon^M(u_\varepsilon) - \phi_\varepsilon^M(u_\varepsilon^h)][u_\varepsilon - u_\varepsilon^h]\,dx\,dt + \tfrac{1}{4}\left|\int_0^T e_\varepsilon^h\,dt\right|^2_{1,\Omega}\\ &\quad = \int_{Q_T} [\phi_\varepsilon^M(u_\varepsilon) - \phi_\varepsilon^M(u_\varepsilon^h)][(I-Q^h)u_\varepsilon]\,dx\,dt + C|\phi_\varepsilon^M(g) - \phi_\varepsilon^M(\pi^h g)|^2_{0,\Omega}\,.\end{aligned}$$

Noting (2.2) and (2.3) yields that

$$\begin{aligned}&\tfrac{\lambda}{2}\varepsilon|u_\varepsilon - u_\varepsilon^h|^2_{0,Q_T} + \tfrac{1}{4}L_M^{-1}|\phi_\varepsilon^M(u_\varepsilon) - \phi_\varepsilon^M(u_\varepsilon^h)|^2_{0,Q_T} + \tfrac{1}{4}\left|\int_0^T e_\varepsilon^h\,dt\right|^2_{1,\Omega}\\ &\quad \le C|(I-Q^h)u_\varepsilon|^2_{0,Q_T} + C|(I-\pi^h)g|^2_{0,\Omega}\,.\end{aligned}$$

The desired result (3.7) now follows from noting (3.2), (2.49), (3.1) and (3.4). ∎

We have the following analogue of Theorem 2.1.

Theorem 2.3 *Assume that (D3) and (B2) hold. Then there exists a unique weak solution u^h of (P^h). Furthermore, for all $\varepsilon \in (0, \varepsilon_0]$ we have that*

$$\varepsilon |u^h - u_\varepsilon^h|^2_{0,Q_T} + |v^h - v_\varepsilon^h|^2_{0,Q_T} \leq C \{ A_\varepsilon^{u^h}(T)\, \varepsilon^{\frac{p+1}{p-1}} + B_\varepsilon\, \varepsilon^{\frac{2p}{p-1}} \} \leq C \varepsilon^{\frac{p+1}{p-1}} . \tag{3.8}$$

Proof. The proof is a direct analogue of the proof of Theorem 2.1. Note however, that since we now have the globally Lipschitz continuous functions ϕ^M and ϕ_ε^M in place of ϕ and ϕ_ε, analogues of (2.47) for u_ε^h and u^h are not needed. ∎

As a consequence of Theorems 2.1, 2.2 and 2.3 we obtain on noting (3.4) and choosing $\varepsilon = C h^{\frac{4(p-1)}{p+1}}$ that

$$|u - u^h|_{0,Q_T} + |u - u_\varepsilon^h|_{0,Q_T} \leq C \varepsilon^{\frac{1}{p-1}} + C \varepsilon^{-\frac{1}{2}} h^2 \leq C h^{\frac{4}{p+1}} , \tag{3.9}$$

$$|v - v^h|_{0,Q_T} + |v - v_\varepsilon^h|_{0,Q_T} \leq C \varepsilon^{\frac{p+1}{2(p-1)}} + C h^2 \leq C h^2 . \tag{3.10}$$

2.3.1 The Effect of Numerical Integration

The approximation (P_ε^h) studied above is not practical since terms like $\partial_t [\phi_\varepsilon^M(u_\varepsilon^h)]\chi$ and $f\chi$ need to be integrated exactly. It is computationally more convenient to consider a scheme, which uses numerical integration on these terms. Below we introduce and analyse such a scheme.

The introduction of numerical integration into a scheme $(\hat{P}_\varepsilon^h)$ which we give below makes this scheme more difficult to analyse than (P_ε^h). In fact, our analysis will require a smoother regularization when $p \in (1,2)$:

For all $\varepsilon \in (0, \varepsilon_0]$ and $s \in \mathbb{R}$ we define

$$\tilde{\phi}_\varepsilon(s) := \left\{ \begin{array}{ll} \lambda[a_\varepsilon s^3 + b_\varepsilon s] , & |s| \leq \varepsilon^{\frac{1}{p-1}} \\ \phi(s) , & \varepsilon^{\frac{1}{p-1}} \leq |s| \\ \phi_\varepsilon(s) , & |s| < \infty \end{array} \right\} \begin{array}{l} \text{for } p \in (1,2) , \\ \\ \text{for } p \in [2, \infty) , \end{array} \tag{3.11}$$

where $a_\varepsilon := \frac{p-1}{2} \varepsilon^{\frac{p-3}{p-1}}$ and $b_\varepsilon := \frac{3-p}{2} \varepsilon$.

It is straightforward to check that for $p \in (1,2)$ we have $\tilde{\phi}_\varepsilon \in C^1(\mathbb{R})$ with $\tilde{\phi}_\varepsilon'(s) \geq \frac{\lambda}{2}\varepsilon$, $\forall s \in \mathbb{R}$ (if we used the cubic regularization for the whole range of p, this would not be true). Hence, $\tilde{\phi}_\varepsilon$ satisfies (2.2) [resp. (2.3)] with λ [resp. L_M] replaced by $\frac{\lambda}{2}$ [resp. $L_M := \lambda p \max\{M^{p-1}, \varepsilon_0\}$]. Similarly to (2.7) we define $\tilde{\phi}_\varepsilon^M$ and note that it again satisfies (2.2) and (2.3) with the aforementioned changes.

Since nearly all results of section 2.2 are based on (2.2) and (2.3) they also hold when ϕ_ε and ϕ_ε^M are replaced by $\tilde{\phi}_\varepsilon$ and $\tilde{\phi}_\varepsilon^M$. A close inspection of the proof of Lemma 2.3 (in particular (2.44), (2.45)) which is the only possible exception shows that it also holds true. Hence, we also have (3.4) and (3.5) for $\tilde{\phi}_\varepsilon$ and $\tilde{\phi}_\varepsilon^M$.

Although we will work in this subsection with $\tilde{\phi}_\varepsilon$ and $\tilde{\phi}_\varepsilon^M$, we can always refer back to the results obtained with ϕ_ε and ϕ_ε^M. The advantage of the smoother regularization however is the following bound which is valid for $p \in (1,2)$:

$$|(\tilde{\phi}_\varepsilon^M)''|_{0,\infty,\mathbb{R}} \leq C\varepsilon^{\frac{p-2}{p-1}}. \tag{3.12}$$

This will allow us to do the same kind of error analysis for $p \in [2,\infty)$ and $p \in (1,2)$.

Now, a quadrature rule is introduced as follows: given a simplex $\kappa \in \mathcal{T}^h$ having vertices $\{a_i\}_{i=1}^{d+1}$, for $\psi \in C(\bar{\kappa})$, define

$$Q_\kappa(\psi) := \int_\kappa \pi^h \psi \, dx \equiv \frac{1}{d+1} meas(\kappa) \sum_{i=1}^{d+1} \psi(a_i).$$

Then on denoting the inner product in $L^2(\Omega)$ by $\langle \cdot, \cdot \rangle$ we have that

$$\int_\Omega \psi_1 \psi_2 \, dx = \langle \psi_1, \psi_2 \rangle \approx \langle \psi_1, \psi_2 \rangle^h := \sum_{\kappa \in \mathcal{T}^h} Q_\kappa(\psi_1 \psi_2) \equiv \sum_{\kappa \in \mathcal{T}^h} \int_\kappa \pi^h(\psi_1 \psi_2) \, dx,$$

for all $\psi_1, \psi_2 \in C(\bar{\Omega})$.

With $|\psi|_h := [\langle \psi, \psi \rangle^h]^{\frac{1}{2}}$ we have the following standard results:

$$|\chi|_{0,\Omega} \leq |\chi|_h \leq C|\chi|_{0,\Omega} \qquad \forall \chi \in S^h \tag{3.13}$$

and, for $m = 0$ or 1

$$|\langle \chi_1, \chi_2 \rangle - \langle \chi_1, \chi_2 \rangle^h| \leq Ch^{1+m} |\chi_1|_{1,\Omega} |\chi_2|_{m,\Omega} \qquad \forall \chi_1, \chi_2 \in S^h. \tag{3.14}$$

A more practical approximation to (P_ε) than (P_ε^h) is then

$(\hat{P}_\varepsilon^h)$ Find $\hat{u}_\varepsilon^h \in L^2(0,T;S_0^h)$ with $\hat{v}_\varepsilon^h \equiv \pi^h[\tilde{\phi}_\varepsilon^M(\hat{u}_\varepsilon^h)] \in H^1(0,T;S_0^h)$ such that for all $\chi \in L^2(0,T;S_0^h)$,

$$\begin{aligned} \int_0^T \{\langle \partial_t \hat{v}_\varepsilon^h, \chi \rangle^h + \langle \nabla \hat{u}_\varepsilon^h, \nabla \chi \rangle - \langle f, \chi \rangle^h\} \, dt &= 0, \\ \hat{v}_\varepsilon^h(\cdot, 0) &= \pi^h[\tilde{\phi}_\varepsilon^M(g(\cdot))]. \end{aligned}$$

Again, $(\hat{P}^h)$, $\hat{u}^h$ and $\hat{v}^h$ are defined analogously. Below we will prove an analogue of Theorem 2.2. Note that because of numerical integration we obtain a more pessimistic bound on the discretization error. We also point out that a further assumption on $\mathcal{T}^h$ is needed for the proof:

(WA) The triangulation is such that (i) for $d = 2$ it is weakly acute; that is, for any pair of adjacent triangles the sum of opposite angles relative to a common side does not exceed π; and (ii) for $d = 3$ the angle between any two faces of the same tetrahedron must not exceed $\frac{\pi}{2}$.

We need the following consequence of (WA):

$$\begin{aligned} &\alpha \in W^{1,\infty}(\mathbb{R}),\ \alpha(0) = 0 \quad \text{and } 0 \le \alpha' \le L_\alpha < \infty \text{ a.e. in } \mathbb{R} \implies \\ &|\nabla \pi^h[\alpha(\chi)]|^2_{0,\Omega} \le L_\alpha \langle \nabla \chi, \nabla \pi^h[\alpha(\chi)] \rangle \qquad \forall \chi \in S^h_0 , \end{aligned}$$

e.g. see §2.4.2 in [34]. Hence, on noting that $0 \le (\tilde{\phi}^M_\varepsilon)' \le L_M$ and that $\chi \in S^h_0 \subset L^\infty(\Omega)$, we have for all $\varepsilon \in (0, \varepsilon_0]$,

$$|\pi^h[\tilde{\phi}^M_\varepsilon(\chi)]|_{1,\Omega} \le L_M |\chi|_{1,\Omega} \qquad \forall \chi \in S^h_0 . \tag{3.15}$$

Theorem 2.4 *Assume that (D3) and (B2) hold. Then for all $\varepsilon \in (0, \varepsilon_0]$ and $h > 0$ there exists a unique solution $\hat{u}^h_\varepsilon$ to $(\hat{P}^h_\varepsilon)$. Furthermore, we have that*

$$\varepsilon^{\frac{1}{2}} |\partial_t \hat{u}^h_\varepsilon|_{0,Q_T} + L_M^{-\frac{1}{2}} |\partial_t \hat{v}^h_\varepsilon|_{0,Q_T} + \|\hat{u}^h_\varepsilon\|_{L^\infty(0,T;H^1(\Omega))} \le C . \tag{3.16}$$

If in addition, (WA) *holds and $f \in L^2(0,T;H^2(\Omega))$, we have that*

$$\varepsilon |u_\varepsilon - \hat{u}^h_\varepsilon|^2_{0,Q_T} + |v_\varepsilon - \hat{v}^h_\varepsilon|^2_{0,Q_T} \le C \begin{cases} \left[\varepsilon^{\frac{p-3}{p-1}} h^4 + h^2\right] & , \text{ if } p \in (1,2) , \\ \left[\varepsilon^{\frac{p+1}{p-1}} + \varepsilon^{-1} h^4 + h^2\right], & \text{if } p \in [2,\infty) . \end{cases} \tag{3.17}$$

Proof. Existence, uniqueness of a solution and the bounds (3.16) follow similarly to (P^h_ε). Let $\hat{e}^h_\varepsilon := Q^h u_\varepsilon - \hat{u}^h_\varepsilon$. On noting (3.4) , subtracting $(\hat{P}^h_\varepsilon)$ from (P_ε), choosing $\chi(x,t) := \int_t^T \hat{e}^h_\varepsilon(x,s)\, ds$ and integrating by parts we obtain that

$$\begin{aligned} &\int_0^T \langle \tilde{\phi}^M_\varepsilon(u_\varepsilon) - \tilde{\phi}^M_\varepsilon(\hat{u}^h_\varepsilon), u_\varepsilon - \hat{u}^h_\varepsilon \rangle^h \, dt + \tfrac{1}{2} \left| \int_0^T \hat{e}^h_\varepsilon \, dt \right|^2_{1,\Omega} \\ &= \int_0^T \langle \tilde{\phi}^M_\varepsilon(u_\varepsilon) - \tilde{\phi}^M_\varepsilon(\hat{u}^h_\varepsilon), u_\varepsilon - Q^h u_\varepsilon \rangle^h + \int_0^T [\langle \tilde{\phi}^M_\varepsilon(u_\varepsilon), \hat{e}^h_\varepsilon \rangle^h - \langle \tilde{\phi}^M_\varepsilon(u_\varepsilon), \hat{e}^h_\varepsilon \rangle] \, dt \\ &\quad + \int_0^T [\langle F, \hat{e}^h_\varepsilon \rangle - \langle F, \hat{e}^h_\varepsilon \rangle^h] \, dt + \left[\langle \tilde{\phi}^M_\varepsilon(g), \int_0^T \hat{e}^h_\varepsilon \, dt \rangle - \langle \tilde{\phi}^M_\varepsilon(\pi^h(g)), \int_0^T \hat{e}^h_\varepsilon \, dt \rangle^h \right] \\ &\equiv T_1 + T_2 + T_3 + T_4 , \end{aligned} \tag{3.18}$$

where $F(x,t) := \int_0^t f(x,s)\, ds$.

First, we bound the left hand side of (3.18) from below. On noting (2.2) for $\tilde{\phi}_\varepsilon^M$ with λ replaced by $\frac{\lambda}{2}$, (2.3), the monotonicity and definition of π^h, (3.13) and (3.4) we have that

$$\begin{aligned}
&\int_0^T \langle \tilde{\phi}_\varepsilon^M(u_\varepsilon) - \tilde{\phi}_\varepsilon^M(\hat{u}_\varepsilon^h), u_\varepsilon - \hat{u}_\varepsilon^h\rangle^h = \int_{Q_T} \pi^h[(\tilde{\phi}_\varepsilon^M(u_\varepsilon) - \tilde{\phi}_\varepsilon^M(\hat{u}_\varepsilon^h))(u_\varepsilon - \hat{u}_\varepsilon^h)]\,dx\,dt \\
&\geq \tfrac{1}{4}\lambda\varepsilon \int_{Q_T} \pi^h[(u_\varepsilon - \hat{u}_\varepsilon^h)^2]\,dx\,dt + \tfrac{1}{2}L_M^{-1}\int_{Q_T} \pi^h[(\tilde{\phi}_\varepsilon^M(u_\varepsilon) - \tilde{\phi}_\varepsilon^M(\hat{u}_\varepsilon^h))^2]\,dx\,dt \\
&= \tfrac{1}{4}\lambda\varepsilon \int_0^T |u_\varepsilon - \hat{u}_\varepsilon^h|_h^2\,dt + \tfrac{1}{2}L_M^{-1}\int_0^T |\tilde{\phi}_\varepsilon^M(u_\varepsilon) - \tilde{\phi}_\varepsilon^M(\hat{u}_\varepsilon^h)|_h^2\,dt \\
&\geq \tfrac{1}{4}\lambda\varepsilon |\pi^h u_\varepsilon - \hat{u}_\varepsilon^h|_{0,Q_T}^2 + \tfrac{1}{2}L_M^{-1}|\pi^h v_\varepsilon - \hat{v}_\varepsilon^h|_{0,Q_T}^2 . \qquad (3.19)
\end{aligned}$$

Now we establish upper bounds for $T_1 - T_4$. On noting Young's inequality, (3.13), (3.2), (3.1) and (2.49) we obtain that

$$\begin{aligned}
T_1 &\equiv \int_0^T \langle \tilde{\phi}_\varepsilon^M(u_\varepsilon) - \tilde{\phi}_\varepsilon^M(\hat{u}_\varepsilon^h), u_\varepsilon - Q^h u_\varepsilon\rangle^h\,dt \\
&\leq \tfrac{1}{4}L_M^{-1}|\pi^h[\tilde{\phi}_\varepsilon^M(u_\varepsilon) - \tilde{\phi}_\varepsilon^M(\hat{u}_\varepsilon^h)]|_{0,Q_T}^2 + C|(\pi^h - Q^h)u_\varepsilon|_{0,Q_T}^2 \\
&\leq \tfrac{1}{4}L_M^{-1}|\pi^h v_\varepsilon - \hat{v}_\varepsilon^h|_{0,Q_T}^2 + Ch^4 . \qquad (3.20)
\end{aligned}$$

From (3.14), Cauchy's and Young's inequalities we have that

$$\begin{aligned}
T_2 &\equiv \int_0^T [\langle \tilde{\phi}_\varepsilon^M(u_\varepsilon), \hat{e}_\varepsilon^h\rangle^h - \langle \tilde{\phi}_\varepsilon^M(u_\varepsilon), \hat{e}_\varepsilon^h\rangle]\,dt \\
&= \int_0^T [\langle \pi^h\tilde{\phi}_\varepsilon^M(u_\varepsilon), \hat{e}_\varepsilon^h\rangle^h - \langle \pi^h\tilde{\phi}_\varepsilon^M(u_\varepsilon), \hat{e}_\varepsilon^h\rangle + \langle (\pi^h - I)\tilde{\phi}_\varepsilon^M(u_\varepsilon), \hat{e}_\varepsilon^h\rangle]\,dt \\
&\leq Ch^2\,\|\pi^h\tilde{\phi}_\varepsilon^M(u_\varepsilon)\|_{L^2(0,T;H^1(\Omega))}\,\|Q^h u_\varepsilon - \hat{u}_\varepsilon^h\|_{L^2(0,T;H^1(\Omega))} \\
&\qquad + C\varepsilon^{-1}|(I - \pi^h)\tilde{\phi}_\varepsilon^M(u_\varepsilon)|_{0,Q_T}^2 + \tfrac{\lambda}{16}\varepsilon|\hat{e}_\varepsilon^h|_{0,Q_T}^2 \qquad (3.21)
\end{aligned}$$

From (3.15), (3.1) and (2.49) we have that

$$\begin{aligned}
&\|\pi^h\tilde{\phi}_\varepsilon^M(u_\varepsilon)\|_{L^2(0,T;H^1(\Omega))} = \|\pi^h\tilde{\phi}_\varepsilon^M(\pi^h u_\varepsilon)\|_{L^2(0,T;H^1(\Omega))} \\
&\leq C\|\pi^h u_\varepsilon\|_{L^2(0,T;H^1(\Omega))} \leq C\|u_\varepsilon\|_{L^2(0,T;H^2(\Omega))} \leq C , \qquad (3.22)
\end{aligned}$$

and from (3.2), (2.48) and (3.16) that

$$\|Q^h u_\varepsilon - \hat{u}_\varepsilon^h\|_{L^2(0,T;H^1(\Omega))} \leq C . \qquad (3.23)$$

In order to obtain a bound on $|(I - \pi^h)\tilde{\phi}_\varepsilon^M(u_\varepsilon)|_{0,Q_T}^2$ we distinguish between $p \in [2,\infty)$ and $p \in (1,2)$. For $p \in [2,\infty)$ we have on noting (3.4) that

$$\begin{aligned}
&|(I - \pi^h)\tilde{\phi}_\varepsilon^M(u_\varepsilon)|_{0,Q_T}^2 = |(I - \pi^h)\phi_\varepsilon(u_\varepsilon)|_{0,Q_T}^2 \\
&\leq C\left[|\phi(u_\varepsilon) - \phi_\varepsilon(u_\varepsilon)|_{0,Q_T}^2 + |\pi^h[\phi(u_\varepsilon) - \phi_\varepsilon(u_\varepsilon)]|_{0,Q_T}^2 + |(I - \pi^h)\phi(u_\varepsilon)|_{0,Q_T}^2\right] \\
&\leq C\left[\varepsilon^{\frac{2p}{p-1}} + |(I - \pi^h)\phi(u_\varepsilon)|_{0,Q_T}^2\right] , \qquad (3.24)
\end{aligned}$$

where we have noted that for all $t \in [0,T]$, $|\phi(u_\varepsilon) - \phi_\varepsilon(u_\varepsilon)|_{0,\infty,\Omega} \leq C\varepsilon^{\frac{p}{p-1}}$, which follows from (2.1).

By (3.1), we have that

$$|(I-\pi^h)\phi(u_\varepsilon)|^2_{0,Q_T} = \int_0^T |(I-\pi^h)\phi(u_\varepsilon)|^2_{0,\Omega}\,dt \le Ch^4 \int_0^T |\phi(u_\varepsilon)|^2_{2,\Omega}\,dt\,. \tag{3.25}$$

On expanding $|\phi(u_\varepsilon)|^2_{2,\Omega}$ and noting that by (2.47) $|\phi'(u_\varepsilon)|_{0,\infty,\Omega}$, $|\phi''(u_\varepsilon)|_{0,\infty,\Omega} \le C$ for $p \ge 2$, we obtain that

$$|\phi(u_\varepsilon)|^2_{2,\Omega} \le C(|u_\varepsilon|^2_{2,\Omega} + |u_\varepsilon|^4_{1,4,\Omega}) \le C\left(\|u_\varepsilon\|^2_{2,\Omega} + \left[\|u_\varepsilon\|^{\frac{1}{2}}_{2,\Omega}|u_\varepsilon|^{\frac{1}{2}}_{0,\infty,\Omega}\right]^4\right),$$

where we have used a Gagliardo-Nirenberg inequality, e.g. see [49], p.1034. Hence, on noting (2.49) and (2.47) we obtain that

$$|(I-\pi^h)\phi(u_\varepsilon)|^2_{0,Q_T} \le Ch^4\,, \quad \text{if } p \in [2,\infty)\,. \tag{3.26}$$

For $p \in (1,2)$ we replace ϕ by $\tilde{\phi}^M_\varepsilon$ in (3.25), note (3.12) and proceed similarly to above to obtain that

$$|(I-\pi^h)\tilde{\phi}^M_\varepsilon(u_\varepsilon)|^2_{0,Q_T} \le C\varepsilon^{\frac{2(p-2)}{p-1}}h^4\,. \tag{3.27}$$

(3.21) – (3.27) yield that

$$T_2 \le \tfrac{\lambda}{16}\varepsilon|\hat{e}^h_\varepsilon|^2_{0,Q_T} + C\begin{cases} \left[\varepsilon^{\frac{p-3}{p-1}}h^4 + h^2\right] & , \text{ if } p \in (1,2)\,, \\ \left[\varepsilon^{\frac{p+1}{p-1}} + \varepsilon^{-1}h^4 + h^2\right], & \text{if } p \in [2,\infty)\,. \end{cases} \tag{3.28}$$

On treating T_3 similarly to T_2 (here we need $f \in L^2(0,T;H^2(\Omega))$) we obtain that

$$T_3 \le \tfrac{\lambda}{16}\varepsilon|\hat{e}^h_\varepsilon|^2_{0,Q_T} + Ch^2 + C\varepsilon^{-1}h^4\,. \tag{3.29}$$

Finally, we have to bound T_4. From (3.14) on noting Cauchy's, Young's and Poincaré's inequalities we obtain that

$$\begin{aligned} T_4 &\equiv \left[\langle\tilde{\phi}^M_\varepsilon(g), \int_0^T \hat{e}^h_\varepsilon\,dt\rangle - \langle\tilde{\phi}^M_\varepsilon(\pi^h g), \int_0^T \hat{e}^h_\varepsilon\,dt\rangle^h\right] \\ &= \left[\langle\pi^h\tilde{\phi}^M_\varepsilon(g), \int_0^T \hat{e}^h_\varepsilon\,dt\rangle - \langle\pi^h\tilde{\phi}^M_\varepsilon(g), \int_0^T \hat{e}^h_\varepsilon\,dt\rangle^h\right] + \langle(I-\pi^h)\tilde{\phi}^M_\varepsilon(g), \int_0^T \hat{e}^h_\varepsilon\,dt\rangle \\ &\le Ch^2|\pi^h\tilde{\phi}^M_\varepsilon(g)|_{1,\Omega}\left|\int_0^T \hat{e}^h_\varepsilon\,dt\right|_{1,\Omega} + |(I-\pi^h)\tilde{\phi}^M_\varepsilon(g)|_{0,\Omega}\left|\int_0^T \hat{e}^h_\varepsilon\,dt\right|_{0,\Omega} \\ &\le Ch^4|\pi^h\tilde{\phi}^M_\varepsilon(g)|^2_{1,\Omega} + C|(I-\pi^h)\tilde{\phi}^M_\varepsilon(g)|^2_{0,\Omega} + \tfrac{1}{4}\left|\int_0^T \hat{e}^h_\varepsilon\,dt\right|^2_{1,\Omega} \\ &\le \tfrac{1}{4}\left|\int_0^T \hat{e}^h_\varepsilon\,dt\right|^2_{1,\Omega} + C\begin{cases} \varepsilon^{\frac{2(p-2)}{p-1}}h^4 & , \text{ if } p \in (1,2)\,, \\ \left[\varepsilon^{\frac{2p}{p-1}} + h^4\right], & \text{if } p \in [2,\infty)\,, \end{cases} \end{aligned} \tag{3.30}$$

where on noting (3.5) for $\tilde{\phi}_\varepsilon$ and $\tilde{\phi}^M_\varepsilon$ we have treated the term $|\pi^h\tilde{\phi}^M_\varepsilon(g)|_{1,\Omega}$ [resp. $|(I-\pi^h)\tilde{\phi}^M_\varepsilon(g)|_{0,\Omega}$] similarly to $|\pi^h\tilde{\phi}^M_\varepsilon(u_\varepsilon)|_{1,\Omega}$ [resp. $|(I-\pi^h)\tilde{\phi}^M_\varepsilon(u_\varepsilon)|_{0,\Omega}$] above.

From (3.18), (3.19), (3.20), (3.28), (3.29) and (3.30) we obtain that

$$\lambda\varepsilon|\pi^h u_\varepsilon - \hat{u}_\varepsilon^h|^2_{0,Q_T} + L_M^{-1}|\pi^h v_\varepsilon - \hat{v}_\varepsilon^h]|^2_{0,Q_T} + \left|\int_0^T \hat{e}_\varepsilon^h\,dt\right|^2_{1,\Omega}$$
$$\leq C\begin{cases} \left[\varepsilon^{\frac{p-3}{p-1}}h^4 + h^2\right] & , \text{ if } p \in (1,2)\,, \\ \left[\varepsilon^{\frac{p+1}{p-1}} + \varepsilon^{-1}h^4 + h^2\right], & \text{if } p \in [2,\infty)\,. \end{cases} \tag{3.31}$$

The desired result (3.17) follows from (3.31) on noting $|(I-\pi^h)u_\varepsilon|^2_{0,Q_T} \leq Ch^4$, (3.27), (3.24) and (3.26). ∎

We have the following analogue of Theorems 2.1 and 2.3.

Theorem 2.5 *Assume that (D3) and (B2) hold. Then there exists a unique weak solution $\hat{u}^h$ of $(\hat{P}^h)$. Furthermore, for all $\varepsilon \in (0,\varepsilon_0]$ we have that*

$$\varepsilon|\hat{u}^h - \hat{u}_\varepsilon^h|^2_{0,Q_T} + |\hat{v}^h - \hat{v}_\varepsilon^h|^2_{0,Q_T} \leq C\varepsilon^{\frac{p+1}{p-1}}\,. \tag{3.32}$$

Proof. The proof proceeds similarly to the proof of Theorem 2.1, with the same remark as in the proof of Theorem 2.3. Note in addition, that for the term $|\pi^h[\tilde{\phi}^M_{\varepsilon_j}(\hat{u}^h_{\varepsilon_i}) - \tilde{\phi}^M_{\varepsilon_i}(\hat{u}^h_{\varepsilon_i})]|_{0,Q_T}$ we cannot use any possible non-degeneracy as in (2.51). ∎

It follows from Theorems 2.1, 2.4 and 2.5 on choosing $\varepsilon = Ch^{\frac{2(p-1)}{p+1}}$ that

$$|u-\hat{u}^h|_{0,Q_T} + |u-\hat{u}_\varepsilon^h|_{0,Q_T} \leq C\left[\varepsilon^{\frac{1}{p-1}} + \varepsilon^{-\frac{1}{2}}h + \begin{cases} \varepsilon^{\frac{-1}{p-1}}h^2\,, & p \in (1,2)\,, \\ 0 & , \ p \in [2,\infty) \end{cases}\right] \leq Ch^{\frac{2}{p+1}}, \tag{3.33}$$

$$|v-\hat{v}^h|_{0,Q_T} + |v-\hat{v}_\varepsilon^h|_{0,Q_T} \leq C\left[\varepsilon^{\frac{p+1}{2(p-1)}} + h + \begin{cases} \varepsilon^{\frac{p-3}{2(p-1)}}h^2\,, & p \in (1,2)\,, \\ \varepsilon^{-\frac{1}{2}}h^2 & , \ p \in [2,\infty) \end{cases}\right] \leq Ch\,. \tag{3.34}$$

On comparing (3.33) [resp. (3.34)] and (3.9) [resp. (3.10)] we see that the order of convergence with quadrature is exactly half as good as without quadrature.

