# Finite element approximation of non-Fickian polymer diffusion

Norbert Bauermeister and Simon Shaw*

www.brunel.ac.uk/bicom
*Brunel Institute of Computational Mathematics*
Brunel University UB8 3PH
England

January 9, 2008 (revised 21 July 2008)

*This paper is dedicated to the memory of John Crank*

**abstract** The problem of nonlinear non-Fickian polymer diffusion as modelled by a diffusion equation with an adjoined spatially local evolution equation for a viscoelastic stress is considered (see, for example, Cohen, White & Witelski, SIAM J. Appl. Math. 55, pp. 348–368, 1995). We present numerical schemes based, spatially, on the Galerkin finite element method and, temporally, on the Crank-Nicolson method. Special attention is paid to linearising the discrete equations by extrapolating the value of the nonlinear term from previous time steps. Optimal *a priori* error estimates are given, based on the assumption that the exact solution possesses certain regularity properties, and numerical experiments are given to support these error estimates.

**keywords** *a priori* error estimates, nonlinear diffusion, non-Fickian diffusion, finite element method

**AMS MSC 2000 subject classifications** 74S05 (FEM), 74S20 (FDM), 76R50, 82D60 (polymers), 45D05 (Volterra equations), 74D10

**acknowledgement** Shaw would like to acknowledge a helpful conversation with Professors William Layton and Béatrice Rivière (both in the *Computational and Applied Mathematics Group, University of Pittsburgh*) in the very early stages of this work. Bauermeister would like to acknowledge the financial support of Brunel University and the EPSRC.

## 1 Introduction

In [23, 22] Thomas & Windle demonstrated by experiment that diffusion of a solvent in a viscoelastic polymer substrate is highly non-Fickian with the solvent developing

*Corresponding author: simon.shaw@brunel.ac.uk

a steep travelling wave front. To model this behavior Cohen *et al.* in [10, 9] (see also the references therein) noted that the solvent causes local stresses which, in polymers, are viscoelastic and therefore described by either a hereditary constitutive law (see [13]), or by an equivalent spatially-local stress-rate equation. As the solvent moves into the substrate its concentration level can rise above a critical value, $u_a$ say, which in turn causes a 'phase change' of the polymer from glass to rubber. The viscoelastic relaxation time ($\gamma^{-1}$ in the equations that follow) in the stress-law is more or less constant in each of these phases, but changes abruptly and significantly across the phase change boundary. This abrupt and significant step change is the basic (nonlinear) mechanism for the development of the steep front. A description of the molecular processes of non-Fickian polymer diffusion, and the requirements for it to occur seems to have been first given in [17].

This article proposes five fully discrete schemes for the partial differential equations that comprise Cohen *et al.*'s model. We employ the Galerkin finite element method for the spatial discretisation, and a version of the Crank-Nicolson method for the temporal discretisation. Special attention is paid to the nonlinear term in that of the five discrete schemes given, three of this family are *linear*. This linearisation is accomplished by extrapolating the value of the nonlinear term from previous time levels, as in [7, 25], with a simpler treatment at the initial time step, as in [18] (see also [28]). The other two schemes give nonlinear discrete equations and, curiously, perform less well. Our *a priori* error estimates cover four of the five methods, although we give a comprehensive set of numerical results that demonstrate all five have errors of optimal order. Further details are given in [5].

The physical model proposed in, for example, [10] consists of the nonlinearly coupled degenerate system,

$$\frac{\partial c}{\partial \tilde{t}} = D\tilde{\nabla}^2 c + K\tilde{\nabla}^2 \tau \qquad \text{and} \qquad \frac{\partial \tau}{\partial \tilde{t}} + \beta(c)\tau = \mu c,$$

where $c$ denotes the solvent's concentration, $\tau$ is a viscoelastic 'stress', $D$, $K$, $\mu$ are positive constants and $\tilde{\nabla}^2 = (\partial^2/\partial\tilde{x}_1^2, \partial^2/\partial\tilde{x}_2^2, \ldots)$.

To keep the presentation below cleaner we prefer to scale out the three constants $D$, $K$ and $\mu$, and for this we define $t = M\tilde{t}$, $\boldsymbol{x} = E\tilde{\boldsymbol{x}}$, $c = u$ and $\sigma = B\tau$ where $B = K/D$, $E = (\mu K)^{1/2}/D$ and $M = \mu K/D$. Lastly, setting $\gamma(u) := D\beta(c)/\mu K$, defining $\mathcal{J} = (0, T]$, and assuming the polymer occupies an open, bounded, connected domain $\Omega \subset \mathbb{R}^d$, we consider the problem in the form: find $u = u(\boldsymbol{x}, t)$ and $\sigma = \sigma(\boldsymbol{x}, t)$ such that,

$$\dot{u} - \nabla^2 u = f + \nabla^2 \sigma \quad \text{in } \Omega \times \mathcal{J}, \tag{1a}$$

$$\dot{\sigma} + \gamma(u)\sigma = u + \hat{f} \quad \text{in } \Omega \times \mathcal{J}, \tag{1b}$$

$$u = 0 \quad \text{on } \Gamma_D \times \mathcal{J}, \tag{1c}$$

$$(\nabla u + \nabla \sigma) \cdot \boldsymbol{n} = g \quad \text{on } \Gamma_N \times \mathcal{J}, \tag{1d}$$

$$u = \breve{u} \quad \text{at } t = 0, \tag{1e}$$

$$\sigma = \breve{\sigma} \quad \text{at } t = 0. \tag{1f}$$

Here: the overdots denote partial time differentiation; we assume the initial data are compatible with the boundary data at $t = 0$; $\partial\Omega = \Gamma_D \cup \Gamma_N$; $\Gamma_D$ is closed and has positive $(d-1)$-dimensional measure; and, $\Gamma_D \cap \Gamma_N = \varnothing$. Furthermore $f = f(\boldsymbol{x}, t)$, $\hat{f} = \hat{f}(\boldsymbol{x}, t)$, $g = g(\boldsymbol{x}, t)$, $\breve{u} = \breve{u}(\boldsymbol{x})$, and $\breve{\sigma} = \breve{\sigma}(\boldsymbol{x})$ are given functions. Note that we have allowed each equation to be 'forced' by $f$ and $\hat{f}$. These are not needed in the model and

are used only to generate artificial exact solutions that we use later to demonstrate the error estimates. The term $\hat{f}$ is in fact purely artificial and so is not included in the estimates that follow.

It is not yet necessary to give a specific form for the nonlinear coupling function $\gamma$ since all we need for the error analysis is contained in the following assumptions. These are realistic in the context of the model we are working from.

**Assumption 1.1** (Properties of $\gamma$)**.** *There are positive constants, $\check{\gamma}$, $\hat{\gamma}$, $C'_\gamma$ and $C''_\gamma$ such that $\gamma : \mathbb{R} \to \mathbb{R}$ satisfies $\gamma \in C^2(\mathbb{R})$ and $\forall u \in \mathbb{R}$:*

$$0 < \check{\gamma} \leqslant \gamma(u) \leqslant \hat{\gamma}, \qquad 0 \leqslant \gamma'(u) \leqslant C'_\gamma \qquad \textit{and} \qquad |\gamma''(u)| \leqslant C''_\gamma.$$

The well-posedness of this problem appears to have been first established by Amann in [1], who then generalised his results a couple of years later in [2]. Since then Hu and Zhang gave an alternative proof of existence and uniqueness in [6] and, most recently, Vorotnikov in [26] has examined the problem from the viewpoint of 'dissipative' solutions (an 'ultra-weak' concept of solution). Each of these results have dealt with various generalisations of (1a) and (1b) with some or all coefficients allowed to depend on $u$. Also, the right hand side of (1b) is sometimes taken in the form $\mu_1 u + \mu_2 \dot{u}$, but we will leave this for another time. Lastly here we note that by replacing $-\nabla^2\sigma$ with $\dot{\sigma}$ in (1a) and $\gamma(u)\sigma - u$ with $k\sigma - k\varphi(u)$ in (1b) we arrive at the porous medium system considered in [4]. This similarity is at best only superficial though since their $\varphi$ is less well behaved than our $\gamma$, and the authors of [4] used deeper techniques of analysis to consider well-posedness rather than just numerical analysis.

For more background on the underlying physics we refer to the original literature on the experiments, [23, 22], and on the development of the mathematical model, [10, 9, 12]. Also, the article [15] gives a comprehensive study of the one-space dimensional problem and illustrates the parameter regions governing the formation of steep wave fronts, as well as whether or not these fronts are mobile.

Note also that solving (1b), assuming $\hat{f} = \breve{\sigma} = 0$, and substituting into (1a) yields an example of a nonlinear parabolic Volterra equation,

$$\dot{u} - \nabla^2 u = f + \nabla^2 \int_0^t \mu e^{-\int_s^t \gamma(u(\xi))\,d\xi}\; u(s)\,ds. \tag{2}$$

Although heat equations with memory, such as the one above, have an extensive numerical analysis literature, we are not aware of any results that can be applied here. For example, linear problems have been studied in [24, 21, 14, 16, 29], with particular attention often being paid to sparse quadrature, weakly singular kernels and non-smooth initial data, and nonlinear problems have been studied in [11, 19, 28, 7]. These results are all for problems written in divergence form where $u_t = \nabla \cdot F(t, u, \nabla u)$ for some $F$, or for problems where the memory term contains only zero or first order spatial derivatives of $u$.

Since (2) is not in divergence form we find it convenient to use the 'inverse Laplacian' (see later in (6) and (19)) and then the key technique in the error analysis is to examine the errors in the quantities $\sigma$ and $u + \sigma$, rather than in $u$ and $\sigma$ directly.

The plan of this paper is as follows. Section 2 contains the weak formulation of the problem, recalls some standard notation and sets the scene for the '$u + \sigma$' error estimates that follow later. The numerical schemes are described in Section 3 and some stability estimates for the discrete solution are given.

The error analysis begins in Section 4 and, in an effort to make it easier to digest, this is broken down into several subsections.

Subsection 4.1 gives some basic observations on the nonlinear function, $\gamma$. Subsection 4.2 outlines familiar tools such as the elliptic projection and some estimates from Taylor's series. The error in the 'linear' part of the equations is dealt with in Subsection 4.3 and then Subsection 4.4 focusses specifically on the errors generated by the nonlinearity. Eventually, Subsection 4.5 synthesises these lemmas into the *a priori* error estimate contained in Theorem 4.12.

We conclude with some numerical experiments in Section 5 and some brief concluding remarks in Section 6.

## 2 Weak formulation and preliminaries

Our notation is standard. We use $W_p^m(X;Y)$ to denote the Sobolev (Banach) space of functions, $v\colon X \to Y$, which together with their first $m$ weak derivatives belong to $L_p(X;Y)$ ($1 \leqslant p \leqslant \infty$). The target space, $Y$, is omitted when $Y = \mathbb{R}$. When $p = 2$ we obtain the Hilbert space $H^m(\Omega) = W_2^m(\Omega)$ and write $\|\cdot\|_{H^m(\Omega)}$ for the norm in $H^m(\Omega)$. Also, because it is used so frequently below, we use the abbrevation $\|\cdot\|_0 := \|\cdot\|_{L_2(\Omega)}$.

If $X$ is a Banach space then notations such as $L_p(0,t;X)$ denote the Banach space of $L_p$-maps from $(0,t)$ into $X$. The norm in $L_p(0,t;X)$ is merely the $L_p(0,t)$ norm of $\|\cdot\|_X$. This is standard, as also is our use of $C$ to denote a generic positive constant that may have different values in different places, and is often dependent on $T$ through Gronwall estimates. We also recall Young's inequality: for $a, b \in \mathbb{R}$ and any $\epsilon > 0$,

$$2ab \leqslant \epsilon\, a^2 + \frac{1}{\epsilon}\, b^2. \tag{3}$$

Noting the essential boundary condition (1c) we define the test space $V$ for the variational formulation of (1a) and (1b) as

$$V := \{v \in H^1(\Omega) \ :\ v = 0 \text{ on } \Gamma_D\}.$$

Furthermore, the norm $\|\cdot\|_V := \sqrt{(\nabla\cdot, \nabla\cdot)}$ is equivalent (on $V$) to $\|\cdot\|_{H^1(\Omega)}$, and we note for use later that there is a constant $C_V > 0$ such that,

$$\|v\|_0 \leqslant C_V \|v\|_V \quad \forall v \in V \qquad \text{and} \qquad \|v\|_{V'} \leqslant C_V \|v\|_0 \quad \forall v \in L_2(\Omega). \tag{4}$$

Now, recall that in (1b) the term $\hat{f}$ is purely artificial in terms of the non-Fickian physics, and is introduced only to aid in the construction of test problems later on. Therefore, if we assume throughout that $\hat{f}|_{\Gamma_D} = \breve{\sigma}|_{\Gamma_D} = 0$ then, by (1c), it follows that $\sigma(t)|_{\Gamma_D} = 0$ a.e. in $\mathcal{J}$.

In conclusion, we write the problem, (1a) and (1b) with boundary data, in weak form as: find $(u,\sigma)\colon \mathcal{J} \to V \times V$ such that,

$$(\dot{u}(t), v) + (\nabla u(t), \nabla v) + (\nabla\sigma(t), \nabla v) = \langle L(t), v\rangle \quad \forall v \in V, \tag{5a}$$

$$(\dot{\sigma}(t), w) + (\gamma(u)\sigma(t), w) = (u(t), w) \quad \forall w \in V, \tag{5b}$$

with (1e) and (1f), with $\breve{\sigma}|_{\Gamma_D} = 0$ and where $L\colon \mathcal{J} \to V'$ is defined by,

$$\langle L(t), v\rangle := (f(t), v) + (g(t), v)_{\Gamma_N}.$$

One of the difficulties in dealing with these equations, either to prove stability, discrete stability or an error estimate, lies in combining the terms $(\nabla\sigma, \nabla v)$ in (5a) and $(u, w)$ in (5b) (and their discrete counterparts) in a way that yields useful estimates.

In [18] the simplification of replacing $\nabla(\gamma(u)\sigma)$ with a term like $\gamma(u)\nabla\sigma$ was made, and this led to *a priori* estimates. However this simplification does not represent the non-Fickian problem which we deal with here.

Firstly, recall from [3] (for example) the 'inverse Laplacian' $\mathcal{G}\colon V' \to V$ defined by,

$$(\nabla\mathcal{G}w, \nabla v) = \langle w, v\rangle \qquad \forall v \in V \tag{6}$$

and for any $w \in V'$. (This is no more than the Riesz representation theorem and so we can immediately note that $\|w\|_{V'} = \|\mathcal{G}w\|_V$). Then, choosing $v = \mathcal{G}u$ in (5a) and $w = \sigma$ in (5b), and adding the equations we see that $(\nabla\sigma, \nabla\mathcal{G}u) - (u, \sigma) = 0$, and so these problematic terms vanish. This can be used to derive stability estimates (see below in Prop. 2.1) but the norm on $u$ is apparently too weak for deriving error estimates.

Secondly, by adding (5a) to (5b) and working with the sum $u + \sigma$ instead of only $u$ we get,

$$(\dot{u} + \dot{\sigma}, v) + (\nabla(u + \sigma), \nabla v) + (\gamma(u)\sigma, v) = (u, v) + \langle L, v\rangle \quad \forall v \in V. \tag{7}$$

It appears that for error estimation the pair $u + \sigma$ and $\sigma$ are easier to work with than the pair $u$ and $\sigma$. Moreover, by the triangle inequality it is equivalent to have bounds for either $\|u\|$ and $\|\sigma\|$ or $\|u + \sigma\|$ and $\|\sigma\|$. This technique also provides a stability estimate (Prop. 2.2) but in stronger norms.

