INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 8, Number 2, Pages 226–251

FINITE ELEMENT APPROXIMATION OF A NON-LOCAL PROBLEM IN NON-FICKIAN POLYMER DIFFUSION

SIMON SHAW

(Communicated by Mark Ainsworth)

Abstract. The problem of non-local nonlinear non-Fickian polymer diffusion as modelled by a diffusion equation with a nonlinearly coupled boundary value problem for a viscoelastic 'pseudostress' is considered (see, for example, DA Edwards in Z. angew. Math. Phys., **52**, 2001, pp. 254—288). We present two numerical schemes using the implicit Euler method and also the Crank-Nicolson method. Each scheme uses a Galerkin finite element method for the spatial discretisation. Special attention is paid to linearising the discrete equations by extrapolating the value of the nonlinear terms from previous time steps. *A priori* error estimates are given, based on the usual assumptions that the exact solution possesses certain regularity properties, and numerical experiments are given to support these error estimates. We demonstrate by example that although both schemes converge at their optimal rates the Euler method may be more robust than the Crank-Nicolson method for problems of practical relevance.

Key words. *a priori* error estimates, nonlinear diffusion, non-Fickian diffusion, finite element method, linearisation, extrapolation, implicit Euler, Crank-Nicolson

1. Introduction and background

In [12, 11] Thomas & Windle demonstrated by experiment that diffusion of a solvent in a viscoelastic polymer matrix is highly non-Fickian with the solvent concentration developing a steep, and possibly travelling, wave front. This front demarcates a concentration-forced phase transition of the polymer from a 'glassy' state to a 'rubbery' state. The viscoelastic time constants in the viscoelastic stress-strain constitutive equation vary sharply across this transition, and this variation is believed to be basic driving mechanism behind the formation of the steep stationary or travelling fronts.

To date most modern attempts at modelling this phenomenon mathematically have been based on introducing a temporal nonlocality into the classical Fickian diffusion law using a hereditary integral for a concentration-induced 'stress'. The motivation is of course from the phenomenological theory of viscoelasticity, e.g. [8], where stress is usually written as a convolution of strain with a 'relaxation function'. The decaying exponential form of this relaxation function then allows the stress in the non-Fickian diffusion law to be represented in terms of an ordinary differential equation (in time). This equation is nonlinearly coupled to the partial differential equation for the concentration. See [4, 5] for more on this and [2, 10] for some related numerical analysis.

In an alternative approach Edwards in [6, 7] argued the need to also permit spatial nonlocality due to the 'long chain' polymer molecules being much larger than the penetrant's molecules. He then proposed a non-Fickian diffusion model based on the introduction of a spatially nonlocal 'viscoelastic *psuedostress*'. The

Received by the editors April 30, 2010 and, in revised form, July 22, 2010.
2000 *Mathematics Subject Classification.* 74S05 (FEM), 74S20 (FDM), 76R50 (diffusion), 74D10 (nonlinear constitutive equations), 82D60 (polymers).

result is still a non-Fickian diffusion law but with this time the 'stress' governed by an elliptic partial differential equation which, again, is nonlinearly coupled.

It is important to realise that at present, in the absence of a 'fundamental theory', these models have been proposed with the aim of developing a mathematical formalism that can capture the experimentally observed behaviour. This type of experimental mathematics requires numerical solution and so, with that motivation, our goal here is to give fully discrete formulations and derive *a priori* stability and error estimates. First we review Edwards' model and then we pose it in a form more suited for our purpose.

The model proposed by Edwards in [6] for the concentration, C, and pseudostress, Θ, takes the form,

$$C_\tau = DC_{yy} + M\Theta_{yy}, \tag{1}$$

$$-(\beta(C)^{-1}\Theta_y)_y + \beta(C)\Theta = \eta C - \varkappa C_y, \tag{2}$$

with D, M, η and $\varkappa$ constant with the first three positive and the last non-negative.

Edwards considers this problem on an unbounded domain, but if we restrict to $(a,b) \subset \mathbb{R}$, the pseudostress is given in [6] by,

$$\Theta(y,t) := -\frac{1}{2}\int_a^b f(C(y',t), C_{y'}(y',t)) \exp\left(-\left|\int_{y'}^{y} \beta(C(z,t))\, dz\right|\right) dy' \tag{3}$$

with $f(C, C_y) := -\eta C + \nu C_y$ for $\eta > 0$, $\nu > 0$ and $\beta(\mathbb{R}) > 0$. In this β^{-1}, the *dependence length*, represents the radius of the smallest sphere, centred at z, that contains a typical polymer chain passing through z. Since these chains will be entangled in a random spaghetti-like manner 'holes' or 'pockets' are formed at their intersections and these provide sites for the penetrant's molecules. The ability of such a molecule to diffuse then depends on the strength (density) of the entanglement, β^{-1}, which in turn is influenced by the degree of penetrant saturation. Indeed the key ingredient in this model is the observation that, due to swelling, β^{-1} in the saturated rubber phase is expected to be much larger than β^{-1} in the drier and more crystalline glassy phase. We will return to this below, but note that it is this effect that generates the nonlinear coupling. The spatial nonlocality arises because a 'path of holes' needs to be formed for the penetrant molecule to move, but we expect the entanglment density far from the molecule to have less influence than that nearby—hence the decay built in to (3).

Although it is not necessary to non-dimensionalise this problem it is convenient to simplify it by scaling out some unneccessary parameters. Setting, $\beta_0 := \sqrt{\eta/D}$, $x = \beta_0 y$, $t = \eta\tau$, $u = \eta C$ and $\sigma = \beta_0\Theta$, with the definitions, $\gamma(u) = \beta(C)/\beta_0$, $\nu = \beta_0\varkappa/\eta$ and $E = \beta_0 M$, and then generalising to many space dimensions (since there is no reason not to), we arrive at our model problem.

Let $\Omega \subset \mathbb{R}^d$ $(d = 1, 2, 3)$ be a bounded (polygonal or polyhedral for $d = 2$ or 3) domain and $I := (0, T]$ a finite time interval. We consider the degenerate problem: find u and σ such that,

$$u_t = \nabla^2 u + E\nabla^2\sigma \tag{4}$$

$$-\nabla\cdot\gamma(u)^{-1}\nabla\sigma + \gamma(u)\sigma = u - \boldsymbol{\nu}\cdot\nabla u, \tag{5}$$

for $E \geqslant 0$ and $\gamma(\mathbb{R}) > 0$. We assume initial and boundary data as follows,

$$u(x,0) = \breve{u}(x) \qquad \text{in } \Omega, \tag{6}$$

$$\hat{\boldsymbol{n}} \cdot \nabla(u + E\sigma) = \lambda(u^\flat - u) \qquad \text{on } \Gamma_u, \tag{7}$$

$$u = 0 \qquad \text{on } \partial\Omega \setminus \Gamma_u, \tag{8}$$

$$\hat{\boldsymbol{n}} \cdot \gamma(u)^{-1} \nabla\sigma + \sigma = 0 \qquad \text{on } \Gamma_\sigma, \tag{9}$$

$$\sigma = 0 \qquad \text{on } \partial\Omega \setminus \Gamma_\sigma, \tag{10}$$

where $\Gamma_u \subseteq \partial\Omega$ and $\Gamma_\sigma \subseteq \partial\Omega$ are time independent (and possibly empty), λ is a positive constant, $\breve{u}$ and $u^\flat$ are given functions and $\hat{\boldsymbol{n}}$ is (a.e.) the unit outward normal to $\partial\Omega$.

Although Edward's model was posed on the whole of $\mathbb{R}$ we have had to restrict to a bounded domain because we want to consider numerical approximations. Therefore we have had to introduce some relevant boundary conditions. To motivate them notice that if we partially differentiate (3) once with respect to y we can derive boundary conditions of Robin type. Specifically, if $\partial/\partial n$ denotes the 'outward' derivative (i.e. $\partial/\partial n = \partial/\partial y$ if $y = b$ and $\partial/\partial n = -\partial/\partial y$ if $y = a$), then,

$$\frac{1}{\beta(C)} \frac{\partial \Theta}{\partial n} + \Theta = 0 \qquad \text{on the boundary.}$$

This motivates (9), and if we (distributionally) differentiate again we arrive at (2)—which we consider in the form (5).

Edwards takes γ to be a piecewise constant idealisation but, to avoid the ensuing numerical difficulties as well as to recognise that the rubber-glass transition is in practice likely to be more nebulous, we follow (in inverted form) the model given in [4] and use the 'smooth step function',

$$\gamma(u) = \left(\frac{\gamma_G + \gamma_R}{2}\right) + \left(\frac{\gamma_G - \gamma_R}{2}\right) \tanh\left(\frac{u - u_c}{\Delta}\right). \tag{11}$$

Here Δ is the width of the transition region around the critical concentration u_c, and $0 < \gamma_R \ll \gamma_G$ where the subscripts refer to the rubber and glass 'phases' (note that this is not a misprint, even though $0 < \gamma_G \ll \gamma_R$ in [4]).