2.4 A Fully Discrete Finite Element Approximation

We now discretize our problem both in space and time. We choose the time step τ in terms of h, the size of the mesh in space: $\tau \equiv Ch^q$, where q will be specified later. Given τ, we define

$$N := \lfloor T/\tau \rfloor \equiv \max\{n \in \mathbb{N}_0 : n \leq T/\tau\} \quad \text{and} \quad T_* := N\tau \leq T\,. \tag{4.1}$$

We assume that (D3) and (B2) hold and consider the following problem:

$(P^h_{\varepsilon,N})$ For $n = 1 \to N$ find $u^{h,n}_{\varepsilon,N} \in S^h_0$ such that for all $\chi \in S^h_0$,

$$\begin{aligned} \int_\Omega \frac{\phi^M_\varepsilon(u^{h,n}_{\varepsilon,N}) - \phi^M_\varepsilon(u^{h,n-1}_{\varepsilon,N})}{\tau}\chi\,dx + \int_\Omega \nabla u^{h,n}_{\varepsilon,N}\cdot\nabla\chi\,dx &= \int_\Omega f^n_N\chi\,dx\,, \\ \phi^M_\varepsilon(u^{h,0}_{\varepsilon,N}) &= \phi^M_\varepsilon(\pi^h g)\,. \end{aligned}$$

Here we remind the reader that we only want to use $\tilde{\phi}^M_\varepsilon$ in the case with numerical integration. The proof of existence and uniqueness of a solution to $(P^h_{\varepsilon,N})$ is similar to $(P^M_{\varepsilon,N})$. Also, analogously to $u^M_{\varepsilon,N}$, $u^{M,\mathcal{L}}_{\varepsilon,N}$, $v^M_{\varepsilon,N}$ and $v^{M,\mathcal{L}}_{\varepsilon,N}$ we define $u^h_{\varepsilon,N}$, $u^{h,\mathcal{L}}_{\varepsilon,N}$, $v^h_{\varepsilon,N}$ and $v^{h,\mathcal{L}}_{\varepsilon,N}$; and we have analogously to (2.19) and (2.20) that

$$\int_{Q_{T_*}} \{[\partial_t v^{h,\mathcal{L}}_{\varepsilon,N}]\chi + \nabla u^h_{\varepsilon,N}\cdot\nabla\chi - f_N\chi\}\,dx\,dt = 0, \quad \forall\chi \in L^2(0,T_*;S^h_0)\,, \tag{4.2}$$

$$\varepsilon^{\frac{1}{2}}|\partial_t u^{h,\mathcal{L}}_{\varepsilon,N}|_{0,Q_{T_*}} + L_M^{-\frac{1}{2}}|\partial_t v^{h,\mathcal{L}}_{\varepsilon,N}|_{0,Q_{T_*}} + \|u^{h,\mathcal{L}}_{\varepsilon,N}\|_{L^\infty(0,T_*;H^1(\Omega))} + \|u^h_{\varepsilon,N}\|_{L^\infty(0,T_*;H^1(\Omega))} \le C\,. \tag{4.3}$$

We note that C in (4.3) is independent of h and τ. This can be seen from the proof of Lemma 2.1. Furthermore, we have analogously to (2.23) – (2.25) that

$$|u^{h,\mathcal{L}}_{\varepsilon,N} - u^h_{\varepsilon,N}|_{0,Q_{T_*}} \le \tau\,|\partial_t u^{h,\mathcal{L}}_{\varepsilon,N}|_{0,Q_{T_*}}\,, \tag{4.4}$$

$$|v^{h,\mathcal{L}}_{\varepsilon,N} - v^h_{\varepsilon,N}|_{0,Q_{T_*}} \le \tau\,|\partial_t v^{h,\mathcal{L}}_{\varepsilon,N}|_{0,Q_{T_*}}\,, \tag{4.5}$$

$$|f - f_N|_{0,Q_{T_*}} \le \tau|\partial_t f|_{0,Q_{T_*}}\,. \tag{4.6}$$

The a priori bounds (4.3) – (4.6) enable us to prove a basic error bound for the time discretization. Despite being basic on combining this result with our bounds for the semidiscrete approximation we can improve on existing error bounds for the fully discrete approximation in the literature, see below.

Lemma 2.7 *Assume that (D3) and (B2) hold. Then for all $\varepsilon \in (0,\varepsilon_0]$, $h > 0$ and τ, N and T_* as above we have that*

$$\varepsilon|u^h_\varepsilon - u^h_{\varepsilon,N}|^2_{0,Q_{T_*}} + |v^h_\varepsilon - v^h_{\varepsilon,N}|^2_{0,Q_{T_*}} \le C\varepsilon^{-1}\tau^2\,. \tag{4.7}$$

Proof. Let $E_u \equiv u^h_\varepsilon - u^h_{\varepsilon,N}$, $E_v \equiv v^h_\varepsilon - v^h_{\varepsilon,N}$, $E^{\mathcal{L}}_v \equiv v^h_\varepsilon - v^{h,\mathcal{L}}_{\varepsilon,N}$ and $E_f \equiv f - f_N$. On setting

$$\chi(\cdot,t) \equiv \begin{cases} \int_t^{T_*} E_u(\cdot,s)\,ds\,, & \text{if } t \in [0,T_*]\,, \\ 0 \qquad\qquad\quad , & \text{if } t \in (T_*,T]\,, \end{cases}$$

subtracting (4.2) from (P^h_ε), and integrating by parts we obtain that

$$\int_0^{T_*}\langle E_v, E_u\rangle\,dt + \tfrac{1}{2}\left|\int_0^{T_*} E_u\,dt\right|^2_{1,\Omega} = \int_0^{T_*}\langle E_v - E^{\mathcal{L}}_v, E_u\rangle\,dt + \int_0^{T_*}\langle\int_0^t E_f\,ds, E_u\rangle\,dt\,.$$

Hence we have on noting (2.2), (2.3) and Young's inequality that

$$\lambda\varepsilon|E_u|^2_{0,Q_{T_*}} + L_M^{-1}|E_v|^2_{0,Q_{T_*}} \leq C\varepsilon^{-1}\left[|v^{h,\mathcal{L}}_{\varepsilon,N} - v^h_{\varepsilon,N}|^2_{0,Q_{T_*}} + \left|\int_0^t E_f\,ds\right|^2_{0,Q_{T_*}}\right]. \tag{4.8}$$

Next we note that by Hölder's inequality and (4.6),

$$\left|\int_0^t E_f\,ds\right|^2_{0,Q_{T_*}} \leq T_*^2|E_f|^2_{0,Q_{T_*}} \leq C\tau^2. \tag{4.9}$$

Now the desired result follows from (4.8) on noting (4.5), (4.3) and (4.9). ∎

It is straightforward to check that the above results remain true when we introduce numerical integration into the scheme $(P^h_{\varepsilon,N})$, i.e. $(\hat{P}^h_{\varepsilon,N})$ with solution $\hat{u}^h_{\varepsilon,N} \in L^\infty(0,T;S^h_0)$ and $\hat{v}^h_{\varepsilon,N} \equiv \pi^h\tilde{\phi}^M_\varepsilon(\hat{u}^h_{\varepsilon,N}) \in L^\infty(0,T;S^h_0)$. With numerical integration the analogous result to (4.7) is:

$$\varepsilon|\hat{u}^h_\varepsilon - \hat{u}^h_{\varepsilon,N}|^2_{0,Q_{T_*}} + |\hat{v}^h_\varepsilon - \hat{v}^h_{\varepsilon,N}|^2_{0,Q_{T_*}} \leq C\varepsilon^{-1}\tau^2, \tag{4.10}$$

We also note that we can discretize (P^h) and $(\hat{P}^h)$ in time without regularization. The analogous results to (3.8) and (3.32) are

$$\varepsilon|u^h_N - u^h_{\varepsilon,N}|^2_{0,Q_{T_*}} + |v^h_N - v^h_{\varepsilon,N}|^2_{0,Q_{T_*}} \leq C\varepsilon^{\frac{p+1}{p-1}}, \tag{4.11}$$

$$\varepsilon|\hat{u}^h_N - \hat{u}^h_{\varepsilon,N}|^2_{0,Q_{T_*}} + |\hat{v}^h_N - \hat{v}^h_{\varepsilon,N}|^2_{0,Q_{T_*}} \leq C\varepsilon^{\frac{p+1}{p-1}}, \tag{4.12}$$

where C is independent of h and τ, which can be seen on inspecting the proof of Theorem 2.1.

We can now summarize our results:

In the case without numerical integration we obtain from (4.11), (3.9), (3.10) and (4.7) on choosing $\varepsilon \equiv Ch^{\frac{4(p-1)}{p+1}}$ and $\tau \equiv Ch^{\frac{4p}{p+1}}$ that for $p > 1$,

$$|u - u^h_N|_{0,Q_{T_*}} + |u - u^h_{\varepsilon,N}|_{0,Q_{T_*}} \leq Ch^{\frac{4}{p+1}} \equiv C\tau^{\frac{1}{p}}, \tag{4.13}$$

$$|v - v^h_N|_{0,Q_{T_*}} + |v - v^h_{\varepsilon,N}|_{0,Q_{T_*}} \leq Ch^2 \equiv C\tau^{\frac{p+1}{2p}}. \tag{4.14}$$

In the case with numerical integration, on additionally assuming (WA) we obtain from (4.12), (3.33), (3.34) and (4.10) on choosing $\varepsilon \equiv Ch^{\frac{2(p-1)}{p+1}}$ and $\tau \equiv Ch^{\frac{2p}{p+1}}$ that for $p > 1$,

$$|u - \hat{u}^h_N|_{0,Q_{T_*}} + |u - \hat{u}^h_{\varepsilon,N}|_{0,Q_{T_*}} \leq Ch^{\frac{2}{p+1}} \equiv C\tau^{\frac{1}{p}}, \tag{4.15}$$

$$|v - \hat{v}^h_N|_{0,Q_{T_*}} + |v - \hat{v}^h_{\varepsilon,N}|_{0,Q_{T_*}} \leq Ch \equiv C\tau^{\frac{p+1}{2p}}. \tag{4.16}$$

We consider the case $p = 2$ for illustration: Due to (4.15) the best order of convergence is $O(h^{\frac{2}{3}})$, if we choose $\tau \equiv Ch^{\frac{4}{3}}$. This result improves on the result in [29], where the best order of convergence is $O(h^{\frac{1}{2}})$ with $\tau \equiv Ch^{\frac{3}{2}}$. Since we consider up to three space dimensions for (P) and a nonlinearity $\phi(s) \equiv \lambda\ \mathrm{sign}(s)\,|s|^p$, our results are also more general than those in [29], where only one space dimension is considered and $p = 2$ only.

2.5 An Improved Error Bound for the Time Discretization

Rulla's paper [39] analyses the backward Euler method for approximating the abstract Cauchy problem

$$\frac{dv(t)}{dt} + A(v(t)) \ni 0\,, \quad v(0) \in D(A)\,, \tag{5.1}$$

where A is assumed to be a maximal monotone (and possibly multivalued) operator on a Hilbert space H. The main result of [39] asserts that the rate of convergence of the backward Euler approximation is $O(\tau)$, τ being the uniform time step, if A is a subgradient.

If $f \equiv 0$, then (P) can be cast into the form (5.1) on setting $A := -\Delta \circ \phi^{-1}$ and $H := H^{-1}(\Omega)$, cf. [39], p. 86, and [9], p. 123.

We will now transfer Rulla's techniques to the regularized semidiscrete problem (P_ε^h) with $f \equiv 0$ and show $O(\tau)$ convergence of $v_{\varepsilon,N}^h$ towards v_ε^h in $L^2(Q_{T_*})$; more precisely, we will show that (4.7) can be improved to

$$\varepsilon |u_\varepsilon^h - u_{\varepsilon,N}^h|_{0,Q_{T_*}}^2 + |v_\varepsilon^h - v_{\varepsilon,N}^h|_{0,Q_{T_*}}^2 \le C\tau^2\,, \tag{5.2}$$

i.e. we can drop the factor ε^{-1} on the right hand side of (4.7). It can be seen that this improvement goes through to the case with numerical integration, $(\hat{P}_\varepsilon^h)$; i.e. (5.2) holds if we put 'hats' on the u's and v's.

It is possible to apply Rulla's techniques to our concrete situation without referring to the theory of maximal monotone operators. We will do this in the following three Lemmas which are adapted and simplified versions of Theorems 3 – 5 in [39]. This is similar to the approach in [23] where Rulla's techniques are applied to the porous medium equation. We point out however that contrary to a remark in [39] we can prove (5.2) and its analogue for numerical integration only when $f \equiv 0$, i.e. for the homogeneous case. The key difference in our application, as opposed to the Stefan and porous medium problem cited in [39] and [23] is that ϕ' is not uniformly bounded below away from zero.

Since in $(P_{\varepsilon,N}^h)$ we apply the backward Euler method to (P_ε^h) which is already discretized in space, we begin with recalling the notion of the "discrete inverse Laplacian". As before we will always assume that $(B2)$ holds.

Let $\mathcal{G}^h \in \mathcal{L}(L^2(\Omega), S_0^h)$ be the "discrete inverse Laplacian" defined by

$$\langle \nabla \mathcal{G}^h \eta, \nabla \chi \rangle = \langle \eta, \chi \rangle \qquad \forall \chi \in S_0^h\,, \tag{5.3}$$

where $\langle \cdot, \cdot \rangle$ again denotes the inner product in $L^2(\Omega)$. For $\eta \in L^2(\Omega)$ we set

$$\|\eta\|_{-1,h} := \langle \eta, \mathcal{G}^h \eta \rangle^{\frac{1}{2}}\,. \tag{5.4}$$

Since $\mathcal{G}^h$ does not depend on t, we have the following well-known (integration by parts) formula for $\eta \in H^1(0,T; L^2(\Omega))$:

$$\|\eta(T)\|_{-1,h}^2 - \|\eta(0)\|_{-1,h}^2 = 2\int_0^T \langle \partial_t \eta, \mathcal{G}^h \eta \rangle \, dt\,. \tag{5.5}$$

Throughout this section we write $\eta(t) \equiv \eta(\cdot, t)$ for notational convenience.

Finally, with τ, N and T_* as in section 2.4, we set $t_n := n\tau$, $n = 0 \to N$, and define a function μ^τ on $(0, T_*]$ by

$$\mu^\tau(t) := \frac{t_n - t}{\tau}, \quad t \in (t_{n-1}, t_n], \quad n = 1 \to N. \tag{5.6}$$

With this definition we have in a generalized sense that

$$v^h_{\varepsilon,N} - v^{h,\mathcal{L}}_{\varepsilon,N} = \tau\mu^\tau \partial_t v^{h,\mathcal{L}}_{\varepsilon,N}. \tag{5.7}$$

Lemma 2.8 *Assume that (D3), (B2) and $f \equiv 0$ hold. Then for all $\varepsilon \in (0, \varepsilon_0]$, $h > 0$ and $n = 1 \to N$ we have that*

$$\begin{aligned} &\|v^h_\varepsilon(t_n) - v^h_{\varepsilon,N}(t_n)\|^2_{-1,h} + 2\int_0^{t_n} \langle v^h_\varepsilon - v^h_{\varepsilon,N}, u^h_\varepsilon - u^h_{\varepsilon,N}\rangle\, dt + \tau\int_0^{t_n} \mu^\tau |u^h_\varepsilon - u^h_{\varepsilon,N}|^2_{1,\Omega}\, dt \\ &\quad = \|v^h_\varepsilon(0) - v^h_{\varepsilon,N}(0)\|^2_{-1,h} + \tau\int_0^{t_n} \mu^\tau |u^h_\varepsilon|^2_{1,\Omega}\, dt - \tfrac{\tau}{2}\int_0^{t_n} |u^h_{\varepsilon,N}|^2_{1,\Omega}\, dt. \end{aligned} \tag{5.8}$$

Proof. On setting $\eta := v^h_\varepsilon - v^{h,\mathcal{L}}_{\varepsilon,N} \in H^1(0,T; L^2(\Omega))$ we obtain from (5.5), (P^h_ε), $(P^h_{\varepsilon,N})$ and (5.3) that

$$\begin{aligned} &\|v^h_\varepsilon(t_n) - v^{h,\mathcal{L}}_{\varepsilon,N}(t_n)\|^2_{-1,h} - \|v^h_\varepsilon(0) - v^{h,\mathcal{L}}_{\varepsilon,N}(0)\|^2_{-1,h} = 2\int_0^{t_n} \langle \partial_t(v^h_\varepsilon - v^{h,\mathcal{L}}_{\varepsilon,N}), \mathcal{G}^h(v^h_\varepsilon - v^{h,\mathcal{L}}_{\varepsilon,N})\rangle\, dt \\ &\quad = -2\int_0^{t_n} \langle \nabla(u^h_\varepsilon - u^h_{\varepsilon,N}), \nabla\mathcal{G}^h(v^h_\varepsilon - v^{h,\mathcal{L}}_{\varepsilon,N})\rangle\, dt = -2\int_0^{t_n} \langle v^h_\varepsilon - v^{h,\mathcal{L}}_{\varepsilon,N}, u^h_\varepsilon - u^h_{\varepsilon,N}\rangle\, dt. \end{aligned}$$

Hence, on noting (5.7) and (4.2) it follows that

$$\begin{aligned} &\|v^h_\varepsilon(t_n) - v^{h,\mathcal{L}}_{\varepsilon,N}(t_n)\|^2_{-1,h} + 2\int_0^{t_n} \langle v^h_\varepsilon - v^h_{\varepsilon,N}, u^h_\varepsilon - u^h_{\varepsilon,N}\rangle\, dt \\ &\quad = \|v^h_\varepsilon(0) - v^{h,\mathcal{L}}_{\varepsilon,N}(0)\|^2_{-1,h} - 2\tau\int_0^{t_n} \mu^\tau \langle \partial_t v^{h,\mathcal{L}}_{\varepsilon,N}, u^h_\varepsilon - u^h_{\varepsilon,N}\rangle\, dt \\ &\quad = \|v^h_\varepsilon(0) - v^{h,\mathcal{L}}_{\varepsilon,N}(0)\|^2_{-1,h} + 2\tau\int_0^{t_n} \mu^\tau \langle \nabla u^h_{\varepsilon,N}, \nabla[u^h_\varepsilon - u^h_{\varepsilon,N}]\rangle\, dt \\ &\quad = \|v^h_\varepsilon(0) - v^{h,\mathcal{L}}_{\varepsilon,N}(0)\|^2_{-1,h} + \tau\int_0^{t_n} \mu^\tau \langle \nabla[u^h_\varepsilon + u^h_{\varepsilon,N}], \nabla[u^h_\varepsilon - u^h_{\varepsilon,N}]\rangle\, dt \\ &\qquad\qquad - \tau\int_0^{t_n} \mu^\tau \langle \nabla[u^h_\varepsilon - u^h_{\varepsilon,N}], \nabla[u^h_\varepsilon - u^h_{\varepsilon,N}]\rangle\, dt \\ &\quad = \|v^h_\varepsilon(0) - v^{h,\mathcal{L}}_{\varepsilon,N}(0)\|^2_{-1,h} + \tau\int_0^{t_n} \mu^\tau |u^h_\varepsilon|^2_{1,\Omega}\, dt - \tau\int_0^{t_n} \mu^\tau |u^h_{\varepsilon,N}|^2_{1,\Omega}\, dt \\ &\qquad\qquad - \tau\int_0^{t_n} \mu^\tau |u^h_\varepsilon - u^h_{\varepsilon,N}|^2_{1,\Omega}\, dt. \end{aligned} \tag{5.9}$$

The desired result (5.8) now follows from (5.9) on noting that $u^h_{\varepsilon,N}$ is piecewise constant in time and that $v^{h,\mathcal{L}}_{\varepsilon,N}(t_n) = v^h_{\varepsilon,N}(t_n)$, $n = 0 \to N$. ∎

Note that all the terms on the left hand side of (5.8) are non-negative; the second because of (2.3).

Next, we want to remove μ^τ in (5.8). To this end we need the following result which states that without source term the energy of the semidiscrete solution is non-increasing in time:

Let $f \equiv 0$. Then for all $\varepsilon > 0$ we have that

$$|u_\varepsilon^h(s_2)|_{1,\Omega}^2 \le |u_\varepsilon^h(s_1)|_{1,\Omega}^2 \qquad \forall\, 0 \le s_1 \le s_2 \le T\,. \tag{5.10}$$

(5.10) is easy to prove: Since S_0^h is finite dimensional, we have from (3.6) that $\partial_t u_\varepsilon^h \in L^2(0,T;S_0^h)$. Furthermore, we have that $\partial_t v_\varepsilon^h \equiv (\phi_\varepsilon^M)'(u_\varepsilon^h)\,\partial_t u_\varepsilon^h$. Hence, on defining $\chi \in L^2(0,T;S_0^h)$ by

$$\chi(s) := \begin{cases} 0 & , \text{ if } \ 0 \le s < s_1\,, \\ \partial_t u_\varepsilon^h & , \text{ if } \ s_1 \le s < s_2\,, \\ 0 & , \text{ if } \ s_2 \le s \le T \end{cases}$$

we obtain from (P_ε^h) with $f \equiv 0$ that

$$\int_{s_1}^{s_2} \langle (\phi_\varepsilon^M)'(u_\varepsilon^h), [\partial_t u_\varepsilon^h]^2 \rangle \, dt = -\tfrac{1}{2} \int_{s_1}^{s_2} \tfrac{d}{dt} |u_\varepsilon^h|_{1,\Omega}^2 \, dt = \tfrac{1}{2} \left[|u_\varepsilon^h(s_1)|_{1,\Omega}^2 - |u_\varepsilon^h(s_2)|_{1,\Omega}^2 \right].$$

Now (5.10) follows from noting that $(\phi_\varepsilon^M)' \ge 0$.

Lemma 2.9 *Assume that (D3), (B2) and $f \equiv 0$ hold. Then for all $\varepsilon \in (0, \varepsilon_0]$, $h > 0$ and $n = 1 \to N$ we have that*

$$\begin{aligned} &\|v_\varepsilon^h(t_n) - v_{\varepsilon,N}^h(t_n)\|_{-1,h}^2 + 2\int_0^{t_n} \langle v_\varepsilon^h - v_{\varepsilon,N}^h, u_\varepsilon^h - u_{\varepsilon,N}^h \rangle \, dt + \tfrac{\tau}{2} \int_0^{t_n} |u_\varepsilon^h - u_{\varepsilon,N}^h|_{1,\Omega}^2 \, dt \\ &\quad \le \|v_\varepsilon^h(\tfrac{\tau}{2}) - v_{\varepsilon,N}^h(0)\|_{-1,h}^2 + \tfrac{\tau^2}{4} |u_\varepsilon^h(0)|_{1,\Omega}^2 + \tau \int_0^{t_n} \left[|u_\varepsilon^h|_{1,\Omega}^2 - |u_{\varepsilon,N}^h|_{1,\Omega}^2 \right] dt\,. \end{aligned} \tag{5.11}$$

Proof. Since $\mu^\tau \ge \frac{1}{2}$ on $(t_{m-1}, t_{m-\frac{1}{2}}]$, $m = 1 \to N$, we have that

$$\tfrac{\tau}{2} \sum_{m=1}^{n} \int_{t_{m-1}}^{t_{m-\frac{1}{2}}} |u_\varepsilon^h - u_{\varepsilon,N}^h|_{1,\Omega}^2 \, dt \le \tau \int_0^{t_n} \mu^\tau |u_\varepsilon^h - u_{\varepsilon,N}^h|_{1,\Omega}^2 \, dt\,. \tag{5.12}$$

If instead of u_ε^h and v_ε^h we consider $\tilde{u}_\varepsilon^h$ and $\tilde{v}_\varepsilon^h$ defined by $\tilde{u}_\varepsilon^h(t) := u_\varepsilon^h(t + \frac{\tau}{2})$ and $\tilde{v}_\varepsilon^h(t) := v_\varepsilon^h(t + \frac{\tau}{2})$ we obtain analogously to Lemma 2.8 that

$$\begin{aligned} &\tau \int_0^{t_n} \mu^\tau |\tilde{u}_\varepsilon^h - u_{\varepsilon,N}^h|_{1,\Omega}^2 \, dt \quad (+ \textit{ non-negative terms}) \\ &\quad \le \|\tilde{v}_\varepsilon^h(0) - v_{\varepsilon,N}^h(0)\|_{-1,h}^2 + \tau \int_0^{t_n} \mu^\tau |\tilde{u}_\varepsilon^h|_{1,\Omega}^2 \, dt - \tfrac{\tau}{2} \int_0^{t_n} |u_{\varepsilon,N}^h|_{1,\Omega}^2 \, dt \end{aligned}$$

and hence, similarly to (5.12), that

$$\begin{aligned} &\tfrac{\tau}{2} \sum_{m=1}^{n} \int_{t_{m-1}}^{t_{m-\frac{1}{2}}} |u_\varepsilon^h(t + \tfrac{\tau}{2}) - u_{\varepsilon,N}^h(t)|_{1,\Omega}^2 \, dt \\ &\quad \le \|v_\varepsilon^h(\tfrac{\tau}{2}) - v_{\varepsilon,N}^h(0)\|_{-1,h}^2 + \tau \int_0^{t_n} \mu^\tau(t) |u_\varepsilon^h(t + \tfrac{\tau}{2})|_{1,\Omega}^2 \, dt - \tfrac{\tau}{2} \int_0^{t_n} |u_{\varepsilon,N}^h(t)|_{1,\Omega}^2 \, dt\,. \end{aligned}$$

Since $u^h_{\varepsilon,N}$ is constant in time on $(t_{m-1}, t_m]$, the change of variable $t \to t + \frac{\tau}{2}$ yields that

$$\begin{aligned}
&\tfrac{\tau}{2} \sum_{m=1}^{n} \int_{t_{m-\frac{1}{2}}}^{t_m} |u^h_\varepsilon(t) - u^h_{\varepsilon,N}(t)|^2_{1,\Omega}\, dt \\
&\quad \le \|v^h_\varepsilon(\tfrac{\tau}{2}) - v^h_{\varepsilon,N}(0)\|^2_{-1,h} + \tau \int_{t_n}^{t_{n+\frac{1}{2}}} \mu^\tau(t-\tfrac{\tau}{2}) |u^h_\varepsilon(t)|^2_{1,\Omega}\, dt + \tau \int_0^{t_n} \mu^\tau(t-\tfrac{\tau}{2}) |u^h_\varepsilon(t)|^2_{1,\Omega}\, dt \\
&\qquad - \tau \int_0^{t_{\frac{1}{2}}} \mu^\tau(t-\tfrac{\tau}{2}) |u^h_\varepsilon(t)|^2_{1,\Omega}\, dt - \tfrac{\tau}{2} \int_0^{t_n} |u^h_{\varepsilon,N}(t)|^2_{1,\Omega}\, dt\,, \qquad (5.13)
\end{aligned}$$

where $\mu^\tau(t) := \mu^\tau(t+\tau)$, if $t \in (-\tau, 0]$.

By (5.10), the fourth term on the right hand side of (5.13) dominates the second. Hence, on noting that $v^h_\varepsilon(0) = v^h_{\varepsilon,N}(0)$ we obtain from Lemma 2.8, (5.12), and (5.13) that

$$\begin{aligned}
&\|v^h_\varepsilon(t_n) - v^h_{\varepsilon,N}(t_n)\|^2_{-1,h} + 2 \int_0^{t_n} \langle v^h_\varepsilon - v^h_{\varepsilon,N}, u^h_\varepsilon - u^h_{\varepsilon,N} \rangle\, dt + \tfrac{\tau}{2} \int_0^{t_n} |u^h_\varepsilon - u^h_{\varepsilon,N}|^2_{1,\Omega}\, dt \\
&\le \|v^h_\varepsilon(\tfrac{\tau}{2}) - v^h_{\varepsilon,N}(0)\|^2_{-1,h} + \tau \int_0^{t_n} [\mu^\tau(t) + \mu^\tau(t-\tfrac{\tau}{2})] |u^h_\varepsilon(t)|^2_{1,\Omega}\, dt - \tau \int_0^{t_n} |u^h_{\varepsilon,N}(t)|^2_{1,\Omega}\, dt. \qquad (5.14)
\end{aligned}$$

It is easy to see that

$$\mu^\tau(t) + \mu^\tau(t - \tfrac{\tau}{2}) = \tfrac{1}{2} + \mu^{\frac{\tau}{2}}(t)\,. \qquad (5.15)$$

On noting (5.10) we have that

$$\begin{aligned}
\int_0^{t_n} \mu^{\frac{\tau}{2}}(t) |u^h_\varepsilon(t)|^2_{1,\Omega}\, dt &= \tfrac{1}{2} \int_0^{t_n} |u^h_\varepsilon(t)|^2_{1,\Omega}\, dt + \int_0^{t_n} [\mu^{\frac{\tau}{2}}(t) - \tfrac{1}{2}] |u^h_\varepsilon(t)|^2_{1,\Omega}\, dt \\
&= \tfrac{1}{2} \int_0^{t_n} |u^h_\varepsilon(t)|^2_{1,\Omega}\, dt + \sum_{j=1}^{2n} \int_{(j-1)\tau/2}^{j\tau/2} [\mu^{\frac{\tau}{2}}(t) - \tfrac{1}{2}] |u^h_\varepsilon(t)|^2_{1,\Omega}\, dt \\
&\le \tfrac{1}{2} \int_0^{t_n} |u^h_\varepsilon(t)|^2_{1,\Omega}\, dt + \sum_{j=1}^{2n} \tfrac{\tau}{2} \tfrac{1}{2} \left[|u^h_\varepsilon((j-1)\tfrac{\tau}{2})|^2_{1,\Omega} - |u^h_\varepsilon(j\tfrac{\tau}{2})|^2_{1,\Omega} \right] \\
&\le \tfrac{1}{2} \int_0^{t_n} |u^h_\varepsilon(t)|^2_{1,\Omega}\, dt + \tfrac{\tau}{4} |u^h_\varepsilon(0)|^2_{1,\Omega}\, dt\,. \qquad (5.16)
\end{aligned}$$

Now the desired result (5.11) follows from (5.14), (5.15) and (5.16). ∎

Finally, the right hand side of (5.11) can be bounded by $C\tau^2$:

Lemma 2.10 *Assume that (D3), (B2) and $f \equiv 0$ hold. Then for all $\varepsilon \in (0, \varepsilon_0]$, $h > 0$ and $n = 1 \to N$ we have that*

$$\begin{aligned}
&\|v^h_\varepsilon(t_n) - v^h_{\varepsilon,N}(t_n)\|^2_{-1,h} + 4 \int_0^{t_n} \langle v^h_\varepsilon - v^h_{\varepsilon,N}, u^h_\varepsilon - u^h_{\varepsilon,N} \rangle\, dt + \tau \int_0^{t_n} |u^h_\varepsilon - u^h_{\varepsilon,N}|^2_{1,\Omega}\, dt \\
&\qquad \le 2\tau^2 |u^h_\varepsilon(0)|^2_{1,\Omega}\,. \qquad (5.17)
\end{aligned}$$

Proof. On noting Cauchy's inequality, (5.4), (P_ε^h), (5.3), and (5.10) we have for the first term on the right hand side of (5.11) that

$$\begin{aligned}
\|v_\varepsilon^h(\tfrac{\tau}{2}) - v_{\varepsilon,N}^h(0)\|_{-1,h}^2 &= \|v_\varepsilon^h(\tfrac{\tau}{2}) - v_\varepsilon^h(0)\|_{-1,h}^2 \\
&= \|\int_0^{\frac{\tau}{2}} \partial_t v_\varepsilon^h \, dt\,\|_{-1,h}^2 \leq \tfrac{\tau}{2}\int_0^{\frac{\tau}{2}} \|\partial_t v_\varepsilon^h\|_{-1,h}^2 \, dt \\
&= \tfrac{\tau}{2}\int_0^{\frac{\tau}{2}} \langle\, \partial_t v_\varepsilon^h, \mathcal{G}^h \partial_t v_\varepsilon^h\rangle \, dt = -\tfrac{\tau}{2}\int_0^{\frac{\tau}{2}} \langle \nabla u_\varepsilon^h, \nabla \mathcal{G}^h \partial_t v_\varepsilon^h\rangle \, dt \\
&= -\tfrac{\tau}{2}\int_0^{\frac{\tau}{2}} \langle \partial_t v_\varepsilon^h, u_\varepsilon^h\rangle \, dt = \tfrac{\tau}{2}\int_0^{\frac{\tau}{2}} |u_\varepsilon^h|_{1,\Omega}^2 \, dt \\
&\leq \tfrac{\tau^2}{4}\, |u_\varepsilon^h(0)|_{1,\Omega}^2 . \qquad (5.18)
\end{aligned}$$

In order to deal with the last integral of (5.11) we define for all $s \in \mathbb{R}$,

$$\Psi_\varepsilon^M(s) := \int_0^s (\phi_\varepsilon^M)^{-1}(\sigma)\, d\sigma . \qquad (5.19)$$

Note that Ψ_ε^M is convex, since $(\phi_\varepsilon^M)^{-1}$ is monotone.