**Proposition 2.1** (Stability in weak norms)**.**

$$\|u(t)\|_{V'}^2 + \|\sigma(t)\|_0^2 + \|u\|_{L_2(0,t;L_2(\Omega))}^2 + 2\check{\gamma}\|\sigma\|_{L_2(0,t;L_2(\Omega))}^2 \leqslant \|\breve{u}\|_{V'}^2 + \|\breve{\sigma}\|_0^2 + C_V^2\|L\|_{L_2(0,t;V')}^2.$$

*Proof.* Choose $v = 2\mathcal{G}u(t)$ in (5a) and $w = 2\sigma(t)$ in (5b) and add. Estimate the duality product; invoke $C_V$ from the right-hand part of (4); integrate over $(0, t)$; kickback $\|u\|_0^2$ using Young's inequality; use the initial data and the lower bound on $\gamma(u)$; and, finally, remove $\mathcal{G}$ by replacing $\|\mathcal{G}\cdot\|_V$ with $\|\cdot\|_{V'}$. □

**Proposition 2.2** (Stability in stronger norms)**.**

$$\begin{aligned}\|u(t)\|_0^2 + \|\sigma(t)\|_0^2 + \|u\|_{L_2(0,t;L_2(\Omega))}^2 + \|\sigma\|_{L_2(0,t;L_2(\Omega))}^2 + \|u + \sigma\|_{L_2(0,t;V)}^2 \\ \leqslant C\left(\|\breve{u}\|_0^2 + \|\breve{\sigma}\|_0^2 + \|L\|_{L_2(0,t;V')}^2\right).\end{aligned} \tag{8}$$

*Proof.* Add (5a) and (5b) and choose $v = w = 2u + 2\sigma$. Then use (1e) and (1f), the Cauchy-Schwarz inequality, several Young's inequalities and (4), and we arrive at,

$$\begin{aligned}\|u + \sigma\|_0^2 + \int_0^t \|u + \sigma\|_V^2\, ds + \check{\gamma}\int_0^t \|\sigma\|_0^2\, ds \\ \leqslant \|\breve{u} + \breve{\sigma}\|_0^2 + \left(\frac{\hat{\gamma}^2}{\check{\gamma}} + 2C_V^2\right)\int_0^t \|u\|_0^2\, ds + 2\int_0^t \|L\|_{V'}^2\, ds.\end{aligned} \tag{9}$$

To get a second inequality with which we can handle the term $\int_0^t \|u\|_0^2\, ds$ on the right-hand side, we choose $v = 2\sigma$ in (5b) and follow the same pattern of estimation. This yields,

$$\|\sigma\|_0^2 + \check{\gamma}\int_0^t \|\sigma\|_0^2\, ds \leqslant \|\breve{\sigma}\|_0^2 + \frac{1}{\check{\gamma}}\int_0^t \|u(s)\|_0^2\, ds. \tag{10}$$

Now, by adding (9) and (10) and then using the triangle inequality, the proof is completed by tidying up the constants, adding $\|u\|^2_{L_2(0,t;L_2(\Omega))}$ to both sides and using Gronwall's lemma. □

The main point to take from these results is that while the inverse Laplacian removes the 'cross terms' involving $u$ and $\sigma$, it is the use of $u+\sigma$ which provides norms which are strong enough to make all the steps of the proof possible. This observation carries over to the error estimates that follow, and allows us to remove a bound on the time step size for the linearized methods.

# 3 The numerical scheme

The spatial discretisation is a standard Galerkin finite element method using piecewise polynomials of degree $r \geqslant 1$. We assume for simplicity that $\Omega$ is polygonal (in 2D) or polyhedral (in 3D) and that it can be discretised into a quasi-uniform family of simplicial subdivisions $\mathcal{E}_h$, depending on a mesh size parameter $h \in (0, \hat{h}]$ for some $\hat{h} > 0$. In the usual way the finite element space is then defined as,

$$V^h := \{\, v \in V \cap C^0(\bar{\Omega}) \;\mid\; v \in \mathbb{P}_r(E) \;\; \forall\, E \in \mathcal{E}_h \,\}.$$

We denote the Lagrange basis functions by $\phi_j$, $j = 1, \ldots, N_\phi$ and the Lagrange nodes by $\boldsymbol{x}_j$, $j = 1, \ldots, N_\phi$. We assume there is an interpolation operator $I^h : C(\Omega) \to V^h$ (i.e. $I^h v = \sum v(\boldsymbol{x}_j)\phi_j$) such that,

$$\|v - I^h v\|_{H^m(\Omega)} \leqslant C h^{r+1-m} \|v\|_{H^{r+1}(\Omega)} \qquad \text{for } m = 0, \ldots, r. \tag{11}$$

For the time discretisation we divide the interval $[0, T]$ into $N$ subintervals with equidistant endpoints $t_n$ such that $0 = t_0 < t_1 < \cdots < t_N = T$. We define the constant time step by $k = T/N$.

To simplify notation, we set $v_n := v(t_n)$ and define,

$$\partial_t v_n := (v_n - v_{n-1})/k, \qquad \partial_t^2 v_n := (v_n - 2v_{n-1} + v_{n-2})/k^2, \qquad \bar{v}_n := (v_n + v_{n-1})/2$$

$$\text{and} \qquad \Delta_n v := \bar{\dot{v}}_n - \partial_t v_n = \frac{\dot{v}(t_n) + \dot{v}(t_{n-1})}{2} - \frac{v(t_n) - v(t_{n-1})}{k}.$$

Observe that $\partial_t v_n$ is an approximation for the derivative $\dot{v}$ at $t = (t_n + t_{n-1})/2$, and $\partial_t^2 v_n$ is an approximation for the second time derivative $\ddot{v}$ at $t = t_{n-1}$. Since we will need it frequently later, we note that,

$$\sum_{n=1}^{m} \|\bar{v}_n\|^2 \leqslant \frac{1}{2}\|v_0\|^2 + \sum_{n=1}^{m-1} \|v_n\|^2 + \frac{1}{2}\|v_m\|^2 \leqslant \sum_{n=0}^{m} \|v_n\|^2. \tag{12}$$

Next we need a few easy consequences of Taylor's theorem for use later in the error analysis (see e.g. [18]).

**Lemma 3.1** (Taylor estimates)**.** *Let $X$ be a Banach space. If $v$ has the indicated regularity, then*

$$\|\partial_t v_n\|_X \leqslant \|\dot{v}\|_{L_\infty(t_{n-1},t_n;X)}, \tag{13}$$
$$\|\partial_t v_n\|_0 \leqslant k^{-1/2}\|\dot{v}\|_{L_2(t_{n-1},t_n;L_2(\Omega))}, \tag{14}$$
$$\|\partial_t^2 v_n\|_X \leqslant \|\ddot{v}\|_{L_\infty(t_{n-2},t_n;X)}, \tag{15}$$
$$\|\partial_t^2 v_n\|_0 \leqslant k^{-1/2}\|\ddot{v}\|_{L_2(t_{n-2},t_n;L_2(\Omega))}, \tag{16}$$
$$\|\Delta_n v\|_0 \leqslant Ck^{3/2}\|\dddot{v}\|_{L_2(t_{n-1},t_n;L_2(\Omega))}. \tag{17}$$

Denoting the discrete approximations to $u$ and $\sigma$ by $u^h$ and $\sigma^h$ the fully discrete approximation to the weak form (5a) and (5b) is: for $n = 0, \dots, N$ find $(u_n^h, \sigma_n^h) \in V^h \times V^h$ such that

$$(\partial_t u_n^h, v) + (\nabla \bar{u}_n^h, \nabla v) + (\nabla \bar{\sigma}_n^h, \nabla v) = \langle \bar{L}_n, v\rangle \qquad \forall\, v \in V^h,\ n = 1, \dots, N \tag{18a}$$
$$(\partial_t \sigma_n^h, v) + (\mathscr{B}_n^Q(u^h, \sigma^h), v) = (\bar{u}_n^h, v) \qquad \forall\, v \in V^h,\ n = 1, \dots, N \tag{18b}$$
$$(u_0^h, v) = (\breve{u}, v) \qquad \forall\, v \in V^h, \tag{18c}$$
$$(\sigma_0^h, v) = (\breve{\sigma}, v) \qquad \forall\, v \in V^h. \tag{18d}$$

Here $\mathscr{B}_n^Q(u^h, \sigma^h)$ is the discrete approximation of the term $\gamma(u)\sigma$ at time $t_{n-1/2}$. To deal with this term we consider five possibilities, indexed by $Q \in \{1, 2, 3, 4, 5\}$. The first three are based on linearly extrapolating $u_n^h$ or $\bar{u}_n^h$ in order to approximate $\gamma(u_n^h)$ or $\gamma(\bar{u}_n^h)$ from the previous two time levels, and these lead to *linear* numerical schemes. The last two, on the other hand, are the *nonlinear* numerical schemes that result from using $\gamma(u_n^h)$ and $\gamma(\bar{u}_n^h)$ directly.

We'll discuss these methods a little more below, but first we give the details. Define $\mathscr{B}_n^Q$ via

$$\begin{aligned}
\mathscr{B}_n^1(u^h, \sigma^h) &:= \tfrac{1}{2}\gamma(\mathscr{E}_n^1 u^h)\sigma_n^h + \tfrac{1}{2}\gamma(u_{n-1}^h)\sigma_{n-1}^h \\
\mathscr{B}_n^2(u^h, \sigma^h) &:= \gamma(\mathscr{E}_n^2 u^h)\, \bar{\sigma}_n^h \\
\mathscr{B}_n^3(u^h, \sigma^h) &:= \tfrac{1}{2}\sum_{j=1}^{N_\phi} \gamma(\mathscr{E}_n^1 u^h(\boldsymbol{x}_j))\sigma_n^h(\boldsymbol{x}_j)\phi_j + \tfrac{1}{2}\sum_{j=1}^{N_\phi} \gamma(u_{n-1}^h(\boldsymbol{x}_j))\sigma_{n-1}^h(\boldsymbol{x}_j)\phi_j \\
\mathscr{B}_n^4(u^h, \sigma^h) &:= \tfrac{1}{2}\gamma(u_n^h)\sigma_n^h + \tfrac{1}{2}\gamma(u_{n-1}^h)\sigma_{n-1}^h \\
\mathscr{B}_n^5(u^h, \sigma^h) &:= \gamma(\bar{u}_n^h)\, \bar{\sigma}_n^h
\end{aligned}$$

where the $\phi_j$'s are Lagrange basis functions and $\boldsymbol{x}_j$ their nodes, i.e. $\phi_i(\boldsymbol{x}_j) = \delta_{ij}$. Here, for methods $Q = 2$ and $5$, we define

$$\mathscr{E}_n^5 u^h := \bar{u}_n^h \qquad \text{and} \qquad \mathscr{E}_n^2 u^h := \begin{cases} \frac{3}{2}u_{n-1}^h - \frac{1}{2}u_{n-2}^h & \text{for } n \geqslant 2, \\ u_{n-1}^h & \text{for } n = 1, \end{cases}$$

and so for $Q \in \{2, 5\}$ we have $\mathscr{B}_n^Q = \gamma(\mathscr{E}_n^Q u^h)\bar{\sigma}_n^h$. Similarly, for methods $Q = 1$ and $4$ we define,

$$\mathscr{E}_n^4 u^h := u_n^h \qquad \text{and} \qquad \mathscr{E}_n^1 u^h := \begin{cases} 2u_{n-1}^h - u_{n-2}^h & \text{for } n \geqslant 2, \\ u_{n-1}^h & \text{for } n = 1, \end{cases}$$

giving, for $Q \in \{1, 4\}$, that $\mathscr{B}_n^Q = \frac{1}{2}\gamma(\mathscr{E}_n^Q u^h)\sigma_n^h + \frac{1}{2}\gamma(u_{n-1}^h)\sigma_{n-1}^h$. We can see that $Q = 1$ is a linearised version of $Q = 4$ and $Q = 2$ is a linearised version of $Q = 5$.

We note that a linear extrapolation of $u_1^h$ is not possible and so at the initial time step we extrapolate as a constant from the initial condition. Optimal *a priori* error estimates are given later for the schemes $Q = 1, 2, 4, 5$. For method $Q = 3$ we have no results at the moment, but we note that it is similar to the so-called *product approximation* described in [8][1].

The discrete inverse Laplacian $\mathcal{G}^h$ is defined for any $w \in V'$ via,

$$(\nabla \mathcal{G}^h w, \nabla v) = \langle w, v \rangle \quad \forall v \in V^h. \tag{19}$$

It is clear that $\mathcal{G}^h$ is linear. The next two lemmas are equally clear.

**Lemma 3.2.** $\|\mathcal{G}^h w\|_V \leqslant \|w\|_{V'} \quad \forall w \in V'$.

**Lemma 3.3.** $2k(\partial_t v_n, \mathcal{G}^h \bar{v}_n) = \|\mathcal{G}^h v_n\|_V^2 - \|\mathcal{G}^h v_{n-1}\|_V^2 \quad \forall v_n, v_{n-1} \in V^h$.

The discrete schemes corresponding to $Q = 1$ and $Q = 2$ possess unique solutions as shown by the following lemma. The situation for $Q = 3$ is, as yet, unclear, while for $Q = 4$ and $Q = 5$ we can show, see [5], that non-uniqueness is possible (at least for some choices of time step).

**Proposition 3.4.** *The schemes $Q = 1$ and $Q = 2$ have unique solutions.*

*Proof.* At each timestep $t_n$ the discrete solution $(u_n^h, \sigma_n^h)$ is determined by a linear system of equations in a finite-dimensional space. Assume that there are (at least) two solutions $(u_n^h, \sigma_n^h)$ and $(\tilde{u}_n^h, \tilde{\sigma}_n^h)$ at a certain timestep $t_n$, for $n > 0$, with all previous solutions being uniquely defined. Define $z := u_n^h - \tilde{u}_n^h$ and $p := \sigma_n^h - \tilde{\sigma}_n^h$. Subtracting (18a) for the tilde-solution from the same equation for the non-tilde solution and repeating this for (18b) gives

$$\tfrac{1}{k}(z, v) + \tfrac{1}{2}(\nabla z, \nabla v) + \tfrac{1}{2}(\nabla p, \nabla v) = 0 \qquad \text{and} \qquad \tfrac{1}{k}(p, w) + (\mathscr{B}_n^Q - \tilde{\mathscr{B}}_n^Q, w) = \tfrac{1}{2}(z, w).$$

Choosing $v = \mathcal{G}^h z$, $w = p$ and adding the equations yields

$$\tfrac{1}{k}\|\mathcal{G}^h z\|_V^2 + \tfrac{1}{k}\|p\|_0^2 + \tfrac{1}{2}\|z\|_0^2 + (\mathscr{B}_n^Q - \tilde{\mathscr{B}}_n^Q, p) = 0.$$

Observe that $\mathscr{B}_n^Q - \tilde{\mathscr{B}}_n^Q$ is $\frac{1}{2}\gamma(\mathscr{E}_n^1 u^h)\, p$ if $Q = 1$ and is $\frac{1}{2}\gamma(\mathscr{E}_n^2 u^h)\, p$ if $Q = 2$ and so, in both cases, $(\mathscr{B}_n^Q - \tilde{\mathscr{B}}_n^Q, p) \geqslant \frac{\check{\gamma}}{2}\|p\|_0^2$. Hence $z = 0$ and $p = 0$ and thus the two solutions are identical. In fact, by subtracting the system for two solutions, we have obtained the homogeneous linear system and have shown that this homogeneous system has only the trivial solution $z = p = 0$. The system matrix is therefore invertible and there is exactly one solution. □

The next step is to establish some stability estimates for the discrete solutions. First we derive upper and lower bounds on the nonlinear term (method $Q = 3$ is not treated here). First, the upper bound (the proof of which is straightforward).

**Lemma 3.5** (Upper bound for $\mathscr{B}_n^Q$)**.** *For methods $Q = 1, 2, 4, 5$ we have*

$$\|\mathscr{B}_n^Q(u^h, \sigma^h)\|_0 \leqslant \frac{\hat{\gamma}}{2}\left(\|\sigma_n^h\|_0 + \|\sigma_{n-1}^h\|_0\right).$$

And, secondly, the lower bound.

[1]The authors are grateful to Andrew Wathen (COMLAB, Oxford University, UK) for pointing this out.