This article is organised as follows. The weak formulation of the problem and a basic stability estimate is given in Section 2, and with that estimate we will see that the main difficulty with this problem is not due to the nonlinearity but to $\boldsymbol{\nu}$. In fact all of our estimates contain conditions that are related to this term. The numerical schemes are given in Section 3. We concentrate on the implicit Euler and Crank-Nicolson methods, each linearised by extrapolation from previous time levels. The advantage of this is a simpler implementation as well as easier-to-prove well-posedness results for the discrete problems (see Prop. 3 below). Unlike in [10, 2] (where there was a σ_t term) we have to make a special effort at the first time step for the Crank-Nicolson method in order to preserve its second-order temporal accuracy. These linearised discretisations are examples of what Lowrie in [9] calls 'lagged' schemes.

The error estimates are contained in Subsections 4.1 and 4.2 of Section 4, and some numerical results are given in Section 5. We conclude in Section 6 with some observations relating to this material as well as to its potential for extension to the model in [7] where 'preferred directions' were eliminated.

The notation we use is fairly standard and is introduced as it is needed. For clarity, we sometimes use an overdot for time derivatives, $\dot{u} = u_t$ etc.

2. Weak formulation and preliminaries

Recalling Green's theorem in the form,

$$-\int_\Omega v\nabla\cdot F\,d\Omega = -\oint_{\partial\Omega} vF\cdot\hat{\boldsymbol{n}}\,d\Gamma + \int_\Omega \nabla v\cdot F\,d\Omega, \tag{12}$$

we arrive at the following weak formulation of the problem (4), (5) with (6), (7), (8), (9), (10) as: find $(u,\sigma)\colon I\to V_u\times V_\sigma$ such that,

$$\begin{aligned}(u_t(t),v)+(\nabla u(t),\nabla v)+(E\nabla\sigma(t),\nabla v)+(\lambda u(t),v)_{\Gamma_u}\\ =(\lambda u^{\flat}(t),v)_{\Gamma_u}\quad\forall v\in V_u,\end{aligned}\tag{13}$$

$$\begin{aligned}(\gamma(u)^{-1}\nabla\sigma(t),\nabla w)+(\gamma(u)\sigma(t),w)+(\sigma(t),w)_{\Gamma_\sigma}\\ -(u(t),w)+(\boldsymbol{\nu}\cdot\nabla u(t),w)=0\quad\forall w\in V_\sigma,\end{aligned}\tag{14}$$

where,

$$\begin{aligned}V_u &:= \{v\in H^1(\Omega)\colon v=0 \text{ on } \partial\Omega\setminus\Gamma_u\},\\ V_\sigma &:= \{v\in H^1(\Omega)\colon v=0 \text{ on } \partial\Omega\setminus\Gamma_\sigma\},\end{aligned}$$

and also: $(\cdot,\cdot)$ denotes the $L_2(\Omega)$ inner product; $(\cdot,\cdot)_\Gamma$ the $L_2(\Gamma)$ inner product; and the dependence of u, σ etc. on $x\in\Omega$ is suppressed.

Our basic assumptions on $\gamma\colon\mathbb{R}\to\mathbb{R}$ are that there exist constants, $\check{\gamma}$, $\hat{\gamma}$ and C'_γ such that,

$$0<\check{\gamma}\leqslant\gamma(x)\leqslant\hat{\gamma}\qquad\text{and}\qquad|\gamma'(x)|\leqslant C'_\gamma\qquad\forall x\in\mathbb{R},\tag{15}$$

so that we also have, $0<\hat{\gamma}^{-1}\leqslant\gamma(x)^{-1}\leqslant\check{\gamma}^{-1}$ for all $x\in\mathbb{R}$. Moreover, it then easily follows from the relationship,

$$\gamma(v)-\gamma(w)=\int_0^1\gamma'(sv+(1-s)w)\,ds\;(v-w),$$

that

$$\|\gamma(v)-\gamma(w)\|_{L_p(\Omega)}\leqslant C'_\gamma\|v-w\|_{L_p(\Omega)}\tag{16}$$

for all $v,w\in L_p(\Omega)$ and for any $p\geqslant 1$.

In what follows, $\|\cdot\|_X$ will always denote the norm on the Banach space X. For simplicity, when $X=H^r(\Omega)$ we abbreviate $\|\cdot\|_{H^r(\Omega)}$ to $\|\cdot\|_r$, and $\|\cdot\|_{H^r(\Gamma)}$ to $\|\cdot\|_{r,\Gamma}$.

From [1, Thm. 6.3.14 & Exmpl. 6.3.16] for example, we have for $\partial\Omega$ Lipschitz,

$$\|v\|_1\leqslant C\left(\|\nabla v\|_0+\|v\|_{L_1(\Gamma)}\right)\qquad\forall v\in H^1(\Omega)$$

and for any non-empty (i.e. $\mathrm{meas}_{n-1}(\Gamma)>0$) open or closed subset $\Gamma\subseteq\partial\Omega$.

It follows that $(\|\nabla\cdot\|_0^2+\|\cdot\|_{L_2(\Gamma)}^2)^{1/2}$ is a norm on V_u (resp. V_σ) equivalent to $\|\cdot\|_1$ in the case $\Gamma=\Gamma_u$ (resp. $\Gamma=\Gamma_\sigma$). Also, by the Poincaré-Friedrich's inequality this equivalence continues to hold even if $\Gamma=\varnothing$ so long as $v=0$ on $\partial\Omega$ (which applies to our set-up).

The next step is to give a basic stability estimate. The method of proof will be used for the stability of the discrete problems in Prop. 2 and also for the error estimates later in Theorems 8 and 12. First, for use here and later we recall Young's inequality,

$$ab\leqslant\frac{a^p}{p\epsilon^p}+\frac{\epsilon^q b^q}{q}\qquad\forall a,b\geqslant 0,\quad\forall\epsilon>0\tag{17}$$

and $1 < p, q < \infty$ such that $p^{-1} + q^{-1} = 1$. And second, we obtain from Green's theorem, (12),

$$(\boldsymbol{\nu} \cdot \nabla u, \sigma) = \oint_{\Gamma_u \cap \Gamma_\sigma} \sigma u \boldsymbol{\nu} \cdot \hat{\boldsymbol{n}} \, d\Gamma - (u, \boldsymbol{\nu} \cdot \nabla \sigma) \tag{18}$$

for all $u \in V_u$ and $\sigma \in V_\sigma$. Now and later we use $\|\cdot\|_{\mathbb{E}}$ to denote the usual 2-norm on $\mathbb{R}^d$ so that $\|\boldsymbol{x}\|_{\mathbb{E}} = \sqrt{(x_1^2 + \cdots + x_d^2)}$.

Proposition 1 (basic stability)**.** *If at least one of the following conditions holds,*

$$\textit{(A) } \boldsymbol{\nu} = \mathbf{0}; \qquad \textit{(B) } \|\boldsymbol{\nu}\|_{\mathbb{E}} < \frac{2}{E}\sqrt{\frac{\check{\gamma}}{\hat{\gamma}}}; \qquad \textit{(C) } \Gamma_u \cap \Gamma_\sigma = \varnothing; \qquad \textit{(D) } \|\boldsymbol{\nu}\|_{\mathbb{E}} < \frac{2}{E}\sqrt{\frac{2\lambda}{\hat{\gamma}}},$$

then there is a constant $C > 0$ such that,

$$\int_0^t \left(\|\gamma(u)^{-1/2} \nabla \sigma(s)\|_0^2 + \|\gamma(u)^{1/2} \sigma(s)\|_0^2 + \|\sigma(s)\|_{0,\Gamma_\sigma}^2 \right) ds + \|u(t)\|_0^2 + \int_0^t \left(\|\nabla u(s)\|_0^2 + \|u(s)\|_{0,\Gamma_u}^2 \right) ds \leqslant C \|\breve{u}\|_0^2 + C \int_0^t \|u^\flat(s)\|_{0,\Gamma_u}^2 \, ds,$$

for all $t \in I$.

Proof. First note that if we can show that

$$\frac{d}{dt}\|u(t)\|_0^2 + \|\nabla u(t)\|_0^2 + \|u(t)\|_{0,\Gamma_u}^2 + \|\gamma(u)^{-1/2}\nabla\sigma(t)\|_0^2 + \|\gamma(u)^{1/2}\sigma(t)\|_0^2 + \|\sigma(t)\|_{0,\Gamma_\sigma}^2 \leqslant C\|u^\flat(t)\|_{0,\Gamma_u}^2 + C\|u(t)\|_0^2, \tag{19}$$

then the result is implied by Grönwall's inequality. So all we need to do is derive (19) for each of conditions (A), (B), (C) and (D).