On noting (5.19) we obtain from (P_ε^h) that

$$\begin{aligned}
\int_0^{t_n} |u_\varepsilon^h|_{1,\Omega}^2 \, dt &= -\int_0^{t_n} \langle \partial_t v_\varepsilon^h, (\phi_\varepsilon^M)^{-1}(v_\varepsilon^h)\rangle \, dt \\
&= -\int_0^{t_n} \tfrac{d}{dt}\langle 1, \Psi_\varepsilon^M(v_\varepsilon^h)\rangle \, dt = \langle 1, \Psi_\varepsilon^M(v_\varepsilon^h(0))\rangle - \langle 1, \Psi_\varepsilon^M(v_\varepsilon^h(t_n))\rangle . \qquad (5.20)
\end{aligned}$$

Next, we have from $(P_{\varepsilon,N}^h)$ on noting (5.19) that

$$\begin{aligned}
-\int_0^{t_n} |u_{\varepsilon,N}^h|_{1,\Omega}^2 \, dt &= \int_0^{t_n} \langle \partial_t v_{\varepsilon,N}^{h,\mathcal{L}}, u_{\varepsilon,N}^h\rangle \, dt = \sum_{m=1}^n \langle u_{\varepsilon,N}^{h,m}, v_{\varepsilon,N}^{h,m} - v_{\varepsilon,N}^{h,m-1}\rangle \\
&= \sum_{m=1}^n \langle u_{\varepsilon,N}^{h,m-1}, v_{\varepsilon,N}^{h,m} - v_{\varepsilon,N}^{h,m-1}\rangle + \sum_{m=1}^n \langle u_{\varepsilon,N}^{h,m} - u_{\varepsilon,N}^{h,m-1}, v_{\varepsilon,N}^{h,m} - v_{\varepsilon,N}^{h,m-1}\rangle \\
&= \sum_{m=1}^n \langle (\Psi_\varepsilon^M)'(v_{\varepsilon,N}^{h,m-1}), v_{\varepsilon,N}^{h,m} - v_{\varepsilon,N}^{h,m-1}\rangle + \sum_{m=1}^n \langle u_{\varepsilon,N}^{h,m} - u_{\varepsilon,N}^{h,m-1}, v_{\varepsilon,N}^{h,m} - v_{\varepsilon,N}^{h,m-1}\rangle . \qquad (5.21)
\end{aligned}$$

By the convexity of Ψ_ε^M we have that

$$\begin{aligned}
\sum_{m=1}^n \langle (\Psi_\varepsilon^M)'(v_{\varepsilon,N}^{h,m-1}), v_{\varepsilon,N}^{h,m} - v_{\varepsilon,N}^{h,m-1}\rangle &\leq \sum_{m=1}^n \langle 1, \Psi_\varepsilon^M(v_{\varepsilon,N}^{h,m})\rangle - \langle 1, \Psi_\varepsilon^M(v_{\varepsilon,N}^{h,m-1})\rangle \\
&= \langle 1, \Psi_\varepsilon^M(v_{\varepsilon,N}^{h,n})\rangle - \langle 1, \Psi_\varepsilon^M(v_{\varepsilon,N}^{h,0})\rangle , \qquad (5.22)
\end{aligned}$$

and by $(P_{\varepsilon,N}^h)$ and Young's inequality,

$$\begin{aligned}
\sum_{m=1}^n \langle u_{\varepsilon,N}^{h,m} - u_{\varepsilon,N}^{h,m-1}, v_{\varepsilon,N}^{h,m} - v_{\varepsilon,N}^{h,m-1}\rangle &= \tau \sum_{m=1}^n \langle u_{\varepsilon,N}^{h,m} - u_{\varepsilon,N}^{h,m-1}, \tfrac{v_{\varepsilon,N}^{h,m} - v_{\varepsilon,N}^{h,m-1}}{\tau}\rangle \\
&= -\tau \sum_{m=1}^n \langle \nabla u_{\varepsilon,N}^{h,m} - \nabla u_{\varepsilon,N}^{h,m-1}, \nabla u_{\varepsilon,N}^{h,m}\rangle \leq \tfrac{\tau}{2}|u_{\varepsilon,N}^{h,0}|_{1,\Omega}^2 - \tfrac{\tau}{2}|u_{\varepsilon,N}^{h,n}|_{1,\Omega}^2 . \qquad (5.23)
\end{aligned}$$

On noting that $v^{h,0}_{\varepsilon,N} = v^h_\varepsilon(0)$ and $u^{h,0}_{\varepsilon,N} = u^h_\varepsilon(0)$ we obtain from (5.20) – (5.23) that

$$\int_0^{t_n} \left[|u^h_\varepsilon|^2_{1,\Omega} - |u^h_{\varepsilon,N}|^2_{1,\Omega} \right] dt \leq \tfrac{\tau}{2}|u^h_\varepsilon(0)|^2_{1,\Omega} - \tfrac{\tau}{2}|u^{h,n}_{\varepsilon,N}|^2_{1,\Omega} + \langle 1, \Psi^M_\varepsilon(v^{h,n}_{\varepsilon,N})\rangle - \langle 1, \Psi^M_\varepsilon(v^h_\varepsilon(t_n))\rangle .$$

By the convexity of Ψ^M_ε, (5.19), (5.3) and Young's inequality we have that

$$\begin{aligned} \langle 1, \Psi^M_\varepsilon(v^{h,n}_{\varepsilon,N})\rangle - \langle 1, \Psi^M_\varepsilon(v^h_\varepsilon(t_n))\rangle &\leq \langle (\Psi^M_\varepsilon)'(v^{h,n}_{\varepsilon,N}), v^{h,n}_{\varepsilon,N} - v^h_\varepsilon(t_n)\rangle \\ &= \langle u^{h,n}_{\varepsilon,N}, v^{h,n}_{\varepsilon,N} - v^h_\varepsilon(t_n)\rangle = \langle \nabla u^{h,n}_{\varepsilon,N}, \nabla \mathcal{G}^h[v^{h,n}_{\varepsilon,N} - v^h_\varepsilon(t_n)]\rangle \\ &\leq \tfrac{\tau}{2}|u^{h,n}_{\varepsilon,N}|^2_{1,\Omega} + \tfrac{1}{2\tau}\|v^{h,n}_{\varepsilon,N} - v^h_\varepsilon(t_n)\|^2_{-1,h} , \end{aligned}$$

and hence that

$$\tau \int_0^{t_n} \left[|u^h_\varepsilon|^2_{1,\Omega} - |u^h_{\varepsilon,N}|^2_{1,\Omega} \right] dt \leq \tfrac{\tau^2}{2}|u^h_\varepsilon(0)|^2_{1,\Omega} + \tfrac{1}{2}\|v^{h,n}_{\varepsilon,N} - v^h_\varepsilon(t_n)\|^2_{-1,h} . \tag{5.24}$$

From (5.11), (5.18) and (5.24) we obtain the desired result (5.17). ∎

The bounds (5.2) now follow from (5.17) on noting (2.2), (2.3) and the non-negativity of the first and third term in (5.17).

Similarly, on defining "the discrete inverse Laplacian with numerical integration" $\widehat{\mathcal{G}}^h \in \mathcal{L}(C(\bar\Omega); S^h_0)$ by

$$\langle \nabla \widehat{\mathcal{G}}^h \eta, \nabla \chi\rangle = \langle \eta, \chi\rangle^h \qquad \forall \chi \in S^h_0 \tag{5.25}$$

and introducing

$$\|\eta\|_{\widehat{-1},h} := \left[\langle \eta, \widehat{\mathcal{G}}^h \eta\rangle^h \right]^{\frac{1}{2}} \tag{5.26}$$

for $\eta \in C(\bar\Omega)$, it is straightforward to check that all the above results go through to the case with numerical integration (where, in addition, we use $\tilde\phi^M_\varepsilon$ instead of ϕ^M_ε) ; in particular, we have that

$$\varepsilon|\hat u^h_\varepsilon - \hat u^h_{\varepsilon,N}|^2_{0,Q_{T_*}} + |\hat v^h_\varepsilon - \hat v^h_{\varepsilon,N}|^2_{0,Q_{T_*}} \leq C\tau^2 . \tag{5.27}$$

In order to summarize our results for the homogeneous case, $f \equiv 0$, we again combine the bounds for the time discretization with those for the space discretization:

In the case without numerical integration we obtain from (4.11), (3.9), (3.10) and (5.2) on choosing $\varepsilon \equiv Ch^{\frac{4(p-1)}{p+1}}$ and $\tau \equiv Ch^2$ that for $p > 1$,

$$|u - u^h_N|_{0,Q_{T_*}} + |u - u^h_{\varepsilon,N}|_{0,Q_{T_*}} \leq Ch^{\frac{4}{p+1}} \equiv C\tau^{\frac{2}{p+1}} , \tag{5.28}$$

$$|v - v^h_N|_{0,Q_{T_*}} + |v - v^h_{\varepsilon,N}|_{0,Q_{T_*}} \leq Ch^2 \equiv C\tau . \tag{5.29}$$

In the case with numerical integration, on additionally assuming (WA) we obtain from (4.12), (3.33), (3.34) and (5.27) on choosing $\varepsilon \equiv Ch^{\frac{2(p-1)}{p+1}}$ and $\tau \equiv Ch$ that for $p > 1$,

$$|u - \hat u^h_N|_{0,Q_{T_*}} + |u - \hat u^h_{\varepsilon,N}|_{0,Q_{T_*}} \leq Ch^{\frac{2}{p+1}} \equiv C\tau^{\frac{2}{p+1}} , \tag{5.30}$$

$$|v - \hat v^h_N|_{0,Q_{T_*}} + |v - \hat v^h_{\varepsilon,N}|_{0,Q_{T_*}} \leq Ch \equiv C\tau . \tag{5.31}$$

2.6 Le Roux's Method for the Discretization in Time

As we have seen in Lemma 2.5 the homogeneous (P) has the important feature that its solution becomes extinct at a certain time T_{ex}. However, it is not clear whether the discrete solution we have considered so far by using the backward Euler method for the time discretization has the same property. Nevertheless, it is possible to give a semidiscretization (P_L) in time which does have finite time extinction. This was done by Le Roux [28]. Here we report on this method and we elaborate on a number of points in [28] which are interesting in our context. The extinction in finite time property goes through to the fully discrete problem $(\hat{P}_L^h)$ which we introduce at the end of this section.

Le Roux's semidiscretization is based on the following idea. Instead of approximating $\partial_t u^p$ by the finite difference $\tau^{-1}[(u^n)^p - (u^{n-1})^p]$ as in the backward Euler method, we first note that

$$\partial_t u^p \equiv \partial_t[(u^{p-1})^{\frac{p}{p-1}}] \equiv \tfrac{p}{p-1} u\, \partial_t u^{p-1}$$

and then approximate $\partial_t u^{p-1}$ by $\tau^{-1}[(u_L^n)^{p-1} - (u_L^{n-1})^{p-1}]$. The subscript L shall remind the reader that we are using Le Roux's method. Now, the semidiscrete approximation of (P) in time with $g \geq 0$, $g \not\equiv 0$ and $f \equiv 0$ is posed as:

(P_L) For $n \geq 1$, given $u_L^{n-1} \in C(\bar{\Omega})$ find $u_L^n \in H_0^1(\Omega)$ such that for all $\eta \in H_0^1(\Omega)$,

$$\lambda \tfrac{p}{p-1} \int_\Omega u_L^n \frac{(u_L^n)^{p-1} - (u_L^{n-1})^{p-1}}{\tau} \eta \, dx + \int_\Omega \nabla u_L^n \cdot \nabla \eta \, dx = 0, \tag{6.1}$$

$$(u_L^0)^p = g^p. \tag{6.2}$$

It is shown in [28] that if (S) of subsection 2.2.3 holds and $g \in H_0^1(\Omega) \cap C(\bar{\Omega})$ is as above, i.e. $g \geq 0$ and $g \not\equiv 0$, then for all $n \geq 1$, (6.1) has a maximal solution $\bar{u}_L^n \in H_0^1(\Omega) \cap C(\bar{\Omega})$ and a minimal solution $\underline{u}_L^n$. Moreover, it is shown that

$$0 \equiv \underline{u}_L^n \leq \bar{u}_L^n \leq |\bar{u}_L^{n-1}|_{0,\infty,\Omega}, \tag{6.3}$$

and there exists an $n_0 \in \mathbb{N}$ such that

$$\bar{u}_L^n > 0 \quad \text{in } \Omega \qquad \text{for} \qquad n = 1 \to n_0 - 1, \tag{6.4}$$

$$\bar{u}_L^n \equiv 0 \quad \text{in } \Omega \qquad \text{for all} \qquad n \geq n_0. \tag{6.5}$$

Furthermore, $\bar{u}_L^n$ is the unique positive solution of (6.1) for $n = 1 \to n_0 - 1$. For completeness we present Le Roux's proof of (6.5) which we have already used in an adapted version for the proof of (2.59) in Lemma 2.5:

Firstly, it suffices to prove that (6.5) holds for a certain $n \equiv n_0$. For, if this is the case, it follows inductively from (6.1) that $u_L^n \equiv 0$ is the unique solution of (6.1) for all $n \geq n_0$ and hence, (6.5) holds for all $n \geq n_0$. Next, we assume for the purpose of contradiction that $\bar{u}_L^n \not\equiv 0$, $\forall n \in \mathbb{N}$. Then from (S) and (2.57) we have that

$$|\bar{u}_L^n|_{0,p+1,\Omega} > 0 \qquad \forall n \in \mathbb{N}. \tag{6.6}$$

On choosing $\eta \equiv \bar{u}_L^n$ (6.1) yields for $n \geq 1$ that

$$\lambda\,|\bar{u}_L^n|_{0,p+1,\Omega}^{p+1} + \tfrac{p-1}{p}\tau|\bar{u}_L^n|_{1,\Omega}^2 = \lambda \int_\Omega (\bar{u}_L^n)^2 (\bar{u}_L^{n-1})^{p-1}\,dx\,.$$

Hence, on noting (S), (2.57) and Hölder's inequality we have that

$$\lambda|\bar{u}_L^n|_{0,p+1,\Omega}^{p+1} + \tfrac{p-1}{p}C_p^{-2}\tau\,|\bar{u}_L^n|_{0,p+1,\Omega}^2 \leq \lambda|\bar{u}_L^n|_{0,p+1,\Omega}^{p+1} + \tfrac{p-1}{p}\tau\,|\bar{u}_L^n|_{1,\Omega}^2 \leq \lambda|\bar{u}_L^n|_{0,p+1,\Omega}^2\,|\bar{u}_L^{n-1}|_{0,p+1,\Omega}^{p-1}.$$

By (6.6) we can divide the last equation by $\lambda\,|\bar{u}_L^n|_{0,p+1,\Omega}^2$ which yields that

$$|\bar{u}_L^n|_{0,p+1,\Omega}^{p-1} + \lambda^{-1}\tfrac{p-1}{p}C_p^{-2}\tau \leq |\bar{u}_L^{n-1}|_{0,p+1,\Omega}^{p-1}.$$

This yields inductively that

$$|\bar{u}_L^n|_{0,p+1,\Omega}^{p-1} \leq |u(\cdot,0)|_{0,p+1,\Omega}^{p-1} - \lambda^{-1}\tfrac{p-1}{p}C_p^{-2}n\tau\,. \tag{6.7}$$

Hence, we have for large n that $|\bar{u}_L^n|_{0,p+1,\Omega}^{p-1} \leq 0$ which contradicts (6.6).

From (6.7) we obtain an upper bound on the extinction time $T_{ex}^\tau \equiv n_0\tau$ for the semidiscrete solution:

$$T_{ex}^\tau \leq \lambda\tfrac{p}{p-1}C_p^2|u(\cdot,0)|_{0,p+1,\Omega}^{p-1}.$$

Note that this is the same bound as (2.60) for the extinction time of the continuous solution. We remark that analogously to [28] a lower bound can also be proved for the extinction time T_{ex}^τ of the semidiscrete solution:

$$\lambda\tfrac{p}{p-1}|u(\cdot,0)|_{0,p+1,\Omega}^{p+1}\,|u(\cdot,0)|_{1,\Omega}^{-2} \leq T_{ex}^\tau\,. \tag{6.8}$$

Moreover, it is shown in [28] that if for fixed $T = N\tau$, $\bar{v}_{L,N}^{\mathcal{L}} \in H^1(0,T;L^1(\Omega))$ is given by

$$\bar{v}_{L,N}^{\mathcal{L}}(t) := \tau^{-1}\left[(t-t_{n-1})\phi(\bar{u}_L^n) + (t_n - t)\phi(\bar{u}_L^{n-1})\right]\,, \quad \text{if } t \in [t_{n-1},t_n]\,,$$

where $t_n \equiv n\tau$, $n = 1 \to N$, then

$$\bar{v}_{L,N}^{\mathcal{L}} \to v \equiv \phi(u) \quad \text{in } C([0,T];L^1(\Omega)) \quad \text{as } N \to \infty\,, \tag{6.9}$$

where u is the weak solution of (P).

Note that in the limit case $p = 1$, i.e. the homogeneous heat equation, T_{ex}^τ from (6.8) becomes infinite. This is consistent with the well-known fact that no extinction occurs in the heat equation.

We also report that Le Roux [28] proves bounds on the decay of $|\bar{u}_L^n|_{0,\infty,\Omega}$: If n_0^* denotes the smallest n_0 such that (6.5) holds and $T_{ex}^\tau \equiv n_0^*\tau$, then for $n = 0 \to n_0^*$ and $t_n \equiv n\tau$ we have that

$$\left[\tfrac{p-1}{\lambda p C_p^2}\,(T_{ex}^\tau - t_n)\right]^{\frac{1}{p-1}} \le |\bar{u}_L^n|_{0,p+1,\Omega} \le \left(1 - \frac{t_n}{T_{ex}^\tau}\right)^{\frac{1}{p-1}} |u(\cdot,0)|_{0,p+1,\Omega}\,.$$

Le Roux points out that these bounds are the same as those proved for separable solutions of (P) in [8], where the factor $(1 - \frac{t}{T_{ex}})^{\frac{1}{p-1}}$ can be rewritten as $T^*(t)/T^*(0)$, see (2.66).

By the above properties Le Roux's method seems to be the appropriate method for the time discretization of (P). However, as far as we know no error estimates exist in the literature for Le Roux's method. (6.9) is shown in [28] using compactness arguments similar to those in our proof of existence of a weak solution u_ε^M of (P_ε^M) in subsection 2.2.1. Unfortunately, we were not able to improve on this situation and derive a priori error bounds for Le Roux's method.

Thus, although the backward Euler method lacks the qualitative properties Le Roux's method shares with the continuous problem (P), we prefer the backward Euler method since we were able to give a full error analysis for it in sections 2.4 and 2.5. Nonetheless, we will present some numerical results for Le Roux's method in section 2.7 and so we also have to report on the computational aspect of it.

In order to compute $\bar{u}_L^n$ Le Roux suggests the following linear iterative method:

Set $\bar{u}_L^{n,0} := |\bar{u}_L^{n-1}|_{0,\infty,\Omega}$, then for $k \ge 0$ find $\bar{u}_L^{n,k+1} \in H_0^1(\Omega)$ such that for all $\eta \in H_0^1(\Omega)$,

$$\begin{aligned} &\lambda \tfrac{p}{(p-1)\tau} \left\langle \left[p|\bar{u}_L^{n-1}|_{0,\infty,\Omega}^{p-1} - (\bar{u}_L^{n-1})^{p-1}\right] \bar{u}_L^{n,k+1}\,,\, \eta \right\rangle + \langle \nabla \bar{u}_L^{n,k+1}\,,\, \nabla \eta \rangle \\ &\qquad = \quad \lambda \tfrac{p}{(p-1)\tau} \left\langle \left[p|\bar{u}_L^{n-1}|_{0,\infty,\Omega}^{p-1} - (\bar{u}_L^{n,k})^{p-1}\right] \bar{u}_L^{n,k}\,,\, \eta \right\rangle, \end{aligned} \tag{6.10}$$

where we remind the reader that in this chapter $\langle \cdot, \cdot \rangle$ denotes the inner product in $L^2(\Omega)$.

Lemma 2.11 *Assume that (B1) holds. Let* $\bar{u}_L^{n,0} = |\bar{u}_L^{n-1}|_{0,\infty,\Omega}$. *Then the iterates* $(\bar{u}_L^{n,k+1})_{k\ge 0}$ *given by (6.10) satisfy*

$$\bar{u}_L^{n,0} \ge \bar{u}_L^{n,k} \ge \bar{u}_L^{n,k+1} \ge \bar{u}_L^n \tag{6.11}$$

(in fact, $\bar{u}_L^{n,k} \searrow \bar{u}_L^n$ *as* $k \to \infty$*).*

Le Roux [28] refers to [24] for the proof of Lemma 2.11. In the following we give some, but not all, of the details of the proof of Lemma 2.11. As in [28] we will assume throughout this subsection that (S) of subsection 2.2.3 holds and that $\bar{u}_L^0 \equiv g \in H_0^1(\Omega) \cap C(\bar{\Omega})$ is non-negative.

First we define a function f as follows: Given $u \in H_0^1(\Omega)$, set

$$f(u) := \lambda_{\frac{p}{(p-1)\tau}} u \left([\bar{u}_L^{n-1}]^{p-1} - u^{p-1} \right) .$$

Lemma 2.12 *f defined as above satisfies*

$$f(v) - f(u) \geq -\lambda_{\frac{p}{(p-1)\tau}} \left(p|\bar{u}_L^{n-1}|_{0,\infty,\Omega}^{p-1} - [\bar{u}_L^{n-1}]^{p-1} \right) (v-u)$$

for $0 \leq u \leq v \leq |\bar{u}_L^{n-1}|_{0,\infty,\Omega}$, i.e.

$$f(v) - f(u) \geq -m\,(v-u)$$

with a non-negative $m \in C(\bar{\Omega})$.

Proof. On noting the Mean Value Theorem we obtain that

$$\begin{aligned} f(v) - f(u) &= \lambda_{\frac{p}{(p-1)\tau}} \left\{ v \left([\bar{u}_L^{n-1}]^{p-1} - v^{p-1} \right) - u \left([\bar{u}_L^{n-1}]^{p-1} - u^{p-1} \right) \right\} \\ &= \lambda_{\frac{p}{(p-1)\tau}} \left\{ [\bar{u}_L^{n-1}]^{p-1} (v-u) - (v^p - u^p) \right\} \\ &= \lambda_{\frac{p}{(p-1)\tau}} \left\{ [\bar{u}_L^{n-1}]^{p-1} (v-u) - p\xi^{p-1}(v-u) \right\} \\ &\geq \lambda_{\frac{p}{(p-1)\tau}} \left([\bar{u}_L^{n-1}]^{p-1} - p|\bar{u}_L^{n-1}|_{0,\infty,\Omega}^{p-1} \right) (v-u) . \end{aligned}$$

Hence, we set $m \equiv \lambda_{\frac{p}{(p-1)\tau}} \left(p|\bar{u}_L^{n-1}|_{0,\infty,\Omega}^{p-1} - [\bar{u}_L^{n-1}]^{p-1} \right)$. Clearly, m is non-negative. Since $\bar{u}_L^{n-1} \subset C(\bar{\Omega})$, we also have that $m \in C(\bar{\Omega})$. ∎

With the help of f and m as in Lemma 2.12 we can rewrite (6.10) as

$$\begin{aligned} \langle \nabla \bar{u}_L^{n,k+1}, \nabla \eta \rangle + \langle m \bar{u}_L^{n,k+1}, \eta \rangle &= \langle f(\bar{u}_L^{n,k}) + m \bar{u}_L^{n,k}, \eta \rangle \qquad (6.12) \\ &\equiv \lambda_{\frac{p}{(p-1)\tau}} \left\langle \bar{u}_L^{n,k} \left(p|\bar{u}_L^{n,k}|_{0,\infty,\Omega}^{p-1} - [\bar{u}_L^{n,k}]^{p-1} \right), \eta \right\rangle \qquad (6.13) \end{aligned}$$

Note that since $\bar{u}_L^{n,0} \equiv \bar{u}_L^{n-1} \in C(\bar{\Omega})$ we obtain inductively from (6.12) on noting elliptic regularity and Sobolev embedding (we assume $(B1)$) that $\bar{u}_L^{n,k} \in H_0^1(\Omega) \cap H^2(\Omega) \subset C(\bar{\Omega})$ for all $k \geq 0$.

We also need the following well-known maximum principle (see Theorem 8.1 in [17]):

The Weak Maximum Principle (WMP):

Let $c \in L^\infty(\Omega)$ with $c \geq 0$. If $\psi \in H^1(\Omega)$ satisfies

$$\langle \nabla \psi, \nabla \eta \rangle + \langle c\psi, \eta \rangle \geq 0 \quad \forall \eta \in H_0^1(\Omega),\ \eta \geq 0,$$

then $\inf_\Omega \psi \geq \inf_{\partial\Omega} [\psi]_-$ in Ω.

We can now give most of the details for the proof of Lemma 2.11: We will prove by induction that

$$\bar{u}_L^{n,0} \geq \bar{u}_L^{n,k} \geq \bar{u}_L^{n,k+1} \geq 0 \qquad \forall k \geq 0 \tag{6.14}$$

holds. We then outline how this implies existence of $\bar{u}_L^n$ and

$$\bar{u}_L^{n,k} \geq \bar{u}_L^n \qquad \forall k \geq 0\,. \tag{6.15}$$

Obviously, (6.11) is an immediate consequence of (6.14) and (6.15).

We begin with the proof of (6.14).

$\underline{k = 0:}$ The first inequality in (6.14) is trivial. Next, we have on noting (6.12) that

$$\langle \nabla(\bar{u}_L^{n,0} - \bar{u}_L^{n,1}), \nabla\eta\rangle = -\langle \nabla \bar{u}_L^{n,1}, \nabla\eta\rangle = \langle m(\bar{u}_L^{n,1} - \bar{u}_L^{n,0}) - f(\bar{u}_L^{n,0}), \eta\rangle$$

and hence that $\langle \nabla(\bar{u}_L^{n,0} - \bar{u}_L^{n,1}), \nabla\eta\rangle + \langle m(\bar{u}_L^{n,0} - \bar{u}_L^{n,1}), \eta\rangle = -\langle f(\bar{u}_L^{n,0}), \eta\rangle \geq 0$ for all $\eta \in H_0^1(\Omega)$ with $\eta \geq 0$. Thus, the second inequality in (6.14) follows for $k = 0$ from noting (WMP). From (6.13) we have that

$$\langle \nabla \bar{u}_L^{n,1}, \nabla\eta\rangle + \langle m\bar{u}_L^{n,1}, \eta\rangle = \lambda \tfrac{p}{(p-1)\tau} \left\langle \bar{u}_L^{n,0} \left(p|\bar{u}_L^{n,0}|_{0,\infty,\Omega}^{p-1} - [\bar{u}_L^{n,0}]^{p-1} \right), \eta \right\rangle \geq 0$$

for all $\eta \in H_0^1(\Omega)$ with $\eta \geq 0$ and the third inequality in (6.14) follows again from (WMP).

$\underline{k \to k+1:}$ We assume that (6.14) holds for a fixed $k \geq 0$ and want to show that

$$\bar{u}_L^{n,0} \geq \bar{u}_L^{n,k+1} \geq \bar{u}_L^{n,k+2} \geq 0\,. \tag{6.16}$$

The first inequality in (6.16) follows from the inductive hypothesis. Next, we have from (6.12) that

$$\begin{aligned} &\langle \nabla(\bar{u}_L^{n,k+1} - \bar{u}^{n,k+2}), \nabla\eta\rangle + \langle m(\bar{u}_L^{n,k+1} - \bar{u}_L^{n,k+2}), \eta\rangle \\ &\quad = \quad \langle f(\bar{u}_L^{n,k}) - f(\bar{u}_L^{n,k+1}) + m(\bar{u}_L^{n,k} - \bar{u}_L^{n,k+1}), \eta\rangle\,, \quad \forall \eta \in H_0^1(\Omega)\,. \end{aligned}$$

By the inductive hypothesis we can apply Lemma 2.12 and obtain that

$$\langle \nabla(\bar{u}_L^{n,k+1} - \bar{u}_L^{n,k+2}), \nabla\eta\rangle + \langle m(\bar{u}_L^{n,k+1} - \bar{u}_L^{n,k+2}), \eta\rangle \geq 0.$$

for all $\eta \in H_0^1(\Omega)$ with $\eta \geq 0$. Hence, on noting (WMP) we obtain the second inequality in (6.16). From (6.13) we have that

$$\langle \nabla \bar{u}_L^{n,k+2} + m\bar{u}_L^{n,k+2}, \nabla\eta\rangle = \lambda \tfrac{p}{(p-1)\tau} \left\langle \bar{u}_L^{n,k+1} \left(p|\bar{u}_L^{n,0}|_{0,\infty,\Omega}^{p-1} - \bar{u}_L^{n,k+1} \right), \eta \right\rangle$$

and hence, by the inductive hypothesis that $\langle \nabla \bar{u}_L^{n,k+2}, \nabla\eta\rangle + \langle m\bar{u}_L^{n,k+2}, \eta\rangle \geq 0$ for all $\eta \in H_0^1(\Omega)$ with $\eta \geq 0$, which on noting (WMP) completes the proof of (6.16) and hence (6.14).