**Lemma 3.6** (Lower bounds for $\mathscr{B}_n^Q$)**.** *For methods $Q \in \{2,5\}$, we have*

$$(\mathscr{B}_n^Q, \bar{\sigma}_n^h) \geqslant \check{\gamma}\|\bar{\sigma}_n^h\|_0^2. \tag{20}$$

*For methods $Q \in \{1,4\}$, and for any $\epsilon > 0$, there is a constant $C_\epsilon$ such that*

$$(\mathscr{B}_n^Q, \bar{\sigma}_n^h) \geqslant \check{\gamma}\|\bar{\sigma}_n^h\|_0^2 - \epsilon\|\sigma_n^h\|_0^2 - C_\epsilon\|\sigma_{n-1}^h\|_0^2. \tag{21}$$

*In fact, $C_\epsilon = \frac{\hat{\gamma}}{2} + \frac{\hat{\gamma}^2}{16\epsilon}$.*

*Proof.* For $Q \in \{2,5\}$ the assertion follows immediately from the definition of $\mathscr{B}_n^Q$. Thus, for the remainder, let $Q \in \{1,4\}$. For both these methods, $\mathscr{B}_n^Q$ has the following form:

$$\mathscr{B}_n^Q = \frac{1}{2}\overset{\dashv}{\gamma}_n\sigma_n^h + \frac{1}{2}\overset{\vdash}{\gamma}_n\sigma_{n-1}^h.$$

with certain functions $\overset{\dashv}{\gamma}_n$ and $\overset{\vdash}{\gamma}_n$, depending on $u^h$, such that $\check{\gamma} \leqslant \overset{\dashv}{\gamma}_n(\boldsymbol{x}) \leqslant \hat{\gamma}$ and $\check{\gamma} \leqslant \overset{\vdash}{\gamma}_n(\boldsymbol{x}) \leqslant \hat{\gamma}$. For later use, we define $G := \overset{\dashv}{\gamma}_n - \overset{\vdash}{\gamma}_n$ and observe that $\|G\|_{L_\infty(\Omega)} \leqslant 2\hat{\gamma}$. We can rearrange

$$\mathscr{B}_n^Q = \frac{1}{2}\overset{\dashv}{\gamma}_n\sigma_n^h + \frac{1}{2}\overset{\vdash}{\gamma}_n\sigma_{n-1}^h = \frac{1}{2}\overset{\dashv}{\gamma}_n(\sigma_n^h + \sigma_{n-1}^h) - \frac{1}{2}(\overset{\dashv}{\gamma}_n - \overset{\vdash}{\gamma}_n)\sigma_{n-1}^h.$$

Thus $(\mathscr{B}_n^Q, \bar{\sigma}_n^h) = (\overset{\dashv}{\gamma}_n\bar{\sigma}_n^h, \bar{\sigma}_n^h) - \frac{1}{2}\big((\overset{\dashv}{\gamma}_n - \overset{\vdash}{\gamma}_n)\sigma_{n-1}^h, \bar{\sigma}_n^h\big) \geqslant \check{\gamma}\|\bar{\sigma}_n^h\|_0^2 - \frac{1}{2}|\big((\overset{\dashv}{\gamma}_n - \overset{\vdash}{\gamma}_n)\sigma_{n-1}^h, \bar{\sigma}_n^h\big)|$ and using Young's inequality we get for any $\epsilon_1 > 0$ that,

$$\frac{1}{2}|\big((\overset{\dashv}{\gamma}_n - \overset{\vdash}{\gamma}_n)\sigma_{n-1}^h, \bar{\sigma}_n^h\big)| \leqslant \frac{\hat{\gamma}}{2}(\|\sigma_{n-1}^h\|_0\,\|\sigma_n^h\|_0 + \|\sigma_{n-1}^h\|_0^2) \leqslant \frac{\hat{\gamma}}{2}\left(\frac{\epsilon_1}{2}\|\sigma_n^h\|_0^2 + (\frac{1}{2\epsilon_1} + 1)\|\sigma_{n-1}^h\|_0^2\right).$$

Choosing $\epsilon_1 = \frac{4}{\hat{\gamma}}\epsilon$ finishes the proof. ∎

**Proposition 3.7** (Discrete stability: inverse Laplacian)**.** *For each of the methods given by $Q \in \{1,2,4,5\}$ we have, for any $m \in \{1,\dots,N\}$,*

$$\|\mathcal{G}^h u_m^h\|_V^2 + \|\sigma_m^h\|_0^2 + k\sum_{n=1}^m\|\bar{u}_n^h\|_0^2 + k\sum_{n=1}^m\|\bar{\sigma}_n^h\|_0^2 \leqslant C\left(\|\breve{u}\|_{V'}^2 + \|\breve{\sigma}\|_0^2 + \|L\|_{L_\infty(0,T;V')}^2\right).$$

*Proof.* Choosing $v = \mathcal{G}^h\bar{u}_n^h$ in (18a), $v = \bar{\sigma}_n^h$ in (18b), adding the results and using Lemma 3.3, multiplying by $2k$ and summing for $n = 1,\dots,m$ we have,

$$\|\mathcal{G}^h u_m^h\|_V^2 + \|\sigma_m^h\|_0^2 + 2k\sum_{n=1}^m\|\bar{u}_n^h\|_0^2 + 2k\sum_{n=1}^m(\mathscr{B}_n^Q, \bar{\sigma}_n^h) = \|\mathcal{G}^h u_0^h\|_V^2 + \|\sigma_0^h\|_0^2 + 2k\sum_{n=1}^m\langle\bar{L}_n, \mathcal{G}^h\bar{u}_n^h\rangle. \tag{22}$$

Now, $\|\sigma_0^h\|_0 \leqslant \|\breve{\sigma}\|_0$, $\|\mathcal{G}^h u_0^h\|_V \leqslant \|\breve{u}\|_{V'}$ and $2\langle\bar{L}_n, \mathcal{G}^h\bar{u}_n^h\rangle \leqslant \epsilon^{-1}\|L\|_{L_\infty(t_{n-1},t_n;V')}^2 + \epsilon\, C_V^2\,\|\bar{u}_n^h\|_0^2$. Choosing $\epsilon := 1/C_V^2$ to obtain,

$$2k\sum_{n=1}^m\langle\bar{L}_n, \mathcal{G}^h\bar{u}_n^h\rangle \leqslant C_V^2\,T\,\|L\|_{L_\infty(0,t_m;V')}^2 + k\sum_{n=1}^m\|\bar{u}_n^h\|_0^2,$$

we can then insert these three bounds in to (22) and arrive at,

$$\|\mathcal{G}^h u_m^h\|_V^2 + \|\sigma_m^h\|_0^2 + k\sum_{n=1}^m\|\bar{u}_n^h\|_0^2 + 2k\sum_{n=1}^m(\mathscr{B}_n^Q, \bar{\sigma}_n^h) \leqslant \|\breve{u}\|_{V'}^2 + \|\breve{\sigma}\|_0^2 + C_V^2\,T\,\|L\|_{L_\infty(0,t_m;V')}^2. \tag{23}$$

For methods $Q = 2$ or $5$ the proof is concluded by using (20).

Now, for methods $Q = 1$ or $4$, we note that,

$$2k\sum_{n=1}^{m}(\epsilon\|\sigma_n^h\|_0^2 + C_\epsilon\|\sigma_{n-1}^h\|_0^2) \leqslant 2T\epsilon\|\sigma_m^h\|_0^2 + 2(\epsilon + C_\epsilon)\,k\sum_{n=1}^{m-1}\|\sigma_n^h\|_0^2 + 2TC_\epsilon\|\breve{\sigma}\|_0^2,$$

and then choosing $\epsilon = 1/(4T)$ along with using (21) in (23) we obtain,

$$\begin{aligned}\|\mathcal{G}^h u_m^h\|_V^2 + \frac{1}{2}\|\sigma_m^h\|_0^2 + k\sum_{n=1}^{m}\|\bar{u}_n^h\|_0^2 + 2\check{\gamma}k\sum_{n=1}^{m}\|\bar{\sigma}_n^h\|_0^2 \\ \leqslant \|\breve{u}\|_{V'}^2 + c_1\|\breve{\sigma}\|_0^2 + C_V^2\,T\,\|L\|_{L_\infty(0,t_m;V')}^2 + c_2\,k\sum_{n=1}^{m-1}\|\sigma_n^h\|_0^2,\end{aligned}$$

where $c_1 = 1 + 2TC_\epsilon = 1 + \hat{\gamma}T + \hat{\gamma}^2T^2/2$ and $c_2 = 2\epsilon + 2C_\epsilon = 1/(2T) + \hat{\gamma} + \hat{\gamma}^2T/2$. The proof is then concluded for $Q \in \{1, 4\}$ by using a discrete Gronwall lemma. □

**Proposition 3.8** (Boundedness of discrete solutions using $u + \sigma$-terms)**.** *For methods $Q = 1, 2, 4, 5$ we have for $m \in \{1, \ldots, N\}$,*

$$\|u_m^h\|_0^2 + \|\sigma_m^h\|_0^2 + k\sum_{n=1}^{m}\|\bar{u}_n^h + \bar{\sigma}_n^h\|_V^2 \leqslant C\left(\|\breve{u}\|_0^2 + \|\breve{\sigma}\|_0^2 + \|L\|_{L_\infty(0,T;V')}^2\right). \tag{24}$$

*Proof.* Adding (18a) and (18b), choosing $v = 2k(\bar{u}_n^h + \bar{\sigma}_n^h)$ and then applying the Cauchy-Schwarz inequality and Young's inequality yields,

$$\|u_n^h + \sigma_n^h\|_0^2 - \|u_{n-1}^h + \sigma_{n-1}^h\|_0^2 + k\|\bar{u}_n^h + \bar{\sigma}_n^h\|_V^2 \leqslant 3k\left(\|\mathscr{B}_n^Q\|_{V'}^2 + \|\bar{L}_n\|_{V'}^2 + \|\bar{u}_n^h\|_{V'}^2\right). \tag{25}$$

Also, choosing $v = 2k\bar{\sigma}_n^h$ in (18b) and estimating similarly gives,

$$\|\sigma_n^h\|_0^2 - \|\sigma_{n-1}^h\|_0^2 \leqslant k\|\mathscr{B}_n^Q\|_0^2 + 2k\|\bar{\sigma}_n^h\|_0^2 + k\|\bar{u}_n^h\|_0^2. \tag{26}$$

Now add 2 of (25) to 3 of (26), use (4), sum for $n = 1, \ldots, m$, use the triangle inequality first to see that $2\|u_m^h + \sigma_m^h\|_0^2 + 3\|\sigma_m^h\|_0^2 \geqslant \|u_m^h\|_0^2 + \|\sigma_m^h\|_0^2$ and then again to split the term $\|u_0^h + \sigma_0^h\|_0^2$, use the bound on $\mathscr{B}_n^Q$ from Lemma 3.5 and we obtain,

$$\begin{aligned}\|u_m^h\|_0^2 + \|\sigma_m^h\|_0^2 + 2k\sum_{n=1}^{m}\|\bar{u}_n^h + \bar{\sigma}_n^h\|_V^2 \\ \leqslant 4\|u_0^h\|_0^2 + 7\|\sigma_0^h\|_0^2 + 6T\|L\|_{L_\infty(0,T;V')}^2 + (6C_V^2 + 3)\hat{\gamma}^2T/2\|\sigma_0^h\|_0^2 \\ + (6C_V^2 + 3)\hat{\gamma}^2k\sum_{n=1}^{m}\|\sigma_n^h\|_0^2 + (6C_V^2 + 3)k\sum_{n=1}^{m}\|\bar{u}_n^h\|_0^2 + 6k\sum_{n=1}^{m}\|\bar{\sigma}_n^h\|_0^2.\end{aligned} \tag{27}$$

Also note that from (18c) and (18d), $\|u_0^h\|_0 \leqslant \|\breve{u}\|_0$ and $\|\sigma_0^h\|_0 \leqslant \|\breve{\sigma}\|_0$. Using these in (27) along with Proposition 3.7 gives,

$$k\sum_{n=1}^{m}\|\sigma_n^h\|_0^2 + k\sum_{n=1}^{m}\|\bar{u}_n^h\|_0^2 + k\sum_{n=1}^{m}\|\bar{\sigma}_n^h\|_0^2 \leqslant C\left(\|\breve{u}\|_{V'}^2 + \|\breve{\sigma}\|_0^2 + \|L\|_{L_\infty(0,T;V')}^2\right),$$

and using (4) then completes the proof. □

As a straightforward corollary we note that the sum of $\|u_n^h\|_0^2$ is also bounded.

**Corollary 3.9** (Boundedness of discrete solutions)**.** *For methods $Q = 1, 2, 4, 5$, and for any $m \in \{1, \dots, N\}$,*

$$\begin{aligned} &\|u_m^h\|_0^2 + \|\sigma_m^h\|_0^2 + k\sum_{n=1}^m \|u_n^h\|_0^2 + k\sum_{n=1}^m \|\sigma_n^h\|_0^2 + k\sum_{n=1}^m \|\bar{u}_n^h + \bar{\sigma}_n^h\|_V^2 \\ &\quad \leqslant C\left(\|\breve{u}\|_0^2 + \|\breve{\sigma}\|_0^2 + \|L\|_{L_\infty(0,T;V')}^2\right). \end{aligned} \tag{28}$$

These stability estimates show that while $u + \sigma$ is controlled in $V$ by the data each of these is only controlled individually in $L_2(\Omega)$. We now move on to give the error analysis.

# 4 Error estimate

In this section we derive *a priori* error bounds for methods $Q \in \{1, 2, 4, 5\}$. The way the error estimate is proved is based on the error analysis for the simplified model in [18]. Here, however, it seems necessary to combine the approaches of calculating with $u+\sigma$ and of using the inverse Laplacian to achieve similar estimates to those in [18].

This section contains many technical lemmas. In an effort to make the material easier to digest we have broken it down into a sequence of subsections.

In Subsection 4.1 we outline some basic properties and inequalities relating to the nonlinear function $\gamma$. This is followed, in Subsection 4.2, with bounds on some terms involving an elliptic projection. These bounds will be needed throughout this section. In Subsection 4.3 two inequalities are proven that deal just with the linear parts of the system; these are general results that still apply to all five methods $Q \in \{1, 2, 3, 4, 5\}$. Then, in Subsection 4.4, we restrict ourselves to the four methods $Q \in \{1, 2, 4, 5\}$ and give bounds for the nonlinearity error: the difference between the nonlinearity $\gamma(u)\sigma$ and its approximation $\mathscr{B}_n^Q$.

Finally, in Subsection 4.5 the error estimate is stated and proved by combining the error inequalities from Section 4.3 with the nonlinearity errors from Section 4.4.

## 4.1 Basic properties of $\gamma$

In this subsection two small lemmas involving differences of function values of $\gamma$ are given. These will be needed later for proving the error estimate. First, though, we get from Assumption 1.1 the following.

**Lemma 4.1.** *If $1 \leqslant p \leqslant \infty$ and $v, w \in L_p(\Omega)$ then $\|\gamma(v) - \gamma(w)\|_{L_p(\Omega)} \leqslant C_\gamma' \|v - w\|_{L_p(\Omega)}$.*

Before we state the second result of this subsection, Lemma 4.3, we need the following lemma which is an easy consequence of Taylor's theorem.