First we choose $v = 2u(t) \in V_u$ in (13) and, for some $\mu > 0$ to be specified later, choose $w = \mu\sigma(t) \in V_\sigma$ in (14). Adding the results gives,

$$\frac{d}{dt}\|u(t)\|_0^2 + 2\|\nabla u(t)\|_0^2 + 2\lambda\|u(t)\|_{0,\Gamma_u}^2 + \mu\|\gamma(u)^{-1/2}\nabla\sigma(t)\|_0^2 + \mu\|\gamma(u)^{1/2}\sigma(t)\|_0^2 + \mu\|\sigma(t)\|_{0,\Gamma_\sigma}^2 = 2\lambda(u^\flat(t), u(t))_{\Gamma_u} + \mu(u(t), \sigma(t)) - 2E(\nabla\sigma(t), \nabla u(t)) - \mu(\boldsymbol{\nu} \cdot \nabla u(t), \sigma(t)).$$

We number the terms on the right as one, two, three and four and apply Young's inequality, (17), to the first three with ϵ subscripted by the number of the term. This results in,

$$\frac{d}{dt}\|u(t)\|_0^2 + \left(2 - \frac{\epsilon_3 E^2 \hat{\gamma}}{\mu}\right)\|\nabla u(t)\|_0^2 + \left(2 - \frac{1}{\epsilon_1}\right)\lambda\|u(t)\|_{0,\Gamma_u}^2 + \left(1 - \frac{1}{\epsilon_3}\right)\mu\|\gamma(u)^{-1/2}\nabla\sigma(t)\|_0^2 + \left(1 - \frac{\epsilon_2}{2}\right)\mu\|\gamma(u)^{1/2}\sigma(t)\|_0^2 + \mu\|\sigma(t)\|_{0,\Gamma_\sigma}^2 \leqslant \epsilon_1\lambda\|u^\flat(t)\|_{0,\Gamma_u}^2 + \frac{\mu}{2\check{\gamma}\epsilon_2}\|u(t)\|_0^2 + \mu|(\boldsymbol{\nu} \cdot \nabla u(t), \sigma(t))|. \tag{20}$$

Assuming condition (A) we now need only choose $\epsilon_1 = \epsilon_2 = 1$ and any $\mu > E^2\hat{\gamma}/2$. This ensures that we can find $\epsilon_3 > 0$ satisfying $1 - 1/\epsilon_3 > 0$ and $2 - \epsilon_3 E^2\hat{\gamma}/\mu > 0$ and we arrive at a form of (19). This completes the proof under condition (A).

Now assume that condition (B) holds and note that,

$$\mu|(\boldsymbol{\nu} \cdot \nabla u(t), \sigma(t))| \leqslant \frac{\mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{2\epsilon_4\check{\gamma}}\|\nabla u(t)\|_0^2 + \frac{\epsilon_4\mu}{2}\|\gamma(u)^{1/2}\sigma(t)\|_0^2.$$

Using this in (20) gives,

$$\frac{d}{dt}\|u(t)\|_0^2+\left(2-\frac{\epsilon_3E^2\hat{\gamma}}{\mu}-\frac{\mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{2\epsilon_4\check{\gamma}}\right)\|\nabla u(t)\|_0^2+\left(2-\frac{1}{\epsilon_1}\right)\lambda\|u(t)\|_{0,\Gamma_u}^2$$
$$+\left(1-\frac{1}{\epsilon_3}\right)\mu\|\gamma(u)^{-1/2}\nabla\sigma(t)\|_0^2+\left(1-\frac{\epsilon_2}{2}-\frac{\epsilon_4}{2}\right)\mu\|\gamma(u)^{1/2}\sigma(t)\|_0^2$$
$$+\mu\|\sigma(t)\|_{0,\Gamma_\sigma}^2\leqslant\epsilon_1\lambda\|u^\flat(t)\|_{0,\Gamma_u}^2+\frac{\mu}{2\check{\gamma}\epsilon_2}\|u(t)\|_0^2.$$

We choose $\epsilon_1=1$ and $\epsilon_4\in(0,2)$ such that condition (B) implies $\epsilon_4^2\check{\gamma}>E^2\hat{\gamma}\|\boldsymbol{\nu}\|_{\mathbb{E}}^2$. Also, setting $\epsilon_2=1-\epsilon_4/2>0$ we have $1-\epsilon_2/2-\epsilon_4/2=1/2-\epsilon_4/4>0$ and then selecting $\mu=2\check{\gamma}\epsilon_4/\|\boldsymbol{\nu}\|_{\mathbb{E}}^2$ we find that,

$$\frac{(4\check{\gamma}\epsilon_4-\mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2)\mu}{2\hat{\gamma}\check{\gamma}E^2\epsilon_4}>\frac{\check{\gamma}\epsilon_4^2}{\hat{\gamma}E^2\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}>1.$$

It is now clear that we can find ϵ_3 satisfying $1<\epsilon_3<(4\check{\gamma}\epsilon_4-\mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2)\mu/2\hat{\gamma}\check{\gamma}E^2\epsilon_4$ and so it follows that $1-1/\epsilon_3>0$ and $2-\epsilon_3E^2\hat{\gamma}/\mu-\mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2/2\check{\gamma}\epsilon_4>0$ and we arrive again at the form (19).

For conditions (C) and (D) we return again to (20) but this time using (18) to get,

$$\mu|(\boldsymbol{\nu}\cdot\nabla u(t),\sigma(t))|\leqslant\frac{\delta\|\boldsymbol{\nu}\|_{\mathbb{E}}^2\mu}{2\epsilon_4}\|u(t)\|_{0,\Gamma_u}^2+\frac{\delta\epsilon_4\mu}{2}\|\sigma(t)\|_{0,\Gamma_\sigma}^2$$
$$+\frac{\mu\hat{\gamma}\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{2}\|u(t)\|_0^2+\frac{\mu}{2}\|\gamma(u)^{-1/2}\nabla\sigma(t)\|_0^2,$$

where we understand that $\delta=0$ if condition (C) holds and $\delta=1$ otherwise. The result is,

$$\frac{d}{dt}\|u(t)\|_0^2+\left(2-\frac{\epsilon_3E^2\hat{\gamma}}{\mu}\right)\|\nabla u(t)\|_0^2+\left(2\lambda-\frac{\lambda}{\epsilon_1}-\frac{\delta\|\boldsymbol{\nu}\|_{\mathbb{E}}^2\mu}{2\epsilon_4}\right)\|u(t)\|_{0,\Gamma_u}^2$$
$$+\left(\frac{1}{2}-\frac{1}{\epsilon_3}\right)\mu\|\gamma(u)^{-1/2}\nabla\sigma(t)\|_0^2+\left(1-\frac{\epsilon_2}{2}\right)\mu\|\gamma(u)^{1/2}\sigma(t)\|_0^2$$
$$+\left(1-\frac{\delta\epsilon_4}{2}\right)\mu\|\sigma(t)\|_{0,\Gamma_\sigma}^2\leqslant\epsilon_1\lambda\|u^\flat(t)\|_{0,\Gamma_u}^2+\left(\frac{\mu}{2\check{\gamma}\epsilon_2}+\frac{\mu\hat{\gamma}\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{2}\right)\|u(t)\|_0^2.$$

Now, for condition (C) we need only choose $\epsilon_1=\epsilon_2=1$ and $\mu>E^2\hat{\gamma}$ so that we can find an ϵ_3 satisfying $2<\epsilon_3<2\mu/E^2\hat{\gamma}$. We then have $1/2-1/\epsilon_3>0$ and $2-\epsilon_3E^2\hat{\gamma}/\mu>0$ and, again, we arrive at (19).

Finally, condition (D) has $\delta=1$ and implies that $\hat{\gamma}E^2<\varepsilon\lambda/\|\boldsymbol{\nu}\|_{\mathbb{E}}^2$ for some $\varepsilon\in(0,8)$ so choosing $\epsilon_2=1$ and $\epsilon_1>0$ so that $2\lambda-\lambda/\epsilon_1=\varepsilon\lambda/4$ we can choose μ satisfying $\hat{\gamma}E^2<\mu<\varepsilon\lambda/\|\boldsymbol{\nu}\|_{\mathbb{E}}^2$ and, therefore, find ϵ_4 such that, $2\mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2/\varepsilon\lambda<\epsilon_4<2$. These then imply that $1-\epsilon_4/2>0$ and $2\lambda-\lambda/\epsilon_1-\mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2/2\epsilon_4>0$.