It follows from (6.14) that there exists a $w \in L^\infty(\Omega)$ such that $\bar{u}_L^{n,k} \searrow w \geq 0$ pointwise as $k \to \infty$. Now define the inverse Laplacian $\mathcal{G} \in \mathcal{L}(L^2(\Omega), H_0^1(\Omega))$ to be such that for all $\xi \in L^2(\Omega)$

$$\langle \nabla \mathcal{G}\xi, \nabla \eta \rangle = \langle \xi, \eta \rangle, \quad \forall \eta \in H_0^1(\Omega)\,.$$

Assuming $(B1)$ it follows that $\mathcal{G} \in \mathcal{L}(L^2(\Omega), H_0^1(\Omega)) \cap H^2(\Omega))$. In addition, we have on noting Sobolev embedding that $\mathcal{G} \in \mathcal{L}(L^2(\Omega), H_0^1(\Omega) \cap C(\bar{\Omega}))$; and it is well-known that this $\mathcal{G}$ is a compact operator. One can now show on noting (6.14) that w solves (6.1). Moreover, it can be shown that w is the maximal solution of (6.1), i.e. $w \equiv \bar{u}_L^n$. This is a rough outline of the proof of existence of $\bar{u}_L^n$ and of (6.3). (6.15) is now obvious, since by the above $\bar{u}_L^{n,k} \searrow \bar{u}_L^n$. ∎

Le Roux's method has a further interesting property which is not mentioned in [28]: Applied to the ODE equivalent of (P) it yields the exact solution at every mesh point $t_n \equiv n\tau$. To see this, we consider for $\alpha > 0$ the ODE

$$\tfrac{d}{dt} u^p + \alpha u = 0\,, \qquad u(0) = g \geq 0 \tag{6.17}$$

which can be regarded as a simple ODE model for (P). It is easy to see that the exact solution of (6.17) is given by

$$u(t) = [g^{p-1} - \alpha \tfrac{p-1}{p} t]_+^{\frac{1}{p-1}}\,, \qquad \forall t \geq 0\,. \tag{6.18}$$

From the Le Roux discretization of (6.17),

$$\tfrac{p}{p-1}\, u_L^n \, \frac{(u_L^n)^{p-1} - (u_L^{n-1})^{p-1}}{\tau} + \alpha u_L^n = 0\,, \tag{6.19}$$

and $u_L^0 = g$ we obtain that for $n \geq 1$ the maximal solution $\bar{u}_L^n$ of (6.19) is given by

$$\bar{u}_L^n = [(\bar{u}_L^{n-1})^{p-1} - \alpha \tfrac{p-1}{p} \tau]_+^{\frac{1}{p-1}}\,. \tag{6.20}$$

It follows inductively that

$$\bar{u}_L^n = [g^{p-1} - \alpha \tfrac{p-1}{p} t_n]_+^{\frac{1}{p-1}}\,, \qquad \forall n \geq 0\,.$$

On comparing the last result with (6.18) we have that

$$\bar{u}_L^n = u(t_n)\,, \qquad \forall n \geq 0\,. \tag{6.21}$$

Finally, we remark that Le Roux's idea can also be used on (6.17) with p = 1, i.e. a linear ODE. The key is to note that

$$\tfrac{d}{dt}u \equiv \tfrac{d}{dt}(e^{\ln u}) \equiv u\,\tfrac{d}{dt}\ln u\,.$$

Hence, on approximating $\frac{d}{dt}\ln u$ by $\tau^{-1}[\ln u^n - \ln u^{n-1}]$ we obtain the numerical scheme

$$u^n\,\frac{\ln u^n - \ln u^{n-1}}{\tau} + \alpha u^n = 0\,, \quad u^0 = g\,. \tag{6.22}$$

Again, with $\bar{u}^n$ denoting the maximal solution of (6.22) it is straightforward to see that

$$\bar{u}^n = u(t_n) \equiv e^{-\alpha t_n} g\,, \quad \forall n \geq 0\,, \tag{6.23}$$

Also note that (6.21) and (6.23) remain valid for non-equidistant time steps τ.

Finally we also discretize in space: the fully discrete Le Roux's method with numerical integration is given for fixed time step τ and fixed mesh size h by

$(\hat{P}_L^h)$ For $n \geq 1$ find $\hat{u}_L^{h,n} \in S_0^h$ such that for all $\chi \in S_0^h$,

$$\lambda\tfrac{p}{p-1}\left\langle \hat{u}_L^{h,n}\,\frac{(\hat{u}_L^{h,n})^{p-1} - (\hat{u}_L^{h,n-1})^{p-1}}{\tau}, \chi \right\rangle^h + \langle \nabla \hat{u}_L^{h,n}, \nabla\chi\rangle = 0\,, \tag{6.24}$$

$$\pi^h\left[(\hat{u}_L^{h,0})^p\right] = \pi^h[g^p]\,. \tag{6.25}$$

For $n \geq 1$ the existence of a maximal solution of (6.24) can be shown similarly to the outlined existence proof for the semidiscrete problem (P_L) above. In this context we note that a discrete weak maximum principle is valid under assumption (WA) of section 2.3 and that we can replace the inverse Laplacian by the discrete inverse Laplacian with numerical integration which we defined in (5.25). In the following $\hat{u}_L^{h,n}$ will denote the maximal solution of (6.24). We also note that $\hat{u}_L^{h,n}$ has the finite time extinction property, i.e. it satisfies (6.5). This can be shown similarly to the semidiscrete situation by replacing the continuous L^p norms by their discrete analogues.

The fully discrete iterative linear method corresponding to (6.10) reads:

Set $\hat{u}_L^{h,n,0} := |\hat{u}_L^{h,n-1}|_{0,\infty,\Omega}$, then for $k \geq 0$ find $\hat{u}_L^{h,n,k+1} \in S_0^h$ such that for all $\chi \in S_0^h$,

$$\tfrac{\mu}{\tau}\left\langle p\left[|\hat{u}_L^{h,n-1}|_{0,\infty,\Omega}^{p-1} - (\hat{u}_L^{h,n-1})^{p-1}\right]\hat{u}_L^{h,n,k+1}, \chi\right\rangle^h + \langle \nabla \hat{u}_L^{h,n,k+1}, \nabla\chi\rangle$$

$$= \tfrac{\mu}{\tau}\left\langle p\left[|\hat{u}_L^{h,n-1}|_{0,\infty,\Omega}^{p-1} - (\hat{u}_L^{h,n,k})^{p-1}\right]\hat{u}_L^{h,n,k}, \chi\right\rangle^h\,, \tag{6.26}$$

where $\mu \equiv \lambda\frac{p}{p-1}$.

2.7 Numerical Experiments

In this section we deal with the practical solution of $(\hat{P}_N^h)$ and $(\hat{P}_L^h)$, the fully discrete problems with numerical integration. To keep matters simple we will assume throughout that

$$f \equiv 0 \quad \text{and} \quad g \geq 0 \quad \text{in } \Omega .$$

In subsection 2.7.1 we describe the system of nonlinear algebraic equations corresponding to $(\hat{P}_N^h)$ and its numerical treatment via the Newton method.

We then report on some numerical results in the case $d = 1$. Most of our computations deal with the case when (P) has a separable solution (cf. end of subsection 2.2.3). In subsection 2.7.2 we report on the discretization error for the fully discrete backward Euler method and compare our numerical results with the a priori error estimates of section 2.5. Our results suggest that the approximation error is determined by the behaviour in time of the exact solution. This will be studied further in subsection 2.7.3 both theoretically and computationally. In subsection 2.7.4 we exhibit some numerical results for $(\hat{P}_L^h)$ with an emphasis on finite time extinction.

Finally, in subsection 2.7.5 we consider the case of non-separable solutions of (P), again for $d = 1$. We will first define the notion of a 'computational extinction time'. Then we will present some computations which exhibit the profile in space of the fully discrete solutions of both the backward Euler and Le Roux's method near the computational extinction time. In doing this our aim is to investigate whether the fully discrete solutions of both $(\hat{P}_N^h)$ and $(\hat{P}_L^h)$ show the kind of asymptotic behaviour which was described in Lemma 2.6 for the exact solution of (P).

2.7.1 The System of Nonlinear Algebraic Equations for $(\hat{P}_N^h)$

Throughout the following we assume that $(D3)$ and $(B2)$ hold and that we are given a partitioning $\mathcal{T}^h$ of Ω which satisfies assumption (WA) of subsection 2.3.1. In practice, one considers a family $\{\mathcal{T}^h\}_h$ and $(\hat{P}_N^h)$ is solved for some selected h and N, i.e. τ.

In this subsection we let $h > 0$ and $N \in \mathbb{N}$ be given and fixed. We begin with recalling that $(\hat{P}_N^h)$ with $f \equiv 0$ is the following problem:

For $n = 1 \to N$ find $\hat{u}_N^{h,n} \in S_0^h$ such that for all $\chi \in S_0^h$

$$\left\langle \frac{\phi^M(\hat{u}_N^{h,n}) - \phi^M(\hat{u}_N^{h,n-1})}{\tau}, \chi \right\rangle^h + \langle \nabla \hat{u}_N^{h,n}, \nabla \chi \rangle = 0 , \tag{7.1}$$

$$\pi^h \phi^M(\hat{u}_N^{h,0}) = \pi^h \phi^M(g) . \tag{7.2}$$

We know from our previous analysis that if $\hat{u}_N^{h,0}$ is given by (7.2) there exists a unique solution $\hat{u}_N^{h,n}$ of (7.1) for $n = 1 \to N$.

We now also fix $n \in \{1, \ldots, N\}$. On choosing the standard basis functions $\{\chi_i\}_{i=1}^R$ for S_0^h we obtain the following nonlinear system for (7.1):

Find $\hat{u}_N^{h,n} = \sum_{j=1}^R U_j^n \chi_j$, i.e. find $U^n = (U_1^n, \ldots, U_R^n)^t \in \mathbb{R}^R$ such that

$$F(U^n) := \tau K U^n + E\Phi(U^n) = E\Phi(U^{n-1}), \tag{7.3}$$

where K and E are $R \times R$ matrices with entries $K_{ij} := \langle \nabla\chi_i, \nabla\chi_j \rangle$, $E_{ij} := \langle 1, \chi_i \rangle \delta_{ij}$ and $\Phi : \mathbb{R}^R \to \mathbb{R}^R$ is given by its component functions

$$\Phi_i(V) := \phi^M(V_i), \qquad i = 1 \to R, \quad \forall V \in \mathbb{R}^R. \tag{7.4}$$

Firstly, we note some properties of F: It is easy to see that F is differentiable with

$$DF(V) = \tau K + ED\Phi(V), \quad \forall V \in \mathbb{R}^R.$$

On noting that $\phi^M \in C^{1,1}(\mathbb{R})$ we have that DF is Lipschitz continuous. Since $ED\Phi(V) \geq 0 \ \forall V \in \mathbb{R}^R$ and K is an M-matrix under assumption (WA) (see §2.4.2 in [34]), $DF(V)$ must also be an M-matrix (see [36], 2.4.11). Noting that ϕ^M is convex on $\mathbb{R}_+$ we obtain that Φ and F are convex on $\mathbb{R}_+^R$.

By the aforementioned properties of F we obtain from the Monotone Newton Theorem (cf. 13.3.4 in [36]) that the Newton iteration

$$U^{n,0} = \begin{cases} \overline{G}, & \text{if } n = 1, \\ U^{n-1}, & \text{if } n > 1, \end{cases} \tag{7.5}$$

$$\left.\begin{aligned} [\tau K + ED\Phi(U^{n,k})]\, \delta U^{n,k} &= [\tau K U^{n,k} + E(\Phi(U^{n,k}) - \Phi(U^{n-1}))], \\ U^{n,k+1} &= U^{n,k} - \delta U^{n,k} \end{aligned}\right\} k = 0, 1, 2, \ldots, \tag{7.6}$$

where $\overline{G}$ denotes $(\overline{g}, \ldots, \overline{g})^t \in \mathbb{R}^R$ with $\overline{g} \in \mathbb{R}$ as defined after $(D2)$, has the properties

(i) $\quad U^{n,k} \searrow U^n \geq 0 \quad \text{as } k \to \infty,$

(ii) $\quad F(U^n) - E\Phi(U^n) = \tau K U^n = -E[\Phi(U^n) - \Phi(U^{n-1})] \geq 0,$

(iii) $\quad$ the convergence in (i) is quadratic,

provided

$$0 \leq U^{n,0} \quad \text{and} \quad F(0) = 0 \leq E\Phi(U^{n-1}) \leq F(U^{n,0}). \tag{7.7}$$

On noting (7.3) and (7.4) we obtain inductively from (i) and (ii) that (7.7) holds for all $n > 1$, if

$$0 \leq U^{1,0} \quad \text{and} \quad F(0) = 0 \leq E\Phi(U^0) \leq F(U^{1,0}). \tag{7.8}$$

Since we have that $0 \leq g \leq \overline{g}$, we also have that $0 \leq \overline{G} = U^{1,0}$. Furthermore, it follows from (7.2) that $0 \leq U^0$ and hence, by (7.3) and (7.4) that $F(0) = 0$. On the other hand, since $K\overline{G} \geq 0$ (as K is diagonally dominant, see e.g. §2.4.2 in [34]) and $\overline{g} \geq g$, we have

$$F_i(\overline{G}) = \tau(K\overline{G})_i + E\Phi^M(\overline{g}) \geq E\Phi^M(U_i^0), \quad i = 1 \to R.$$

Hence, (7.8) is satisfied.

When solving (7.3) in practice (n fixed), one needs a criterion for which $k > 0$ the Newton iteration (7.6) is to be stopped. In order to give such a criterion we define for notational convenience functions $\Psi : \mathbb{R}^R \to \mathbb{R}^R$ by

$$\Psi(V) := (\tau K)^{-1} E \Phi(V), \qquad \forall V \in \mathbb{R}^R.$$

With these definitions, (7.3) can be rewritten as

$$U^n + \Psi(U^n) = \Psi(U^{n-1}) =: R^n. \tag{7.9}$$

In order to see how well a Newton iterate $U^{n,k}$ from (7.6) solves (7.9), we set

$$U^{n,k} + \Psi(U^{n,k}) =: R^{n,k}. \tag{7.10}$$

On subtracting (7.9) from (7.10) we obtain that

$$[U^{n,k} - U^n] + [\Psi(U^{n,k}) - \Psi(U^n)] = R^{n,k} - R^n \equiv \mathrm{Res}^{n,k}. \tag{7.11}$$

Note that the first term on the left hand side of (7.11) is non-negative by (i). The second term is non-negative by (i), the monotonicity of Φ^M and since K is an M-matrix. Hence, we have that

$$0 \leq U^{n,k} - U^n \leq \mathrm{Res}^{n,k}. \tag{7.12}$$

In our practical computations we will therefore stop the Newton iteration (7.6) if

$$\max\{\mathrm{Res}_i^{n,k} : i = 1 \to R\} \equiv \max\{U_i^{n,k} + \Psi_i(U^{n,k}) - \Psi_i(U^{n-1}) : i = 1 \to R\} \leq Tol, \tag{7.13}$$

where Tol is a prescribed (small) tolerance. (7.13) is computationally not impractical for $d = 1$ as the linear system which has to be solved can be dealt with by fast standard methods, see below. If (7.13) is satisfied for a certain k, we set

$$U^n \equiv U^{n,k}. \tag{7.14}$$

We point out that what has been said in this subsection is valid for $d = 1$, 2 and 3. However, in the following subsections all our computations were only done for $d = 1$ and $\Omega \equiv (0,1)$. We quickly mention the practical peculiarities:

We chose equidistant meshes $\mathcal{T}^h$ on $[0,1]$, i.e. for fixed h there exists an $R \in \mathbb{N}$ s.t. $h = \frac{1}{R+1}$ and the mesh points are given by $x_i = ih$, $i = 0 \to R+1$. Straightforward calculations show that K is given by

$$K_{ij} = \left\{ \begin{array}{ll} \frac{2}{h}, & \text{if } i = j, \\ -\frac{1}{h}, & \text{if } |i-j| = 1, \\ 0, & \text{if } |i-j| > 0 \end{array} \right\} \quad \text{for } i,j = 1 \to R$$

and that

$$E_{ij} = h\,\delta_{ij}, \qquad i,j = 1 \to R.$$

We used a direct method to solve (7.6). Since $\tau K + ED\Phi^n(V)$ is an M matrix, we can use Gaussian elimination without pivoting. As it is also tridiagonal, one can store it as an $R \times 3$ or $3 \times R$ matrix. Algorithms which are adapted to this storage are available and solve (7.6) with $O(R)$ multiplications. We used the FORTRAN 77 code for an algorithm given in [31], p.424. Alternatively, on noting that $\tau K + ED\Phi^n(V)$ is symmetric and positive definite one could also use Cholesky factorization in order to solve the linear system (7.6).

For n fixed we chose $h = 2^{-l}$ in our computations, i.e. $R(l) = 2^l - 1$, $l = 3 \to L$. For fixed h we stopped the Newton iteration, if (7.13) was satisfied with $Tol := 10^{-8}$.

2.7.2 Numerical Results for $(\hat{P}_N^h)$ in the Separable Case

In this subsection we present some numerical results for the problem mentioned at the end of subsection 2.2.3:

Let $d = 1$, $\Omega \equiv (0,1)$, $\lambda = 1$, $p > 1$ and consider the problem

$$\partial_t u^p = \partial_{xx} u \quad \text{in } Q_T\,, \tag{7.15}$$

$$u(\cdot,0) = kw(\cdot) \quad \text{in } \Omega\,, \tag{7.16}$$

where $k = [(p-1)T_{ex}]^{\frac{1}{p-1}}$ and $w \in H_0^1(\Omega)$ is given by

$$-w'' = pw^p \quad \text{and} \quad w > 0 \quad \text{in } \Omega\,. \tag{7.17}$$

As mentioned previously, the exact solution of (7.15), (7.16) is separable and given by

$$u(x,t) = w(x)T^*(t), \quad \text{where } T^*(t) := \begin{cases} [(p-1)(T_{ex}-t)]^{\frac{1}{p-1}}, & t \in [0, T_{ex}), \\ 0 \qquad\qquad\qquad\quad , & t \in [T_{ex}, T]. \end{cases} \tag{7.18}$$

As there is no explicit solution to (7.17), we solved it numerically by using COLSYS, a standard boundary value problems package for ordinary differential equations.

Our main aim in this section is to check the error estimates of (5.30) and (5.31) computationally. Since there is no practical way of monitoring the errors $|u - \hat{u}_N^h|_{0,Q_{T_*}}$ and $|v - \hat{v}_N^h|_{0,Q_{T_*}}$ as they involve continuous L^2 norms, we resort to monitoring discrete errors which is a standard practice, see e.g. [5]. This is done as follows.

We recall that for a given discrete time step τ we set $t_n \equiv n\tau$, $n = 1 \to N$. Let

$$\begin{aligned} \Pi^{\tau,h} : C(\bar{Q}_T) \longrightarrow \{ w \in L^\infty(Q_T) : &\text{ For } n = 1 \to N,\ w \in C([0,1] \times (t_{n-1}, t_n])\,, \\ &\text{and for } n = 1 \to N \text{ and } i = 1 \to R+1\,, \\ &w \text{ linear in space and constant in time on } [x_{i-1}, x_i] \times (t_{n-1}, t_n] \} \end{aligned}$$

be given by

$$(\Pi^{\tau,h}\psi)(x_i,t) := \psi(x_i,t_n)\,, \qquad i = 1 \to R+1,\ t \in (t_{n-1}, t_n],\ n = 1 \to N\,.$$

On noting (3.13) we have that

$$|u - \hat{u}_N^h|_{0,Q_{T_*}} \leq |u - \Pi^{\tau,h}u|_{0,Q_{T_*}} + E_u^\tau(h)\,,$$

where

$$E_u^\tau(h) := \left[\tau \sum_{n=1}^{N} |u - \hat{u}_N^h|_h^2\right]^{\frac{1}{2}} \equiv \left[h\tau \sum_{n=1}^{N}\sum_{i=1}^{R} |(u - \hat{u}_N^h)(x_i, t_n)|^2\right]^{\frac{1}{2}}.$$

On assuming that $E_u^\tau(h) \approx Ch^{q_u}$, $q_u > 0$, we obtain that $2^{q_u}E_u^\tau(h) - E_u^\tau(2h) \approx 0$ and hence that $q_u \approx \log_2 \frac{E_u^\tau(2h)}{E_u^\tau(h)}$. For given $h > 0$ we therefore define the computational order of convergence as

$$q_u(h) := \log_2 \frac{E_u^\tau(2h)}{E_u^\tau(h)}\,.$$

Similarly, on defining $E_v^\tau(h)$ analogously to $E_u^\tau(h)$ and assuming that $E_v^\tau(h) \approx Ch^{q_v}$,

$$|v - \hat{v}_N^h|_{0,Q_{T_*}} \leq |v - \Pi^{\tau,h}v|_{0,Q_{T_*}} + E_v^\tau(h)\,, \quad q_v \approx q_v(h) := \log_2 \frac{E_v^\tau(2h)}{E_v^\tau(h)}\,.$$

Note that our computational orders of convergence $q_u(h)$ and $q_v(h)$ do not take into account the approximation errors $|u - \Pi^{\tau,h}u|_{0,Q_{T_*}}$ and $|v - \Pi^{\tau,h}v|_{0,Q_{T_*}}$. Theoretically these errors are also difficult to bound as we do not know the (existence of the) derivatives of u with respect to time and of v with respect to the space variable(s).

In Tables 2.1 and 2.2 we show $E_u^\tau(h)$ and $q_u(h)$ for $p = 2$ and 5, $h = 2^{-3}, 2^{-4}, \ldots$. The extinction time T_{ex} in (7.18) was chosen to be 0.25. We chose $T = 0.5$ and $\tau = h$ and h^2 from which it follows that T_*, the time end of the computation, coincides with a mesh point in time: $T_* = T = 2^{l-1} \times 2^{-l}$. For $\tau = h$ we computed up to $h = 2^{-12}$ (i.e. $\tau = 2^{-12} \approx 2.4 \times 10^{-4}$) and for $\tau = h^2$ up to $h = 2^{-8}$ (i.e. $\tau = 2^{-16} \approx 1.5 \times 10^{-5}$). In both cases the total number of mesh points on $[0,1] \times [0, 0.5]$ is about $2^{23} \approx 8.4 \times 10^6$.

	$\tau = h$		$\tau = h^2$	
h	$E_u^\tau(h)$	$q_u(h)$	$E_u^\tau(h)$	$q_u(h)$
2^{-3}	0.12035673	—	0.02226208	—
2^{-4}	0.06551429	0.8774	0.00605239	1.8790
2^{-5}	0.03543913	0.8865	0.00157847	1.9390
2^{-6}	0.01891377	0.9059	0.00040235	1.9720
2^{-7}	0.00994855	0.9269	0.00010142	1.9881
2^{-8}	0.00516550	0.9456	0.00002544	1.9952
2^{-9}	0.00265390	0.9608	—	—
2^{-10}	0.00135249	0.9725	—	—
2^{-11}	0.00068514	0.9812	—	—
2^{-12}	0.00034559	0.9873	—	—

Table 2.1: $E_u^\tau(h)$ and $q_u(h)$ for $p = 2$ with $T_{ex} = 0.25$ and $T_* = T = 0.5$

Table 2.1 suggests that for $p = 2$ the order of convergence is $O(\tau)$ whereas Table 2.2 suggests that for $p = 5$ the order of convergence is $O(\tau^{0.7})$. Note that both rates of convergence are better than predicted: (5.30) yields $O(\tau^{0.67})$ for $p = 2$ and $O(\tau^{0.33})$ for $p = 5$. We note however, that (5.30) bounds $|u - \hat{u}_N^h|_{0,Q_{T_*}}$; $E_u^\tau(h)$ may be better.

	$\tau = h$		$\tau = h^2$	
h	$E_u^\tau(h)$	$q_u(h)$	$E_u^\tau(h)$	$q_u(h)$
2^{-3}	0.20194757	—	0.05863779	—
2^{-4}	0.12685315	0.6708	0.02273196	1.3671
2^{-5}	0.07960638	0.6722	0.00871865	1.3825
2^{-6}	0.04992887	0.6730	0.00331439	1.3954
2^{-7}	0.03128963	0.6742	0.00125330	1.4030
2^{-8}	0.01958286	0.6761	0.00047234	1.4079
2^{-9}	0.01223322	0.6788	—	—
2^{-10}	0.00762385	0.6822	—	—
2^{-11}	0.00473867	0.6860	—	—
2^{-12}	0.00293754	0.6899	—	—

Table 2.2: $E_u^\tau(h)$ and $q_u(h)$ for $p = 5$ with $T_{ex} = 0.25$ and $T_* = T = 0.5$

The following Tables 2.3 and 2.4 are analogues of Tables 2.1 and 2.2 with u replaced by v; all other parameters are kept the same: $T_{ex} = 0.25$ and $T_* = 0.5$.

	$\tau = h$		$\tau = h^2$	
h	$E_v^\tau(h)$	$q_v(h)$	$E_v^\tau(h)$	$q_v(h)$
2^{-3}	0.07484528	—	0.01394176	—
2^{-4}	0.04041038	0.8892	0.00354062	1.9773
2^{-5}	0.02101655	0.9432	0.00088871	1.9942
2^{-6}	0.01072872	0.9700	0.00022241	1.9986
2^{-7}	0.00542223	0.9845	0.00005561	1.9996
2^{-8}	0.00272592	0.9921	0.00001390	1.9999
2^{-9}	0.00136670	0.9961	—	—
2^{-10}	0.00068428	0.9980	—	—
2^{-11}	0.00034238	0.9990	—	—
2^{-12}	0.00017125	0.9995	—	—

Table 2.3: $E_v^\tau(h)$ and $q_v(h)$ for $p = 2$ with $T_{ex} = 0.25$ and $T_* = T = 0.5$

	$\tau = h$		$\tau = h^2$	
h	$E_v^\tau(h)$	$q_v(h)$	$E_v^\tau(h)$	$q_v(h)$
2^{-3}	0.08191118	—	0.02366867	—
2^{-4}	0.03618808	1.1785	0.00547222	2.1128
2^{-5}	0.01701051	1.0891	0.00136223	2.0062
2^{-6}	0.00823919	1.0459	0.00034023	2.0014
2^{-7}	0.00405542	1.0227	0.00008504	2.0003
2^{-8}	0.00201256	1.0108	0.00002126	2.0001
2^{-9}	0.00100277	1.0050	—	—
2^{-10}	0.00050058	1.0023	—	—
2^{-11}	0.00025011	1.0010	—	—
2^{-12}	0.00012501	1.0005	—	—

Table 2.4: $E_v^\tau(h)$ and $q_v(h)$ for $p = 5$ with $T_{ex} = 0.25$ and $T_* = T = 0.5$

From Tables 2.3 and 2.4 we infer that the order of convergence is $O(\tau)$ as predicted by (5.31). Table 2.5 shows that in general we see the predicted order of convergence for $\hat{v}_N^h$ but a better order of convergence for $\hat{u}_N^h$ when we compare the computational results with our theory. As before the orders of convergence are expressed in terms of a power of h in Table 2.5 and $T_{ex} = 0.25$, $T_* = T = 0.5$.

	$\tau = h$		$\tau = h^2$		Theory $(\tau = Ch)$	
p	$q_u(2^{-12})$	$q_v(2^{-8})$	$q_u(2^{-12})$	$q_v(2^{-8})$	(5.30)	(5.31)
1.5	0.9984	0.9990	1.9998	1.9999	0.8000	1.0000
2.0	0.9873	0.9995	1.9952	1.9999	0.6667	1.0000
2.5	0.9306	0.9997	1.9213	1.9999	0.5714	1.0000
3.0	0.8571	0.9998	1.7719	2.0000	0.5000	1.0000
3.5	0.7969	1.0000	1.6399	2.0000	0.4444	1.0000
4.0	0.7522	1.0001	1.5396	2.0000	0.4000	1.0000
4.5	0.7176	1.0003	1.4641	2.0000	0.3636	1.0000
5.0	0.6899	1.0005	1.4079	2.0001	0.3333	1.0000

Table 2.5: Computational and theoretical orders of convergence for different values of p

Next, we want to remark on a special feature of the computations above: We chose T_{ex} and T so that they coincided with mesh points in time.

The Tables 2.6 and 2.7 are analogues of Tables 2.1 and 2.2 but with $T_{ex} = 0.2$ and $T = 0.4$. Now these points do not coincide with a mesh point in time. We have included the values of T_* according to (4.1) in separate columns.

	$\tau = h$			$\tau = h^2$		
h	T_*	$E_u^\tau(h)$	$q_u(h)$	T_*	$E_u^\tau(h)$	$q_u(h)$
2^{-3}	0.375000	0.08188825	—	0.390625	0.01856099	—
2^{-4}	0.375000	0.05260402	0.6385	0.398438	0.00517114	1.8437
2^{-5}	0.375000	0.02899146	0.8595	0.399414	0.00131920	1.9248
2^{-6}	0.390625	0.01619551	0.8400	0.399902	0.00034868	1.9656
2^{-7}	0.398438	0.00860460	0.9124	0.399963	0.00008806	1.9854
2^{-8}	0.398438	0.00454044	0.9223	0.399994	0.00002210	1.9941
2^{-9}	0.398438	0.00234278	0.9546	—	—	—
2^{-10}	0.399414	0.00120074	0.9643	—	—	—
2^{-11}	0.399902	0.00060957	0.9781	—	—	—
2^{-12}	0.399902	0.00030817	0.9841	—	—	—

Table 2.6: $E_u^\tau(h)$ and $q_u(h)$ for $p = 2$ with $T_{ex} = 0.2$ and $T = 0.4$

	$\tau = h$			$\tau = h^2$		
h	T_*	$E_u^\tau(h)$	$q_u(h)$	T_*	$E_u^\tau(h)$	$q_u(h)$
2^{-3}	0.375000	0.13761134	—	0.390625	0.05006703	—
2^{-4}	0.375000	0.05566873	1.3057	0.398438	0.01482894	1.7554
2^{-5}	0.375000	0.04781176	0.2195	0.399414	0.00772589	0.9406
2^{-6}	0.390625	0.04326621	0.1443	0.399902	0.00239636	1.6889
2^{-7}	0.398438	0.02400518	0.8498	0.399963	0.00113886	1.0733
2^{-8}	0.398438	0.01186907	1.0161	0.399994	0.00036664	1.6352
2^{-9}	0.398438	0.00861822	0.4617	—	—	—
2^{-10}	0.399414	0.00679117	0.3437	—	—	—
2^{-11}	0.399902	0.00384617	0.8202	—	—	—
2^{-12}	0.399902	0.00205132	0.9069	—	—	—

Table 2.7: $E_u^\tau(h)$ and $q_u(h)$ for $p = 5$ with $T_{ex} = 0.2$ and $T = 0.4$

Obviously, the fact that T_* does not have the same value for all time steps has an effect on the computational order of convergence. Because of the oscillatory behaviour especially for $\tau = h$ and $p = 5$ it is no longer clear what the computational order of convergence actually is. As far as analogues of Tables 2.6 and 2.7 for the computation of $\hat{v}_N^h$ are concerned, we report that although the numerical results showed an oscillatory behaviour for the computational orders of convergence, this was far less pronounced: for $\tau = h$ and $p = 5$ the values of $q_v(h)$ ranged between 1.0276 and 0.9897 for h ranging between 2^{-6} and 2^{-12}.