**Lemma 4.2.** *If $f : \mathbb{R} \to \mathbb{R}$ is $C^2(\mathbb{R})$ with $|f''(x)| \leqslant \alpha \; \forall\, x \in \mathbb{R}$ then for all $a, b \in \mathbb{R}$,*

$$|f(\tfrac{1}{2}a + \tfrac{1}{2}b) - \tfrac{1}{2}f(a) - \tfrac{1}{2}f(b)| \leqslant \frac{\alpha}{8}\,|b - a|^2.$$

**Lemma 4.3.** *Let $v, w \in L_4(\Omega)$. Then*

$$\|\gamma(\tfrac{1}{2}v + \tfrac{1}{2}w) - \tfrac{1}{2}\gamma(v) - \tfrac{1}{2}\gamma(w)\|_{L_2(\Omega)} \leqslant \frac{C_\gamma''}{8}\|v - w\|_{L_4(\Omega)}^2.$$

*Proof.* Apply Lemma 4.2 to $\gamma$ and use Assumptions 1.1. □

## 4.2 Preparations for the error estimate

For proving the error estimate we invoke the elliptic projections (see [27]), $u^*, \sigma^* \in V^h$, of the exact solutions which are defined via $(\nabla u^*, \nabla v) = (\nabla u, \nabla v)$ and $(\nabla \sigma^*, \nabla v) = (\nabla \sigma, \nabla v)$ each for all $v \in V^h$. Define also $\chi_n := u_n^h - u_n^*$, $\eta_n := \sigma_n^h - \sigma_n^*$, $\xi := u - u^*$ and $\theta := \sigma - \sigma^*$. To render much of what follows a little more compact we introduce the following notation as an expedient.

**Definition 4.4** ('$\mathcal{E}_q$ holds')**.** *For $q = 0$ or $q = 1$ we will say that '$\mathcal{E}_q$ holds' whenever, $\|\zeta\|_{H^{1-q}(\Omega)} \leqslant Ch^{r+q}\|\psi\|_{H^{r+1}(\Omega)}$ for the pair $(\zeta, \psi) = (\xi, u)$ or for the pair $(\zeta, \psi) = (\theta, \sigma)$. The case $q = 0$ is a standard energy estimate while the case $q = 1$ requires elliptic regularity of a dual problem.*

What we have in mind here is that, because the 'natural norm' for the heat equation involves temporally pointwise $L_2(\Omega)$ norms and $L_2(0, T; V)$ norms, an error estimate can easily turn out to be non-optimal in $L_2(\Omega)$. The '$\mathcal{E}_q$' notation will allow us to exploit the '$h^{r+q}$ superconvergence' of the spatial error components, $\chi$ and $\eta$ (see later in (68)), and quote an *a priori* error estimate that is optimal in $V$ and also, elliptic regularity permitting, in $L_2(\Omega)$.

**Lemma 4.5** (Approximation error for elliptic projection)**.** *If $\mathcal{E}_q$ holds and $v^* \in V^h$ is the elliptic projection of $v \in V$ then, for any integers $m \geqslant 0$ and $1 \leqslant s \leqslant r$,*

$$\left\| \frac{\partial^m}{\partial t^m}(v(t) - v^*(t)) \right\|_V \leqslant Ch^s \left\| \frac{\partial^m}{\partial t^m} v(t) \right\|_{H^{s+1}(\Omega)},$$

*provided $\frac{\partial^m}{\partial t^m} v(t) \in H^{s+1}(\Omega)$.*

**Lemma 4.6** (Bounds involving $\xi_n$ and $\theta_n$)**.** *If $\mathcal{E}_q$ holds then, whenever the exact solutions $u$ and $\sigma$ have the indicated regularity, we have for $(\kappa, \zeta, \psi) = (\chi, \xi, u)$ or $(\kappa, \zeta, \psi) = (\eta, \theta, \sigma)$ that,*

$$\|\zeta_n\|_{H^{1-q}(\Omega)} \leqslant Ch^{r+q}\|\psi_n\|_{H^{r+1}(\Omega)}, \tag{29}$$

$$k\sum_{n=1}^{m} \|\bar{\zeta}_n\|^2_{H^{1-q}(\Omega)} \leqslant Ch^{2r+2q}\|\psi\|^2_{L_\infty(0,T,H^{r+1}(\Omega))}, \tag{30}$$

$$\|\kappa_0\|_0 \leqslant \|\zeta_0\|_0 \leqslant Ch^{r+q}\|\breve{\psi}\|_{H^{r+1}(\Omega)}, \tag{31}$$

$$k\sum_{n=1}^{m} \|\partial_t \zeta_n\|_0^2 \leqslant Ch^{2r+2q}\|\dot{\psi}\|^2_{L_\infty(0,T,H^{r+1}(\Omega))},$$

$$k\sum_{n=1}^{m} \|\Delta_n \psi\|_0^2 \leqslant Ck^4\|\dddot{\psi}\|^2_{L_2(0,T,L_2(\Omega))},$$

The proofs of these are straightforward applications of well known techniques. See, for example, [18].

## 4.3 Error inequalities

In this section we establish two inequalities which will form the basis for the error estimate. One inequality is derived by using the inverse Laplacian and the other by using the $u + \sigma$ approach. In both cases we do not yet examine the term that contains the nonlinearity and leave it as it is until the next section.

**Lemma 4.7** (Error inequalities). *If $\mathcal{E}_q$ holds, $u, \sigma \in W^1_\infty(0,T;H^{r+1}(\Omega)) \cap H^3(0,T;L_2(\Omega))$ and $\breve{u}, \breve{\sigma} \in H^{r+1}(\Omega)$ then, for any $\epsilon > 0$ and for any $m \in \{1,\dots,N\}$,*

$$\begin{aligned}\|\mathcal{G}^h\chi_m\|_V^2 + \|\eta_m\|_0^2 + k\sum_{n=1}^m \|\bar{\chi}_n\|_0^2 + 2k\sum_{n=1}^m (\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, \bar{\eta}_n) \\ \leqslant C\Big(1+\frac{1}{\epsilon}\Big)(k^4 + h^{2r+2q}) + \epsilon k\sum_{n=1}^m \|\bar{\eta}_n\|_0^2.\end{aligned} \tag{32}$$

*Furthermore, for all $m \in \{1,\dots,N\}$,*

$$\begin{aligned}\|\chi_m + \eta_m\|_0^2 + k\sum_{n=1}^m \|\bar{\chi}_n + \bar{\eta}_n\|_V^2 + 2k\sum_{n=1}^m (\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, \bar{\chi}_n + \bar{\eta}_n) \\ \leqslant C(k^4 + h^{2r+2q}) + 6C_V^2 k\sum_{n=1}^m \|\bar{\chi}_n\|_0^2.\end{aligned} \tag{33}$$

*Proof.* Subtract the average of (5a) between $t_n$ and $t_{n-1}$ from (18a) and then subtract the average of (5b) between $t_n$ and $t_{n-1}$ from (18b). Observe that $(\nabla\xi_n, \nabla v) = 0$ and $(\nabla\theta_n, \nabla v) = 0$ for any $v \in V^h$ and each $n$. This gives us,

$$(\partial_t\chi_n, v) = (\partial_t\xi_n, v) + (\Delta_n u, v) - (\nabla\bar{\chi}_n, \nabla v) - (\nabla\bar{\eta}_n, \nabla v) \qquad \forall v \in V^h, \tag{34a}$$

$$(\partial_t\eta_n, v) = (\partial_t\theta_n, v) + (\Delta_n\sigma, v) + (\bar{\chi}_n, v) - (\bar{\xi}_n, v) - (\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, v) \quad \forall v \in V^h. \tag{34b}$$

In order to prove (32), choose $v = 2k\mathcal{G}^h\bar{\chi}_n$ in (34a) and $v = 2k\bar{\eta}_n$ in (34b), add the resulting equations, take the sum over $n = 1,\dots,m$ and then apply the Cauchy-Schwarz and Young's inequalities to get,

$$\begin{aligned}\|\mathcal{G}^h\chi_m\|_V^2 + \|\eta_m\|_0^2 + 2k\sum_{n=1}^m \|\bar{\chi}_n\|_0^2 + 2k\sum_{n=1}^m (\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, \bar{\eta}_n) \\ \leqslant \|\mathcal{G}^h\chi_0\|_V^2 + \|\eta_0\|_0^2 + \frac{1}{\epsilon_1}k\sum_{n=1}^m \|\partial_t\xi_n\|_0^2 + \frac{1}{\epsilon_2}k\sum_{n=1}^m \|\partial_t\theta_n\|_0^2 \\ + \frac{1}{\epsilon_3}k\sum_{n=1}^m \|\Delta_n u\|_0^2 + \frac{1}{\epsilon_4}k\sum_{n=1}^m \|\Delta_n\sigma\|_0^2 + \frac{1}{\epsilon_5}k\sum_{n=1}^m \|\bar{\xi}_n\|_0^2 \\ + (\epsilon_1+\epsilon_3)k\sum_{n=1}^m \|\mathcal{G}^h\bar{\chi}_n\|_0^2 + (\epsilon_2+\epsilon_4+\epsilon_5)k\sum_{n=1}^m \|\bar{\eta}_n\|_0^2.\end{aligned}$$

Since by (4), Lemma 3.2 and (4) again $\|\mathcal{G}^h\bar{\chi}_n\|_0 \leqslant C_V\|\mathcal{G}^h\bar{\chi}_n\|_V \leqslant C_V\|\bar{\chi}_n\|_{V'} \leqslant C_V^2\|\bar{\chi}_n\|_0$, we can choose $\epsilon_1 = \epsilon_3 = 1/(2C_V^4)$ and $\epsilon_2 = \epsilon_4 = \epsilon_5 = \epsilon/3$, where $\epsilon > 0$ is arbitrary. Then, using Lemma 3.2 and (4), $\|\mathcal{G}^h\chi_0\|_V \leqslant C_V\|\chi_0\|_0$ and applying $\mathcal{E}_q$ and Lemma 4.6 yields (32).

To prove the second inequality (33), add (34a) and (34b) and choose $v = 2k(\bar{\chi}_n + \bar{\eta}_n)$. Using the Cauchy-Schwarz inequality, $\|v\|_0 \leqslant C_V\|v\|_V$, and then a Young's inequality with $\epsilon = 1/6$ to eliminate the term $\|\bar{\chi}_n + \bar{\eta}_n\|_V^2$ that arises on the right hand side, we sum over $n = 1,\dots,m$ to obtain,

$$\begin{aligned}\|\chi_m + \eta_m\|_0^2 - \|\chi_0 + \eta_0\|_0^2 + k\sum_{n=1}^m \|\bar{\chi}_n + \bar{\eta}_n\|_V^2 + 2k\sum_{n=1}^m (\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, \bar{\chi}_n + \bar{\eta}_n) \\ \leqslant 6C_V^2 k\sum_{n=1}^m \Big(\|\partial_t\xi_n\|_0^2 + \|\partial_t\theta_n\|_0^2 + \|\Delta_n u\|_0^2 + \|\Delta_n\sigma\|_0^2 + \|\bar{\chi}_n\|_0^2 + \|\bar{\xi}_n\|_0^2\Big).\end{aligned} \tag{35}$$

To complete the proof of the second inequality we use the bounds for the terms involving $\xi_n$, $\theta_n$, $\Delta_n u$, $\Delta_n\sigma$, $\chi_0$ and $\eta_0$ from Lemma 4.6 and $\mathcal{E}_q$. ∎

### 4.4 Nonlinearity errrors

The error inequalities which were derived in Lemma 4.7 in the previous section contain terms with the nonlinearity and its approximation. These terms have the following form

$$2k\sum_{n=1}^{m}(\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, \bar{v}_n) \tag{36}$$

where $v_n = \eta_n$ in the case of (32) or $v_n = \chi_n + \eta_n$ in the case of (33). In this section we give bounds for these terms and have to distinguish between the various methods.

Since the terms (36) appear on the left-hand side of (32) and (33) we bound them from below instead of taking the absolute value and bounding them from above. Another reason for bounding below is that for (63), later, (36) contains helpful terms.

This section contains several lemmas. Lemma 4.8 gives a bound for the term (36) for methods $Q \in \{2,5\}$, going as far as possible without distinguishing between $Q = 2$ and $Q = 5$. Then Lemma 4.9 does the same but for methods $Q \in \{1,4\}$. Next, Lemma 4.10 has bounds for the remaining terms from the previous lemmas. These terms contain the extrapolation errors. As opposed to the previous lemmas, here, the linearized methods $Q \in \{1,2\}$ can be dealt with simultaneously, as can the nonlinear methods $Q \in \{4,5\}$ (which do not use extrapolation). Lastly, Lemma 4.11 combines the previous three lemmas and presents a unified nonlinearity error. This result will be used in Section 4.5.

For the next three lemmas, we make no specific choice for $v_n$ in (36) and just assume that $v_n \in L_2(\Omega)$ for $n = 0, \ldots, N$. We will make the choice of either $v_n = \eta_n$ or $v_n = \chi_n + \eta_n$ only in Lemma 4.11.

**Lemma 4.8** (Intermediate nonlinearity error for methods $Q \in \{2,5\}$)**.** *Suppose that $\mathcal{E}_q$ holds and that we have $u \in W_\infty^1(0,T;L_4(\Omega))$ and $\sigma \in L_\infty(0,T;H^{r+1}(\Omega)) \cap W_\infty^1(0,T;L_4(\Omega))$. Let $v_n \in L_2(\Omega)$ for $n = 0, \ldots, N$ and $Q \in \{2,5\}$, then for any $\epsilon > 0$ and any $m \in \{1, \ldots, N\}$,*

$$\begin{aligned}2k\sum_{n=1}^{m}(\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, \bar{v}_n) \geqslant 2k\sum_{n=1}^{m}(\gamma(\mathscr{E}_n^Q u^h)\bar{\eta}_n, \bar{v}_n) - \epsilon k\sum_{n=1}^{m}\|\bar{v}_n\|_0^2 - \frac{C}{\epsilon}(k^4 + h^{2r+2q}) \\ - 2c_1 k\sum_{n=1}^{m}\|\mathscr{E}_n^Q u^h - \bar{u}_n\|_0 \cdot \|\bar{v}_n\|_0,\end{aligned} \tag{37}$$

*where $c_1 = C'_\gamma \|\sigma\|_{L_\infty(0,T;L_\infty(\Omega))}$.*

*Proof.* For $Q \in \{2,5\}$, setting $\tilde{\gamma}_n = \gamma(\mathscr{E}_n^Q u^h)$,

$$\begin{aligned}\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n &= \tilde{\gamma}_n\bar{\sigma}_n^h - \overline{\gamma(u)\sigma}_n \\ &= (\tilde{\gamma}_n\bar{\sigma}_n^h - \tilde{\gamma}_n\bar{\sigma}_n) + (\tilde{\gamma}_n\bar{\sigma}_n - \overline{\gamma(u)}_n\bar{\sigma}_n) + \big(\overline{\gamma(u)}_n\bar{\sigma}_n - \overline{\gamma(u)\sigma}_n\big) \\ &= \tilde{\gamma}_n\bar{\eta}_n - \tilde{\gamma}_n\bar{\theta}_n + \big(\tilde{\gamma}_n - \overline{\gamma(u)}_n\big)\,\bar{\sigma}_n - \tfrac{1}{4}\big(\gamma(u_n) - \gamma(u_{n-1})\big)(\sigma_n - \sigma_{n-1}),\end{aligned}$$

where we have used the fact that for any real numbers $a, b, c, d$,

$$\tfrac{1}{2}(a+b)\cdot\tfrac{1}{2}(c+d) - \tfrac{1}{2}(ac+bd) = -\tfrac{1}{4}(a-b)(c-d).$$

Thus, if for convenience of notation we put,

$$A := \tilde{\gamma}_n\bar{\theta}_n - (\tilde{\gamma}_n - \overline{\gamma(u)}_n)\,\bar{\sigma}_n + \tfrac{1}{4}(\gamma(u_n) - \gamma(u_{n-1}))(\sigma_n - \sigma_{n-1}),$$

then $\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n = \tilde{\gamma}_n\bar{\eta}_n - A$, and taking the scalar product with $\bar{v}_n$, summing over $n$ and multiplying by $2k$ gives us,

$$2k\sum_{n=1}^{m}(\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, \bar{v}_n) \geqslant 2k\sum_{n=1}^{m}(\tilde{\gamma}_n\bar{\eta}_n, \bar{v}_n) - 2k\sum_{n=1}^{m}\|A\|_0\|\bar{v}_n\|_0, \tag{38}$$

and for the norm of $A$ we get

$$\|A\|_0 \leqslant \hat{\gamma}\|\bar{\theta}_n\|_0 + \|\tilde{\gamma}_n - \overline{\gamma(u)}_n\|_0 \cdot \|\bar{\sigma}_n\|_{L_\infty(\Omega)} + \tfrac{1}{4}\|\gamma(u_n) - \gamma(u_{n-1})\|_{L_4(\Omega)} \cdot \|\sigma_n - \sigma_{n-1}\|_{L_4(\Omega)}. \tag{39}$$