Also, since $\hat{\gamma}E^2<\mu$ we can find an ϵ_3 satisfying $2<\epsilon_3<2\mu/\hat{\gamma}E^2$ and this results in $1/2-1/\epsilon_3>0$ and $2-\epsilon_3\hat{\gamma}E^2/\mu>0$. Once again we arrive at the form (19) and this concludes the proof of the lemma under all four conditions. □

We close this section with a comment on the four conditions given in Prop. 1 since versions of them will appear later as well. We can see from (5) that $\boldsymbol{\nu}$ controls the degree to which the psuedostress is driven by the concentration gradient (flux) as opposed to the level of concentration. Condition (A) therefore refers to cases where σ is not flux-driven. Condition (C) is self-explanatory and conditions (B) and (D) place limits of the magnitude of the convective influence exerted by ∇u in

a manner not unlike that of standard approaches to convection-diffusion problems. Which one of (B) and (D) is more useful will depend on the problem at hand. For example, if in (7) we assume near-perfect insulation on Γ_u then we can take λ as small as we please and (D) tends (A). On the other hand (B) and (D) both reveal that as the magnitude of influence of σ on u is decreased through making E smaller, then the effect of $\boldsymbol{\nu}\cdot\nabla u$ can be larger (and *vice versa*). This seems reasonable.

3. The numerical schemes

As usual we define finite dimensional subspaces $V_u^h \subset V_u$ and $V_\sigma^h \subset V_\sigma$ where each of V_u^h and V_σ^h is built with piecewise polynomials of degree $r \geqslant 1$ using the same member of a non-degenerate and quasi-uniform family, $\{\mathscr{T}^h\}_h$, of subdivisions of Ω. Also, for $N \in \mathbb{N}$ we define the time step $k := T/N$ and set $t_i = ik$. In general, we write $v_i := v(t_i)$ and in particular we write the approximate solution to (13) and (14) as $u_i^h \approx u(t_i)$ and $\sigma_i^h \approx \sigma(t_i)$.

We study two schemes, the first is an implicit Euler method and the second a Crank-Nicolson method. Both are linear. The linearisation is achieved for the Euler method by evaluating the nonlinearity at the previous (in time) solution, while for the Crank-Nicolson method we extrapolate linearly from the previous two time levels. This needs a starting value and for this the Euler method is used in a predictor-corrector fashion.

The notation used for the time discretisation is:

$$\partial_t v_i := \frac{v_i - v_{i-1}}{k}, \qquad \mathscr{E}_i v := \frac{3}{2}v_{i-1} - \frac{1}{2}v_{i-2},$$

$$\bar{v}_i := \frac{v_i + v_{i-1}}{2}, \qquad \Delta_i v := \frac{v_t(t_i) + v_t(t_{i-1})}{2} - \frac{v(t_i) - v(t_{i-1})}{k},$$

for $i \in \{1, 2, \ldots, N\}$ and where v_{-1} will be defined appropriately below. In terms of deriving estimates related to these operators, notice that Δ_i is related to the trapezium rule, whereas $\mathscr{E}_i$ is a linear extrapolation of v to its value at the midpoint $t_{i-1/2}$.

The discrete schemes are defined by: for $i = 1, 2, \ldots, N$ in turn, find $(u_i^h, \sigma_i^h) \in V_u^h \times V_\sigma^h$ such that,

for the **implicit Euler scheme:**

$$\begin{aligned}(\partial_t u_i^h, v) + (\nabla u_i^h, \nabla v) + (E\nabla\sigma_i^h, \nabla v) + (\lambda u_i^h, v)_{\Gamma_u} \\ = (\lambda u_i^\flat, v)_{\Gamma_u} \quad \forall v \in V_u^h, \end{aligned} \tag{21}$$

$$\begin{aligned}(\gamma(u_{i-1}^h)^{-1}\nabla\sigma_i^h, \nabla w) + (\gamma(u_{i-1}^h)\sigma_i^h, w) + (\sigma_i^h, w)_{\Gamma_\sigma} \\ - (u_i^h, w) + (\boldsymbol{\nu}\cdot\nabla u_i^h, w) = 0 \quad \forall w \in V_\sigma^h, \end{aligned} \tag{22}$$

or for the **Crank-Nicolson scheme**

$$\begin{aligned}(\partial_t u_i^h, v) + (\nabla \bar{u}_i^h, \nabla v) + (E\nabla\bar{\sigma}_i^h, \nabla v) + (\lambda \bar{u}_i^h, v)_{\Gamma_u} \\ = (\lambda \bar{u}_i^\flat, v)_{\Gamma_u} \quad \forall v \in V_u^h, \end{aligned} \tag{23}$$

$$\begin{aligned}(\gamma(\mathscr{E}_i u^h)^{-1}\nabla\bar{\sigma}_i^h, \nabla w) + (\gamma(\mathscr{E}_i u^h)\bar{\sigma}_i^h, w) + (\bar{\sigma}_i^h, w)_{\Gamma_\sigma} \\ - (\bar{u}_i^h, w) + (\boldsymbol{\nu}\cdot\nabla \bar{u}_i^h, w) = 0 \quad \forall w \in V_\sigma^h, \end{aligned} \tag{24}$$

$$u_{-1}^h := 2u_0^h - \hat{u}_1^h \tag{25}$$

with, in both cases,

$$(u_0^h, v) = (\breve{u}, v) \qquad \forall v \in V_u^h, \tag{26}$$

and $\hat{u}_1^h$ is given by the first step of the implicit Euler method. Note that in the above, for the sake of clarity later on, we do not attempt to distinguish u_i^h between the schemes (like, say, u_i^{ie} and u_i^{cn}). It will always be clear in the sequel which scheme is being discussed.

We notice that these equations are coupled but linear, and that the starting condition for the Crank-Nicolson method is easy to implement. We also note that for the Crank-Nicolson method only the $\bar{\sigma}_i^h$ and not the σ_i^h are needed. A variant Crank-Nicolson method could easily be constructed whereby a stationary problem is solved for σ_0^h at $t = 0$ (using $\breve{u}$) and then the subsequent σ_1^h σ_2^h, ... are solved for.

The first goal is to derive stability estimates.

Proposition 2 (**basic discrete stability**). *If at least one of the following conditions holds,*

$$\textit{(A)}\ \boldsymbol{\nu} = \mathbf{0}; \qquad \textit{(B)}\ \|\boldsymbol{\nu}\|_{\mathbb{E}} < \frac{2}{E}\sqrt{\frac{\tilde{\gamma}}{\hat{\gamma}}}; \qquad \textit{(C)}\ \Gamma_u \cap \Gamma_\sigma = \varnothing; \qquad \textit{(D)}\ \|\boldsymbol{\nu}\|_{\mathbb{E}} < \frac{2}{E}\sqrt{\frac{2\lambda}{\hat{\gamma}}},$$

then there are constants $\hat{k} > 0$ and $C > 0$ such that for $k < \hat{k}$,

$$\|u_j^h\|_0^2 + k\sum_{i=1}^{j}\Bigg(\|\nabla u_i^h\|_0^2 + \|u_i^h\|_{0,\Gamma_u}^2 + \|\gamma(u_{i-1}^h)^{-1/2}\nabla\sigma_i^h\|_0^2 + \|\gamma(u_{i-1}^h)^{1/2}\sigma_i^h\|_0^2 + \|\sigma_i^h\|_{0,\Gamma_\sigma}^2\Bigg) \leqslant C\|\breve{u}\|_0^2 + Ck\sum_{i=1}^{j}\|u_i^\flat\|_{0,\Gamma_u}^2$$

for the implicit Euler method and,

$$\|u_j^h\|_0^2 + k\sum_{i=1}^{j}\Bigg(\|\nabla \bar{u}_i^h\|_0^2 + \|\bar{u}_i^h\|_{0,\Gamma_u}^2 + \|\gamma(\mathscr{E}_i u^h)^{-1/2}\nabla\bar{\sigma}_i^h\|_0^2 + \|\gamma(\mathscr{E}_i u^h)^{1/2}\bar{\sigma}_i^h\|_0^2 + \|\bar{\sigma}_i^h\|_{0,\Gamma_\sigma}^2\Bigg) \leqslant C\|\breve{u}\|_0^2 + Ck\sum_{i=1}^{j}\|\bar{u}_i^\flat\|_{0,\Gamma_u}^2$$

for the Crank-Nicolson method. Each of these holds for every $j \in \{1, 2, \ldots, N\}$.