One can of course argue that in Tables 2.6 and 2.7 T_* depends on h, i.e. on τ, and one should therefore normalize $E_u^\tau(h)$ and $q_u(h)$:

$$\tilde{E}_u^\tau(h) := (\tfrac{T}{T_*})^{\frac{1}{2}} E_u^\tau(h)\,, \qquad \tilde{q}_u(h) := \log_2 \frac{\tilde{E}_u^\tau(2h)}{\tilde{E}_u^\tau(h)}\,.$$

However, as the comparison of Tables 2.8 and 2.7 shows the effect of this normalization is small compared to the oscillations in the computational order of convergence:

	$\tau = h$		$\tau = h^2$	
h	$\tilde{E}_u^\tau(h)$	$\tilde{q}_u(h)$	$\tilde{E}_u^\tau(h)$	$\tilde{q}_u(h)$
2^{-3}	0.14212438	—	0.05066428	—
2^{-4}	0.05749442	1.3057	0.01485798	1.7697
2^{-5}	0.04937977	0.2195	0.00773156	0.9424
2^{-6}	0.04378233	0.1736	0.00239665	1.6897
2^{-7}	0.02409941	1.0604	0.00113891	1.0734
2^{-8}	0.01189131	1.0190	0.00036664	1.6352
2^{-9}	0.00863510	0.4617	—	—
2^{-10}	0.00679615	0.3455	—	—
2^{-11}	0.00384664	0.8211	—	—
2^{-12}	0.00205157	0.9069	—	—

Table 2.8: $\tilde{E}_u^\tau(h)$ and $\tilde{q}_u(h)$ for $p = 5$ with $T_{ex} = 0.2$ and $T = 0.4$

We now remark upon the fact that above we computed a separable solution of (P) for $d = 1$. The generic form of the equation in time obtained from separating the variables in (7.15) is

$$\frac{d}{dt}v + \alpha u = 0 \quad \text{in } (0,T]\,, \quad v \equiv u^p\,. \tag{7.19}$$

For $\alpha = \frac{\beta p}{p-1}$ with $\beta > 0$ and the initial value

$$u(0) = g \geq 0 \tag{7.20}$$

the exact solution of (7.19), (7.20) is given by

$$u(t) = [g^{p-1} - \beta t]_+^{\frac{1}{p-1}}\,, \quad t \in [0,T]\,, \tag{7.21}$$

which is analogous to $T^*(t)$ in (7.18). Note that independently of $p > 1$, $u(t)$ vanishes for all $t \in [\frac{g^{p-1}}{\beta}, T]$. Hence, for

$$g = 1 \tag{7.22}$$

and $\beta = 4$ and $T = 0.5$ [resp. $\beta = 5$ and $T = 0.4$] we have a situation analogous to the computations for Tables 2.1 – 2.5 [resp. Tables 2.6 and 2.7]. We also note that in our practical computations of the ODE (7.19) the analogue of (7.13) reads as

$$U^{n,k} + \tfrac{1}{\alpha\tau}\left[(U^{n,k})^p - (U^{n-1})^p\right] \leq Tol\,. \tag{7.23}$$

In Table 2.9 we compare the computational orders of convergence of $(\hat{P}_N^h)$ for the PDE (7.15) with those obtained from applying the backward Euler method to the ODE (7.19). For the PDE the initial values are given by (7.16) and (7.17), for the ODE by (7.20) and (7.22). Note that $T_{ex} = 0.25$ and $T = 0.5$ coincide with mesh points. q_u and q_v in the fourth and fifth column are defined analogously to the PDE case. The numbers for the PDE are in good agreement with those for the ODE.

	PDE (7.15)		ODE (7.19)	
p	$q_u(2^{-12})$	$q_v(2^{-12})$	$q_u(2^{-14})$	$q_v(2^{-14})$
1.5	0.9984	0.9990	0.9996	0.9997
2.0	0.9873	0.9995	0.9946	0.9999
2.5	0.9306	0.9997	0.9467	0.9999
3.0	0.8571	0.9998	0.8708	0.9999
3.5	0.7969	1.0000	0.8064	0.9999
4.0	0.7522	1.0001	0.7589	0.9999
4.5	0.7176	1.0003	0.7239	0.9999
5.0	0.6899	1.0005	0.6964	1.0000

Table 2.9: Comp. orders of convergence for PDE and ODE with $T_{ex} = 0.25$ and $T = 0.5$

The following Table 2.10 is an analogue of Tables 2.6 and 2.7 for the ODE case where the extinction time $T_{ex} = 0.2$ is not a mesh point. $E_u(\tau)$ is defined analogously to $E_u^\tau(h)$ of the PDE case. It is quite remarkable how well the computational orders of convergence $q_u(\tau)$ for $p = 2$ below agree with the corresponding figures in Table 2.6.

		$p = 2$		$p = 4$	
τ	T_*	$E_u(\tau)$	$q_u(\tau)$	$E_u(\tau)$	$q_u(\tau)$
2^{-6}	0.390625	0.01986583	—	0.04209009	—
2^{-7}	0.398438	0.01056667	0.9108	0.02292756	0.8763
2^{-8}	0.398438	0.00557896	0.9215	0.01204525	0.9286
2^{-9}	0.398438	0.00287942	0.9542	0.00794035	0.6012
2^{-10}	0.399414	0.00147598	0.9641	0.00575887	0.4631
2^{-11}	0.399902	0.00074929	0.9780	0.00320243	0.8468
2^{-12}	0.399902	0.00037885	0.9840	0.00174532	0.8756
2^{-13}	0.399902	0.00019071	0.9902	0.00109421	0.6736
2^{-14}	0.399963	0.00009579	0.9933	0.00073306	0.5778

Table 2.10: $E_u(\tau)$ and $q_u(\tau)$ for ODE case with $T_{ex} = 0.2$ and $T = 0.4$

Similarly to the remark after Table 2.7 we report that in the computation of v_N in the ODE case the oscillatory behaviour of $q_v(\tau)$ is again far less pronounced: for $p = 4$ the values of $q_v(\tau)$ range between 0.9826 and 0.9999 for τ ranging between 2^{-6} and 2^{-14}.

2.7.3 Error Analysis for the ODE

Because of the remarkable similarities between the numerical results for the PDE (7.15) and the ODE (7.19) we now analyse the backward Euler method applied to (7.19). Although one could do this using Rulla's techniques [39] which we have employed in section 2.4, our analysis starts from first principles and is very straightforward.

With $\hat{\phi}(s) = \text{sign}(s)\,|s|^p$, $s \in \mathbb{R}$, (7.19) can be rewritten more generally as

$$\frac{d}{dt}\hat{\phi}(u) + \alpha u = 0 \quad \text{on } [0,T]\,, \tag{7.24}$$

where we remind the reader that $\alpha > 0$. On choosing $\tau = \frac{T}{N}$, $N \in \mathbb{N}$, we set $t_n := n\tau$, $n = 1 \to N$, and approximate $u^n \equiv u(t_n)$ via the backward Euler method

$$\frac{\hat{\phi}(U^n) - \hat{\phi}(U^{n-1})}{\tau} + \alpha U^n = 0\,, \quad n = 1 \to N\,. \tag{7.25}$$

We first show that for non-negative initial values the numerical solution lies above the exact solution. This will be crucial in our error analysis.

Lemma 2.13 *Let $U^0 = u^0 = g \geq 0$ and $p > 1$. Then we have that*

$$U^n \geq u^n \geq 0\,, \quad n = 1 \to N\,. \tag{7.26}$$

Proof. Introducing for notational convenience a function $\psi : \mathbb{R} \to \mathbb{R}$ by

$$\psi(s) = \alpha\tau s + \hat{\phi}(s)\,, \quad s \in \mathbb{R}\,, \tag{7.27}$$

(7.25) can be rewritten as

$$\psi(U^n) = \hat{\phi}(U^{n-1})\,, \quad n = 1 \to N\,. \tag{7.28}$$

On integrating (7.24) on $[t_{n-1}, t_n]$ we have that

$$\hat{\phi}(u^n) + \alpha \int_{t_{n-1}}^{t_n} u\,dt = \hat{\phi}(u^{n-1})\,. \tag{7.29}$$

The exact solution u of (7.24) is given by (7.21), in particular u is non-increasing. Hence, we have that

$$\hat{\phi}(u^n) + \alpha\tau u^n \leq \hat{\phi}(u^n) + \alpha \int_{t_{n-1}}^{t_n} u\,dt\,. \tag{7.30}$$

On noting (7.27) it follows from (7.29) and (7.30) that

$$\psi(u^n) \leq \hat{\phi}(u^{n-1})\,, \quad n = 1 \to N\,. \tag{7.31}$$

Next, we note that $\hat{\phi}$ is strictly monotone. Since $\psi' = \alpha\tau + \hat{\phi}' > 0$, ψ is invertible with $(\psi^{-1})' > 0$, i.e. ψ^{-1} is also strictly monotone. The first inequality in (7.26) now follows inductively from (7.28) and (7.31); the second is an immediate consequence of (7.21). ■

In order to prove error bounds for the backward Euler method applied to (7.19) we need the following well-known 'summation by parts' formulae (where we use superscripts):

$$\sum_{n=1}^{m}[(a^n - a^{n-1})\sum_{j=n}^{m} b^j] = \sum_{n=1}^{m}(a^n - a^0)b^n\,, \tag{7.32}$$

$$\sum_{n=1}^{m}[a^n \sum_{j=n}^{m} a^j] = \tfrac{1}{2}\left(\sum_{n=1}^{m} a^n\right)^2 + \tfrac{1}{2}\sum_{n=1}^{m}(a^n)^2\,, \tag{7.33}$$

$$\sum_{n=1}^{m}[(a^n - a^{n-1})a^n] = \tfrac{1}{2}[\,(a^m)^2 - (a^0)^2 + \sum_{n=1}^{m}(a^n - a^{n-1})^2\,]\,. \tag{7.34}$$

Also needed is the following standard bound: Let $p \geq 1$ and $a,\, b \geq 0$. Then

$$|a-b|^{p+1} \leq (a^p - b^p)(a-b)\,. \tag{7.35}$$

We introduce some more notation: Let U, $\bar{u}$ be the piecewise constant functions given by

$$\begin{aligned} U(t) &:= U^n\,, \quad \bar{u}(t) := u^n\,, \qquad \text{if } t \in (t_{n-1}, t_n]\,, \\ U(0) &:= U^0\,, \quad \bar{u}(0) := u^0\,. \end{aligned}$$

We define for $r \in [1, \infty]$ discrete L^r errors and computational orders of convergence as

$$\begin{aligned} E_{u,r}(\tau) &:= \left[\tau \sum_{n=1}^{N} |u^n - U^n|^r\right]^{\frac{1}{r}}, \quad \text{if } r < \infty\,; \quad E_{u,\infty}(\tau) := \max_{n=1\to N} |u^n - U^n|\,, \\ q_{u,r}(\tau) &:= \log_2 \frac{E_{u,r}(2\tau)}{E_{u,r}(\tau)}\,. \end{aligned}$$

Analogous definitions hold for v and V.

Lemma 2.14 *Let $U^0 = u^0 = g \geq 0$ and $p > 1$. Then we have that*

$$|\bar{u} - U|_{L^1(0,T)} \equiv E_{u,1}(\tau) \leq C\tau\,, \tag{7.36}$$

$$|\bar{u} - U|_{L^2(0,T)} \equiv E_{u,2}(\tau) \leq C\tau^{\frac{p+1}{2p}}\,, \tag{7.37}$$

$$|\bar{u} - U|_{L^{p+1}(0,T)} \equiv E_{u,p+1}(\tau) \leq C\tau^{\frac{2}{p+1}}\,, \tag{7.38}$$

$$|\bar{u} - U|_{L^\infty(0,T)} \equiv E_{u,\infty}(\tau) \leq C\tau^{\frac{1}{p}}\,, \tag{7.39}$$

$$|\bar{v} - V|_{L^\infty(0,T)} \equiv E_{v,\infty}(\tau) \leq C\tau\,. \tag{7.40}$$

Proof. On noting that u is non-increasing we obtain from integrating (7.24) on $(t_{n-1}, t_n]$ that

$$\frac{\hat{\phi}(u^n) - \hat{\phi}(u^{n-1})}{\tau} + \alpha u^n = \tfrac{\alpha}{\tau} \int_{t_{n-1}}^{t_n} (u^n - u)\, dt \leq 0\,. \tag{7.41}$$

Let $E^n \equiv \hat{\phi}(u^n) - \hat{\phi}(U^n)$, $e^n \equiv u^n - U^n$ and $\delta^n \equiv \frac{1}{\tau}\int_{t_{n-1}}^{t_n}(u^n - u)\, dt \leq 0$. Subtracting (7.25) from (7.41) yields that

$$\frac{E^n - E^{n-1}}{\tau} + \alpha e^n = \alpha \delta^n\,. \tag{7.42}$$

Hence, for $1 \leq m \leq N$ we have that

$$\tau \sum_{n=1}^{m} \left[\left(\frac{E^n - E^{n-1}}{\tau} \right) \left(\tau \sum_{j=n}^{m} e^j \right) \right] + \alpha\tau \sum_{n=1}^{m} \left[e^n \left(\tau \sum_{j=n}^{m} e^j \right) \right] = \alpha\tau \sum_{n=1}^{m} \delta^n \left(\tau \sum_{j=n}^{m} e^j \right). \tag{7.43}$$

From (7.43) we obtain on noting(7.32), (7.33) and $e^0 = E^0 = 0$ that

$$\tau \sum_{n=1}^{m} E^n e^n + \tfrac{\alpha}{2} \left(\tau \sum_{n=1}^{m} e^n \right)^2 + \tfrac{\alpha}{2}\tau^2 \sum_{n=1}^{m} (e^n)^2 = \alpha\tau \sum_{n=1}^{m} \delta^n \left(\tau \sum_{j=n}^{m} e^j \right). \tag{7.44}$$

As $\delta^n \leq 0$ and by (7.26) $e^n \leq 0$, it follows from (7.44) that

$$\tau \sum_{n=1}^{m} E^n e^n + \tfrac{\alpha}{2} \left(\tau \sum_{n=1}^{m} e^n \right)^2 + \tfrac{\alpha}{2}\tau^2 \sum_{n=1}^{m} (e^n)^2 \leq \alpha \left(\tau \sum_{n=1}^{m} \delta^n \right) \left(\tau \sum_{n=1}^{m} e^n \right).$$

Applying Young's inequality we obtain that

$$\tau \sum_{n=1}^{m} E^n e^n + \tfrac{\alpha}{4} \left(\tau \sum_{n=1}^{m} e^n \right)^2 + \tfrac{\alpha}{2}\tau^2 \sum_{n=1}^{m} (e^n)^2 \leq C \left(\tau \sum_{n=1}^{m} \delta^n \right)^2 = C \left(\int_0^{t_m} (\bar{u} - u)\, dt \right)^2 \leq C\tau^2 \tag{7.45}$$

since from (7.21) it is easily seen that $\frac{du}{dt} \in L^1(0,T)$ for all $p > 1$.

The desired result (7.36) now is an immediate consequence of the bound on the second term in (7.45) and (7.26). (7.38) follows from the bound on the first term in (7.45) on noting (7.35). The desired bound (7.37) can be obtained from (7.36) and (7.38) via interpolation: With $\gamma \in (0, 1]$ we have on noting Hölder's inequality that

$$|\bar{u} - U|^2_{L^2(0,T)} = \int_0^T |\bar{u} - U|^{\gamma} |\bar{u} - U|^{2-\gamma}\, dt \leq \left(\int_0^T |\bar{u} - U|\, dt \right)^{\gamma} \left(\int_0^T |\bar{u} - U|^{\frac{2-\gamma}{1-\gamma}}\, dt \right)^{1-\gamma}. \tag{7.46}$$

We can bound the first factor on the right hand side of (7.46) on noting (7.36). In order to use (7.38) to bound the second factor, we choose $\frac{2-\gamma}{1-\gamma} = p + 1$ and hence $\gamma = \frac{p-1}{p}$ and $1 - \gamma = \frac{1}{p}$. Thus, we obtain from (7.46) that

$$|\bar{u} - U|^2_{L^2(0,T)} \leq |\bar{u} - U|^{\frac{p-1}{p}}_{L^1(0,T)} |\bar{u} - U|^{\frac{p+1}{p}}_{L^{p+1}(0,T)} \leq C\tau^{\frac{p-1}{p}} \tau^{\frac{2}{p}} = C\tau^{\frac{p+1}{p}}$$

which proves (7.37).

Next, we prove (7.40). It follows from (7.42) that

$$\tau \sum_{n=1}^{m} \left(\frac{E^n - E^{n-1}}{\tau} \right) E^n + \tau\alpha \sum_{n=1}^{m} e^n E^n = \tau\alpha \sum_{n=1}^{m} \delta^n E^n . \tag{7.47}$$

On noting (7.34) and $E^0 = 0$ we obtain that

$$\tfrac{1}{2}|E^m|^2 + \tfrac{1}{2} \sum_{n=1}^{m} |E^n - E^{n-1}|^2 + \tau\alpha \sum_{n=1}^{m} e^n E^n = \tau\alpha \sum_{n=1}^{m} \delta^n E^n .$$

On applying Hölder's and Young's inequality we have that

$$\begin{aligned} \tfrac{1}{2}|E^m|^2 + \tfrac{1}{2} \sum_{n=1}^{m} |E^n - E^{n-1}|^2 + \tau\alpha \sum_{n=1}^{m} e^n E^n &\leq \tfrac{1}{4} \max_{n=1\to m} |E^n|^2 + C \left(\tau \sum_{n=1}^{m} |\delta^n| \right)^2 \\ &\leq \tfrac{1}{4} \max_{n=1\to m} |E^n|^2 + C\tau^2 . \end{aligned} \tag{7.48}$$

As (7.48) is valid for all $1 \leq m \leq N$, it also holds if we replace the left hand side by its maximum over $m = 1 \to N$. Hence, on noting that the second and third term in (7.48) are non-negative we obtain that

$$\max_{n=1\to N} |E^n|^2 \leq C\tau^2$$

which proves (7.40). Obviously, (7.39) follows from (7.40) on noting (7.35). ∎

We note that it is not possible to adapt the proof above in order to derive bounds analogous to (7.36) – (7.40) for the PDE case, i.e. when α in (7.19) is replaced by $-\Delta$: Proceeding similarly as in the proof of Lemma 2.14 the equation analogous to (7.44) reads

$$\tau \sum_{n=1}^{m} \langle E^n, e^n \rangle + \tfrac{1}{2} \left| \nabla \left(\tau \sum_{n=1}^{m} e^n \right) \right|_{0,\Omega}^2 + \tfrac{1}{2}\tau^2 \sum_{n=1}^{m} |\nabla e^n|_{0,\Omega}^2 = \tau \sum_{n=1}^{m} \left\langle \nabla \delta^n, \nabla \left(\tau \sum_{j=n}^{m} e^j \right) \right\rangle . \tag{7.49}$$

However, we have now no control over the sign of $\langle \nabla \delta^n, \nabla e^n \rangle$ and so we cannot proceed further to prove bounds analogous to (7.36) – (7.38). Similarly, because of the appearance of ∇ in the analogue of (7.47), it is not possible to obtain analogues of the bounds (7.39) and (7.40). Note that the same is true for the proof of the first inequality in (7.26).

Table 2.11 shows the computational orders $q_{u,1}(2^{-14})$, $q_{u,2}(2^{-14})$, $q_{u,p+1}(2^{-14})$ and $q_{u,\infty}(2^{-14})$ as compared to the corresponding theoretical orders from (7.36), (7.37), (7.38) and (7.39) for different values of p. The computation was done with $T_{ex} = 0.25$ and $T = 0.5$, i.e. both points coincide with mesh points.

p	$q_{u,1}(2^{-14})$	1.0000	$q_{u,2}(2^{-14})$	$\frac{p+1}{2p}$	$q_{u,p+1}(2^{-14})$	$\frac{2}{p+1}$	$q_{u,\infty}(2^{-14})$	$\frac{1}{p}$
1.5	0.9999	1.0000	0.9996	0.8333	0.9995	0.8000	0.9994	0.6667
2.0	1.0000	1.0000	0.9946	0.7500	0.9875	0.6666	0.8708	0.5000
2.5	1.0010	1.0000	0.9467	0.7000	0.8340	0.5714	0.5845	0.4000
3.0	1.0038	1.0000	0.8708	0.6667	0.6750	0.5000	0.4406	0.3333
3.5	1.0080	1.0000	0.8064	0.6429	0.5659	0.4444	0.3537	0.2857
4.0	1.0125	1.0000	0.7589	0.6250	0.4879	0.4000	0.2956	0.2500
4.5	1.0167	1.0000	0.7239	0.6111	0.4293	0.3636	0.2540	0.2222
5.0	1.0201	1.0000	0.6963	0.6000	0.3839	0.3333	0.2227	0.2000

Table 2.11: Computational and theoretical orders of convergence for ODE (7.19)

On comparing the numbers of the second and third column in Table 2.11 we see that the bound (7.36) is sharp. We recall from the proof of Lemma 2.14 that (7.36) was obtained from the bound on the second term in (7.45).

The bound (7.38) was obtained from the bound on the first term in (7.45) on noting (7.35). From the sixth and seventh column in Table 2.11 we see that the computational order of convergence is better than the theoretical.

In Table 2.12 we show the computational order of convergence for the first term in (7.45), where we define

$$\hat{q}(\tau) := \log_2 \frac{\hat{E}(2\tau)}{\hat{E}(\tau)} \quad \text{with} \quad \hat{E}(\tau) := \tau \sum_{n=1}^{N} [\hat{\phi}(u^n) - \hat{\phi}(U^n)][u^n - U^n] .$$

p	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
$\hat{q}(2^{-14})$	1.9994	1.9993	1.9985	1.9976	1.9983	2.0012	2.0056	2.0105

Table 2.12: Computational orders for the first term in (7.45)

Table 2.12 suggests that the bound on the first term in (7.45) is sharp. Obviously, since we have obtained the bound (7.37) from interpolating (7.36) and the suboptimal (7.38), it is no surprise that (7.37) is also suboptimal, as is seen on comparing the fourth and fifth column of Table 2.11.

The bound (7.39) was obtained from the bound (7.40) on noting (7.35). From the eighth and nineth column in Table 2.11 we see that the computational order of convergence is better than the theoretical. However, Table 2.13 shows that the bound (7.40) is sharp (the computations were again done with $T_{ex} = 0.25$ and $T = 0.5$):

p	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
$q_{v,\infty}(2^{-14})$	0.9996	0.9998	0.9997	0.9997	0.9997	0.9996	0.9995	0.9994

Table 2.13: $q_{v,\infty}(2^{-14})$ for different values of p.

We conclude this subsection with the remark that if we choose $T_{ex} = 0.2$ and $T = 0.4$, we again see the 'oscillatory behaviour' of the computational orders of convergence we have seen before in this situation (cf. Tables 2.7, 2.8 and 2.10). As an example of this we show in Table 2.14 the numerical results for $p = 4$.

τ	$q_{u,1}(\tau)$	$q_{u,2}(\tau)$	$q_{u,p+1}(\tau)$	$q_{u,\infty}(\tau)$
2^{-7}	1.1477	0.8763	0.6055	0.4056
2^{-8}	1.1034	0.9286	0.8097	0.6449
2^{-9}	0.9526	0.6012	0.2245	-0.0063
2^{-10}	0.8772	0.4631	0.1825	-0.0113
2^{-11}	1.0612	0.8468	0.5806	0.3785
2^{-12}	1.0405	0.8756	0.6912	0.5151
2^{-13}	0.9821	0.6736	0.3517	0.1345
2^{-14}	0.9515	0.5778	0.2821	0.1053

Table 2.14: Comput. orders of conv. for ODE case with $p = 4$, $T_{ex} = 0.2$ and $T = 0.4$

Although the computational orders of convergence in Table 2.14 'oscillate', it is interesting to observe that on average they (nearly) yield the corresponding computational orders $q_{u,r}(2^{-14})$ from Table 2.11 for $p = 4$: With

$$\bar{q}_{u,r} := \tfrac{1}{8} \sum_{n=7}^{14} q_{u,r}(2^{-n})$$

we have that $\bar{q}_{u,1} = 1.0145$, $\bar{q}_{u,2} = 0.7304$, $\bar{q}_{u,p+1} = 0.4660$ and $\bar{q}_{u,\infty} = 0.2708$ as compared to 1.0125, 0.7589, 0.4879 and 0.2956 from Table 2.11.

2.7.4 Some Numerical Results for Le Roux's Method

In the previous two subsections we presented some numerical results for the backward Euler method applied to the ODE (7.19) and (7.20). This was motivated by the fact that we have so far considered the case where (P) has a separable solution. However, as we pointed out in section 2.6 (7.19) and (7.20) can also be considered as a simple ODE model for (P) generally.

We have seen in section 2.6 that Le Roux's method applied to (7.19) and (7.20) yields a discrete solution which is exact at every mesh point and which can be given explicitly, see (6.18) and (6.21). Hence, there is no need for computations. Nonetheless, it seems to us that the ODE

$$\tfrac{d}{dt}u^p + \alpha u = 0\,, \quad u(0) = g \geq 0 \tag{7.50}$$

and its Le Roux's discretization

$$\tfrac{p}{p-1}\, u_L^n \, \frac{(u_L^n)^{p-1} - (u_L^{n-1})^{p-1}}{\tau} + \alpha u_L^n = 0 \tag{7.51}$$

provide a good test case for the iterative solver (6.10). After some simple algebraic manipulations we can rewrite the iteration in its ODE adapted form as:

Set $\bar{u}_L^{n,0} = \bar{u}_L^{n-1}$, then for $k \geq 0$ find $\bar{u}_L^{n,k+1}$ such that

$$\left[(p-1)(\bar{u}_L^{n-1})^{p-1} + \alpha\tfrac{p-1}{p}\tau\right] \bar{u}_L^{n,k+1} \;=\; \left[p(\bar{u}_L^{n-1})^{p-1} - (\bar{u}_L^{n,k})^{p-1}\right] \bar{u}_L^{n,k}\,. \tag{7.52}$$

Firstly, we note that a stopping criterion is needed for the iteration. To this end we rewrite (7.51) notationally somewhat simpler as:

Let $a \equiv \bar{u}_L^{n-1}$ find $s_* \equiv \bar{u}_L^n \in [0, a]$ such that

$$F(s_*) = 0 \quad \text{where} \quad F(s) := s(s^{p-1} - a^{p-1}) + \mu s = 0 \tag{7.53}$$

and $\mu := \frac{\alpha\tau(p-1)}{p} > 0$.

We have that $s_* \equiv [a^{p-1} - \mu]_+^{\frac{1}{p-1}}$ is a root of (7.53), i.e. $F(s_*) = 0$. Next, we have that $F'(s) = ps^{p-1} + [\mu - a^{p-1}]$. Hence, on letting $D \equiv a^{p-1} - \mu$ we obtain that

$$\begin{aligned} F'(s_*) &= \begin{cases} (p-1)D\,, & \text{if } D \geq 0\,, \\ -D\quad\;\;, & \text{if } D \leq 0\,, \end{cases} \\ F''(s) &= p(p-1)s^{p-2} > 0 \quad \text{for } s > 0\,. \end{aligned}$$

and consequently

$$F'(s) \geq F'(s_*) \geq 0 \tag{7.54}$$

provided

$$s \geq s_*\,. \tag{7.55}$$

The iterates $(\bar{u}_L^{n,k})_{k\geq 0}$ of (7.52) satisfy $\bar{u}_L^{n,k} \geq s_*$ as we have seen for the more general PDE case in Lemma 2.11. Hence, (7.55) is satisfied for $s \equiv u_L^{n,k}$. Thus we obtain on noting the Mean Value Theorem and (7.54) that

$$0 \leq F'(s_*)\,(s - s_*) \leq F'(\eta)(s - s_*) = F(s) - F(s_*) = F(s)$$

where $s_* \leq \eta \leq s$. Hence, on assuming that

$$D \equiv a^{p-1} - \mu \neq 0 \tag{7.56}$$

we obtain that

$$s_* \leq s \leq s_* + Tol$$

provided

$$F(s) \leq F'(s_*)\; Tol = \left\{ \begin{array}{ll} (p-1)D\,, & \text{if } D > 0\,, \\ -D \quad\;\; , & \text{if } D < 0 \end{array} \right\} Tol\,. \tag{7.57}$$

Note that (7.56) translated back means that extinction does not occur at a mesh point in time, see the above notation and (6.20). Also note that the stopping criterion (7.57) uses s_*, the root of (7.53), which is unknown in the PDE case.

The data for all the following numerical results are as before: For the exact solution u of (7.50) which is given by (6.18) we choose $g = 1$ and $\alpha = \beta\frac{p}{p-1}$, i.e. $u(t) = [1 - \beta t]_+^{\frac{1}{p-1}}$. We set $\beta \equiv 5$ and obtain the extinction time of the exact solution as $T_{ex} = 0.2$.

In the following Table 2.15 we show for different values of p and τ the behaviour of the Le Roux approximation $\bar{u}_L^n$ to (7.50) around T_{ex} when $\bar{u}_L^n$ is obtained from (7.52) with stopping criterion (7.57) and

$$Tol \equiv 10^{-8}\,.$$

For each time step τ the values of t in Table 2.15 are those seven mesh points t_n which are closest to T_{ex}. By $e_{u-\bar{u}_L}(t)$ we denote the error $|u(t_n) - \bar{u}_L^n|$ and by k_{it} the number of iterations after which (7.52) was stopped, i.e $\bar{u}_L^n = \bar{u}_L^{n,k_{it}}$.

Firstly, we note from Table 2.15 that the values of $e_{u-\bar{u}_L}(t)$ are smaller than $Tol \equiv 10^{-8}$ for all $t > T_{ex}$; i.e. within the precision of our computations we can say that $\bar{u}_L^n$ becomes extinct immediately after $T_{ex} = 0.2$ as expected. Secondly, when t is closest to T_{ex} (the values of such t have a '*' next to them) a large number of iterations is required. Finally not surprisingly , we note a certain kind of 'error propagation' since $\bar{u}_L^{n-1} \equiv \bar{u}_L^{n-1,k_{it}} \neq u(t_{n-1})$: Before the extinction time is reached the error $e_{u-\bar{u}_L}(t)$ is larger than $Tol \equiv 10^{-8}$.