Next we bound each term on the right of (39). First, $\|\bar{\sigma}_n\|_{L_\infty(\Omega)} \leqslant \|\sigma\|_{L_\infty(0,T;L_\infty(\Omega))}$. Next, by (13), $\|\sigma_n - \sigma_{n-1}\|_{L_4(\Omega)} \leqslant k\|\dot{\sigma}\|_{L_\infty(t_{n-1},t_n;L_4(\Omega))}$ and, similarly, using Lemma 4.1 and (13), $\|\gamma(u_n) - \gamma(u_{n-1})\|_{L_4(\Omega)} \leqslant C'_\gamma\, k\|\dot{u}\|_{L_\infty(t_{n-1},t_n;L_4(\Omega))}$. Also, using Lemmas 4.1 and 4.3,

$$\begin{aligned}\|\gamma(\mathscr{E}_n^Q u^h) - \overline{\gamma(u)}_n\|_0 &\leqslant \|\gamma(\mathscr{E}_n^Q u^h) - \gamma(\bar{u}_n)\|_0 + \|\gamma(\bar{u}_n) - \overline{\gamma(u)}_n\|_0,\\ &\leqslant C'_\gamma\,\|\mathscr{E}_n^Q u^h - \bar{u}_n\|_0 + \frac{C''_\gamma}{8}k^2\|\dot{u}\|^2_{L_\infty(t_{n-1},t_n;L_4(\Omega))}.\end{aligned}$$

Combining all these and inserting them in (39), we get

$$\|A\|_0 \leqslant \hat{\gamma}\|\bar{\theta}_n\|_0 + c_1\|\mathscr{E}_n^Q u^h - \bar{u}_n\|_0 + c_2\, k^2, \tag{40}$$

where $c_2 = \frac{C''_\gamma}{8}\,\|\dot{u}\|^2_{L_\infty(0,T;L_4(\Omega))}\|\sigma\|_{L_\infty(0,T;L_\infty(\Omega))} + \frac{C'_\gamma}{4}\,\|\dot{u}\|_{L_\infty(0,T;L_4(\Omega))}\|\dot{\sigma}\|_{L_\infty(0,T;L_4(\Omega))}$. Inserting (40) in (38) gives

$$2k\sum_{n=1}^{m}(\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, \bar{v}_n) \geqslant 2k\sum_{n=1}^{m}(\tilde{\gamma}_n\bar{\eta}_n, \bar{v}_n) - 2k\sum_{n=1}^{m}\big(\hat{\gamma}\|\bar{\theta}_n\|_0 + c_1\|\mathscr{E}_n^Q u^h - \bar{u}_n\|_0 + c_2\, k^2\big)\|\bar{v}_n\|_0,$$

and using Young's inequality and rearranging finally gives us

$$\begin{aligned}2k\sum_{n=1}^{m}(\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, \bar{v}_n) \geqslant 2k\sum_{n=1}^{m}(\tilde{\gamma}_n\bar{\eta}_n, \bar{v}_n) &- \frac{2\hat{\gamma}^2}{\epsilon}k\sum_{n=1}^{m}\|\bar{\theta}_n\|_0^2 - \frac{2c_2^2T}{\epsilon}k^4\\ &- \epsilon k\sum_{n=1}^{m}\|\bar{v}_n\|_0^2 - 2c_1k\sum_{n=1}^{m}\|\mathscr{E}_n^Q u^h - \bar{u}_n\|_0 \cdot \|\bar{v}_n\|_0,\end{aligned}$$

and an application of (30) from Lemma 4.6 now yields inequality (37). To conclude we read off the regularity requirements on $u$ and $\sigma$ from the norms appearing in (30) and in the definitions of $c_1$ and $c_2$. Note that $\sigma \in L_\infty(0,T;H^{r+1}(\Omega))$ implies $\sigma \in L_\infty(0,T;L_\infty(\Omega))$, since $r \geqslant 1$ and $d \leqslant 3$. □

**Lemma 4.9** (Intermediate nonlinearity error for methods $Q \in \{1,4\}$)**.** *Suppose that $\mathcal{E}_q$ holds and also that we have $\breve{u}, \breve{\sigma} \in H^{r+1}(\Omega)$ and $u, \sigma \in L_\infty(0,T;H^{r+1}(\Omega))$. Let also $v_n \in L_2(\Omega)$ for $n = 0,\ldots,N$ and $Q \in \{1,4\}$. Then for any $\epsilon > 0$ and for any $m \in \{1,\ldots,N\}$,*

$$\begin{aligned}2k\sum_{n=1}^{m}(\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, \bar{v}_n) \geqslant 2k\sum_{n=1}^{m}(\gamma(\mathscr{E}_n^Q u^h)\bar{\eta}_n, \bar{v}_n) &- \epsilon k\sum_{n=1}^{m}\|\bar{v}_n\|_0^2 - \frac{C}{\epsilon}\,h^{2r+2q}\\ -\frac{2c_1^2}{\epsilon}k\sum_{n=1}^{m-1}\|\chi_n\|_0^2 - \frac{4\hat{\gamma}^2}{\epsilon}k\sum_{n=1}^{m-1}\|\eta_n\|_0^2 &- c_1k\sum_{n=1}^{m}\|\mathscr{E}_n^Q u^h - u_n\|_0 \cdot \|\bar{v}_n\|_0,\end{aligned} \tag{41}$$

*where $c_1 = C'_\gamma\|\sigma\|_{L_\infty(0,T;L_\infty(\Omega))}$.*

*Proof.* For $Q \in \{1,4\}$ we have, setting $\tilde{\gamma}_n := \gamma(\mathscr{E}_n^Q u^h)$,

$$\begin{aligned}\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n &= \tfrac{1}{2}\tilde{\gamma}_n\sigma_n^h + \tfrac{1}{2}\gamma(u_{n-1}^h)\sigma_{n-1}^h - \tfrac{1}{2}\gamma(u_n)\sigma_n - \tfrac{1}{2}\gamma(u_{n-1})\sigma_{n-1}\\ &= \tfrac{1}{2}\tilde{\gamma}_n(\sigma_n^h - \sigma_n) + \tfrac{1}{2}\gamma(u_{n-1}^h)(\sigma_{n-1}^h - \sigma_{n-1})\\ &\quad + \tfrac{1}{2}\big(\tilde{\gamma}_n - \gamma(u_n)\big)\sigma_n + \tfrac{1}{2}\big(\gamma(u_{n-1}^h) - \gamma(u_{n-1})\big)\sigma_{n-1}. \end{aligned} \tag{42}$$

We can further rearrange,

$$\tfrac{1}{2}\tilde{\gamma}_n(\sigma_n^h - \sigma_n) + \tfrac{1}{2}\gamma(u_{n-1}^h)(\sigma_{n-1}^h - \sigma_{n-1}) = \tilde{\gamma}_n(\bar{\eta}_n - \bar{\theta}_n) - \tfrac{1}{2}\big(\tilde{\gamma}_n - \gamma(u_{n-1}^h)\big)(\eta_{n-1} - \theta_{n-1}).$$

Inserting this in (42) gives

$$\begin{aligned}\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n &= \tilde{\gamma}_n(\bar{\eta}_n - \bar{\theta}_n) - \tfrac{1}{2}\big(\tilde{\gamma}_n - \gamma(u_{n-1}^h)\big)(\eta_{n-1} - \theta_{n-1})\\ &\quad + \tfrac{1}{2}\big(\tilde{\gamma}_n - \gamma(u_n)\big)\sigma_n + \tfrac{1}{2}\big(\gamma(u_{n-1}^h) - \gamma(u_{n-1})\big)\sigma_{n-1}.\end{aligned}$$

Thus, taking the inner product with $\bar{v}_n$, multiplying by $2k$, summing up this equation for $n = 1, \dots, m$ and using the Cauchy-Schwarz inequality gives,

$$\begin{aligned}2k\sum_{n=1}^m(\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, \bar{v}_n) &\geqslant 2k\sum_{n=1}^m(\tilde{\gamma}_n\bar{\eta}_n, \bar{v}_n) - 2k\sum_{n=1}^m\|\tilde{\gamma}_n\bar{\theta}_n\|_0 \cdot \|\bar{v}_n\|_0\\ &\quad - k\sum_{n=1}^m\|(\tilde{\gamma}_n - \gamma(u_{n-1}^h))(\eta_{n-1} - \theta_{n-1})\|_0 \cdot \|\bar{v}_n\|_0\\ &\quad - k\sum_{n=1}^m\|(\tilde{\gamma}_n - \gamma(u_n))\sigma_n + (\gamma(u_{n-1}^h) - \gamma(u_{n-1}))\sigma_{n-1}\|_0 \cdot \|\bar{v}_n\|_0.\end{aligned} \tag{43}$$

We bound some of the terms in this inequality now. Firstly,

$$\|(\tilde{\gamma}_n - \gamma(u_{n-1}^h))(\eta_{n-1} - \theta_{n-1})\|_0 \leqslant 2\hat{\gamma}(\|\eta_{n-1}\|_0 + \|\theta_{n-1}\|_0). \tag{44}$$

Secondly,

$$\begin{aligned}&\|(\tilde{\gamma}_n - \gamma(u_n))\sigma_n + (\gamma(u_{n-1}^h) - \gamma(u_{n-1}))\sigma_{n-1}\|_0\\ &\quad\leqslant \|\tilde{\gamma}_n - \gamma(u_n)\|_0 \cdot \|\sigma_n\|_{L_\infty(\Omega)} + \|\gamma(u_{n-1}^h) - \gamma(u_{n-1})\|_0 \cdot \|\sigma_{n-1}\|_{L_\infty(\Omega)}\\ &\quad\leqslant \|\sigma\|_{L_\infty(0,T,L_\infty(\Omega))}\left(\|\gamma(\mathscr{E}_n^Q u^h) - \gamma(u_n)\|_0 + \|\gamma(u_{n-1}^h) - \gamma(u_{n-1})\|_0\right)\\ &\quad\leqslant C'_\gamma\,\|\sigma\|_{L_\infty(0,T,L_\infty(\Omega))}\left(\|\mathscr{E}_n^Q u^h - u_n\|_0 + \|u_{n-1}^h - u_{n-1}\|_0\right),\end{aligned}$$

where we have used Lemma 4.1. Defining $c_1 = C'_\gamma\,\|\sigma\|_{L_\infty(0,T,L_\infty(\Omega))}$, we obtain

$$\|(\tilde{\gamma}_n - \gamma(u_n))\sigma_n + (\gamma(u_{n-1}^h) - \gamma(u_{n-1}))\sigma_{n-1}\|_0 \leqslant c_1\|\mathscr{E}_n^Q u^h - u_n\|_0 + c_1\left(\|\chi_{n-1}\|_0 + \|\xi_{n-1}\|_0\right). \tag{45}$$

Using (44) and (45) in (43) and applying several Young's inequalities yields

$$\begin{aligned}2k\sum_{n=1}^m(\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, \bar{v}_n) &\geqslant 2k\sum_{n=1}^m(\tilde{\gamma}_n\bar{\eta}_n, \bar{v}_n) - (\epsilon_1 + \epsilon_2 + \epsilon_3)k\sum_{n=1}^m\|\bar{v}_n\|_0^2\\ &\quad - \frac{\hat{\gamma}^2}{\epsilon_1}k\sum_{n=1}^m\|\bar{\theta}_n\|_0^2 - \frac{\hat{\gamma}^2}{\epsilon_2}k\sum_{n=1}^m(\|\eta_{n-1}\|_0 + \|\theta_{n-1}\|_0)^2\\ &\quad - \frac{c_1^2}{4\epsilon_3}k\sum_{n=1}^m(\|\chi_{n-1}\|_0 + \|\xi_{n-1}\|_0)^2 - c_1k\sum_{n=1}^m\|\mathscr{E}_n^Q u^h - u_n\|_0 \cdot \|\bar{v}_n\|_0.\end{aligned}$$

Let $\epsilon > 0$ be arbitrary and choose $\epsilon_1 = \epsilon_3 = \epsilon/4$ and $\epsilon_2 = \epsilon/2$, and then an application of (30) and (31) from Lemma 4.6 finishes the proof. This last step also requires the regularity assumptions on $u$, $\sigma$, $\breve{u}$ and $\breve{\sigma}$. ∎

**Lemma 4.10** (Extrapolation errors). *Let $\mathcal{E}_q$ hold and assume that $\breve{u}, \breve{\sigma} \in H^{r+1}(\Omega)$ and $u \in L_\infty(0,T;H^{r+1}(\Omega))$. For methods $Q \in \{1,2\}$ we also suppose that $u \in H^2(0,T;L_2(\Omega))$. Let also $v_n \in L_2(\Omega)$ for $n = 0,\dots,N$ such that for the first element $\|v_0\|_0^2 \leqslant Ch^{2r+2q}$. Then there is a constant $C > 0$ such that for all $\epsilon, \epsilon_3, \epsilon_4 > 0$ and for all $m \in \{1,\dots,N\}$ these inequalities hold:*
*for methods $Q = 1$ and $Q = 2$,*

$$\left.\begin{aligned} k\sum_{n=1}^{m}\|\mathscr{E}_n^1 u^h - u_n\|_0 \cdot \|\bar{v}_n\|_0 \\ 2k\sum_{n=1}^{m}\|\mathscr{E}_n^2 u^h - \bar{u}_n\|_0 \cdot \|\bar{v}_n\|_0 \end{aligned}\right\} \leqslant C\cdot(1+\tfrac{1}{\epsilon_3}+\tfrac{1}{\epsilon_4})\cdot(k^4+h^{2r+2q}) + \epsilon_4\|v_1\|_0^2 + \epsilon_3 k\sum_{n=2}^{m}\|\bar{v}_n\|_0^2 + \frac{10}{\epsilon_3}k\sum_{n=1}^{m-1}\|\chi_n\|_0^2, \tag{46}$$

*and for methods $Q = 4$ and $Q = 5$,*

$$\left.\begin{aligned} k\sum_{n=1}^{m}\|\mathscr{E}_n^4 u^h - u_n\|_0 \cdot \|\bar{v}_n\|_0 \\ 2k\sum_{n=1}^{m}\|\mathscr{E}_n^5 u^h - \bar{u}_n\|_0 \cdot \|\bar{v}_n\|_0 \end{aligned}\right\} \leqslant \epsilon k\sum_{n=1}^{m}\|\bar{v}_n\|_0^2 + \frac{C}{\epsilon}h^{2r+2q} + \frac{2}{\epsilon}k\sum_{n=1}^{m-1}\|\chi_n\|_0^2 + \frac{4}{\epsilon}k\|\bar{\chi}_m\|_0^2. \tag{47}$$

*Proof.* To have a common notation for all methods, we define for this proof,

$$E_n^Q := \begin{cases} \|\mathscr{E}_n^Q u^h - u_n\|_0 & \text{if } Q \in \{1,4\}, \\ 2\|\mathscr{E}_n^Q u^h - \bar{u}_n\|_0 & \text{if } Q \in \{2,5\}. \end{cases} \tag{48}$$