Proof. Choose $v = 2ku_i^h \in V_u^h$ in (21) and, for some $\mu > 0$ to be specified later, $w = \mu k\sigma_i^h \in V_\sigma^h$ in (22) and add to get,

$$\begin{aligned} k\partial_t\|u_i^h\|_0^2 + k^2\|\partial_t u_i^h\|_0^2 + 2k\|\nabla u_i^h\|_0^2 + 2k\lambda\|u_i^h\|_{0,\Gamma_u}^2 \\ + \mu k\|\gamma(u_{i-1}^h)^{-1/2}\nabla\sigma_i^h\|_0^2 + \mu k\|\gamma(u_{i-1}^h)^{1/2}\sigma_i^h\|_0^2 + \mu k\|\sigma_i^h\|_{0,\Gamma_\sigma}^2 \\ = 2k\lambda(u_i^\flat, u_i^h)_{\Gamma_u} + \mu k(u_i^h, \sigma_i^h) - 2kE(\nabla\sigma_i^h, \nabla u_i^h) - \mu k(\boldsymbol{\nu}\cdot\nabla u_i^h, \sigma_i^h), \end{aligned}$$

where we used the identity $2k(\partial_t w_i, w_i) = k\partial_t\|w_i\|_0^2 + k^2\|\partial_t w_i\|_0^2$. Summing over $i = 1, 2, \ldots, j$ then gives,

$$\begin{aligned}\|u_j^h\|_0^2 + 2k\sum_{i=1}^{j}\|\nabla u_i^h\|_0^2 + 2k\lambda\sum_{i=1}^{j}\|u_i^h\|_{0,\Gamma_u}^2 + k^2\sum_{i=1}^{j}\|\partial_t u_i^h\|_0^2 \\ + \mu k\sum_{i=1}^{j}\|\gamma(u_{i-1}^h)^{-1/2}\nabla\sigma_i^h\|_0^2 + \mu k\sum_{i=1}^{j}\|\gamma(u_{i-1}^h)^{1/2}\sigma_i^h\|_0^2 + \mu k\sum_{i=1}^{j}\|\sigma_i^h\|_{0,\Gamma_\sigma}^2 \\ = \|u_0^h\|_0^2 + 2k\lambda\sum_{i=1}^{j}(u_i^\flat, u_i^h)_{\Gamma_u} + \mu k\sum_{i=1}^{j}(u_i^h, \sigma_i^h) \\ - 2kE\sum_{i=1}^{j}(\nabla\sigma_i^h, \nabla u_i^h) - \mu k\sum_{i=1}^{j}(\boldsymbol{\nu}\cdot\nabla u_i^h, \sigma_i^h).\end{aligned}$$

Labelling the terms on the right as $0,\ 1, \ldots, 4$ we apply Young's inequality to each with, when necessary, an ϵ subscipted with the term's label and also note that $\|u_0^h\|_0 \leqslant \|\breve{u}\|_0$ to obtain,

$$\begin{aligned}\left(1 - \frac{k\mu}{2\breve{\gamma}\epsilon_2}\right)\|u_j^h\|_0^2 + \left(2 - \frac{\epsilon_3 E^2\hat{\gamma}}{\mu}\right)\sum_{i=1}^{j}k\|\nabla u_i^h\|_0^2 + \left(2 - \frac{1}{\epsilon_1}\right)\lambda\sum_{i=1}^{j}k\|u_i^h\|_{0,\Gamma_u}^2 \\ + k^2\sum_{i=1}^{j}\|\partial_t u_i^h\|_0^2 + \left(1 - \frac{1}{\epsilon_3}\right)\mu\sum_{i=1}^{j}k\|\gamma(u_{i-1}^h)^{-1/2}\nabla\sigma_i^h\|_0^2 \\ + \left(1 - \frac{\epsilon_2}{2}\right)\mu\sum_{i=1}^{j}k\|\gamma(u_{i-1}^h)^{1/2}\sigma_i^h\|_0^2 + \mu\sum_{i=1}^{j}k\|\sigma_i^h\|_{0,\Gamma_\sigma}^2 \\ \leqslant \|\breve{u}\|_0^2 + \epsilon_1\lambda\sum_{i=1}^{j}k\|u_i^\flat\|_{0,\Gamma_u}^2 + \frac{\mu}{2\breve{\gamma}\epsilon_2}\sum_{i=1}^{j-1}k\|u_i^h\|_0^2 + \mu\sum_{i=1}^{j}|k(\boldsymbol{\nu}\cdot\nabla u_i^h, \sigma_i^h)|.\end{aligned}$$

The remainder of the proof consists in using Young's inequality on the last term and tailoring the estimates to conditions (A), (B), (C) or (D). The method is similar to that in Prop. 1, and identical (except that $\hat{k}$ is different) to this proof for the Crank-Nicolson method (see (28) below), full details of which now follow.

First we note that if we can derive an inequality of the form,

$$\begin{aligned}(27)\quad k\sum_{i=1}^{j}\Big(\|\nabla\bar{u}_i^h\|_0^2 + \|\bar{u}_i^h\|_{0,\Gamma_u}^2 + \|\gamma(\mathscr{E}_i u^h)^{-1/2}\nabla\bar{\sigma}_i^h\|_0^2 + \|\gamma(\mathscr{E}_i u^h)^{1/2}\bar{\sigma}_i^h\|_0^2 \\ + \|\bar{\sigma}_i^h\|_{0,\Gamma_\sigma}^2\Big) + \|u_j^h\|_0^2 \leqslant C\|u_0^h\|_0^2 + Ck\sum_{i=1}^{j}\|\bar{u}_i^\flat\|_{0,\Gamma_u}^2 + Ck\sum_{i=0}^{j-1}\|u_i^h\|_0^2\end{aligned}$$

then the result follows from (26) and the discrete Grönwall lemma.

Towards this end we choose $v = 2\bar{u}_i^h \in V_u^h$ in (23) and $w = \mu\bar{\sigma}_i^h \in V_\sigma^h$ in (24), for some $\mu > 0$ to be specified later, and add the results to get,

$$\begin{aligned}\partial_t\|u_i^h\|_0^2 + 2\|\nabla\bar{u}_i^h\|_0^2 + 2\lambda\|\bar{u}_i^h\|_{0,\Gamma_u}^2 + \mu\|\gamma(\mathscr{E}_i u^h)^{-1/2}\nabla\bar{\sigma}_i^h\|_0^2 + \mu\|\gamma(\mathscr{E}_i u^h)^{1/2}\bar{\sigma}_i^h\|_0^2 \\ + \mu\|\bar{\sigma}_i^h\|_{0,\Gamma_\sigma}^2 = 2\lambda(\bar{u}_i^\flat, \bar{u}_i^h)_{\Gamma_u} + \mu(\bar{u}_i^h, \bar{\sigma}_i^h) - 2E(\nabla\bar{\sigma}_i^h, \nabla\bar{u}_i^h) - \mu(\boldsymbol{\nu}\cdot\nabla\bar{u}_i^h, \bar{\sigma}_i^h).\end{aligned}$$

We number the terms on the right as one, two, three and four and apply Young's inequality, (17), to the first three with ϵ subscripted by the number of the term.

Multiplying the result by k, summing over $i = 1, 2, \ldots, j$ and noting that,

$$\frac{\mu}{2\check{\gamma}\epsilon_2}\sum_{i=1}^{j} k\|\bar{u}_i^h\|_0^2 \leqslant \frac{\mu k}{4\check{\gamma}\epsilon_2}\|u_j^h\|_0^2 + \frac{\mu}{2\check{\gamma}\epsilon_2}\sum_{i=0}^{j-1} k\|u_i^h\|_0^2$$

then results in,

$$\begin{aligned}(28)\quad &\left(1-\frac{k\mu}{4\check{\gamma}\epsilon_2}\right)\|u_j^h\|_0^2 + \left(2-\frac{\epsilon_3 E^2\hat{\gamma}}{\mu}\right)\sum_{i=1}^{j} k\|\nabla\bar{u}_i^h\|_0^2 \\ &+\left(2-\frac{1}{\epsilon_1}\right)\lambda\sum_{i=1}^{j} k\|\bar{u}_i^h\|_{0,\Gamma_u}^2 + \left(1-\frac{1}{\epsilon_3}\right)\mu\sum_{i=1}^{j} k\|\gamma(\mathscr{E}_i u^h)^{-1/2}\nabla\bar{\sigma}_i^h\|_0^2 \\ &+\left(1-\frac{\epsilon_2}{2}\right)\mu\sum_{i=1}^{j} k\|\gamma(\mathscr{E}_i u^h)^{1/2}\bar{\sigma}_i^h\|_0^2 + \mu\sum_{i=1}^{j} k\|\bar{\sigma}_i^h\|_{0,\Gamma_\sigma}^2 \\ &\leqslant \|u_0^h\|_0^2 + \epsilon_1\lambda\sum_{i=1}^{j} k\|\bar{u}_i^\flat\|_{0,\Gamma_u}^2 + \frac{\mu}{2\check{\gamma}\epsilon_2}\sum_{i=0}^{j-1} k\|u_i^h\|_0^2 + \sum_{i=1}^{j} k|\mu(\boldsymbol{\nu}\cdot\nabla\bar{u}_i^h, \bar{\sigma}_i^h)|,\end{aligned}$$

and the proof for conditions (A) and (B) follows in exactly the same way as for the proof of Prop. 1 except that we now also have to insist that $k < \hat{k} := 4\check{\gamma}\epsilon_2/\mu$.