		$p=2$		$p=3$		$p=4$	
τ	t	$e_{u-\bar{u}_L}(t)$	k_{it}	$e_{u-\bar{u}_L}(t)$	k_{it}	$e_{u-\bar{u}_L}(t)$	k_{it}
2^{-10}	0.19628906250000	0.6698×10^{-6}	12	0.2202×10^{-5}	11	0.2868×10^{-5}	11
	0.19726562500000	0.6796×10^{-6}	14	0.2575×10^{-5}	13	0.3520×10^{-5}	13
	0.19824218750000	0.6890×10^{-6}	19	0.3219×10^{-5}	18	0.4733×10^{-5}	17
	0.19921875000000	0.6969×10^{-6}	36	0.4837×10^{-5}	33	0.8136×10^{-5}	31
	*0.20019531250000	0.9631×10^{-8}	94	0.9256×10^{-8}	185	0.9602×10^{-8}	264
	0.20117187500000	0.9631×10^{-8}	0	0.9256×10^{-8}	0	0.9602×10^{-8}	0
	0.20214843750000	0.9631×10^{-8}	0	0.9256×10^{-8}	0	0.9602×10^{-8}	0
2^{-11}	0.19824218750000	0.7043×10^{-6}	12	0.5214×10^{-5}	11	0.1832×10^{-4}	11
	0.19873046875000	0.7130×10^{-6}	14	0.6141×10^{-5}	14	0.2276×10^{-4}	13
	0.19921875000000	0.7202×10^{-6}	20	0.7849×10^{-5}	19	0.3147×10^{-4}	18
	0.19970703125000	0.7281×10^{-6}	43	0.1280×10^{-4}	41	0.6051×10^{-4}	39
	*0.20019531250000	0.9653×10^{-8}	37	0.8816×10^{-8}	73	0.9861×10^{-8}	103
	0.20068359375000	0.9653×10^{-8}	0	0.8816×10^{-8}	0	0.9861×10^{-8}	0
	0.20117187500000	0.9653×10^{-8}	0	0.8816×10^{-8}	0	0.9861×10^{-8}	0
2^{-12}	0.19921875000000	0.2546×10^{-5}	12	0.2198×10^{-4}	12	0.3815×10^{-4}	11
	0.19946289062500	0.2553×10^{-5}	15	0.2651×10^{-4}	15	0.4898×10^{-4}	14
	0.19970703125000	0.2561×10^{-5}	23	0.3590×10^{-4}	23	0.7336×10^{-4}	22
	*0.19995117187500	0.2571×10^{-5}	101	0.8774×10^{-4}	105	0.2415×10^{-3}	101
	0.20019531250000	0.5972×10^{-8}	9	0.6438×10^{-8}	17	0.6520×10^{-8}	23
	0.20043945312500	0.5972×10^{-8}	0	0.6438×10^{-8}	0	0.6520×10^{-8}	0
	0.20068359375000	0.5972×10^{-8}	0	0.6438×10^{-8}	0	0.6520×10^{-8}	0
2^{-13}	0.19958496093750	0.3227×10^{-5}	11	0.2337×10^{-4}	11	0.3981×10^{-4}	11
	0.19970703125000	0.3236×10^{-5}	13	0.2781×10^{-4}	14	0.5021×10^{-4}	13
	0.19982910156250	0.3245×10^{-5}	19	0.3641×10^{-4}	20	0.7191×10^{-4}	19
	*0.19995117187500	0.3254×10^{-5}	52	0.6803×10^{-4}	55	0.1655×10^{-3}	53
	0.20007324218750	0.7804×10^{-8}	17	0.7580×10^{-8}	35	0.8786×10^{-8}	49
	0.20019531250000	0.7804×10^{-8}	0	0.7580×10^{-8}	0	0.8786×10^{-8}	0
	0.20031738281250	0.7804×10^{-8}	0	0.7580×10^{-8}	0	0.8786×10^{-8}	0
2^{-14}	0.19976806640625	0.2663×10^{-5}	9	0.2000×10^{-4}	10	0.3497×10^{-4}	10
	0.19982910156250	0.2671×10^{-5}	11	0.2330×10^{-4}	12	0.4287×10^{-4}	12
	0.19989013671875	0.2679×10^{-5}	15	0.2906×10^{-4}	16	0.5754×10^{-4}	16
	0.19995117187500	0.2689×10^{-5}	27	0.4357×10^{-4}	30	0.9872×10^{-4}	29
	*0.20001220703125	0.9599×10^{-8}	74	0.9299×10^{-8}	172	0.9792×10^{-8}	254
	0.20007324218750	0.9599×10^{-8}	0	0.9299×10^{-8}	0	0.9792×10^{-8}	0
	0.20013427734375	0.9599×10^{-8}	0	0.9299×10^{-8}	0	0.9792×10^{-8}	0

Table 2.15: Le Roux's method: behaviour around extinction time $T_{ex} = 0.2$

In Table 2.16 we present a test of the computations above. Instead of starting the iteration (7.52) at time level t_n with the *computational* result of the previous time level t_{n-1}, we start it with the known *exact* solution of (7.50) of the previous time level and replace $\bar{u}_L^{n-1}$ in (7.52) with $u(t_{n-1})$. We only present the results for $\tau = 2^{-11}$ and 2^{-12}.

		$p=2$		$p=3$		$p=4$	
τ	t	$e_{u-\bar{u}_L}(t)$	k_{it}	$e_{u-\bar{u}_L}(t)$	k_{it}	$e_{u-\bar{u}_L}(t)$	k_{it}
2^{-11}	0.19824218750000	0.4576×10^{-8}	12	0.8033×10^{-8}	11	0.4642×10^{-8}	11
	0.19873046875000	0.8716×10^{-8}	14	0.4853×10^{-8}	14	0.7607×10^{-8}	13
	0.19921875000000	0.7132×10^{-8}	20	0.7395×10^{-8}	19	0.8312×10^{-8}	18
	0.19970703125000	0.8007×10^{-8}	43	0.8425×10^{-8}	41	0.8295×10^{-8}	39
	*0.20019531250000	0.9543×10^{-8}	37	0.8623×10^{-8}	73	0.9298×10^{-8}	103
	0.20068359375000	0.0000×10^{-8}	0	0.0000×10^{-8}	0	0.0000×10^{-8}	0
	0.20117187500000	0.0000×10^{-8}	0	0.0000×10^{-8}	0	0.0000×10^{-8}	0
2^{-12}	0.19921875000000	0.5476×10^{-8}	12	0.4643×10^{-8}	12	0.9932×10^{-8}	11
	0.19946289062500	0.7184×10^{-8}	15	0.6164×10^{-8}	15	0.9943×10^{-8}	14
	0.19970703125000	0.8447×10^{-8}	23	0.7305×10^{-8}	23	0.7898×10^{-8}	22
	*0.19995117187500	0.9863×10^{-8}	102	0.9079×10^{-8}	106	0.8999×10^{-8}	102
	0.20019531250000	0.5472×10^{-8}	9	0.5593×10^{-8}	17	0.5504×10^{-8}	23
	0.20043945312500	0.0000×10^{-8}	0	0.0000×10^{-8}	0	0.0000×10^{-8}	0
	0.20068359375000	0.0000×10^{-8}	0	0.0000×10^{-8}	0	0.0000×10^{-8}	0

Table 2.16: Test of Le Roux's method: behaviour around extinction time $T_{ex} = 0.2$

Firstly note from Table 2.16 that $e_{u-\bar{u}_L}(t)$ is always less than $Tol \equiv 10^{-8}$ which suggests that our FORTRAN code is correct. Next note that from the second time level after extinction onward $e_{u-\bar{u}_L}(t)$ is identically 0, since the iteration was started with 0, the exact solution at the previous time level. Finally, note that the number of iterations k_{it} needed is hardly different from the corresponding number in Table 2.15 which we interpret as the two computations being consistent.

Before we turn to the fully discrete Le Roux's method for (P) we consider the same computations as above but for the backward Euler method in order to see whether the computational results also capture extinction in finite time. The following Table 2.17 is an analogue of Table 2.15 for the backward Euler method applied to (7.50) instead of Le Roux's method. As before in the case of the backward Euler method we use Newton's Method to solve the nonlinear discrete equation in practice. For $t = t_n$, a mesh point, we set $e_{u-u_{Euler}}(t) := |u(t_n) - U^n|$, where U^n denotes the backward Euler approximation of (7.50) at $t = t_n$. Here, k_{it} denotes the number of iterations needed before the stopping criterion (7.23) is satisfied where we set again $Tol \equiv 10^{-8}$. Again, we show the results for $\tau = 2^{-11}$ and $\tau = 2^{-12}$ only.

		$p=2$		$p=3$		$p=4$	
τ	t	$e_{u-u_{Euler}}(t)$	k_{it}	$e_{u-u_{Euler}}(t)$	k_{it}	$e_{u-u_{Euler}}(t)$	k_{it}
2^{-11}	0.19824218750000	0.5210×10^{-2}	3	0.1359×10^{-1}	3	0.1362×10^{-1}	4
	0.19873046875000	0.5421×10^{-2}	3	0.1635×10^{-1}	3	0.1739×10^{-1}	4
	0.19921875000000	0.5671×10^{-2}	3	0.2089×10^{-1}	4	0.2427×10^{-1}	4
	0.19970703125000	0.5978×10^{-2}	3	0.3066×10^{-1}	4	0.4196×10^{-1}	4
	*0.20019531250000	0.5391×10^{-2}	3	0.5170×10^{-1}	4	0.1187×10^{-0}	4
	0.20068359375000	0.3477×10^{-2}	3	0.3021×10^{-1}	4	0.5762×10^{-1}	5
	0.20117187500000	0.1807×10^{-2}	4	0.7414×10^{-2}	4	0.3388×10^{-2}	3
2^{-12}	0.19921875000000	0.3033×10^{-2}	3	0.1166×10^{-1}	3	0.1339×10^{-1}	4
	0.19946289062500	0.3140×10^{-2}	3	0.1414×10^{-1}	3	0.1739×10^{-1}	4
	0.19970703125000	0.3266×10^{-2}	3	0.1851×10^{-1}	4	0.2532×10^{-1}	4
	*0.19995117187500	0.3421×10^{-2}	3	0.3056×10^{-1}	4	0.5378×10^{-1}	4
	0.20019531250000	0.2642×10^{-2}	3	0.3342×10^{-1}	4	0.8307×10^{-1}	4
	0.20043945312500	0.1690×10^{-2}	3	0.1747×10^{-1}	4	0.2883×10^{-1}	4
	0.20068359375000	0.8641×10^{-3}	3	0.2897×10^{-2}	4	0.4246×10^{-3}	3

Table 2.17: Backward Euler method: behaviour around extinction time $T_{ex} = 0.2$

From Table 2.17 we have two main results. Firstly, the number of Newton iterations needed is between 3 and 5, i.e. considerably less than the number of Le Roux iterations in Table 2.15. Moreover, for a given time step τ the number of iterations hardly varies for different mesh points. Secondly, for t as in Table 2.17, i.e. in a very close vicinity of the extinction time $T_{ex} = 0.2$ of the exact solution, the error in U^n is too large (compared to the computational tolerance of $Tol \equiv 10^{-8}$) in order to speak of computational extinction.

Now we turn to the fully discrete Le Roux's method. In order to simplify (6.26) recall that the parameter h is fixed; L simply denotes Le Roux's method. We will therefore suppress these two indices in the following. On choosing the standard basis functions $\{\chi_j\}_{j=1}^R$ for S_0^h and writing $\hat{u}_L^{h,n-1} = \sum_{j=1}^R U_j^{n-1}\chi_j$ we obtain the following linear system for (6.26):
Find $\hat{u}_L^{h,n,k+1} = \sum_{j=1}^R U_j^{n,k+1}\chi_j$, i.e. find $U^{n,k+1} = (U_1^{n,k+1}, \ldots, U_R^{n,k+1})^t \in \mathbb{R}^R$ such that

$$(\tau K + \mu\tilde{E})U^{n,k+1} = \mu\tilde{\tilde{E}}U^{n,k}, \quad k = 0, 1, 2, \ldots \tag{7.58}$$

where K, $\tilde{E}$ and $\tilde{\tilde{E}}$ are $R \times R$ matrices with entries

$$K_{ij} = \langle \nabla\chi_i, \nabla\chi_j \rangle, \quad \tilde{E}_{ij} = \langle 1, \chi_i \rangle \left[p \max_{l=1\to R} |U_l^{n-1}|^{p-1} - (U_j^{n-1})^{p-1} \right] \delta_{ij}$$

and

$$\tilde{\tilde{E}}_{ij} = \langle 1, \chi_i \rangle \left[p \max_{l=1\to R} |U_l^{n-1}|^{p-1} - (U_j^{n,k})^{p-1} \right] \delta_{ij}\,.$$

Similarly to the argument in subsection 2.7.1 one can see that $\tau K + \mu\tilde{E}$ is an M matrix under assumption (WA) of section 2.3. We will therefore use again Gaussian elimination without pivoting when we solve (7.58) for fixed k. As an ad hoc stopping criterion for the iteration (7.58) we use

$$\max_{j=1\to R} |U_j^{n,k+1} - U_j^{n,k}| \leq Tol \tag{7.59}$$

with $Tol \equiv 10^{-8}$.

The following Table (2.18) is an analogue of Table (2.15) for the fully discrete Le Roux's method applied to the separable problem (7.15) – (7.17) with exact solution u given by (7.18). We choose $h = \tau$, and for a mesh point $t = t_n$ the discretization error $e_{u-\bar{u}_L}(t)$ is defined in analogy to the ODE case:

$$e_{u-\bar{u}_L}(t) := \max_{j=1\to R} |u(x_j, t_n) - U_j^n|,$$

where $\{x_j\}_{j=1}^R$ denotes the set of space mesh points. k_{it} denotes the number of iterations before (6.26) was stopped, i.e.

$$U^n \equiv U^{n,k_{it}}. \tag{7.60}$$

		$p=2$		$p=3$		$p=4$	
τ	t	$e_{u-\bar{u}_L}(t)$	k_{it}	$e_{u-\bar{u}_L}(t)$	k_{it}	$e_{u-\bar{u}_L}(t)$	k_{it}
2^{-10}	0.19628906250	0.1459×10^{-5}	23	0.7170×10^{-5}	38	0.1212×10^{-4}	59
	0.19726562500	0.1475×10^{-5}	24	0.8380×10^{-5}	37	0.1488×10^{-4}	54
	0.19824218750	0.1495×10^{-5}	28	0.1049×10^{-4}	38	0.1999×10^{-4}	54
	0.19921875000	0.1533×10^{-5}	43	0.1579×10^{-4}	52	0.3437×10^{-4}	61
	*0.20019531250	0.8927×10^{-7}	91	0.1593×10^{-6}	210	0.2474×10^{-6}	334
	0.20117187500	0.7769×10^{-16}	2	0.1889×10^{-28}	2	0.1652×10^{-40}	2
	0.20214843750	0.1545×10^{-29}	1	0.2587×10^{-83}	1	0.3814×10^{-160}	1
2^{-11}	0.19824218750	0.1044×10^{-5}	22	0.7732×10^{-5}	37	0.1776×10^{-4}	57
	0.19873046875	0.1059×10^{-5}	23	0.9117×10^{-5}	36	0.2208×10^{-4}	53
	0.19921875000	0.1076×10^{-5}	28	0.1164×10^{-4}	38	0.3054×10^{-4}	51
	*0.19970703125	0.1116×10^{-5}	49	0.1907×10^{-4}	60	0.5877×10^{-4}	70
	0.20019531250	0.2818×10^{-7}	40	0.6154×10^{-7}	88	0.9092×10^{-7}	139
	0.20068359375	0.9772×10^{-17}	2	0.6509×10^{-30}	2	0.5985×10^{-43}	2
	0.20117187500	0.4890×10^{-31}	1	0.2117×10^{-87}	1	0.1314×10^{-169}	1

Table 2.18: Fully discrete Le Roux's method: behaviour around extinction time $T_{ex} = 0.2$

Again we obtain from Table 2.18 that immediately after the extinction time of the exact solution the computational solution becomes very small. We also note again that a large number of computations is needed close to extinction.

2.7.5 The Nonseparable Case and Asymptotic Profiles

So far we have only shown numerical results for the case when (P) has a separable solution, see (7.15) – (7.18), or for the model ODE (7.19), (7.20). In this subsection we deal with the nonseparable case where we do not know the exact solution.

In the following computations we choose the time step to be equal to the mesh size, i.e. $\tau \equiv h$, and we define the 'computational extinction time allowing tolerance' $T^{\tau}_{ex,Tol}$ to be the first mesh step t_n in time where

$$\max_{i=1\to R} |U_i^n| \leq Tol$$

with U^n as in (7.14) for the fully discrete backward Euler method and U^n as in (7.60) for the fully discrete Le Roux's method.

For the domain $\Omega \times (0,T)$ on which we want to solve $(\hat{P}_N^h)$ and $(\hat{P}_L^h)$ we choose

$$\Omega \times (0,T) \equiv (0,1) \times (0, T^{\tau}_{ex,Tol})$$

and for the initial value function g we consider

Case (a):

$$g(x) := \begin{cases} e^{1-\frac{0.25^2}{0.25^2-(x-0.5)^2}}, & 0.25 < x < 0.75, \\ 0, & 0 < x \leq 0.25 \text{ or } 0.75 \leq x < 1 \end{cases}$$

and Case (b):

$$g(x) := \begin{cases} e^{1-\frac{0.25^2}{0.25^2-(x-0.25)^2}}, & 0 < x < 0.5, \\ 0, & 0.5 \leq x < 1, \end{cases}$$

see Figure 2.1 below.

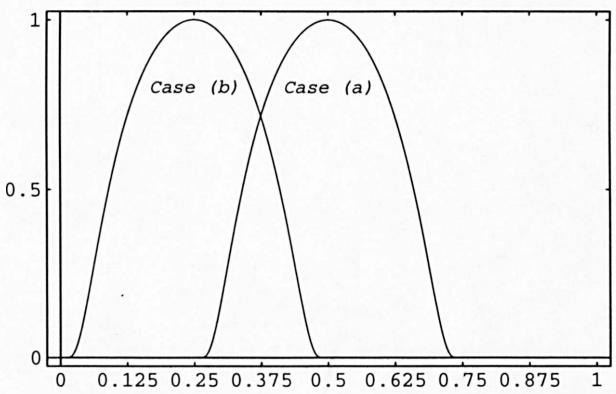


Figure 2.1: Two different initial value functions: Case (a) and Case (b).

We begin our computations with the fully discrete Le Roux's method. Table 2.19 shows the computational extinction times for different time steps τ in Cases (a) and (b). The tolerance Tol for both (7.59) and $T^{\tau}_{ex,Tol}$ was chosen to be 10^{-8}. k_{max} denotes the maximal number of iterations before (7.58) was stopped and $T^{\tau}_{k_{max}}$ the time level at which this maximal number was needed.

		Case (a)		*Case (b)*	
τ		$p=2$	$p=5$	$p=2$	$p=5$
2^{-6}	$T^{\tau}_{ex,10^{-8}}$	0.140625	0.078125	0.125000	0.06250
	k_{max}	216	149	87	204
	$T^{\tau}_{k_{max}}$	0.125000	0.0468750	0.093750	0.015625
2^{-7}	$T^{\tau}_{ex,10^{-8}}$	0.132813	0.06250	0.109375	0.046875
	k_{max}	397	181	107	1386
	$T^{\tau}_{k_{max}}$	0.125000	0.054688	0.007813	0.039063
2^{-8}	$T^{\tau}_{ex,10^{-8}}$	0.130813	0.058594	0.113282	0.046875
	k_{max}	82	257	170	559
	$T^{\tau}_{k_{max}}$	0.003906	0.003906	0.003906	0.003906
2^{-9}	$T^{\tau}_{ex,10^{-8}}$	0.130859	0.054688	0.111328	0.042969
	k_{max}	126	429	277	972
	$T^{\tau}_{k_{max}}$	0.001953	0.001953	0.001953	0.001953
2^{-10}	$T^{\tau}_{ex,10^{-8}}$	0.129883	0.052734	0.111328	0.041992
	k_{max}	200	11838	457	1710
	$T^{\tau}_{k_{max}}$	0.000977	0.051758	0.000977	0.000977
2^{-11}	$T^{\tau}_{ex,10^{-8}}$	0.130371	0.052734	0.111328	0.041504
	k_{max}	322	1260	758	3025
	$T^{\tau}_{k_{max}}$	0.000488	0.000488	0.000488	0.000488

Table 2.19: Fully discrete Le Roux's method: computational extinction times $T^{\tau}_{ex,10^{-8}}$

We make the following three observations on Table 2.19. Firstly, the computational extinction times $T^{\tau}_{ex,Tol}$ seem to converge as τ becomes smaller. This justifies our definition of a 'computational extinction time'. Secondly, it should be noted that the maximal number of iterations needed can be extremely large from a practical point of view: for instance, it took over one hour on an IBM RS6000 to compute from $t=0$ to $T^{2^{-11}}_{ex,10^{-8}} = 0.041504$ (i.e. for 85 time steps) in Case (b) with $p=5$. This is clearly not practical. Also note in this context the extreme number 11838 in Case (a) with $\tau = 2^{-10}$ and $p=5$. Finally, we observe that the maximal number of iterations needed only occurs either at the first time step or close to the computational extinction time. Note that the second case is consistent with the behaviour of the number k_{it} in Table 2.15. The first case can be explained by the

infinite speed of propagation which causes a compactly supported initial function to be spread out over the whole domain Ω immediately. This effect is also seen for the discrete solution, cf. Figures 2.2 – 2.5 below.

Although we have seen above that the fully discrete Le Roux's method is not very practical, we will now use it to present the profiles in space of the computational solution at mesh point t_n in time near the computational extinction time. In Figures 2.2 – 2.5 we show some profiles $\bar{W}$ of the fully discrete solution factored by the time part T^* of the separable solutions:

$$\bar{W}(x, t_n) := \hat{u}_L^{h,n}(x)/T^*(t_n)\,, \qquad x \in \Omega\,,$$

where T^* is as in (2.66) with T_{ex} replaced by $T^{\tau}_{ex,Tol}$. The time step $\tau(\equiv h$ mesh size) was chosen to be $2^{-11} = 0.00048828125 \approx 0.0005$. The computational extinction times were obtained from Table 2.19. The time labels in the Figures contain rounded numbers for easier legibility. By 'true asymptotic profile' we mean the profile of w given by (2.67) which we obtained by using COLSYS and which by Lemma 2.6 is the asymptotic profile of the exact solution after factorization by T^*.

Our first observation from Figures 2.2 – 2.5 is that the profiles for $t = 0.1$ in Figures 2.2 and 2.4 and for $t = 0.04$ in Figure 2.3 are very close to the respective true asymptotic profiles. In Figure 2.5 there is a slightly larger gap between the profile for $t = 0.04$ and the true asymptotic profile than one may have expected after Figure 2.3. It should be noted however that $t = 0.04$ is only three time steps away from the computational extinction time. This may suggest that the (lack of) precision of our computations is reflected in the gap between the two profiles.

We also note that the factored profiles seem to approximate the true asymptotic profile rather quickly. Note in particular, how much the profiles for $t = 0.001$ (i.e. after only two time steps) differ from the profile at $t = 0.000$ at points where the latter is equal to 0. As mentioned above this explains why the maximal number of iterations needed occurs at the start of the computations, cf. Table 2.19.

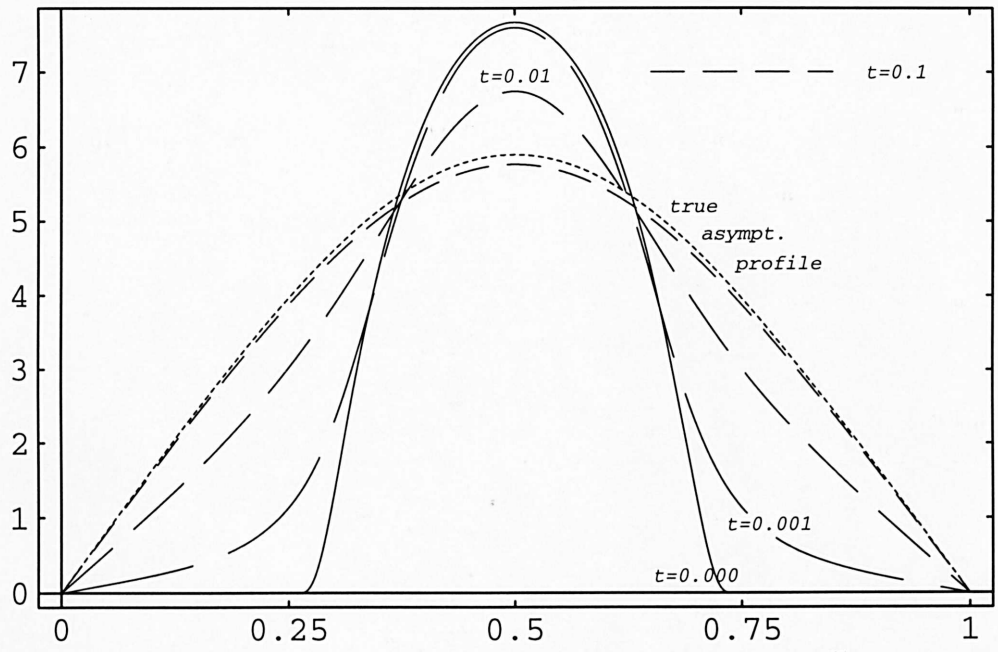


Figure 2.2: Case (a): $p = 2$: Profiles of factored solution as $t \to T^{2^{-11}}_{ex,10^{-8}} = 0.130371$

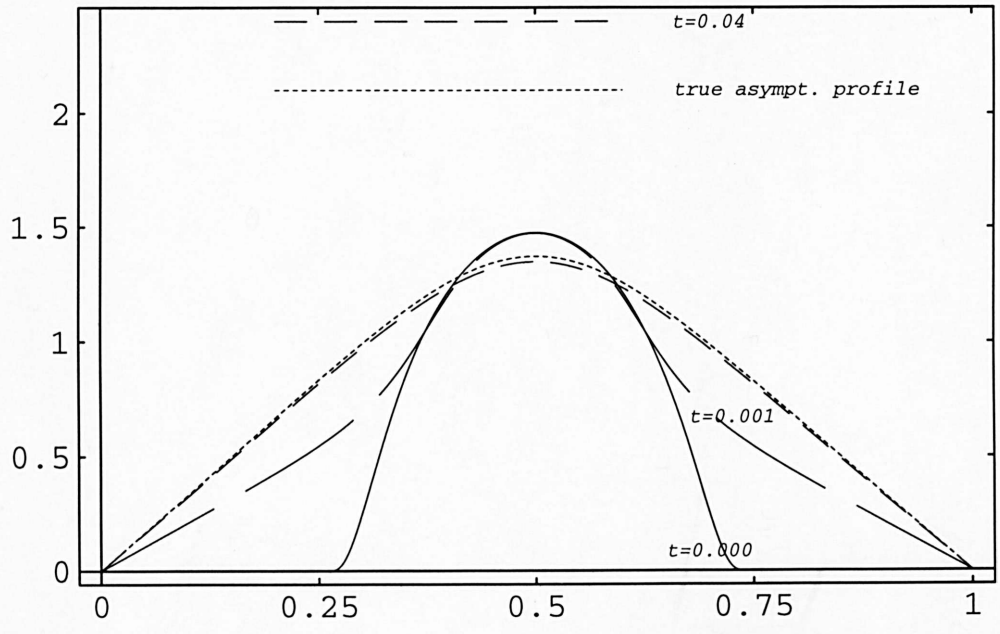


Figure 2.3: Case (a): $p = 5$: Profiles of factored solution as $t \to T^{2^{-11}}_{ex,10^{-8}} = 0.052734$

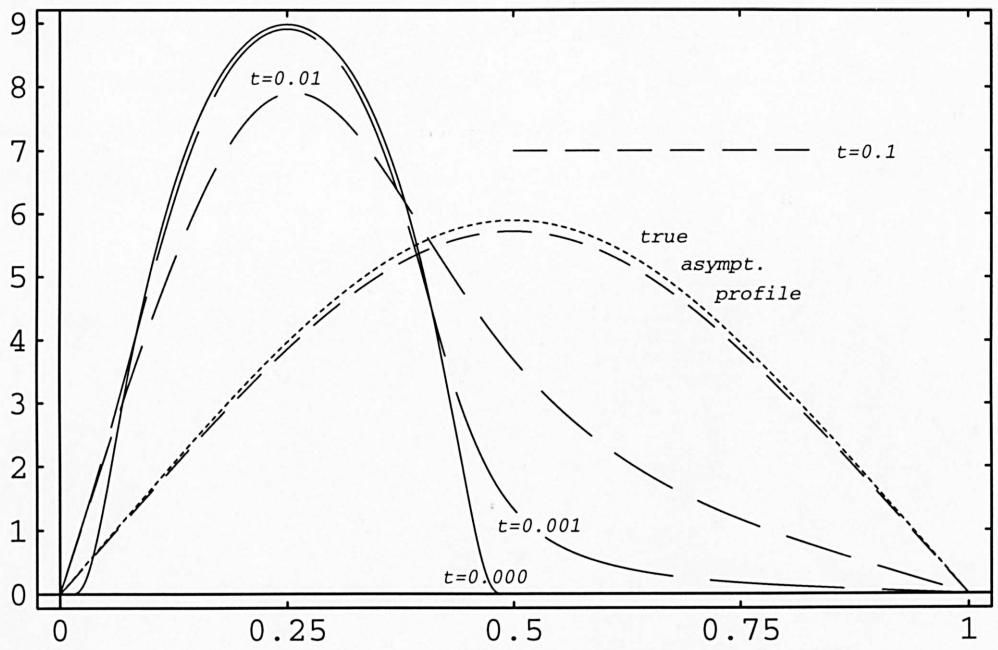


Figure 2.4: Case (b): $p = 2$: Profiles of factored solution as $t \to T_{ex,10^{-8}}^{2^{-11}} = 0.111328$

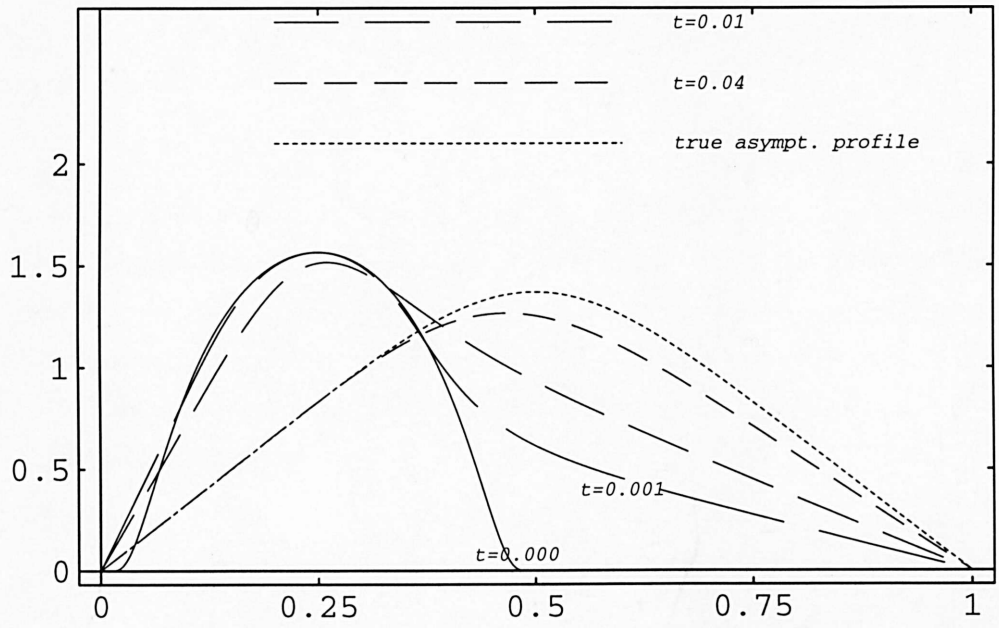


Figure 2.5: Case (b): $p = 5$: Profiles of factored solution as $t \to T_{ex,10^{-8}}^{2^{-11}} = 0.041504$

In the following we will repeat the computations above for the backward Euler method. To begin with we convince ourselves that the notion of 'computational extinction time allowing tolerance' is also appropriate for the backward Euler method. To do this we first consider the backward Euler method applied to the model ODE (7.50). Table 2.20 shows the computational extinction times for different time steps τ which are defined analogously to the PDE case.