To start, we examine $E_n^Q$ for each method separately. Firstly, for $Q = 1$ and $n \geqslant 2$, using the Taylor estimate (16):

$$\begin{aligned} \|\mathscr{E}_n^1 u^h - u_n\|_0 &= \|2u_{n-1}^h - u_{n-2}^h - u_n\|_0 \\ &= \|2(u_{n-1}^h - u_{n-1}) - (u_{n-2}^h - u_{n-2}) - (u_n - 2u_{n-1} + u_{n-2})\|_0 \\ &\leqslant 2\|u_{n-1}^h - u_{n-1}\|_0 + \|u_{n-2}^h - u_{n-2}\|_0 + \|u_n - 2u_{n-1} + u_{n-2}\|_0 \\ &\leqslant 2\|\chi_{n-1}\|_0 + 2\|\xi_{n-1}\|_0 + \|\chi_{n-2}\|_0 + \|\xi_{n-2}\|_0 + Ck^{3/2}\|\ddot{u}\|_{L_2(t_{n-2},t_n;L_2(\Omega))}, \end{aligned} \tag{49}$$

while for $Q = 1$ and $n = 1$, using the Taylor estimate (13) and (31),

$$\begin{aligned} \|\mathscr{E}_1^1 u^h - u_1\|_0 = \|u_0^h - u_1\|_0 = \|(u_0^h - u_0) - (u_1 - u_0)\|_0 = \|(\chi_0 - \xi_0) - (u_1 - u_0)\|_0 \\ \leqslant \|\chi_0\|_0 + \|\xi_0\|_0 + k\|\partial_t u_1\| \leqslant 2\|\xi_0\|_0 + k\|\dot{u}\|_{L_\infty(0,t_1;L_2(\Omega))}. \end{aligned} \tag{50}$$

Secondly, for $Q = 2$ and $n \geqslant 2$, using the Taylor estimate (16),

$$\begin{aligned} \|\mathscr{E}_n^2 u^h - \bar{u}_n\|_0 &= \|\tfrac{3}{2}u_{n-1}^h - \tfrac{1}{2}u_{n-2}^h - \tfrac{1}{2}u_n - \tfrac{1}{2}u_{n-1}\|_0 \\ &= \|\tfrac{3}{2}(u_{n-1}^h - u_{n-1}) - \tfrac{1}{2}(u_{n-2}^h - u_{n-2}) - \tfrac{1}{2}(u_n - 2u_{n-1} + u_{n-2})\|_0 \\ &\leqslant \|\tfrac{3}{2}(\chi_{n-1} - \xi_{n-1}) - \tfrac{1}{2}(\chi_{n-2} - \xi_{n-2})\|_0 + \tfrac{1}{2}\|u_n - 2u_{n-1} + u_{n-2}\|_0 \\ &\leqslant \tfrac{3}{2}\|\chi_{n-1}\|_0 + \tfrac{1}{2}\|\chi_{n-2}\|_0 + \tfrac{3}{2}\|\xi_{n-1}\|_0 \\ &\quad + \tfrac{1}{2}\|\xi_{n-2}\|_0 + Ck^{3/2}\|\ddot{u}\|_{L_2(t_{n-2},t_n;L_2(\Omega))}, \end{aligned} \tag{51}$$

while for $Q = 2$ and $n = 1$, using the Taylor estimate (13) and (31),

$$\begin{aligned} \|\mathscr{E}_1^2 u^h - \bar{u}_1\|_0 = \|u_0^h - \bar{u}_1\|_0 = \|u_0^h - u_0 - \tfrac{1}{2}(u_1 - u_0)\|_0 = \|\chi_0 - \xi_0 - \tfrac{1}{2}k\partial_t u_1\|_0 \\ \leqslant \|\chi_0\|_0 + \|\xi_0\|_0 + \tfrac{1}{2}k\|\partial_t u_1\|_0 \leqslant 2\|\xi_0\|_0 + \tfrac{1}{2}k\|\dot{u}\|_{L_\infty(0,t_1;L_2(\Omega))}. \end{aligned} \tag{52}$$

For $Q = 4$, we obtain

$$\|\mathscr{E}_n^4 u^h - u_n\|_0 = \|u_n^h - u_n\|_0 = \|\chi_n - \xi_n\|_0 \leqslant \|\chi_n\|_0 + \|\xi_n\|_0. \tag{53}$$

Finally, for $Q = 5$, we obtain

$$\|\mathscr{E}_n^5 u^h - \bar{u}_n\|_0 = \|\bar{u}_n^h - \bar{u}_n\|_0 = \|\bar{\chi}_n - \bar{\xi}_n\|_0 \leqslant \|\bar{\chi}_n\|_0 + \|\bar{\xi}_n\|_0. \tag{54}$$

In order to prove the first inequality, (46), we observe that for $Q = 1$ and $Q = 2$, from (49), (50), (51) and (52),

$$E_n^Q \leqslant \begin{cases} 3\|\chi_{n-1}\|_0 + \|\chi_{n-2}\|_0 + Ch^{r+q} + Ck^{3/2}\|\ddot{u}\|_{L_2(t_{n-2},t_n;L_2(\Omega))} & \text{if } n \geqslant 2, \\ Ch^{r+q} + k\|\dot{u}\|_{L_\infty(0,t_1;L_2(\Omega))} & \text{if } n = 1, \end{cases} \tag{55}$$

where the bounds for $\xi_n$ from (29) have also been used. Now we form the term on the left-hand side of (46), split the sum, and use the fact that $\|\bar{v}_1\|_0^2 \leqslant \frac{1}{2}(\|v_1\|_0^2 + \|v_0\|_0^2)$,

$$k\sum_{n=1}^m E_n^Q\|\bar{v}_n\|_0 \leqslant k\sum_{n=2}^m E_n^Q\|\bar{v}_n\|_0 + \frac{k}{2}E_1^Q(\|v_1\|_0 + \|v_0\|_0).$$

Using Young's inequality gives us

$$k\sum_{n=1}^m E_n^Q\|\bar{v}_n\|_0 \leqslant \frac{1}{4\epsilon_3}k\sum_{n=2}^m (E_n^Q)^2 + \epsilon_3 k\sum_{n=2}^m\|\bar{v}_n\|_0^2 + \Big(\frac{1}{4} + \frac{1}{4\epsilon_4}\Big)(\tfrac{k}{2}E_1^Q)^2 + \epsilon_4\|v_1\|_0^2 + \|v_0\|_0^2.$$

Inserting the bounds for $E_n^Q$ from (55), we get, using norm equivalence in $\mathbb{R}^n$,

$$\begin{aligned} k\sum_{n=1}^m E_n^Q\cdot\|\bar{v}_n\|_0 \leqslant \frac{1}{4\epsilon_3}k\sum_{n=2}^m \Bigg(&36\|\chi_{n-1}\|_0^2 + 4\|\chi_{n-2}\|_0^2 + Ch^{2r+2q} \\ &+ Ck^3\|\ddot{u}\|_{L_2(t_{n-2},t_n;L_2(\Omega))}^2\Bigg) + \epsilon_3 k\sum_{n=2}^m\|\bar{v}_n\|_0^2 \\ &+ C(1+\tfrac{1}{\epsilon_4})k^2\big(Ch^{2r+2q} + 2k^2\|\dot{u}\|_{L_\infty(0,t_1;L_2(\Omega))}^2\big) + \epsilon_4\|v_1\|_0^2 + \|v_0\|_0^2. \end{aligned}$$

Now since $\sum_{n=2}^m\|\ddot{u}\|_{L_2(t_{n-2},t_n;L_2(\Omega))}^2 \leqslant 2\|\ddot{u}\|_{L_2(0,T;L_2(\Omega))}^2$,

$$\begin{aligned} k\sum_{n=1}^m E_n^Q\cdot\|\bar{v}_n\|_0 \leqslant{}& C(1+\tfrac{1}{\epsilon_3}+\tfrac{1}{\epsilon_4})h^{2r+2q} + \tfrac{C}{\epsilon_3}k^4\|\ddot{u}\|_{L_2(0,T;L_2(\Omega))}^2 + C(1+\tfrac{1}{\epsilon_4})k^4\|\dot{u}\|_{L_\infty(0,t_1;L_2(\Omega))}^2 \\ &+ \frac{10}{\epsilon_3}k\sum_{n=0}^{m-1}\|\chi_n\|_0^2 + \epsilon_3 k\sum_{n=2}^m\|\bar{v}_n\|_0^2 + \epsilon_4\|v_1\|_0^2 + \|v_0\|_0^2. \end{aligned}$$

To conclude the proof of (46) note that by assumption $\|v_0\|_0^2 \leqslant Ch^{2r+2q}$ and that by (31) the summation of $\|\chi_n\|_0^2$ can start at $n = 1$, thus we arrive at (46).

To prove the second inequality, (47) for $Q = 4$ and $Q = 5$, we also form the term on the left-hand side of (47) and use Young's inequality (3) to get

$$k\sum_{n=1}^m E_n^Q\cdot\|\bar{v}_n\|_0 \leqslant \frac{1}{4\epsilon}k\sum_{n=1}^m (E_n^Q)^2 + \epsilon k\sum_{n=1}^m\|\bar{v}_n\|_0^2. \tag{56}$$

Now for $Q = 4$ from (53) we have $E_n^4 \leqslant \|\chi_n\|_0 + \|\xi_n\|_0$ and since $\chi_m = 2\bar{\chi}_m - \chi_{m-1}$ we have,

$$\sum_{n=1}^m (E_n^4)^2 \leqslant 2\sum_{n=1}^m\|\chi_n\|_0^2 + 2\sum_{n=1}^m\|\xi_n\|_0^2 \leqslant 16\|\bar{\chi}_m\|_0^2 + 6\sum_{n=1}^{m-1}\|\chi_n\|_0^2 + 2\sum_{n=1}^m\|\xi_n\|_0^2.$$

For $Q=5$ we have from (54) that $E_n^5 \leqslant 2(\|\bar{\chi}_n\|_0 + \|\bar{\xi}_n\|_0)$, and so, using (12)

$$\sum_{n=1}^{m}(E_n^5)^2 \leqslant 8\sum_{n=1}^{m}\|\bar{\chi}_n\|_0^2 + 8\sum_{n=1}^{m}\|\bar{\xi}_n\|_0^2 \leqslant 8\|\bar{\chi}_m\|_0^2 + 8\sum_{n=1}^{m-1}\|\chi_n\|_0^2 + 4\|\chi_0\|_0^2 + 8\sum_{n=1}^{m}\|\bar{\xi}_n\|_0^2.$$

To combine methods $Q=4$ and $Q=5$ we observe that for both these methods, after multiplying by $k$ and using (31) and (29)

$$k\sum_{n=1}^{m}(E_n^Q)^2 \leqslant 16k\|\bar{\chi}_m\|_0^2 + 8k\sum_{n=1}^{m-1}\|\chi_n\|_0^2 + Ch^{2r+2q},$$

and inserting this into (56) gives (47). The regularity assumptions arise because we have used (31) and (29), as well as (13) and (16) for $Q=1$ and $Q=2$. We also recalled the embedding $H^2(0,T;L_2(\Omega)) \hookrightarrow W_\infty^1(0,T;L_2(\Omega))$. □

Now the results from the previous lemmas can be combined to get the lower bounds on the term (36). Note that in the next lemma the nonlinear methods $Q=4$ and $Q=5$ require one extra term in the bound (57), as compared to the linearized methods. This term will later give a maximum value for the timestep size $k$ for $Q=4$ and $Q=5$.

**Lemma 4.11** (Nonlinearity errors for methods $Q \in \{1,2,4,5\}$)**.** *If $\breve{u}, \breve{\sigma} \in H^{r+1}(\Omega)$ and $u \in L_\infty(0,T;H^{r+1}(\Omega)) \cap W_\infty^1(0,T;L_4(\Omega)) \cap H^2(0,T;L_2(\Omega))$ along with $\sigma \in L_\infty(0,T;H^{r+1}(\Omega)) \cap W_\infty^1(0,T;L_4(\Omega))$ then, if $\mathcal{E}_q$ holds, for any method $Q \in \{1,2,4,5\}$ there are constants $C, c_3, c_4, c_5^Q, c_6, c_7, c_8, c_9$ such that for any $m \in \{1,\dots,N\}$ we have,*

$$\begin{aligned} 2k\sum_{n=1}^{m}(\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, \bar{\eta}_n) \geqslant \check{\gamma}k\sum_{n=1}^{m}\|\bar{\eta}_n\|_0^2 - C(k^4 + h^{2r+2q}) - c_3\,k\sum_{n=1}^{m-1}\|\chi_n\|_0^2 \\ - c_4\,k\sum_{n=1}^{m-1}\|\eta_n\|_0^2 - c_5^Q\,k\|\chi_m\|_0^2 - \frac{1}{2}\|\eta_1\|_0^2, \end{aligned} \tag{57}$$

*and also*

$$\begin{aligned} 2k\sum_{n=1}^{m}(\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, \bar{\chi}_n + \bar{\eta}_n) \geqslant -C(k^4 + h^{2r+2q}) - \frac{1}{2}\|\chi_1 + \eta_1\|_0^2 - c_6\,k\sum_{n=1}^{m-1}\|\chi_n\|_0^2 \\ - c_7\,k\sum_{n=1}^{m-1}\|\eta_n\|_0^2 - c_8\,k\|\bar{\chi}_m\|_0^2 - c_9\,k\|\bar{\eta}_m\|_0^2, \end{aligned} \tag{58}$$

*where the constants depend on $u$ and $\sigma$, but are independent of $h$ and $k$. In fact $c_1 = C'_\gamma\|\sigma\|_{L_\infty(0,T;L_\infty(\Omega))}$, $c_3 = 28c_1^2/\check{\gamma}$, $c_4 = 8\hat{\gamma}^2/\check{\gamma}$, $c_6 = 4 + 24c_1^2$, $c_7 = 4 + 9\hat{\gamma}^2$, $c_8 = 4 + 8c_1^2$, $c_9 = 4 + \hat{\gamma}^2$ and $c_5^Q = 0$ if $Q \in \{1,2\}$ or $c_5^Q = \frac{4(C'_\gamma)^2}{\check{\gamma}}\|\sigma\|_{L_\infty(0,T;L_\infty(\Omega))}^2$ if $Q \in \{4,5\}$.*

*Proof.* To combine the bounds for all methods, we firstly compare (37) and (41) and thus obtain for any $Q \in \{1,2,4,5\}$ and any $\epsilon_1 > 0$,

$$\begin{aligned} 2k\sum_{n=1}^{m}(\mathscr{B}_n^Q - \overline{\gamma(u)\sigma}_n, \bar{v}_n) \geqslant 2k\sum_{n=1}^{m}(\gamma(\mathscr{E}_n^Q u^h)\bar{\eta}_n, \bar{v}_n) - \epsilon_1 k\sum_{n=1}^{m}\|\bar{v}_n\|_0^2 - \frac{C}{\epsilon_1}(k^4 + h^{2r+2q}) \\ - \frac{2c_1^2}{\epsilon_1}k\sum_{n=1}^{m-1}\|\chi_n\|_0^2 - \frac{4\hat{\gamma}^2}{\epsilon_1}k\sum_{n=1}^{m-1}\|\eta_n\|_0^2 - c_1 k\sum_{n=1}^{m}E_n^Q\|\bar{v}_n\|_0, \end{aligned} \tag{59}$$

where $E_n^Q$ is defined as in (48). Secondly, by comparing (46) and (47) we get

$$k\sum_{n=1}^{m} E_n^Q \cdot \|\bar{v}_n\|_0 \leqslant C\cdot(1+\tfrac{1}{\epsilon_3}+\tfrac{1}{\epsilon_4})\cdot(k^4+h^{2r+2q})+\epsilon_4\|v_1\|_0^2 \\ +\epsilon_3 k\sum_{n=1}^{m}\|\bar{v}_n\|_0^2+\frac{10}{\epsilon_3}k\sum_{n=1}^{m-1}\|\chi_n\|_0^2+\frac{4I^Q}{\epsilon_3}k\|\bar{\chi}_m\|_0^2,$$

where $I^Q$ is an indicator for the method, which we define as $I^Q=1$ for $Q\in\{4,5\}$ and $I^Q=0$ for $Q\in\{1,2\}$. Inserting this in (59) gives

$$2k\sum_{n=1}^{m}(\mathscr{B}_n^Q-\overline{\gamma(u)\sigma}_n,\bar{v}_n)\geqslant 2k\sum_{n=1}^{m}(\gamma(\mathscr{E}_n^Q u^h)\bar{\eta}_n,\bar{v}_n)-(\epsilon_1+c_1\epsilon_3)k\sum_{n=1}^{m}\|\bar{v}_n\|_0^2 \\ -C(1+\tfrac{1}{\epsilon_1}+\tfrac{1}{\epsilon_3}+\tfrac{1}{\epsilon_4})(k^4+h^{2r+2q}) \\ -\Big(\frac{2c_1^2}{\epsilon_1}+\frac{10c_1}{\epsilon_3}\Big)k\sum_{n=1}^{m-1}\|\chi_n\|_0^2-\frac{4\hat{\gamma}^2}{\epsilon_1}k\sum_{n=1}^{m-1}\|\eta_n\|_0^2-\frac{4c_1I^Q}{\epsilon_3}k\|\bar{\chi}_m\|_0^2-c_1\epsilon_4\|v_1\|_0^2.$$