For conditions (C) and (D) we use (18) and Young's inequality to get,

$$\begin{aligned}\mu|(\boldsymbol{\nu}\cdot\nabla\bar{u}_i^h, \bar{\sigma}_i^h)| \leqslant \frac{\delta\mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{2\epsilon_4}\|\bar{u}_i^h\|_{0,\Gamma_u}^2 + \frac{\delta\epsilon_4\mu}{2}\|\bar{\sigma}_i^h\|_{0,\Gamma_\sigma}^2 \\ + \frac{\mu\hat{\gamma}\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{2}\|\bar{u}_i^h\|_0^2 + \frac{\mu}{2}\|\gamma(\mathscr{E}_i u^h)^{-1/2}\nabla\bar{\sigma}_i^h\|_0^2,\end{aligned}$$

where we understand that $\delta = 0$ for condition (C) and $\delta = 1$ for condition (D). Noting that,

$$\frac{\mu\hat{\gamma}\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{2}\sum_{i=1}^{j} k\|\bar{u}_i^h\|_0^2 \leqslant \frac{\mu\hat{\gamma}\|\boldsymbol{\nu}\|_{\mathbb{E}}^2 k}{4}\|u_j^h\|_0^2 + \frac{\mu\hat{\gamma}\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{2}\sum_{i=0}^{j-1} k\|u_i^h\|_0^2,$$

we incorporate these into (28) and get,

$$\begin{aligned}(29)\quad &\left(1-\frac{k\mu}{4\check{\gamma}\epsilon_2}-\frac{\mu\hat{\gamma}\|\boldsymbol{\nu}\|_{\mathbb{E}}^2 k}{4}\right)\|u_j^h\|_0^2 + \left(2-\frac{\epsilon_3 E^2\hat{\gamma}}{\mu}\right)\sum_{i=1}^{j} k\|\nabla\bar{u}_i^h\|_0^2 \\ &+\left(2\lambda-\frac{\lambda}{\epsilon_1}-\frac{\delta\mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{2\epsilon_4}\right)\sum_{i=1}^{j} k\|\bar{u}_i^h\|_{0,\Gamma_u}^2 + \left(\frac{1}{2}-\frac{1}{\epsilon_3}\right)\mu\sum_{i=1}^{j} k\|\gamma(\mathscr{E}_i u^h)^{-1/2}\nabla\bar{\sigma}_i^h\|_0^2 \\ &+\left(1-\frac{\epsilon_2}{2}\right)\mu\sum_{i=1}^{j} k\|\gamma(\mathscr{E}_i u^h)^{1/2}\bar{\sigma}_i^h\|_0^2 + \left(1-\frac{\delta\epsilon_4}{2}\right)\mu\sum_{i=1}^{j} k\|\bar{\sigma}_i^h\|_{0,\Gamma_\sigma}^2 \leqslant \|u_0^h\|_0^2 \\ &+\epsilon_1\lambda\sum_{i=1}^{j} k\|\bar{u}_i^\flat\|_{0,\Gamma_u}^2 + \left(\frac{\mu}{2\check{\gamma}\epsilon_2}+\frac{\mu\hat{\gamma}\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{2}\right)\sum_{i=0}^{j-1} k\|u_i^h\|_0^2.\end{aligned}$$

For condition (C) we have $\delta = 0$ in (29) and we can choose ϵ_1, ϵ_2 and ϵ_3 the same as in the proof of Prop. 1. Noting that

$$1-\frac{k\mu}{4\check{\gamma}\epsilon_2}-\frac{\mu\hat{\gamma}\|\boldsymbol{\nu}\|_{\mathbb{E}}^2 k}{4} > 0 \qquad \text{is guaranteed if} \qquad k < \frac{4\check{\gamma}}{(1+\hat{\gamma}\check{\gamma}\|\boldsymbol{\nu}\|_{\mathbb{E}}^2)E^2\hat{\gamma}},$$

we can obtain a value for $\hat{k}$ and we once again find ourselves at (27).

For condition (D) we use exactly the same choices as in the proof for condition (D) of Prop. 1 and require k bounded in the same way as for condition (C) above. This again gets us back to (27) and completes the proof for all four conditions for the Crank-Nicolson method. The proof for the implicit Euler method makes essentially the same choices in Young's inequality, but uses a different $\hat{k}$. □

Prop. 2 also provides a uniqueness result for the discrete solution and so (due to the 'linearised nonlinearities') we can infer existence as well.

Corollary 3. *The discrete solution exists and is unique.*

4. Error estimates

We begin by recalling some standard results. From, for example, [3, Thm. (4.8.7), Cor. (4.8.9)], we have with $V = V_u$ or $V = V_\sigma$ that if $v \in V \cap W_p^s(\Omega)$, for $0 \leqslant s \leqslant r+1$ and $1 \leqslant p \leqslant \infty$, then there exists a map $\pi\colon W_p^s(\Omega) \to V^h$, for $V^h = V_u^h$ or $V^h = V_\sigma^h$, such that,

$$\|v - \pi v\|_{W_p^n(\Omega)} \leqslant Ch^{s-n}|v|_{W_p^s(\Omega)}, \qquad \text{for } 0 \leqslant n \leqslant s, \tag{30}$$

$$\|\pi v\|_{W_p^s(\Omega)} \leqslant C|v|_{W_p^s(\Omega)}, \tag{31}$$

where $|\cdot|_{W_p^s(\Omega)}$ denotes the semi-norm.

We define the elliptic projection, see [13], of u as $u^* \in V_u^h$ where,

$$(\nabla(u^* - u), \nabla v) + \lambda(u^* - u, v)_{\Gamma_u} = 0 \qquad \forall v \in V_u^h, \tag{32}$$

and define also $\sigma^* := \pi\sigma \in V_\sigma^h$. Setting:

$$\begin{aligned} \psi_i &:= u_i^h - u^*(t_i), & \xi(t) &:= u(t) - u^*(t), \\ \zeta_i &:= \sigma_i^h - \sigma^*(t_i), & \vartheta(t) &:= \sigma(t) - \sigma^*(t), \end{aligned}$$

we have $u_i^h - u(t_i) = \psi_i - \xi(t_i)$ and $\sigma_i^h - \sigma(t_i) = \zeta_i - \vartheta(t_i)$, and it follows by the approximation estimates above and standard techniques that,

$$\|\nabla\xi(t)\|_0^2 + \lambda\|\xi(t)\|_{0,\Gamma_u}^2 \leqslant Ch^{2r}\|u(t)\|_{r+1}^2 \tag{33}$$

$$\|\nabla\xi_t(t)\|_0^2 + \lambda\|\xi_t(t)\|_{0,\Gamma_u}^2 \leqslant Ch^{2r}\|u_t(t)\|_{r+1}^2 \tag{34}$$

(where we noted that (32) can be partially differentiated with respect to t).

The main goal is to estimate ψ_i and ζ_i in terms of ξ and ϑ. For this we need the following estimates.

Lemma 4. *For $p = 1$ or $p = 2$ we have for each $i \in \{1, 2, \ldots\}$ that,*

$$\|\partial_t\xi_i\|_0^2 \leqslant k^{p-3}\|\xi_t\|_{L_p(t_{i-1},t_i;L_2(\Omega))}^2 \leqslant Ch^{2r}k^{p-3}\|u_t\|_{L_p(t_{i-1},t_i;H^{r+1}(\Omega))}^2$$

whenever $u_t \in L_p(t_{i-1}, t_i; H^{r+1}(\Omega))$.

Proof. After noting first that,

$$\|\partial_t\xi_i\|_0^2 \leqslant \left(\frac{1}{k}\int_{t_{i-1}}^{t_i} \|\xi_s(s)\|_0\, ds\right)^2 \leqslant \frac{1}{k}\int_{t_{i-1}}^{t_i} \|\xi_s(s)\|_0^2\, ds$$

the results follow from (30), (33) and (34). □

Lemma 5. *Whenever v has the indicated regularity we have,*

$$\begin{aligned} \|v_t(t_i) - \partial_t v_i\|_0 &\leqslant \|v_{tt}\|_{L_1(t_{i-1},t_i;L_2(\Omega))}, \\ \|\Delta_i v\|_0 &\leqslant Ck^{3/2}\|v_{ttt}\|_{L_2(t_{i-1},t_i;L_2(\Omega))}, \\ \|v_{i-1/2} - \bar{v}_i\|_0 &\leqslant Ck^{3/2}\|v_{tt}\|_{L_2(t_{i-1},t_i;L_2(\Omega))}, \\ \|v_{i-1/2} - \mathscr{E}_i v\|_0 &\leqslant Ck^{3/2}\|v_{tt}\|_{L_2(t_{i-2},t_{i-1/2};L_2(\Omega))}, \end{aligned}$$

for $i \in \{1, 2, \ldots\}$ except for the last where $i = 1$ is disallowed.