	$p=2$	$p=3$	$p=4$	$p=5$
τ	$T^{\tau}_{ex,10^{-8}}$	$T^{\tau}_{ex,10^{-8}}$	$T^{\tau}_{ex,10^{-8}}$	$T^{\tau}_{ex,10^{-8}}$
2^{-6}	0.28125000	0.26562500	0.25000000	0.25000000
2^{-7}	0.24218750	0.23437500	0.22656250	0.22656250
2^{-8}	0.22265625	0.21484375	0.21484337	0.21093750
2^{-9}	0.21093750	0.20898438	0.20703125	0.20703125
2^{-10}	0.20605468	0.20410156	0.20410156	0.20312500
2^{-11}	0.20312500	0.20214848	0.20214843	0.20166016
2^{-12}	0.20166015	0.20117187	0.20092773	0.20092773
2^{-13}	0.20092773	0.20056152	0.20056152	0.20043945
2^{-14}	0.20043945	0.20031738	0.20025635	0.20025635
2^{-15}	0.20022583	0.20016479	0.20013428	0.20010376
2^{-16}	0.20011902	0.20008850	0.20007324	0.20005798

Table 2.20: Backward Euler method: Computational extinction time allowing 10^{-8}

From Table 2.20 it is seen that $T^{\tau}_{ex,10^{-8}} \to T_{ex} = 0.2$ as $\tau \to 0$. The rate of this convergence seems to be (nearly) linear, independently of p.

We can summarize Tables 2.17 and 2.20 as follows: The backward Euler method has the disadvantage of a noticeable discretization error which for fixed time step τ prevents its approximation from becoming extinct immediately after extinction of the exact solution (as opposed to Le Roux's method). However, as the time step becomes smaller one can see the computational extinction time approaching the exact extinction time. This shows that after all the backward Euler method is not impractical for the time discretization of the fast diffusion equation.

The good practical behaviour of the backward Euler method can also be seen from the following Table 2.21 where we return to the PDE case and show the analogous results to Table 2.19. Firstly, note from Table 2.21 that the computational extinction times seem to converge albeit at a slower rate than for the fully discrete Le Roux's method. Secondly we observe that the maximal number of Newton iterations needed is small compared to the maximal number of iterations in Table 2.19. The computations from $t = 0$ to $T^{2^{-11}}_{ex,10^{-8}} = 0.042967$ took only about five minutes on an IBM RS6000 in Case (b) with $p = 5$ (as compared to over one hour for Le Roux's method). Finally, it is also seen that the maximal number of iterations is always needed at the first time step.

		Case (a)		*Case (b)*	
τ		$p = 2$	$p = 5$	$p = 2$	$p = 5$
2^{-6}	$T^{\tau}_{ex,10^{-8}}$	0.203125	0.093750	0.187500	0.093750
	k_{max}	8	11	8	10
	$T^{\tau}_{k_{max}}$	0.015625	0.015625	0.015625	0.015625
2^{-7}	$T^{\tau}_{ex,10^{-8}}$	0.171875	0.078125	0.148438	0.062500
	k_{max}	9	14	9	13
	$T^{\tau}_{k_{max}}$	0.007813	0.007813	0.007813	0.007813
2^{-8}	$T^{\tau}_{ex,10^{-8}}$	0.152343	0.062500	0.132813	0.054688
	k_{max}	10	17	11	16
	$T^{\tau}_{k_{max}}$	0.003906	0.003906	0.003906	0.003906
2^{-9}	$T^{\tau}_{ex,10^{-8}}$	0.142578	0.058594	0.123047	0.046875
	k_{max}	12	20	12	20
	$T^{\tau}_{k_{max}}$	0.001953	0.001953	0.001953	0.001953
2^{-10}	$T^{\tau}_{ex,10^{-8}}$	0.136719	0.054688	0.117188	0.043945
	k_{max}	13	24	14	24
	$T^{\tau}_{k_{max}}$	0.000977	0.000977	0.000977	0.000977
2^{-11}	$T^{\tau}_{ex,10^{-8}}$	0.133789	0.053711	0.114746	0.042967
	k_{max}	15	27	16	27
	$T^{\tau}_{k_{max}}$	0.000488	0.000488	0.000488	0.000488

Table 2.21: Fully discrete backward Euler method: computational extinction times $T^{\tau}_{ex,10^{-8}}$

The Figures 2.6 – 2.9 are analogues of Figures 2.2 – 2.5. At first sight the Figures seem to be very similar to their analogues. Note however that the profiles closest to the true asymptotic profile are not as close as their analogues for the fully discrete Le Roux's method. This is presumably due to larger discretization errors which directly effect the profiles. However, there is also an indirect effect in the Figures below: larger discretization errors causing larger computational extinction times which in turn effect the scaling of the profiles when factorized by T^*.

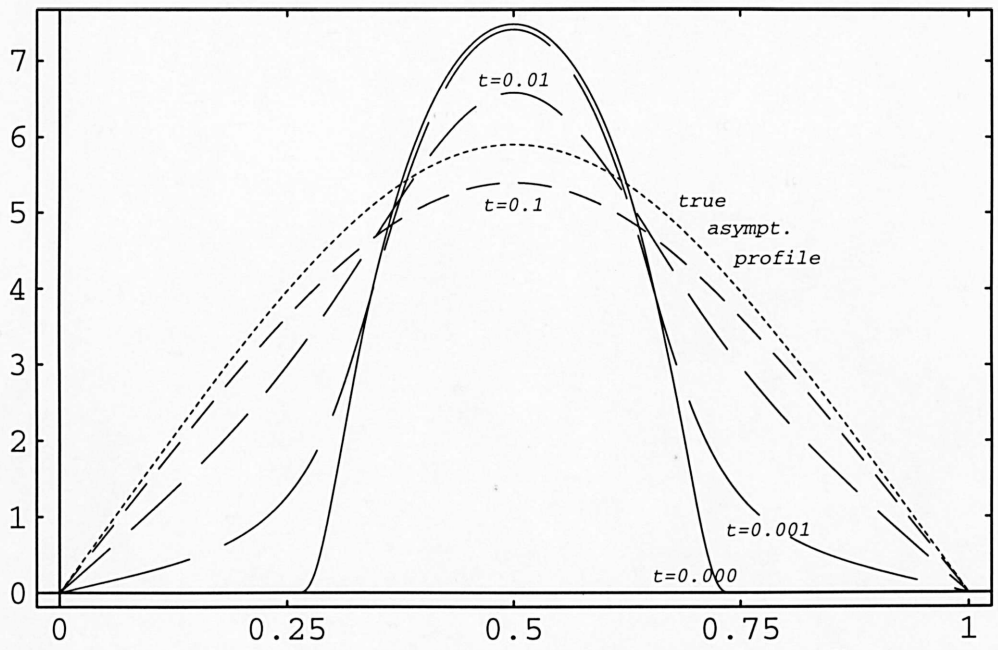


Figure 2.6: Case (a): $p = 2$: Profiles of factored solution as $t \to T^{2^{-11}}_{ex,10^{-8}} = 0.133789$

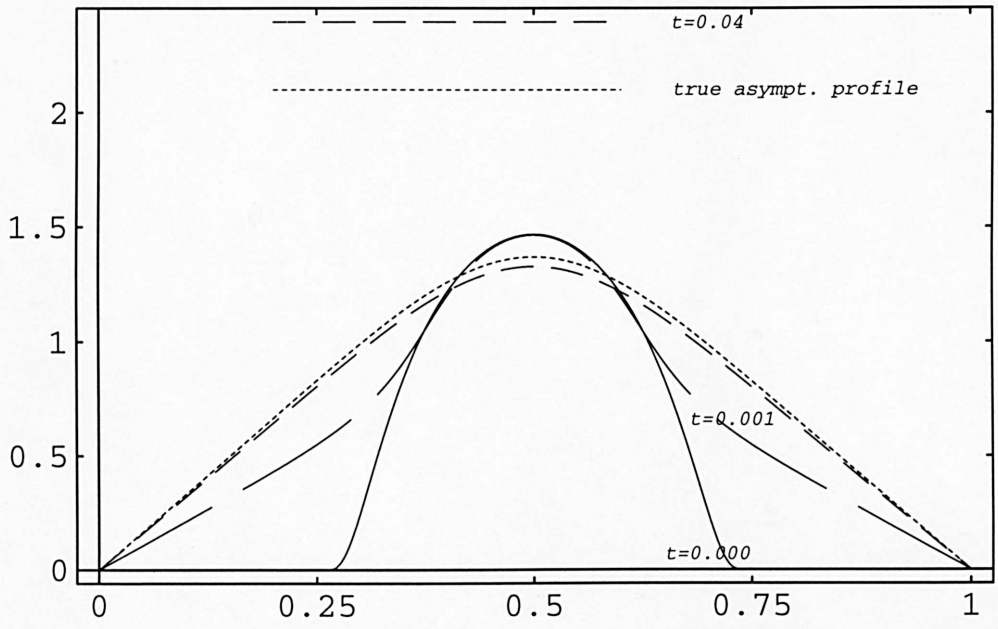


Figure 2.7: Case (a): $p = 5$: Profiles of factored solution as $t \to T^{2^{-11}}_{ex,10^{-8}} = 0.053711$

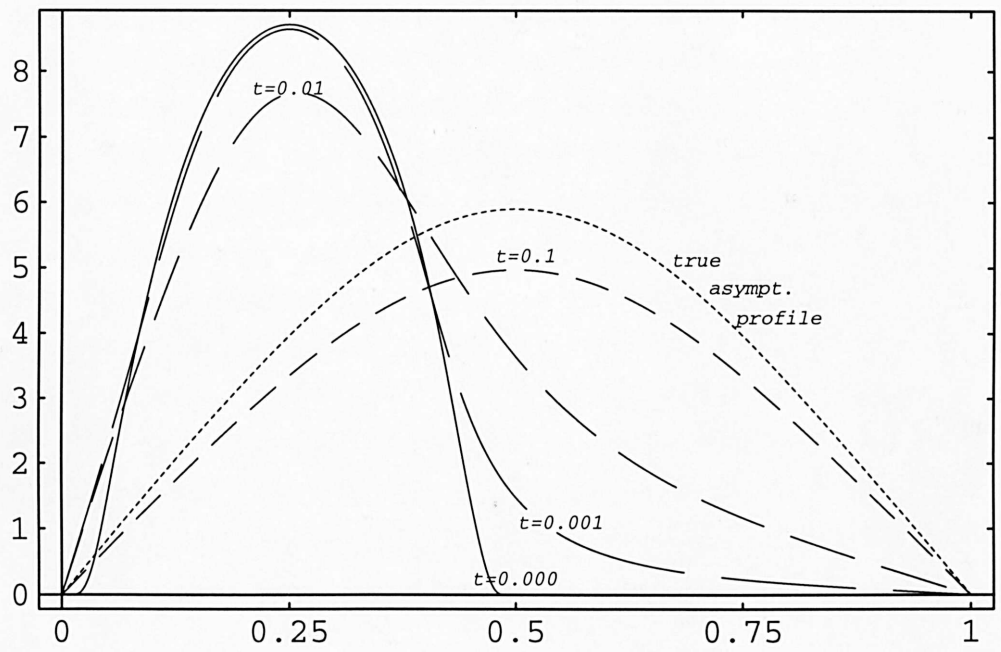


Figure 2.8: Case (b): $p = 2$: Profiles of factored solution as $t \to T^{2^{-11}}_{ex,10^{-8}} = 0.114746$

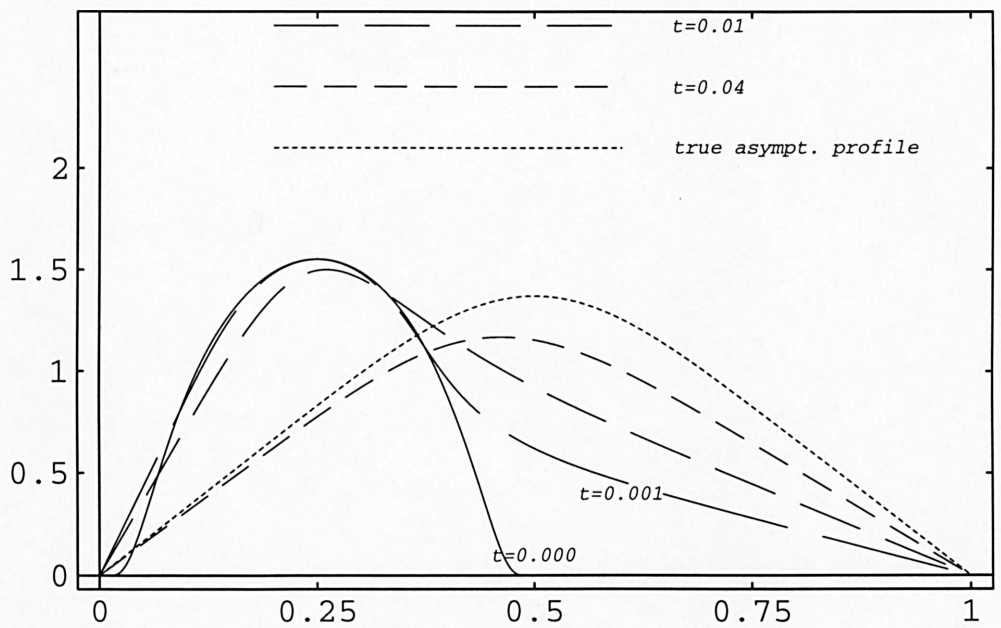


Figure 2.9: Case (b): $p = 5$: Profiles of factored solution as $t \to T^{2^{-11}}_{ex,10^{-8}} = 0.042967$

2.8 Linear Schemes

In this second chapter the emphasis of our analysis for the time discretization of (P) has been on the backward Euler method. For this method we have given a full error analysis in sections 2.4 and 2.5. However, the backward Euler method yields a system of nonlinear algebraic equations which we solved in section 2.7 on using Newton's method. Le Roux's method, for which we have not given a full error analysis, was implemented on using an iteration of a linear scheme, (6.10). In section 2.7 we saw that this iterative scheme performed very well as far as accuracy was concerned. On the other hand it proved to be less practical than solving the nonlinear scheme for the backward Euler method via Newton's iteration as it required far more iterations (especially near extinction time).

In this last section we will consider two linear schemes without iteration. The following however is not meant to be a thorough investigation and we will point out where further research could be done. To keep matters simple we will assume as in the last section that

$$f \equiv 0\,,$$

and we will only deal with model ODEs. We will only remark on the PDE case.

2.8.1 A Scheme based on Extrapolation

Our first linear scheme is based on extrapolation, which Nochetto [33] considered for the Stefan problem. We start by considering the regularized model ODE

$$\tfrac{d}{dt}v + \alpha u = 0\,, \qquad u(0) = g \tag{8.1}$$

with

$$v \equiv \phi(u) \tag{8.2}$$

and where, for notational convenience, we have set

$$u \equiv u_\varepsilon^M\,, \quad \phi \equiv \phi_\varepsilon^M\,, \quad v \equiv v_\varepsilon^M\,. \tag{8.3}$$

On choosing $\tau = \frac{T}{N}$, $N \in \mathbb{N}$, we set again $t_n := n\tau$, $n = 1 \to N$, and approximate $u^n \equiv u(t_n)$ via the backward Euler method

$$\frac{V^n - V^{n-1}}{\tau} + \alpha U^n = 0\,, \qquad n = 1 \to N\,. \tag{8.4}$$

The main point is the following. In the previous sections we required (8.2) to be satisfied exactly at every discrete time level t_n, i.e. $V^n \equiv \phi(U^n)$. This is now replaced by the extrapolation scheme

$$V^n := \phi(U^{n-1}) + L_M(U^n - U^{n-1}) \tag{8.5}$$

where L_M is the Lipschitz constant of $\phi \equiv \phi_\varepsilon^M$ as after (2.7).

On using the standard notation $\partial y^n := \frac{y^n - y^{n-1}}{\tau}$ we obtain from (8.4) and (8.5) the scheme

$$L_M \partial U^n + \alpha U^n = L_M \partial U^{n-1} - \partial\phi(U^{n-1}), \qquad n = 1 \to N. \tag{8.6}$$

At every time level (8.6) is linear in U^n and it requires the results U^{n-1} and U^{n-2} of the two previous time levels. When $n = 1$ a value for U^{-1} is required which we define as

$$U^{-1} := U^0 := g, \tag{8.7}$$

i.e. at the first time level we apply the backward Euler method to $\frac{d}{dt}[L_M u] + \alpha u = 0$.

Our first aim is to show stability of the scheme (8.6).

Lemma 2.15 *For all $\varepsilon \in (0, \varepsilon_0]$ and τ as above we have that*

$$\varepsilon\tau \sum_{n=1}^{N} |\partial U^n|^2 + \max_{n=1\to N} |U^n|^2 \le C. \tag{8.8}$$

Proof. On multiplying (8.6) by $2\tau\partial U^n$ and noting the elementary identity $a(a-b) = \frac{1}{2}a^2 + \frac{1}{2}(a-b)^2 - \frac{1}{2}b^2$ and the Mean Value Theorem we have that

$$2\tau L_M |\partial U^n|^2 + \alpha|U^n|^2 + \alpha|U^n - U^{n-1}|^2 = \alpha|U^{n-1}|^2 + 2\tau[L_M - \phi'(\xi)]\partial U^{n-1}\partial U^n,$$

where $\xi \in (U^{n-1}, U^{n-2})$ or (U^{n-2}, U^{n-1}). Noting that $\phi' \equiv (\phi_\varepsilon^M)' \ge \lambda\varepsilon$ and Young's inequality for the product $\partial U^{n-1}\partial U^n$ yield that

$$\begin{aligned} \tau[L_M - \lambda\varepsilon]|\partial U^n|^2 + 2\lambda\varepsilon\tau|\partial U^n|^2 + \alpha|U^n|^2 + \alpha|U^n - U^{n-1}|^2 \\ \le \alpha|U^{n-1}|^2 + \tau[L_M - \lambda\varepsilon]|\partial U^{n-1}|^2. \end{aligned}$$

On summing over n from 1 to $m \le N$ and noting (8.7) we obtain that

$$\tau[L_M - \lambda\varepsilon]|\partial U^m|^2 + 2\lambda\varepsilon\tau \sum_{n=1}^{m} |\partial U^n|^2 + \alpha|U^m|^2 + \alpha \sum_{n=1}^{m} |U^n - U^{n-1}|^2 \le \alpha|U^0|^2$$

from which the desired result (8.8) follows. ∎

We note that for $\varepsilon = 0$, i.e. no regularization, the method of the proof above does not yield a bound on the term $\tau \sum_{n=1}^{N} |\partial U^n|^2$. The existence of such a bound is crucial in the following error analysis. Before we do this analysis, we also note that under our convention (8.3) the ODE analogue of (2.26) reads

$$\varepsilon^{\frac{1}{2}} |\tfrac{d}{dt}u|_{L^2(0,T)} + L_M^{-\frac{1}{2}} |\tfrac{d}{dt}v|_{L^2(0,T)} + |u|_{L^\infty(0,T)} \le C. \tag{8.9}$$

Lemma 2.16 *Let u be the solution of (8.1) and U be the piecewise constant function given by $U(t) := U^n$ as in (8.6) for $t \in (t_{n-1}, t_n]$ and $U(0) := U^0$ as in (8.7). Then we have that*

$$\varepsilon |u - U|^2_{L^2(0,T)} + |\phi(u) - \phi(U)|^2_{L^2(0,T)} \leq C\varepsilon^{-2}\tau^2 + C\tau\,. \tag{8.10}$$

Proof. We follow along the lines of [33]. On integrating (8.1) on $I^i := (t_{i-1}, t_i]$ we obtain that

$$\partial v^i + \tau^{-1}\alpha \int_{I^i} u\,dt = 0\,, \tag{8.11}$$

where $v^i := \phi(u^i) := \phi(u(i\tau))$. On the other hand we define a function ψ by $\psi(s) := L_M s - \phi(s)$, $s \in \mathbb{R}$, and rewrite (8.6) as

$$\partial\phi(U^n) + \alpha U^n = \partial[\psi(U^{n-1}) - \psi(U^n)]\,. \tag{8.12}$$

Subtracting (8.12) from (8.11) and summing over i from 1 to $n \leq N$ yields that

$$\sum_{i=1}^{n}[\partial v^i - \partial\phi(U^i)] + \alpha\sum_{i=1}^{n}\left[\left(\tau^{-1}\int_{I^i} u\,dt\right) - U^i\right] = \sum_{i=1}^{n}\partial[\psi(U^i) - \psi(U^{i-1})]\,.$$

On setting $e_u := u - U$, multiplying by τ and noting (8.7) the last equation can be rewritten as

$$[\phi(u^n) - \phi(U^n)] + \alpha\sum_{i=1}^{n}\int_{I^i} e_u\,dt = [\phi(u^0) - \phi(U^0)] + \tau\partial\psi(U^n)\,.$$

On multiplying by $\eta \equiv \int_{I^n} e_u\,dt$, summing over n from 1 to N and noting that $U^0 = u^0$ we obtain that

$$\begin{aligned}
&\sum_{n=1}^{N}\int_{I^n}[\phi(u^n) - \phi(U^n)][u - U^n]\,dt && (I)\\
&+ \alpha\sum_{n=1}^{N}\left[\int_{I^n} e_u\,dt\right]\sum_{i=1}^{n}\left[\int_{I^i} e_u\,dt\right] && (II)\\
= &\sum_{n=1}^{N}\tau[\partial\psi(U^n)]\left[\int_{I^n} e_u\,dt\right]\,. && (III)
\end{aligned}$$

We now bound the terms (I) – (III) separately.

$$\begin{aligned}
(I) \equiv &\sum_{n=1}^{N}\int_{I^n}[\phi(u) - \phi(U^n)][u - U^n]\,dt && (I_1)\\
&+ \sum_{n=1}^{N}\int_{I^n}[\phi(u^n) - \phi(u)][u - U^n]\,dt\,. && (I_2)
\end{aligned}$$

Since $\phi \equiv \phi_\varepsilon^M$, we have from (2.2) and (2.3) that

$$(I_1) \geq \tfrac{1}{2}L_M^{-1}|\phi(u) - \phi(U)|^2_{L^2(0,T)} + \tfrac{1}{2}\lambda\varepsilon|e_u|^2_{L^2(0,T)}\,.$$

In order to bound (I_2) we note (8.9) and (8.8) and obtain that

$$|(I_2)| \le \sum_{n=1}^{N} \int_{I^n} |\phi(u^n) - \phi(u)||u - U^n| \, dt \le \sum_{n=1}^{N} \int_{I^n} \left| \int_t^{t_n} \tfrac{dv}{ds} \, ds \right| (|u| + |U^n|) \, dt$$
$$\le C \sum_{n=1}^{N} \int_{I^n} \left[\int_{I^n} |\tfrac{dv}{ds}| \, ds \right] dt = C\tau |\tfrac{dv}{dt}|_{L^1(0,T)} \le C\tau.$$

To bound (II) we use the identity (where we use again superscripts)

$$2 \sum_{n=1}^{N} a^n \left[\sum_{i=1}^{n} a^i \right] = \left[\sum_{n=1}^{N} a^n \right]^2 + \sum_{n=1}^{N} (a^n)^2$$

which yields that

$$(II) \equiv \tfrac{1}{2}\alpha \left| \sum_{n=1}^{N} \int_{I^n} e_u \, dt \right|^2 + \tfrac{1}{2}\alpha \sum_{n=1}^{N} \left[\int_{I^n} e_u \, dt \right]^2 \ge \tfrac{1}{2}\alpha \left| \int_0^T e_u \, dt \right|^2 .$$

Since $|\psi'|_{0,\infty,\mathbb{R}} \le L_M$ we can bound (III) on noting the Mean Value Theorem as follows:

$$(III) \le L_M \sum_{n=1}^{N} \int_{I^n} \tau |\partial U^n| |e_u| \, dt .$$

Now we use Cauchy's and Young's inequalities and the fact that ∂U^n is constant on I^n to obtain that

$$\begin{aligned} (III) &\le L_M \sum_{n=1}^{N} \left[\int_{I^n} |e_u|^2 \, dt \right]^{\frac{1}{2}} \left[\int_{I^n} \tau^2 |\partial U^n|^2 \, dt \right]^{\frac{1}{2}} \\ &\le \tfrac{1}{4}\lambda\varepsilon |e_u|^2_{L^2(0,T)} + C\varepsilon^{-1}\tau^2 \left[\tau \sum_{n=1}^{N} |\partial U^n|^2 \right] \\ &\le \tfrac{1}{4}\lambda\varepsilon |e_u|^2_{L^2(0,T)} + C\varepsilon^{-2}\tau^2 , \end{aligned}$$

where for the last inequality we have noted (8.8). On collecting all the above bounds we obtain the desired result (8.10). ∎

Next we want to obtain an a priori error bound for the *regularized* model ODE (8.1), (8.2). To this end we rewrite (8.10) for clarity as

$$\varepsilon |u_\varepsilon^M - U_\varepsilon^M|^2_{L^2(0,T)} + |\phi_\varepsilon^M(u_\varepsilon^M) - \phi_\varepsilon^M(U_\varepsilon^M)|^2_{L^2(0,T)} \le C\varepsilon^{-2}\tau^2 + C\tau \tag{8.13}$$

and note the ODE analogue of (2.50):

$$\varepsilon |u - u_\varepsilon^M|^2_{L^2(0,T)} + |\phi(u) - \phi_\varepsilon^M(u_\varepsilon^M)|^2_{L^2(0,T)} \le C\varepsilon^{\frac{p+1}{p-1}} \tag{8.14}$$

(recall that it was proved in Lemma 2.4 that for M sufficiently large, $u_\varepsilon \equiv u_\varepsilon^M$ and $\phi_\varepsilon(u_\varepsilon) \equiv \phi_\varepsilon^M(u_\varepsilon^M)$).

From equating the bounds in (8.13) and (8.14) we obtain that

$$|u - U_\varepsilon^M|_{L^2(0,T)} \leq C \begin{cases} \tau^{\frac{1}{p+1}}, \text{ if } \varepsilon = C\tau^{\frac{p-1}{p+1}} \text{ and } p \in (1,3], \\ \tau^{\frac{2}{3p-1}}, \text{ if } \varepsilon = C\tau^{\frac{2(p-1)}{3p-1}} \text{ and } p \in [3,\infty), \end{cases} \tag{8.15}$$

$$|\phi(u) - \phi_\varepsilon^M(U_\varepsilon^M)|_{L^2(0,T)} \leq C \begin{cases} \tau^{\frac{1}{2}}, \text{ if } \varepsilon = C\tau^{\frac{p-1}{p+1}} \text{ and } p \in (1,3], \\ \tau^{\frac{p+1}{3p-1}}, \text{ if } \varepsilon = C\tau^{\frac{2(p-1)}{3p-1}} \text{ and } p \in [3,\infty). \end{cases} \tag{8.16}$$

We note that when equating the bounds (8.13) and (8.14) it turns out that the term $C\tau$ in (8.13) is the dominant term for $p \in (1,3]$. We recall that this is the bound on (I_2) in the proof of Lemma 2.16. Any attempt to improve on the above error bounds for $p \in (1,3]$ should therefore improve on the bound on (I_2) (or try to avoid this term altogether).

We also remark that unlike for the backward Euler method (cf. Theorems 2.3 and 2.5) we were unable to prove a discrete analogue of (8.14). Hence we cannot give error bounds on $|u - U|_{L^2(0,T)}$ and $|\phi(u) - \phi(U)|_{L^2(0,T)}$.

Next, we present some numerical results for the ODE (8.1), (8.2) when solved by the scheme (8.6), (8.7) with $\phi \equiv \phi_\varepsilon^M$. In the following computations λ in the definition of ϕ_ε^M is chosen to be equal to 1, $g \equiv 1$ and ε is chosen according to (8.15) (and (8.16)) with $C \leq 1$. Since $f \equiv 0$, we can choose $M \equiv 1$ by (2.38), and hence by (2.7) $L_M \equiv \lambda p M^{p-1} = p$ in our computations. α in (8.1) is chosen to be equal to $\frac{\beta p}{p-1}$ with $\beta = 5$ and hence by (7.21) the exact solution of (8.1), (8.2) is given by

$$u(t) \equiv [1 - 5t]_+^{\frac{1}{p-1}}, \quad t \in [0,T], \tag{8.17}$$

with extinction time $T_{ex} = 0.2$.

We begin with showing some qualitative behaviour of the pointwise error

$$Error(t_n) = (U_\varepsilon^M)^n - u^n.$$

In Figure 2.10 we show the piecewise linear interpolant of $Error$ for $p = 2$. T was chosen to be equal 0.8 and $\tau = 2^{-10}$ (which is fairly small so that the graphs appear to be nearly smooth). The Figure shows two graphs for comparison. The dashed graph shows the result for $C = 1$ and the continuous graph for $C = 0.001$.