Let $\epsilon>0$ be arbitrary and choose $\epsilon_1=\epsilon/2$, $\epsilon_3=\epsilon/(2c_1)$ and $\epsilon_4=1/(2c_1)$. This yields

$$2k\sum_{n=1}^{m}(\mathscr{B}_n^Q-\overline{\gamma(u)\sigma}_n,\bar{v}_n)\geqslant 2k\sum_{n=1}^{m}(\gamma(\mathscr{E}_n^Q u^h)\bar{\eta}_n,\bar{v}_n)-\epsilon k\sum_{n=1}^{m}\|\bar{v}_n\|_0^2-C(1+\tfrac{1}{\epsilon})(k^4+h^{2r+2q}) \\ -\frac{24c_1^2}{\epsilon}k\sum_{n=1}^{m-1}\|\chi_n\|_0^2-\frac{8\hat{\gamma}^2}{\epsilon}k\sum_{n=1}^{m-1}\|\eta_n\|_0^2-\frac{8c_1^2I^Q}{\epsilon}k\|\bar{\chi}_m\|_0^2-\frac{1}{2}\|v_1\|_0^2. \tag{60}$$

To prove (57) choose $v_n=\eta_n$ in (60) and use $(\gamma(\mathscr{E}_n^Q u^h)\bar{\eta}_n,\bar{\eta}_n)\geqslant\check{\gamma}\|\bar{\eta}_n\|_0^2$. At the same time, to get a term with $\|\chi_m\|_0^2$ instead of $\|\bar{\chi}_m\|_0^2$, we also notice that $\|\bar{\chi}_m\|_0^2\leqslant\frac{1}{2}(\|\chi_m\|_0^2+\|\chi_{m-1}\|_0^2)$ and hide the new $\|\chi_{m-1}\|_0^2$ in the sum of $\|\chi_n\|_0^2$. Choosing $\epsilon=\check{\gamma}$ gives (57).

Next, choose $v_n=\chi_n+\eta_n$ and $\epsilon=1$ in (60) and obtain, since $(\gamma(\mathscr{E}_n^Q u^h)\bar{\eta}_n,\bar{\chi}_n+\bar{\eta}_n)\geqslant -\hat{\gamma}\|\bar{\eta}_n\|_0\cdot\|\bar{\chi}_n+\bar{\eta}_n\|_0$,

$$2k\sum_{n=1}^{m}(\mathscr{B}_n^Q-\overline{\gamma(u)\sigma}_n,\bar{\chi}_n+\bar{\eta}_n)\geqslant -2\hat{\gamma}k\sum_{n=1}^{m}\|\bar{\eta}_n\|_0\cdot\|\bar{\chi}_n+\bar{\eta}_n\|_0-k\sum_{n=1}^{m}\|\bar{\chi}_n+\bar{\eta}_n\|_0^2 \\ -C(k^4+h^{2r+2q})-24c_1^2k\sum_{n=1}^{m-1}\|\chi_n\|_0^2-8\hat{\gamma}^2k\sum_{n=1}^{m-1}\|\eta_n\|_0^2-8c_1^2I^Qk\|\bar{\chi}_m\|_0^2-\frac{1}{2}\|\chi_1+\eta_1\|_0^2. \tag{61}$$

Now using Young's inequality (3) gives $2\hat{\gamma}\|\bar{\eta}_n\|_0\|\bar{\chi}_n+\bar{\eta}_n\|_0\leqslant\hat{\gamma}^2\|\bar{\eta}_n\|_0^2+\|\bar{\chi}_n+\bar{\eta}_n\|_0^2$, so we can write for the first two terms on the right-hand side of (61), using also (12),

$$2\hat{\gamma}k\sum_{n=1}^{m}\|\bar{\eta}_n\|_0\cdot\|\bar{\chi}_n+\bar{\eta}_n\|_0+k\sum_{n=1}^{m}\|\bar{\chi}_n+\bar{\eta}_n\|_0^2\leqslant\hat{\gamma}^2k\sum_{n=1}^{m}\|\bar{\eta}_n\|_0^2+2k\sum_{n=1}^{m}\|\bar{\chi}_n+\bar{\eta}_n\|_0^2 \\ \leqslant(\hat{\gamma}^2+4)k\Big(\|\bar{\eta}_m\|_0^2+\sum_{n=1}^{m-1}\|\eta_n\|_0^2+\tfrac{1}{2}\|\eta_0\|_0^2\Big)+4k\Big(\|\bar{\chi}_m\|_0^2+\sum_{n=1}^{m-1}\|\chi_n\|_0^2+\tfrac{1}{2}\|\chi_0\|_0^2\Big).$$

Inserting this into (61) and using (31) we finally obtain (58), the second inequality we wanted to prove. The regularity requirements are obtained from those in Lemmas 4.8, 4.9 and 4.10, but from now on we do not distinguish between the methods anymore and just take the maximum requirements. □

### 4.5 Error estimate

In this section we combine the error inequalities in Lemma 4.7 from Section 4.3 with the nonlinearity errors in Lemma 4.11 from Section 4.4. As it turns out, a combined approach that uses both the inverse Laplacian and the $u+\sigma$ arguments will remove the condition on the timestep size $k$ for the linearized methods.

**Theorem 4.12** (*a priori* error estimate)**.** *Suppose that $\mathcal{E}_q$ holds, that $\breve{u}, \breve{\sigma} \in H^{r+1}(\Omega)$, and that the exact solution satisfies $u, \sigma \in W^1_\infty(0,T;H^{r+1}(\Omega)) \cap H^3(0,T;L_2(\Omega))$ with $\breve{\sigma}|_{\Gamma_D} = 0$. For the nonlinear methods $Q=4$ and $Q=5$, we also suppose that $k$ is sufficiently small:*

$$k < \hat{k} = \check{\gamma} \cdot \Big(128 c_1^2 \max\big\{2 + 4c_1^2 + 3C_V^2, (4+\hat{\gamma}^2)/\check{\gamma}\big\}\Big)^{-1},$$

*where $c_1 = C'_\gamma \|\sigma\|_{L_\infty(0,T;L_\infty(\Omega))}$. Then for methods $Q \in \{1,2,4,5\}$, we have the following error estimate,*

$$\begin{aligned} &\max_{1\leqslant m\leqslant N} \big\{\|u(t_m) - u_m^h\|_0 + \|\sigma(t_m) - \sigma_m^h\|_0\big\} \\ &\qquad + h^q \left(k \sum_{n=1}^{N} \|\bar{u}_n + \bar{\sigma}_n - \bar{u}_n^h - \bar{\sigma}_n^h\|_V^2\right)^{1/2} \leqslant C(k^2 + h^{r+q}). \end{aligned} \tag{62}$$

*Proof.* Firstly, use (57) in (32), omit the $\|\mathcal{G}^h \chi_m\|_V$-term and choose $\epsilon = \check{\gamma}/2$ to get,

$$\begin{aligned} &\|\eta_m\|_0^2 + k\sum_{n=1}^{m} \|\bar{\chi}_n\|_0^2 + \frac{\check{\gamma}}{2} k \sum_{n=1}^{m} \|\bar{\eta}_n\|_0^2 \\ &\quad \leqslant C(k^4 + h^{2r+2q}) + c_3 k \sum_{n=1}^{m-1} \|\chi_n\|_0^2 + c_4 k \sum_{n=1}^{m-1} \|\eta_n\|_0^2 + c_5^Q k \|\chi_m\|_0^2 + \frac{1}{2}\|\eta_1\|_0^2. \end{aligned} \tag{63}$$

For $m=1$, the term $\frac{1}{2}\|\eta_1\|_0^2$ in (63) can be absorbed into the left-hand side, and, omitting the other terms on the left-hand side, we obtain

$$\frac{1}{2}\|\eta_1\|_0^2 \leqslant C(k^4 + h^{2r+2q}) + c_5^Q k \|\chi_1\|_0^2.$$

Inserting this again into (63) yields

$$\begin{aligned} &\|\eta_m\|_0^2 + k\sum_{n=1}^{m} \|\bar{\chi}_n\|_0^2 + \frac{\check{\gamma}}{2} k \sum_{n=1}^{m} \|\bar{\eta}_n\|_0^2 \\ &\quad \leqslant C(k^4 + h^{2r+2q}) + c_3 k \sum_{n=1}^{m-1} \|\chi_n\|_0^2 + c_4 k \sum_{n=1}^{m-1} \|\eta_n\|_0^2 + c_5^Q k \|\chi_m\|_0^2 + c_5^Q k \|\chi_1\|_0^2. \end{aligned}$$

Depending on whether $m=1$ or $m>1$, this last term can either be combined with the $\|\chi_m\|_0^2$-term or with the sum of $\|\chi_n\|_0^2$. Hence for any $m \geqslant 1$ we can write:

$$\begin{aligned} &\|\eta_m\|_0^2 + k\sum_{n=1}^{m} \|\bar{\chi}_n\|_0^2 + \frac{\check{\gamma}}{2} k \sum_{n=1}^{m} \|\bar{\eta}_n\|_0^2 \\ &\quad \leqslant C(k^4 + h^{2r+2q}) + (c_3 + c_5^Q) k \sum_{n=1}^{m-1} \|\chi_n\|_0^2 + c_4 k \sum_{n=1}^{m-1} \|\eta_n\|_0^2 + 2c_5^Q k \|\chi_m\|_0^2. \end{aligned} \tag{64}$$

To get a second inequality that can be combined with this one we use (58) and (31) in (33) to get,

$$\begin{aligned}\|\chi_m+\eta_m\|_0^2+k\sum_{n=1}^{m}\|\bar{\chi}_n+\bar{\eta}_n\|_V^2 \leqslant C(k^4+h^{2r+2q})+c_{10}k\sum_{n=1}^{m-1}\|\chi_n\|_0^2+c_7k\sum_{n=1}^{m-1}\|\eta_n\|_0^2\\ +c_{11}k\|\bar{\chi}_m\|_0^2+c_9k\|\bar{\eta}_m\|_0^2+\frac{1}{2}\|\chi_1+\eta_1\|_0^2,\end{aligned} \tag{65}$$

where the $\|\bar{\chi}_n\|_0$-term from (33) has been split up and added to the other terms, giving $c_{10}=c_6+6C_V^2$ and $c_{11}=c_8+6C_V^2$.

For $m=1$, the term $\frac{1}{2}\|\chi_1+\eta_1\|_0^2$ in (65) cancels with the left-hand side, and, omitting the other term on the left-hand side, we obtain

$$\frac{1}{2}\|\chi_1+\eta_1\|_0^2 \leqslant C(k^4+h^{2r+2q})+c_{11}k\|\bar{\chi}_1\|_0^2+c_9k\|\bar{\eta}_1\|_0^2.$$

Inserting this into (65) and adding extra terms (for convenience of writing) to the right hand side gives,

$$\begin{aligned}\|\chi_m+\eta_m\|_0^2+k\sum_{n=1}^{m}\|\bar{\chi}_n+\bar{\eta}_n\|_V^2 \leqslant C(k^4+h^{2r+2q})+c_{10}k\sum_{n=1}^{m-1}\|\chi_n\|_0^2+c_7k\sum_{n=1}^{m-1}\|\eta_n\|_0^2\\ +2c_{11}k\sum_{n=1}^{m}\|\bar{\chi}_n\|_0^2+2c_9k\sum_{n=1}^{m}\|\bar{\eta}_n\|_0^2.\end{aligned} \tag{66}$$

Now we define $c_{12}=\max\{2c_{11},4c_9/\check{\gamma}\}$. Multiplying (64) by $c_{12}$ and adding it to (66) gives, after noting that $2c_{11}\leqslant c_{12}$ and $2c_9\leqslant c_{12}\check{\gamma}/2$:

$$\begin{aligned}&\|\chi_m+\eta_m\|_0^2+c_{12}\|\eta_m\|_0^2+k\sum_{n=1}^{m}\|\bar{\chi}_n+\bar{\eta}_n\|_V^2\\ &\leqslant C(k^4+h^{2r+2q})+Ck\sum_{n=1}^{m-1}(\|\chi_n\|_0^2+\|\eta_n\|_0^2)+2c_{12}c_5^Qk\|\chi_m\|_0^2.\end{aligned} \tag{67}$$

If we set $c_{13}=\max\{4,5/c_{12}\}$, then

$$\begin{aligned}2\|\chi_m\|_0^2+\|\eta_m\|_0^2 &\leqslant 2(\|\chi_m+\eta_m\|_0+\|\eta_m\|_0)^2+\|\eta_m\|_0^2 \leqslant 4\|\chi_m+\eta_m\|_0^2+5\|\eta_m\|_0^2\\ &\leqslant 4\|\chi_m+\eta_m\|_0^2+\frac{5}{c_{12}}c_{12}\|\eta_m\|_0^2 \leqslant c_{13}(\|\chi_m+\eta_m\|_0^2+c_{12}\|\eta_m\|_0^2),\end{aligned}$$

and hence $\|\chi_m+\eta_m\|_0^2+c_{12}\|\eta_m\|_0^2 \geqslant c_{13}^{-1}(2\|\chi_m\|_0^2+\|\eta_m\|_0^2)$. Using this in (67) and multiplying by $c_{13}$ gives

$$\begin{aligned}&2\|\chi_m\|_0^2+\|\eta_m\|_0^2+c_{13}k\sum_{n=1}^{m}\|\bar{\chi}_n+\bar{\eta}_n\|_V^2\\ &\leqslant C(k^4+h^{2r+2q})+Ck\sum_{n=1}^{m-1}(\|\chi_n\|_0^2+\|\eta_n\|_0^2)+2c_{13}c_{12}c_5^Qk\|\chi_m\|_0^2.\end{aligned}$$

For the methods $Q=4$ and $Q=5$, where $c_5^Q>0$, we now need that $k\leqslant\hat{k}=1/(2c_{13}c_{12}c_5^Q)$, so that the term $\|\chi_m\|_0^2$ on the right-hand side can be absorbed into the left-hand side. Hence, for all $Q=1,2,4,5$ we get

$$\|\chi_m\|_0^2+\|\eta_m\|_0^2+c_{13}k\sum_{n=1}^{m}\|\bar{\chi}_n+\bar{\eta}_n\|_V^2 \leqslant C(k^4+h^{2r+2q})+Ck\sum_{n=1}^{m-1}(\|\chi_n\|_0^2+\|\eta_n\|_0^2).$$

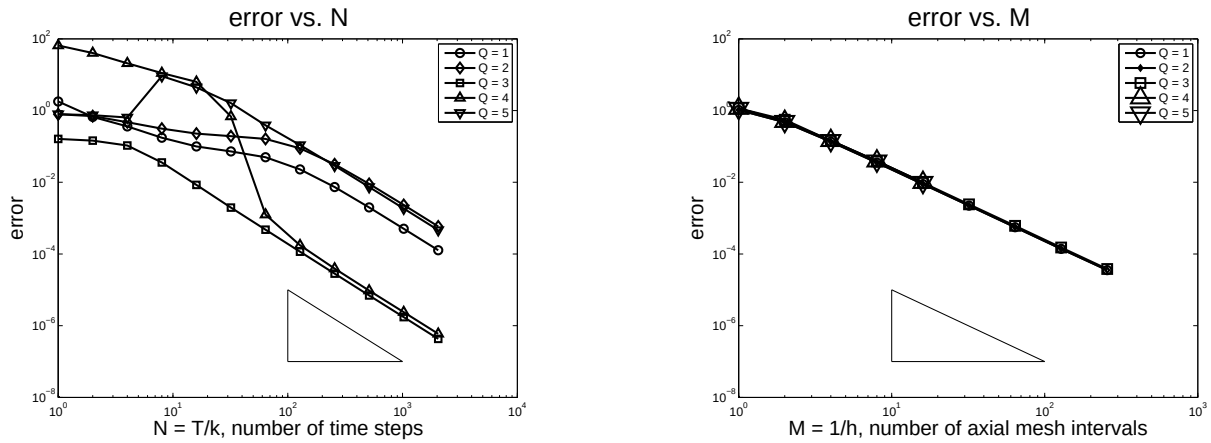


Figure 1: Convergence curves showing the value of the error on the right in Theorem 4.12 with the temporal (left) and spatial (right) discretisation parameters. Those curves on the left are based on Tables 7.11, 7.13, 7.15, 7.17, 7.19 in [5] and those on the right on Tables 7.55, 7.57, 7.59, 7.61, 7.63.