Proof. These follow from Taylor's theorem with integral remainder. □

The proof of the error estimates will follow in much the same way as the proof of the stability estimates given earlier except that there are more terms to deal with.

4.1. Error bound for the implicit Euler method. Our goal in this section is an *a priori* error bound for the implicit Euler method. Since this is used as a starting solution for the Crank-Nicolson method we need to be careful in tracking the u and σ dependencies of the 'constants'. Indeed, Corollary 9 relies on being able to modify these constants and is also the key to obtaining an optimal order of k for the Crank-Nicolson method.

We begin by estimating the error in the nonlinear terms.

Lemma 6 ('nonlinearity error')**.** *There is a constant, $C > 0$, independent of u, σ, h and k such that,*

$$\begin{aligned}\big|(\gamma(u_i)^{-1}\nabla\sigma_i - \gamma(u_{i-1}^h)^{-1}\nabla\sigma_i^*, \nabla\zeta_i) + (\gamma(u_i)\sigma_i - \gamma(u_{i-1}^h)\sigma_i^*, \zeta_i)\big| \\ \leqslant \frac{1}{2\epsilon_A}\|\gamma(u_{i-1}^h)^{-1/2}\nabla\zeta_i\|_0^2 + \frac{1}{2\epsilon_B}\|\gamma(u_{i-1}^h)^{1/2}\zeta_i\|_0^2 \\ + (\epsilon_A + \epsilon_B)Ck\|\sigma\|^2_{L_\infty(t_{i-1},t_i;W^1_\infty(\Omega))}\|u_t\|^2_{L_2(t_{i-1},t_i;L_2(\Omega))} + (\epsilon_A + \epsilon_B)C\|\vartheta_i\|_1^2 \\ + (\epsilon_A + \epsilon_B)C\|\sigma\|^2_{L_\infty(t_{i-1},t_i;W^1_\infty(\Omega))}\|\xi_{i-1}\|_0^2 \\ + (\epsilon_A + \epsilon_B)C\|\sigma\|^2_{L_\infty(t_{i-1},t_i;W^1_\infty(\Omega))}\|\psi_{i-1}\|_0^2,\end{aligned}$$

for all $i \in \{1, 2, \ldots, N\}$ and for all $\epsilon_A,\ \epsilon_B > 0$.

Proof. We have by the Cauchy-Schwarz inequality that,

$$\begin{aligned}\big|(\gamma(u_i)^{-1}\nabla\sigma_i - \gamma(u_{i-1}^h)^{-1}\nabla\sigma_i^*, \nabla\zeta_i) + (\gamma(u_i)\sigma_i - \gamma(u_{i-1}^h)\sigma_i^*, \zeta_i)\big| \\ \leqslant \hat\gamma^{1/2}\|\gamma(u_{i-1}^h)^{-1/2}\nabla\zeta_i\|_0\ \|\gamma(u_i)^{-1}\nabla\sigma_i - \gamma(u_{i-1}^h)^{-1}\nabla\sigma_i^*\|_0 \\ + \check\gamma^{-1/2}\|\gamma(u_{i-1}^h)^{1/2}\zeta_i\|_0\ \|\gamma(u_i)\sigma_i - \gamma(u_{i-1}^h)\sigma_i^*\|_0.\end{aligned}$$

Dealing first with the last term on the right we have,

$$\begin{aligned}\|\gamma(u_i)\sigma_i - \gamma(u_{i-1}^h)\sigma_i^*\|_0 &\leqslant \|(\gamma(u_i) - \gamma(u_{i-1}))\sigma_i\|_0 \\ &\quad + \|(\gamma(u_{i-1}) - \gamma(u_{i-1}^h))\sigma_i\|_0 + \|\gamma(u_{i-1}^h)(\sigma_i - \sigma_i^*)\|_0, \\ &\leqslant C'_\gamma k^{1/2}\|\sigma_i\|_{L_\infty(\Omega)}\|u_t\|_{L_2(t_{i-1},t_i;L_2(\Omega))} + \hat\gamma\|\vartheta_i\|_0 \\ &\quad + C'_\gamma\|\sigma_i\|_{L_\infty(\Omega)}\|\xi_{i-1}\|_0 + C'_\gamma\|\sigma_i\|_{L_\infty(\Omega)}\|\psi_{i-1}\|_0,\end{aligned}$$

where we noted that $\|u_i - u_{i-1}\|_0 \leqslant k^{1/2}\|u_t\|_{L_2(t_{i-1},t_i;L_2(\Omega))}$.

For the first term on the right the procedure is similar but we begin by first removing the denominators,

$$\|\gamma(u_i)^{-1}\nabla\sigma_i - \gamma(u_{i-1}^h)^{-1}\nabla\sigma_i^*\|_0 \leqslant \check\gamma^{-2}\|\gamma(u_{i-1}^h)\nabla\sigma_i - \gamma(u_i)\nabla\sigma_i^*\|_0.$$

Then, almost as before,

$$\begin{aligned}\|\gamma(u_{i-1}^h)\nabla\sigma_i - \gamma(u_i)\nabla\sigma_i^*\|_0 &\leqslant \|(\gamma(u_{i-1}^h) - \gamma(u_{i-1}))\nabla\sigma_i\|_0 \\ &\quad + \|(\gamma(u_{i-1}) - \gamma(u_i))\nabla\sigma_i\|_0 + \|\gamma(u_i)(\nabla\sigma_i - \nabla\sigma_i^*)\|_0 \\ &\leqslant C'_\gamma k^{1/2}\|\nabla\sigma_i\|_{L_\infty(\Omega)}\|u_t\|_{L_2(t_{i-1},t_i;L_2(\Omega))} + \hat\gamma\|\nabla\vartheta_i\|_0 \\ &\quad + C'_\gamma\|\nabla\sigma_i\|_{L_\infty(\Omega)}\|\xi_{i-1}\|_0 + C'_\gamma\|\nabla\sigma_i\|_{L_\infty(\Omega)}\|\psi_{i-1}\|_0.\end{aligned}$$

The proof is then completed by merging these and then using Young's inequality along with obvious estimates. □

The next lemma deals with most of the technical details in deriving the error bound.

Lemma 7. *For $j = 1, 2, \ldots, N$ we have for the implicit Euler method that,*

$$\begin{aligned}
&\max_{1\leqslant i\leqslant j}\|\psi_i\|_0^2 + 2k\sum_{i=1}^{j}\Big(\|\nabla\psi_i\|_0^2 + \lambda\|\psi_i\|_{0,\Gamma_u}^2 + k\|\partial_t\psi_i\|_0^2\Big)\\
&+\mu k\sum_{i=1}^{j}\Big(\|\gamma(u_{i-1}^h)^{-1/2}\nabla\zeta_i\|_0^2 + \|\gamma(u_{i-1}^h)^{1/2}\zeta_i\|_0^2 + \|\zeta_i\|_{0,\Gamma_\sigma}^2\Big)\\
&\leqslant C_{1,j}h^{2r} + C_{2,j}k^2 + Ck\|\sigma\|_{L_\infty(0,t_j;W_\infty^1(\Omega))}^2\sum_{i=0}^{j-1}\|\psi_i\|_0^2\\
&+4kE\sum_{i=1}^{j}|(\nabla\zeta_i,\nabla\psi_i)| + 2\mu k\sum_{i=1}^{j}|(\psi_i,\zeta_i)| + 2\mu k\sum_{i=1}^{j}|(\boldsymbol{\nu}\cdot\nabla\psi_i,\zeta_i)|,
\end{aligned}$$

where,

$$\begin{aligned}
C_{1,j} &\leqslant C\Big(\|\breve{u}\|_{r+1}^2 + \|u_t\|_{L_1(0,t_j;H^{r+1}(\Omega))}^2 + t_j\|\sigma\|_{L_\infty(0,t_j;H^{r+1}(\Omega))}^2\\
&\qquad + \Big(\|\sigma\|_{L_\infty(0,t_j;W_\infty^1(\Omega))}^2 + 1\Big)t_j\|u\|_{L_\infty(0,t_j;H^{r+1}(\Omega))}^2\Big)\\
C_{2,j} &\leqslant C\Big(\|\sigma\|_{L_\infty(0,t_j;W_\infty^1(\Omega))}^2\ \|u_t\|_{L_2(0,t_j;L_2(\Omega))}^2 + \|u_{tt}\|_{L_1(0,t_j;L_2(\Omega))}^2\Big)
\end{aligned}$$

for constants $\mu, C > 0$ both independent of u, σ, h, k and j, and where μ is arbitrary.