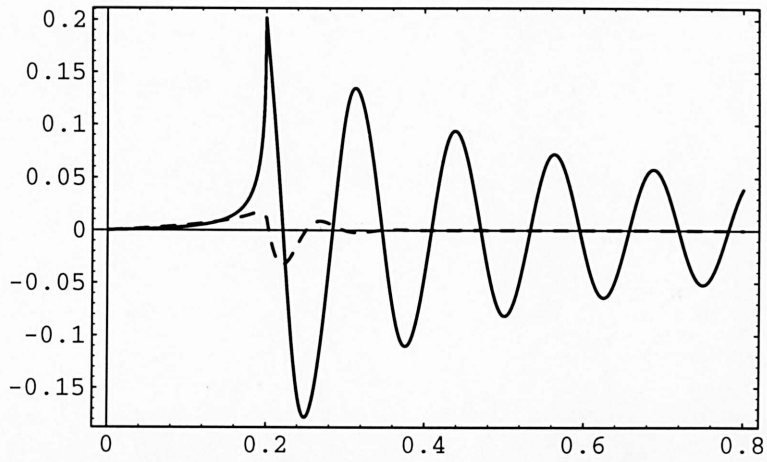


Figure 2.10: *Error* for extrapolation method, $p = 2$; $C = 0.001$: ——, $C = 1$: – – –

We observe two interesting features of the extrapolation method in Figure 2.10 which we do not try to explain. Firstly, after extinction time the discrete solution of the extrapolation method shows an oscillation with decaying amplitude. Secondly, the modulus of *Error* is smaller for larger regularization parameter ε.

We now check our error bound (8.15) from the first part of this subsection. In the following Table 2.22 we show the discrete L^2 error and the computational order of convergence for $p = 2$. We set

$$E^{\varepsilon}_{u,2}(\tau) := \left[\tau \sum_{n=1}^{N} |u^n - (U^M_{\varepsilon})^n|^2\right]^{\frac{1}{2}}, \quad q^{\varepsilon}_{u,2}(\tau) := \log_2 \frac{E^{\varepsilon}_{u,2}(2\tau)}{E^{\varepsilon}_{u,2}(\tau)}.$$

		$C = 0.001$		$C = 1$	
τ	T_*	$E^{\varepsilon}_{u,2}(\tau)$	$q^{\varepsilon}_{u,2}(\tau)$	$E^{\varepsilon}_{u,2}(\tau)$	$q^{\varepsilon}_{u,2}(\tau)$
2^{-6}	0.390625	0.05465931	—	0.05018574	—
2^{-7}	0.398438	0.03588186	0.6072	0.03121130	0.6852
2^{-8}	0.398438	0.02394927	0.5832	0.01888206	0.7251
2^{-9}	0.398438	0.01557446	0.6208	0.01091032	0.7913
2^{-10}	0.399414	0.00993606	0.6484	0.00610044	0.8387
2^{-11}	0.399902	0.00625522	0.6676	0.00331455	0.8801
2^{-12}	0.399902	0.00389416	0.6837	0.00175643	0.9161
2^{-13}	0.399902	0.00239790	0.6995	0.00091585	0.9394
2^{-14}	0.399963	0.00146384	0.7121	0.00048209	0.9258

Table 2.22: $E^{\varepsilon}_{u,2}(\tau)$ and $q^{\varepsilon}_{u,2}(\tau)$ for ODE case with $T_{ex} = 0.2$ and $T = 0.4$

We recall that T_* was defined in (4.1). As in Figure 2.10 the numbers for $C = 1$ are better than those for $C = 0.001$. This may well be due to our suboptimal error analysis which not only produced a presumably suboptimal error bound (8.10) but also determined the balancing of terms for the final error bound (8.15) and the choice of ε as a power of τ which we used in our computations. In both cases however, the computational order of convergence is better than predicted by (8.15) which yields $O(\tau^{\frac{1}{3}})$ for $p = 2$.

The following Table 2.23 is an analogue of Table 2.22 in order to check our error bound (8.16). It shows $E^{\varepsilon}_{v,2}(\tau)$ and $q^{\varepsilon}_{v,2}(\tau)$ for $p = 2$ where we define $E^{\varepsilon}_{v,2}(\tau)$ and $q^{\varepsilon}_{v,2}(\tau)$ by

$$E^{\varepsilon}_{v,2}(\tau) := \left[\tau \sum_{n=1}^{N} |\phi(u^n) - \phi^M_{\varepsilon}((U^M_{\varepsilon})^n)|^2\right]^{\frac{1}{2}}, \qquad q^{\varepsilon}_{v,2}(\tau) := \log_2 \frac{E^{\varepsilon}_{v,2}(2\tau)}{E^{\varepsilon}_{v,2}(\tau)}.$$

		$C = 0.001$		$C = 1$	
τ	T_*	$E^{\varepsilon}_{v,2}(\tau)$	$q^{\varepsilon}_{v,2}(\tau)$	$E^{\varepsilon}_{v,2}(\tau)$	$q^{\varepsilon}_{v,2}(\tau)$
2^{-6}	0.390625	0.02259265	—	0.02425153	—
2^{-7}	0.398438	0.01250653	0.8531	0.01333004	0.8634
2^{-8}	0.398438	0.00666611	0.9078	0.00704993	0.9190
2^{-9}	0.398438	0.00345826	0.9468	0.00363087	0.9572
2^{-10}	0.399414	0.00176380	0.9714	0.00183832	0.9819
2^{-11}	0.399902	0.00089054	0.9859	0.00092139	0.9965
2^{-12}	0.399902	0.00044711	0.9940	0.00045937	1.0041
2^{-13}	0.399902	0.00022383	0.9982	0.00022853	1.0073
2^{-14}	0.399963	0.00011191	1.0001	0.00011367	1.0075

Table 2.23: $E^{\varepsilon}_{v,2}(\tau)$ and $q^{\varepsilon}_{v,2}(\tau)$ for ODE case with $T_{ex} = 0.2$ and $T = 0.4$

The results in Figure 2.23 for $C = 0.001$ and $C = 1$ are in good agreement with each other. Again the computational order of convergence is better than predicted: (8.16) yields $O(\tau^{\frac{1}{2}})$ for $p = 2$.

We also note that the errors in Table 2.22 both for $C = 1$ and $C = 0.001$ are worse than the corresponding errors for the backward Euler method in Table 2.10. We report that the same is true for the errors in Table 2.23 and the corresponding errors for the backward Euler method.

We conclude our exposition of the extrapolation method by some short remarks on the PDE case. As mentioned before, the extrapolation scheme (8.6) was adapted from a paper by Nochetto [33] where it is used on the Stefan problem. Given a (regular) partitioning $\mathcal{T}^h$ of Ω polyhedral, i.e. $\Omega \equiv \bigcup_{\kappa \in \mathcal{T}^h} \bar{\kappa}$, it is a simple matter to adapt Nochetto's fully discrete variational formulation to our problem (P^M_{ε}):

For $n = 1 \to N$ find $U^n \in S^h_0$ such that

$$L_M \langle \partial \mathcal{P}^h U^n, \chi \rangle + \langle \nabla U^n, \nabla \chi \rangle = \langle \partial [L_M \mathcal{P}^h U^{n-1} - \phi(\mathcal{P}^h U^{n-1})], \chi \rangle \qquad \forall \chi \in S^h_0, \quad (8.18)$$

where $U^n \equiv (U^M_{\varepsilon})^n$, $\phi \equiv \phi^M_{\varepsilon}$ and $\mathcal{P}^h$ denotes the L^2 projection operator onto $\bar{S}^h \equiv \{\xi \in L^{\infty}(\Omega) : \xi|_{\kappa}$ is constant $\forall \kappa \in \mathcal{T}^h\}$. The projection operator $\mathcal{P}^h$ is given by

$$\langle \mathcal{P}^h z, \xi \rangle = \langle z, \xi \rangle \qquad \forall z \in L^2(\Omega),\ \xi \in \bar{S}^h.$$

$\mathcal{P}^h$ occurs in (8.18) in order to include quadrature in space.

With $\{\chi_i\}_{i=1}^R$ being the canonical basis of S_0^h and φ_j denoting the characteristic function of supp χ_j multiplied by $\frac{1}{d+1}$ we can rewrite (8.18) in matrix form as

$$AU^n := [L_M\mathcal{M} + \tau K]U^n = b\,, \tag{8.19}$$

where K and $\mathcal{M}$ are $R \times R$ matrices with entries $K_{ij} = \langle \nabla\chi_i, \nabla\chi_j\rangle$, $\mathcal{M}_{ij} = \langle \chi_i, \varphi_j\rangle$ and b is a vector depending on U^{n-1} and U^{n-2}.

A is a symmetric, positive definite matrix and hence (8.19) is uniquely solvable. Moreover, A is independent of n which suggests that in practice one uses a Cholesky factorization just once at the beginning of the computations. Obviously, this is a computationally nice feature of (8.18). However, here we will not present any computational results. Instead we remark that a straightforward adaptation of Nochetto's error analysis yields the following analogue of (8.10) (where for clarity we have included the indices ε and M):

$$\varepsilon|u_\varepsilon^M - U_\varepsilon^M|^2_{L^2(Q_T)} + |\phi_\varepsilon^M(u_\varepsilon^M) - \phi_\varepsilon^M(\mathcal{P}^h U_\varepsilon^M)|^2_{L^2(Q_T)} \le C\varepsilon^{-2}\tau^2 + C\tau + Ch^2\,. \tag{8.20}$$

2.8.2 A Scheme based on Relaxation

In this subsection we consider the *unregularized* model ODE (8.1) where instead of (8.2) we now require that for arbitrary but fixed $k \in \mathbb{R}_+$,

$$k^{-1}\tfrac{d}{dt}v_k = u_k - \phi^{-1}(v_k)\,. \tag{8.21}$$

We will show further below how (8.21) can be used to derive a linear discretization of (8.1). First we try to explain the reasoning behind (8.21). As $k \to \infty$ one expects the left hand side of (8.21) to converge towards 0 and the right hand side to converge towards $u - \phi^{-1}(v)$, which is equivalent to (8.2). Thus (8.21) is a *relaxation* of (8.1) and we expect it to be a feasible approximation for large k. The parameter k is called the relaxation parameter. We remark that in the slow diffusion case relaxations of the form (8.21) are in fact used in the literature to replace constitutive relations of the form (8.2) in order to improve on the modelling of such phase transition effects as superheating and undercooling, e.g. [45], [46]. But here we do not attempt to give any physical meaning to (8.21).

Alternatively, one can relax (8.2) by replacing the right hand side of (8.21) by $\phi(u_k) - v_k$, see e.g. [5] for the porous medium equation. The advantage of (8.21) as stated above however is that we can write down the exact solution of the relaxed problem (8.1) (with $u \equiv u_k$, $v \equiv v_k$) and (8.21): From equating (8.1) and (8.21) we obtain that

$$\alpha u_k = -k(u_k - \phi^{-1}(v_k))$$

and hence that

$$v_k = \phi([1 + \alpha k^{-1}]u_k)\,.$$

Thus we obtain from (8.1) that

$$\tfrac{d}{dt}[\phi([1+\alpha k^{-1}]u_k)] + \alpha u_k = 0\,, \qquad u_k(0) = g\,. \tag{8.22}$$

On setting

$$\tilde{u}_k := [1+\alpha k^{-1}]u_k\,, \quad \tilde{\alpha}_k := \frac{\alpha}{1+\alpha k^{-1}}\,, \quad \tilde{g} := [1+\alpha k^{-1}]g \tag{8.23}$$

(8.22) is equivalent to

$$\tfrac{d}{dt}\phi(\tilde{u}_k) + \tilde{\alpha}_k \tilde{u}_k = 0\,, \quad \tilde{u}_k(0) = \tilde{g}\,. \tag{8.24}$$

Equation (8.24) is of the same form as (8.1), (8.2) whose exact solution is given by

$$u(t) = \Big[g^{p-1} - \alpha\tfrac{p-1}{p}t\Big]_+^{\frac{1}{p-1}}\,, \quad t \in [0,T]\,, \tag{8.25}$$

see (7.21). Hence, on noting (8.23) we obtain that

$$u_k(t) = \Big[g^{p-1} - \tfrac{\alpha}{[1+\alpha k^{-1}]^p}\tfrac{p-1}{p}t\Big]_+^{\frac{1}{p-1}}\,, \quad t \in [0,T]\,, \tag{8.26}$$

with extinction time

$$T_{ex,k} = [1+\alpha k^{-1}]^p \tfrac{p}{\alpha(p-1)} g^{p-1} = [1+\alpha k^{-1}]^p T_{ex}\,.$$

It is easy to derive from (8.25) and (8.26) the maximal relaxation error: Assuming $T \geq T_{ex}$ we have that

$$\begin{aligned} \max_{t\in[0,T]} |u(t) - u_k(t)| &= \max_{t\in[0,T]} [u_k(t) - u(t)] = u_k(T_{ex}) \\ &= g[1-(1+\alpha k^{-1})^{-p}]^{\frac{1}{p-1}} \approx Ck^{-1}\,. \end{aligned} \tag{8.27}$$

We now derive a linear discretization for (8.1) (with $u \equiv u_k$, $v \equiv v_k$) and (8.21). Analogously to (8.4) we first approximate (8.1) by

$$\frac{V_k^n - V_k^{n-1}}{\tau} + \alpha U_k^n = 0\,, \qquad n = 1 \to N\,, \tag{8.28}$$

where we set

$$V_k^0 := \phi(g)\,. \tag{8.29}$$

Now we approximate (8.21) by

$$\frac{V_k^n - V_k^{n-1}}{\tau} - kU_k^n = -k\phi^{-1}(V_k^{n-1})\,, \qquad n = 1 \to N\,. \tag{8.30}$$

Clearly, if V_k^{n-1} is given, then (8.28) and (8.30) form a system of linear equations in the unknowns U_k^n and V_k^n which are easily obtained as

$$\left.\begin{aligned} U_k^n &= \tfrac{k}{\alpha+k}\phi^{-1}(V_k^{n-1})\,, \\ V_k^n &= V_k^{n-1} - \tfrac{\alpha k}{\alpha+k}\tau\phi^{-1}(V_k^{n-1}) \end{aligned}\right\} \qquad n = 1 \to N\,. \tag{8.31}$$

Below we present some numerical results for the ODE (8.1), (8.2) when solved by the relaxation method (8.29), (8.31). As in the previous subsection we choose $\lambda = 1$, $g = 1$ and $\alpha = \frac{\beta p}{p-1}$ with $\beta = 5$ which again means that the exact solution is given by (8.17). As for the choice of k in our computations we set $k = \frac{1}{\tau}$.

Figures 2.11 and 2.12 show the piecewise linear interpolant of the discrete solution U_k^n for $p = 2$ and $p = 4$ when τ is equal to $2^{-7} \approx 0.0078$.

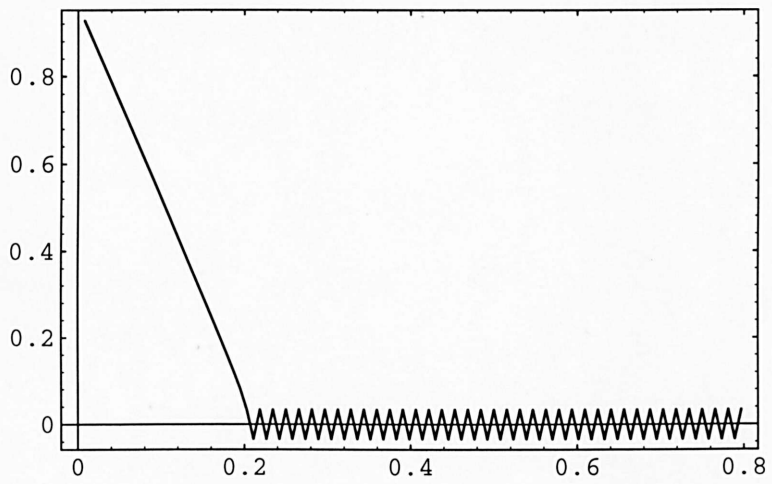


Figure 2.11: Relaxation method applied to ODE, $p = 2$

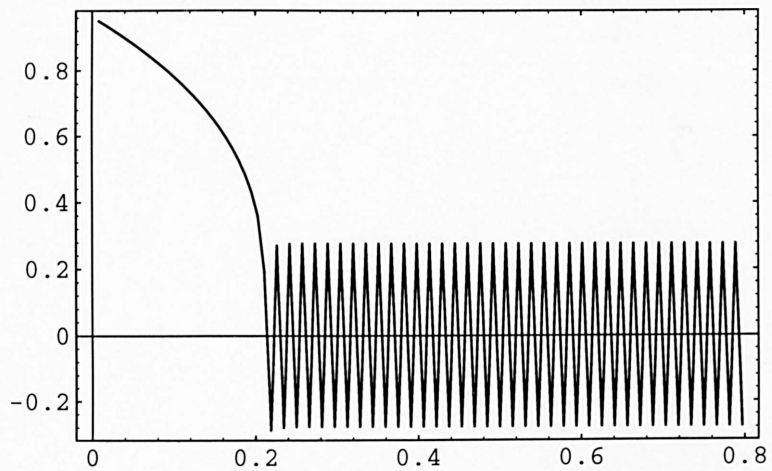


Figure 2.12: Relaxation method applied to ODE, $p = 4$

Obviously the most striking feature of Figures 2.11 and 2.12 is that after extinction time we observe an oscillation developing which seems to be of the form

$$\ldots, U_k^{n-1} = -A,\ U_k^n = A,\ U_k^{n+1} = -A, \ldots, \tag{8.32}$$

where $A > 0$. Although one may at first think that this oscillation is an artifact, we are in fact able to derive an exact expression for the amplitude A above: On assuming (8.32) and noting (8.31) and that ϕ is an odd function we obtain the following equivalences

$$\begin{aligned}
& -A = U_k^{n+1} &&= -U_k^n \\
\Longleftrightarrow \quad & \tfrac{k}{\alpha+k}\phi^{-1}(V_k^n) &&= -\tfrac{k}{\alpha+k}\phi^{-1}(V_k^{n-1}) \\
\Longleftrightarrow \quad & \phi^{-1}\left(V_k^{n-1} - \tfrac{\alpha k}{\alpha+k}\tau\phi^{-1}(V_k^{n-1})\right) &&= -\phi^{-1}(V_k^{n-1}) \\
\Longleftrightarrow \quad & V_k^{n-1} - \tfrac{\alpha k}{\alpha+k}\tau\phi^{-1}(V_k^{n-1}) &&= -V_k^{n-1} \\
\Longleftrightarrow \quad & V_k^{n-1} - [\tfrac{\alpha k}{\alpha+k}\tfrac{\tau}{2}](V_k^{n-1})^{\frac{1}{p}} &&= 0\,.
\end{aligned} \tag{8.33}$$

The positive root of (8.33) is given by

$$V_k^{n-1} = [\tfrac{\alpha k}{\alpha+k}\tfrac{\tau}{2}]^{\frac{p}{p-1}}\,. \tag{8.34}$$

Hence, we obtain from (8.34), (8.31) and (8.32) that

$$A = \frac{k}{\alpha+k}\left[\frac{\alpha k}{\alpha+k}\frac{\tau}{2}\right]^{\frac{1}{p-1}}\,. \tag{8.35}$$

Note that A can be bounded independently of the choice of $k > 0$:

$$A \leq C(\alpha, p)\,\tau^{\frac{1}{p-1}}\,.$$

We point out that we have not proved that the iteration (8.31) must always develop into the periodic pattern (8.32), but simply derived (8.35) on assuming (8.32). Nonetheless, we have observed (8.32) in all our computations. In Table 2.24 below we show A according to (8.35) for different values of p and τ with $k = \frac{1}{\tau}$. n_A denotes the number of mesh points greater than $T_{ex} = 0.2$ that were needed before $\big||U_k^n| - A\big| < 10^{-8}$.

	$\tau = 2^{-7}$		$\tau = 2^{-10}$		$\tau = 2^{-13}$	
p	A	n_A	A	n_A	A	n_A
2	0.03360639	4	0.00478882	3	0.00060886	3
2.5	0.09177669	11	0.02514544	9	0.00636107	8
3	0.15715082	13	0.05985656	14	0.02136606	13
3.5	0.21997327	20	0.10217980	19	0.04484891	18
4	0.27700615	24	0.14692710	22	0.07402161	21
4.5	0.32768700	28	0.19111778	27	0.10625071	23
5	0.37241709	32	0.23325673	31	0.13962108	27

Table 2.24: A according to (8.35) and computational convergence

We conclude with the following Table 2.25 which shows the maximum error

$$max.error := \max_{n=1 \to N} |U_k^n - u^n|$$

(where $N = \lfloor T/\tau \rfloor$ and $T = 0.8$ as for Figures 2.11 and 2.12) and the relaxation error given by (8.27). As before we set $k = \frac{1}{\tau}$.

	$\tau = 2^{-7}$		$\tau = 2^{-10}$		$\tau = 2^{-13}$	
p	*max.error*	(8.27)	*max.error*	(8.27)	*max.error*	(8.27)
2	0.05783768	0.13967654	0.00714217	0.01924883	0.00089147	0.00243694
2.5	0.12127306	0.09177669	0.02536342	0.07382499	0.00639335	0.01860933
3	0.21261473	0.15715082	0.06892732	0.14715528	0.02193903	0.05235991
3.5	0.29254198	0.21997327	0.12629536	0.22329932	0.04657279	0.09771847
4	0.35982493	0.27700615	0.18383033	0.29481179	0.07713922	0.14810348
4.5	0.41614794	0.32768700	0.23795308	0.35917382	0.10765121	0.19913288
5	0.46353915	0.37241709	0.28744482	0.41606342	0.14393377	0.24838003

Table 2.25: maximum error and relaxation error

From Table 2.25 we obtain that in most cases the maximum error is smaller than the relaxation error. Since $U_k^n - u^n = (U_k^n - u_k^n) + (u_k^n - u^n)$ this means that in practice we observe a partial cancellation of the relaxation error by the discretization error which is obviously a rather good property. On comparing the results for $\tau = 2^{-13}$ in Tables 2.24 and 2.25 we see that A is a good approximation for the maximum error.

We summarize section 2.8 as follows: Although we possess a full preliminary error analysis for the extrapolation method introduced in subsection 2.8.1, this method shows a rather unpractical behaviour. On the other hand we do not possess a full error analysis (not even in the ODE case) for the relaxation method introduced in this subsection. Nonetheless its good practical behaviour for the ODE case seems to warrant further research on its application to the PDE case. For instance, one could for a start try to adapt the error analysis given in [44], [23] and [5] which deal with slow diffusion. The recently published article [23] would be of particular interest as it applies Rulla's error analysis for the time discretization [39] to relaxation models with extrapolation. Unlike [44] and [5] however, it only deals with discretization in time, but not in space.

Bibliography

[1] Alt, H.W., 1992, *Lineare Funktionalanalysis.* Springer, Berlin.

[2] Alt, H.W., & Luckhaus, S., 1983, *Quasilinear elliptic-parabolic differential equations.* Math. Z. **183**, pp. 311-341.

[3] Amann, H., 1971, *On the existence of positive solutions of nonlinear elliptic boundary value problems.* Indiana Univ. Math. J. **21**, p.125.

[4] Barrett, J.W., 1992, *Finite element approximation of a non-Lipschitz nonlinear eigenvalue problem.* M^2AN **26**, pp.627-656.

[5] Barrett, J.W., & Knabner, P., 1997, *Finite element approximation of transport in porous media, Part 1: Error estimates for non-equilibrium adsorption processes.* SIAM J. Numer. Anal. **34**, pp. 201-227.

[6] Barrett, J.W., & Knabner, P., 1997, *Finite element approximation of transport in porous media, Part 2: Error estimates for equilibrium adsorption processes.* To appear in SIAM J. Numer. Anal. **34** (2).

[7] Berryman, J.G., 1977, *Evolution of a stable profile for a class of nonlinear diffusion equations with fixed boundaries.* J. Mathematical Phys. **18**, pp. 2108-2115.

[8] Berryman, J.G., & Holland, C.J., 1980, *Stability of the separable solution for fast diffusion.* Arch. Rational Mech. Anal. **74**, pp. 379-388.

[9] Brézis, H., 1971, *Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations*, in: "Contributions to Nonlinear Functional Analysis", Zarantonello, E.H. (ed.), Academic Press, New York – London.

[10] Callegari, A. & Nachman, A., 1978, *Some singular nonlinear differential equations arising in boundary layer theorey.* J. Math. Anal. Appl **64**, pp.96-105.

[11] Ciarlet, P.G., 1978, *The Finite Element Method for Elliptic Problems.* North–Holland, Amsterdam.

[12] Crandall, M.G., Rabinowitz, P.H. & Tartar, L., 1977, *On a Dirichlet problem with a singular nonlinearity.* Comm. Partial Differential Equations **2**, pp.193-222.

[13] Del Pino, M.A., 1992, *A global estimate for the gradient in a singular elliptic boundary value problem.* Proc. Roy. Soc. Edinburgh Sect.A **122**, pp.341-355.

[14] Diaz, J.I., & De Thelin, F., 1994, *On a nonlinear parabolic problem arising in some models related to turbulent flows.* SIAM J. Math. Anal. **25**, pp. 1085-1111.

[15] Elliott, C.M., 1987, *Error analysis of the enthalpy method for the Stefan problem.* IMA J. Numer. Anal. **7**, pp. 61-71.

[16] Feistauer, M., 1989, *On the finite element approximation of functions with noninteger derivatives*, Numer. Funct. Anal. and Optimiz. **10**, pp.91-110.

[17] Gilbarg, D. & Trudinger, N.S., 1983, *Elliptic Partial Differential Equations of Second Order.* 2nd Ed., Springer, Berlin.

[18] Girault, V., & Raviart, P.-A., 1979, *Finite Element Approximation of the Navier-Stokes Equations.* Lecture Notes in Mathematics **749**, Springer, Berlin.

[19] Grabmüller, H. & Pirner, R., 1992, *Existence theorems for some boundary problems in the nonlinear theory of annular elastic membranes.* Euro. J. Appl. Math. **3**, pp.299-317.

[20] Grisvard, P., 1992, *Singularities in Boundary Value Problems.* Research Notes in Applied Mathematics **22**, Masson, Paris.

[21] Gui, C. & Lin, F.H., 1993, *Regularity of an elliptic problem with a singular nonlinearity.* Proc. Roy. Soc. Edinburgh Sect.A **123**, pp.1023-1029.

[22] Jiang, X. & Nochetto, R.H. , 1994, *Effect of numerical integration for elliptic obstacle problems.* Numer. Math. **67**, pp.501-512.

[23] Jiang, X. & Nochetto, R.H. , 1997, *Optimal error estimates for semidiscrete phase relaxation models.* M^2AN **31**, pp. 91-120.

[24] Keller, H.B., 1969, *Elliptic boundary value problems suggested by nonlinear diffusion processes.* Arch. Rational Mech. Anal. **35**, pp. 363-388.

[25] Kufner, A., John, O. & Fučik, S., 1977, *Function Spaces.* Nordhoff, Leyden.

[26] Kwong, Y.C., 1988, *Interior and boundary regularity of solutions to a plasma problem.* Proc. AMS **104**, pp. 472-478.

[27] Lazer, A.C. & McKenna, P.J., 1991, *On a singular nonlinear elliptic boundary–value problem.* Proc. Amer. Math. Soc. **111**, pp.721-730.

[28] Le Roux, M.N., 1989, *Semi-discretization in time of a fast diffusion equation*, J. Math. Anal. Appl. **137**, pp. 354-370.

[29] Lesaint, P., & Pousin, J., 1992, *Error estimates for a nonlinear degenerate parabolic problem.* Math. Comp. **59**, pp. 339-358.

[30] Lions, J.L., 1969, *Quelques Méthodes de Résolution des Problèmes aux Limites Non-linéaires.* Dunod, Paris.

[31] Meis, Th. & Marcowitz, U., 1978, *Numerische Behandlung Partieller Differentialgleichungen.* Springer, Berlin.

[32] Nochetto, R.H., 1988, *Sharp L^∞ error estimates for semilinear elliptic problems with free boundaries.* Numer. Math. **54**, pp.243-255.

[33] Nochetto, R.H, 1989, *A stable extrapolation method for multidimensional degenerate parabolic problems.* Math. Comp. **53**, pp. 455-470.

[34] Nochetto, R.H., 1991, *Finite element methods for parabolic free boundary problems.* Chapter 2 in "Advances in Numerical Analysis", Vol I., Light, W. (ed.), pp.34-95. Oxford Univ. Press, New York.

[35] Nochetto, R.H., & Verdi, C., 1988, *Approximation of degenerate parabolic problems using numerical integration.* SIAM J. Numer. Anal. **25**, pp. 784-814.

[36] Ortega, J.M. & Rheinboldt, W.C., 1970, *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York.

[37] Raviart, P.-A., 1970, *Sur la résolution de certaines équations paraboliques non-linéaires.* J. Funct. Anal. **5**, pp. 299-328.

[38] Rodrigues, J.-F., 1987, *Obstacle Problems in Mathematical Physics.* North Holland, Amsterdam.

[39] Rulla, J., 1996, *Error analysis for implicit approximations to Cauchy problems.* SIAM J. Numer. Anal. **33**, pp. 68-87.

[40] Sabinina, E.S., 1962, *A class of nonlinear degenerating parabolic equations.* Soviet. Math. Dokl. **143**, pp.495-498.

[41] Schatz, A.H., & Wahlbin, L.B., 1982, *On the quasi–optimality in L^∞ of the $\overset{\circ}{H}{}^1$–projection into finite element spaces.* Math. Comp. **38**, pp.1-22.

[42] Shangbin, C., 1993, *Positive solutions for Dirichlet problems associated to semilinear elliptic equations with singular nonlinearity.* Nonlinear Analysis **21**, pp.181-190.

[43] Stein, E.M., 1970, *Singular Integrals and Differentiability Properties of Functions.* Princeton Univ. Press,

[44] Verdi, C., & Visintin, A., 1988, *Error analysis for a semiexplicit numerical scheme for Stefan-type problems.* Numer. Math. **52**, 1988, pp. 165-185.

[45] Visintin, A., 1985, Stefan problems with phase relaxation. IMA J. Appl. Math. **34**, pp. 225-245.

[46] Visintin, A., 1985, Supercooling and superheating effects in phase transitions. IMA J. Appl. Math. **35**, pp. 233-256.

[47] Watanabe, M., 1988, *Solutions with compact support of the porous medium equation in arbitrary dimensions.* Proc. Amer. Math. Soc. **103**, pp. 149-152.

[48] Zeidler, E., 1990, *Nonlinear Functional Analysis and its Applications II/A: Linear Monotone Operators.* Springer, Berlin.

[49] Zeidler, E., 1990, *Nonlinear Functional Analysis and its Applications II/B: Nonlinear Monotone Operators.* Springer, Berlin.