Using the discrete Gronwall lemma, and multiplying by a suitable constant, we get

$$\|\chi_m\|_0^2 + \|\eta_m\|_0^2 + k\sum_{n=1}^{m}\|\bar{\chi}_n + \bar{\eta}_n\|_V^2 \leqslant C(k^4 + h^{2r+2q}). \tag{68}$$

Finally since $u_n - u_n^h = \xi_n - \chi_n$ and $\sigma_n - \sigma_n^h = \theta_n - \eta_n$ we use the triangle inequality, (68), $\mathcal{E}_q$ and Lemma 4.6 and arrive at the error estimate. Tracking the constants in $\hat{k}$ results in the expression given in the theorem and this completes the proof. □

# 5 Numerical results

In this section we give some numerical evidence for the convergence rates claimed in Theorem 4.12, and also some computed solutions indicating why $u + \sigma$ is a well-behaved quantity as compared with $u$ and $\sigma$ individually. All results have been obtained with code based on the Alberta library, [20].

These numerical experiments are confined to two spatial dimensions and piecewise linear finite element approximation. Our numerics are a sampling of the extensive set of results given in [5], which contains results in one, two and three spatial dimensions, for polynomial degrees up to and including four, for two forms of $\gamma$ and for imposed and non-imposed boundary conditions on $\sigma$. Following [9, 10] we take $\gamma$ as

$$\gamma(u) = \frac{1}{2}(\hat{\gamma} + \check{\gamma}) + \frac{1}{2}(\hat{\gamma} - \check{\gamma})\tanh\left(\frac{u - u_a}{\Delta}\right), \tag{69}$$

with $\check{\gamma} = 1$, $\hat{\gamma} = 100$, $u_a = 0.5$ and $\Delta = 0.05$. It is easy to see that Assumption 1.1 is satisfied.

For the first set of results we take $\Omega = (0,1)^2$ and $T = 1$ so that the time step is given by $k = N^{-1}$. The mesh on $\Omega$ was generated by subdividing each coordinate direction into $M$ equal intervals along each axis to form a mesh of $M^2$ squares of side-length $h = M^{-1}$. The triangulation was produced by partitioning each square into four triangles, as illustrated by ⊠.

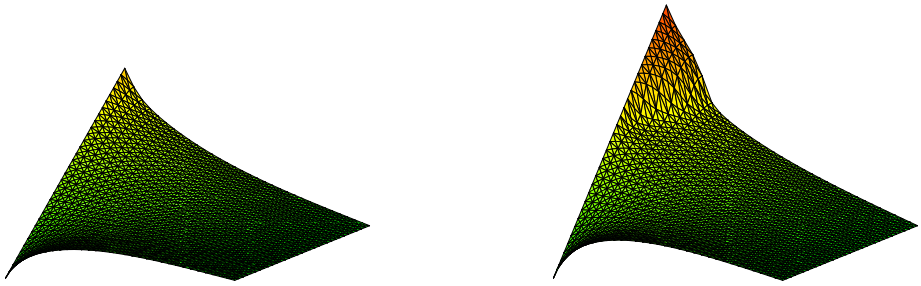

Figure 2: Sharp front in 2D: Numerical solution $u$ at $t = 1.8$ (left) and $t = 3.6$ (right)

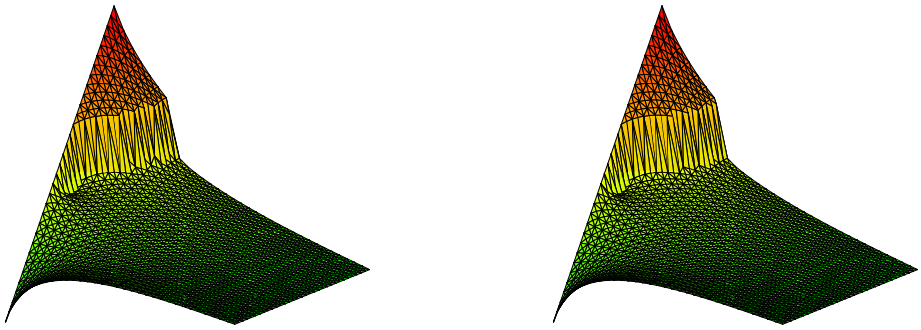

Figure 3: Sharp front in 2D: Numerical solution $u$ at $t = 15$ (left) and $t = 30$ (right).

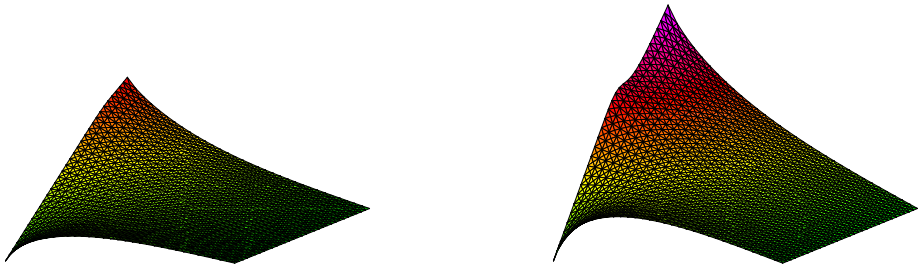

Figure 4: Sharp front in 2D: Numerical solution $u+\sigma$ at $t = 1.8$ (left) and $t = 3.6$ (right).

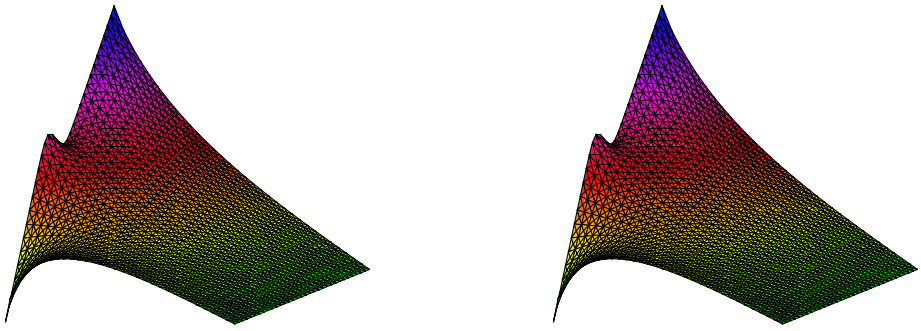

Figure 5: Sharp front in 2D: Numerical solution $u + \sigma$ at $t = 15$ (left) and $t = 30$ (right).

To demonstrate the convergence rates we add a term $\hat{f}$ as in (1b) and then substitute an exact solution into the differential equations to find $f$, $\hat{f}$ and the values of the boundary and initial data. In these convergence tests Dirichlet boundary data are specified on $\{(x, y) \in \partial\Omega \colon x = 0\}$ with the remainder of the boundary having Neumann data.

For the methods $Q \in \{1, 2, 4, 5\}$ the approximation of the exact solution given by $u = (x + y)\cos(t)$ and $\sigma = (2x + y)t^3$ has no spatial discretisation error and is used to illustrate the temporal convergence rate. For $Q = 3$ we use the simpler form for $u$ given by $u = \cos(t)$.

On the other hand, the exact solution $u = \sin(\pi x)\,\sin(\pi y)$ and $\sigma = (1 - t)\exp(x + y)$ is used to demonstrate the spatial convergence rate.

The error plots are shown in Figure 1, with the inscribed triangles showing the slope corresponding to a convergence rate of $k^2 \propto N^{-2}$ (left) and $h^2 \propto M^{-2}$ (right). For the $k$ convergence we took $M = 1$ while for the $h$ convergence $k$ was taken as $1$ for methods $Q \in \{1, 2, 3\}$ and $k = 0.0001$ for $Q \in \{4, 5\}$. This is because the nonlinear systems appear to require a very small value of $k/h$ in order to become meaningfully solvable. (The nonlinear system was solved by a fixed point iteration until either $1000$ iterations were reached or the Euclidean norm of the residual fell below $10^{-13}$, see [5] for more on this.) It is clear that in all cases there is good agreement with Theorem 4.12 and that, as expected, the $h$-convergence is independent of the method chosen.

The second set of results is included only to illustrate the possibility of very sharp fronts evolving in both $u$ and $\sigma$ while the sum, $u + \sigma$, apparently remains smooth. Note that we have no exact solution in this case and so impose no Dirichlet boundary conditions on $\sigma$ (strictly, therefore, the foregoing error analysis does not apply). The domain $\Omega$ remains as $(0, 1)^2$ but now the Dirichlet boundary is $\{(x, y) \in \partial\Omega \colon x = 0$ or $x = 1\}$. The function $u_D = (1 - e^{-t/2})y$ is prescribed on $\{x = 0\}$ and $u_D = 0$ on $\{x = 1\}$, the remainder of the boundary has homogeneous Neumann data.

The numerical solution is computed up to $T = 30$ with method $Q = 3$. Snapshots of the computed $u$ are shown in Figures 2 and 3, and of the computed $u + \sigma$ in Figures 4 and 5. Although we do not show it, it is apparent that $\sigma$ must also have a sharp front that additively cancels the one in $u$.

# 6 Conclusion

For the non-Fickian polymer diffusion problem as described in [9, 10] we have presented five fully discrete approximations, derived *a priori* stability and error estimates for four of them and given numerical results for all five.

Three of these schemes are particularly attractive in that the discrete problems are linear. This is achieved through extrapolation. The numerical evidence suggests that these extrapolation schemes are superior to the nonlinear discrete schemes that arise from the intuitively 'more accurate' non-extrapolated schemes.

# List of references

[1] Herbert Amann. Global existence for a class of highly degenerate parabolic systems. *Japan J. Indust. Appl. Math.*, 8:143—151, 1991.

[2] Herbert Amann. Nonhomogeneous linear and quasilinear elliptic and parabolic boundary value problems. In H.J. Schmeisser and H. Triebel, editors, *Function Spaces, Differential Operators and Nonlinear Analysis*, pages 9—126. Teubner, Stuttgart, Leipzig, 1993.

[3] John W Barrett, Harald Garcke, and Robert Nürnberg. Finite element approximation of surfactant spreading on a thin film. *SIAM J. Numer. Anal.*, 41:1427—1464, 2003.

[4] John W Barrett and Peter Knabner. Finite element approximation of the transport of reactive solutes in porous media. Part 1: error estimates for nonequilibrium adsorption processes. *SIAM J. Numer. Anal.*, 34:201—227, 1997.

[5] Norbert Bauermeister. *Finite element methods for non-Fickian diffusion in viscoelastic polymers*. PhD thesis, Brunel University, England, 2007. (See `www.brunel.ac.uk/bicom`).

[6] Hu Bei and Zhang Jianhua. Global existence for a class of non-Fickian polymer-penetrant systems. *J. Partial Diff. Eqs.*, 9:193—208, 1996.

[7] J. R. Cannon and Y. Lin. *A priori* $L^2$ error estimates for finite-element methods for nonlinear diffusion equations with memory. *SIAM J. Numer. Anal.*, 27:595—607, 1990.

[8] I. Christie, D. F. Griffiths, A. R. Mitchell, and J. M. Sanz-Serna. Product approximation for nonlinear problems in the finite element method. *IMA J. Numer. Anal.*, 1(3):253–266, 1981.

[9] D. S. Cohen and A. B. White Jr. Sharp fronts due to diffusion and viscoelastic relaxation in polymers. *SIAM J. Appl. Math.*, 51:472—483, 1991.

[10] D. S. Cohen, A. B. White Jr., and T. P. Witelski. Shock formation in a multidimensional viscoelastic diffusive system. *SIAM J. Appl. Math.*, 55:348—368, 1995.

[11] J. Douglas and B. F. Jones. Numerical methods for integro-differential equations of parabolic and hyperbolic types. *Numer. Math.*, 4:96—102, 1962.

[12] David A. Edwards. Constant front speed in weakly diffusive non-Fickian systems. *SIAM J. Appl. Math.*, 55:1039—1058, 1995.

[13] J. D. Ferry. *Viscoelastic properties of polymers*. John Wiley and Sons Inc., 1970.

[14] W. Mclean and V. Thomée. Numerical solution of an evolution equation with a positive-type memory term. *J. Austral. Math. Soc.*, 35:23—70, 1993.

[15] David J. Needham, John A. Leach, John Merkin, Norbert Bauermeister, and Simon Shaw. On the properties of degenerate nonlinear non-Fickian polymer diffusion systems. (in preparation).

[16] Amiya K. Pani and Todd E. Peterson. Finite element methods with numerical quadrature for parabolic integrodifferential equations. *SIAM J. Numer. Anal.*, 33:1084—1105, 1996.

[17] G. Rehage, O. Ernst, and J. Fuhrmann. Fickian and non-Fickian diffusion in high polymer systems. *Discuss. Faraday Soc.*, 49:208—221, 1970.

[18] Beatrice Riviere and Simon Shaw. Discontinuous Galerkin finite element approximation of nonlinear non-Fickian diffusion in viscoelastic polymers. *SIAM Journal on Numerical Analysis*, 44(6):2650–2670, 2006.

[19] M.-N. Le Roux and V. Thomée. Numerical solution of semilinear integrodifferential equations of parabolic type with nonsmooth data. *SIAM J. Numer. Anal.*, 26:1291—1309, 1989.

[20] Alfred Schmidt and Kunibert G. Siebert. *Design of adaptive finite element software: the finite element toolbox ALBERTA*, volume 42 of *Lecture notes in computational science and engineering*. Springer-Verlag, 2005. (See also `www.alberta-fem.de`.).

[21] I. H. Sloan and V. Thomée. Time discretization of an integro-differential equation of parabolic type. *SIAM J. Numer. Anal.*, 23:1052—1061, 1986.

[22] N. L. Thomas and A. H. Windle. A theory of Case II diffusion. *Polymer*, 23:529—542, 1982.

[23] Noreen Thomas and A. H. Windle. Transport of methanol in poly(methyl-methacrylate). *Polymer*, 19:255—265, 1978.

[24] V. Thomée and L. B. Wahlbin. Long-time numerical solution of a parabolic equation with memory. *Math. Comp.*, 62:477—496, 1994.

[25] Vidar Thomée. *Galerkin finite element methods for parabolic problems*. Number 1054 in Lecture Notes in Mathematics. Springer-Verlag, Heidelberg, 1984.

[26] Dmitry A. Vorotnikov. Dissipative solutions for equations of viscoelastic diffusion in polymers. *J. Math. Anal. Appl.*, 339(2):876–888, 2008.

[27] M. F. Wheeler. A priori $L_2$ error estimates for Galerkin approximations to parabolic partial differential equations. *SIAM J. Numer. Anal.*, 10:723—759, 1973.

[28] E. G. Yanik and G. Fairweather. Finite element methods for parabolic and hyperbolic partial integro-differential equations. *Nonlinear Analysis, Theory, Methods & Applications*, 12:785—809, 1988.

[29] V. Thomée & N.-Y. Zhang. Error estimates for semidiscrete finite element methods for parabolic integro-differential equations. *Math. Comp.*, 53:121—139, 1989.