Proof. From (13), (14), (21) and (22) we get,

$$\begin{aligned}
&(\partial_t\psi_i, v) + (\nabla\psi_i,\nabla v) + \lambda(\psi_i, v)_{\Gamma_u} + (\zeta_i, w)_{\Gamma_\sigma} + (\gamma(u_{i-1}^h)^{-1}\nabla\zeta_i,\nabla w)\\
&\quad + (\gamma(u_{i-1}^h)\zeta_i, w) = (\dot{u}_i - \partial_t u_i, v) + (\partial_t\xi_i, v) + (\nabla\xi_i, v) + E(\nabla\vartheta_i,\nabla v)\\
&\qquad + \lambda(\xi_i, v)_{\Gamma_u} + (\vartheta_i, w)_{\Gamma_\sigma} - (\xi_i, w) + (\boldsymbol{\nu}\cdot\nabla\xi_i, w)\\
&\qquad - E(\nabla\zeta_i,\nabla v) + (\psi_i, w) - (\boldsymbol{\nu}\cdot\nabla\psi_i, w)\\
&\qquad + (\gamma(u_i)^{-1}\nabla\sigma_i - \gamma(u_{i-1}^h)^{-1}\nabla\sigma_i^*,\nabla w) + (\gamma(u_i)\sigma_i - \gamma(u_{i-1}^h)\sigma_i^*, w).
\end{aligned}$$

Choosing $v = 2k\psi_i \in V_u^h$ and, for some $\mu > 0$, $w = \mu k\zeta_i \in V_\sigma^h$ we again use the identity $2k(\partial_t\psi_i,\psi_i) = k\partial_t\|\psi_i\|_0^2 + k^2\|\partial_t\psi_i\|_0^2$ and obtain,

$$\begin{aligned}
&k\partial_t\|\psi_i\|_0^2 + k^2\|\partial_t\psi_i\|_0^2 + 2k\|\nabla\psi_i\|_0^2 + 2\lambda k\|\psi_i\|_{0,\Gamma_u}^2 + \mu k\|\zeta_i\|_{0,\Gamma_\sigma}^2\\
&+\mu k\|\gamma(u_{i-1}^h)^{-1/2}\nabla\zeta_i\|_0^2 + \mu k\|\gamma(u_{i-1}^h)^{1/2}\zeta_i\|_0^2 = 2k(\dot{u}_i - \partial_t u_i,\psi_i) + 2k(\partial_t\xi_i,\psi_i)\\
&\quad + 2k(\nabla\xi_i,\nabla\psi_i) + 2kE(\nabla\vartheta_i,\nabla\psi_i) + 2k\lambda(\xi_i,\psi_i)_{\Gamma_u} + \mu k(\vartheta_i,\zeta_i)_{\Gamma_\sigma} - \mu k(\xi_i,\zeta_i)\\
&\qquad + \mu k(\boldsymbol{\nu}\cdot\nabla\xi_i,\zeta_i) - 2kE(\nabla\zeta_i,\nabla\psi_i) + \mu k(\psi_i,\zeta_i) - \mu k(\boldsymbol{\nu}\cdot\nabla\psi_i,\zeta_i)\\
&\qquad + \mu k(\gamma(u_i)^{-1}\nabla\sigma_i - \gamma(u_{i-1}^h)^{-1}\nabla\sigma_i^*,\nabla\zeta_i) + \mu k(\gamma(u_i)\sigma_i - \gamma(u_{i-1}^h)\sigma_i^*,\zeta_i).
\end{aligned}$$

Now: eliminate the third and fifth terms on the right using (32); sum over $i = 1, 2, \ldots, j$, noting from (26) that $\|\psi_0\|_0 \leqslant \|\xi_0\|_0$; number the resulting terms (with

the one just referred to as first) on the right as I, II, ... and use the following estimates. For I, II and III, using (33), (34) with Lemmas 4 and 5 we have,

$$|I+II+III| \leqslant Ch^{2r}\|\breve{u}\|_{r+1}^2 + \epsilon_3 Ch^{2r}\|u_t\|_{L_1(0,t_j;H^{r+1}(\Omega))}^2 + \epsilon_2 k^2\|u_{tt}\|_{L_1(0,t_j;L_2(\Omega))}^2 + \left(\frac{1}{\epsilon_2}+\frac{1}{\epsilon_3}\right)\max_{1\leqslant i\leqslant j}\|\psi_i\|_0^2,$$

and for IV and V,

$$|IV+V| \leqslant \frac{2k}{\epsilon_4}\sum_{i=1}^{j}\|\nabla\psi_i\|_0^2 + \frac{\mu k}{2\epsilon_5}\sum_{i=1}^{j}\|\zeta_i\|_{0,\Gamma_\sigma}^2 + (\epsilon_4+\epsilon_5)Ct_j h^{2r}\|\sigma\|_{L_\infty(0,t_j;H^{r+1}(\Omega))}^2.$$

For VI and VII we have,

$$|VI+VII| \leqslant \left(\frac{\mu k}{2\epsilon_6}+\frac{\mu k}{2\epsilon_7}\right)\sum_{i=1}^{j}\|\gamma(u_{i-1}^h)^{1/2}\zeta_i\|_0^2 + (\epsilon_6+\epsilon_7)Ct_j h^{2r}\|u\|_{L_\infty(0,t_j;H^{r+1}(\Omega))}^2,$$

and so with these, Lemma 6 and the fact that the right-hand side that results is non-decreasing, we arrive at,

$$\begin{aligned}
&\left(1-\frac{1}{\epsilon_2}-\frac{1}{\epsilon_3}\right)\left(\max_{1\leqslant i\leqslant j}\|\psi_j\|_0^2\right) + 2\left(1-\frac{1}{\epsilon_4}\right)k\sum_{i=1}^{j}\|\nabla\psi_i\|_0^2 + 2\lambda k\sum_{i=1}^{j}\|\psi_i\|_{0,\Gamma_u}^2 \\
&\quad + \left(1-\frac{1}{2\epsilon_5}\right)\mu k\sum_{i=1}^{j}\|\zeta_i\|_{0,\Gamma_\sigma}^2 + \left(1-\frac{1}{2\epsilon_A}\right)\mu k\sum_{i=1}^{j}\|\gamma(u_{i-1}^h)^{-1/2}\nabla\zeta_i\|_0^2 \\
&\quad + \left(1-\frac{1}{2\epsilon_6}-\frac{1}{2\epsilon_7}-\frac{1}{2\epsilon_B}\right)\mu k\sum_{i=1}^{j}\|\gamma(u_{i-1}^h)^{1/2}\zeta_i\|_0^2 + \sum_{i=1}^{j}k^2\|\partial_t\psi_i\|_0^2 \\
&\leqslant Ch^{2r}\Big(\|\breve{u}\|_{r+1}^2 + (\epsilon_6+\epsilon_7)t_j\|u\|_{L_\infty(0,t_j;H^{r+1}(\Omega))}^2 + \epsilon_3\|u_t\|_{L_1(0,t_j;H^{r+1}(\Omega))}^2 \\
&\quad + (\epsilon_4+\epsilon_5)t_j\|\sigma\|_{L_\infty(0,t_j;H^{r+1}(\Omega))}^2 + (\epsilon_A+\epsilon_B)t_j\|\sigma\|_{L_\infty(0,t_j;H^{r+1}(\Omega))}^2 \\
&\quad + (\epsilon_A+\epsilon_B)t_j\|\sigma\|_{L_\infty(0,t_j;W_\infty^1(\Omega))}^2\|u\|_{L_\infty(0,t_j;H^{r+1}(\Omega))}^2\Big) \\
&\quad + k^2\Big(\epsilon_2\|u_{tt}\|_{L_1(0,t_j;L_2(\Omega)}^2 + (\epsilon_A+\epsilon_B)C\|\sigma\|_{L_\infty(0,t_j;W_\infty^1(\Omega))}^2\ \|u_t\|_{L_2(0,t_j;L_2(\Omega))}^2\Big) \\
&\quad + (\epsilon_A+\epsilon_B)Ck\|\sigma\|_{L_\infty(0,t_j;W_\infty^1(\Omega))}^2\sum_{i=0}^{j-1}\|\psi_i\|_0^2 \\
&\quad + 2kE\sum_{i=1}^{j}|(\nabla\zeta_i,\nabla\psi_i)| + \mu k\sum_{i=1}^{j}|(\psi_i,\zeta_i)| + \mu k\sum_{i=1}^{j}|(\boldsymbol{\nu}\cdot\nabla\psi_i,\zeta_i)|.
\end{aligned}$$

To complete the proof we choose $\epsilon_2=\epsilon_3=4$, $\epsilon_4=2$, $\epsilon_5=\epsilon_A=1$ and $\epsilon_6=\epsilon_7=\epsilon_B=3$, and then multiply the resulting inequality by two. □

We can now state the error estimate.

Theorem 8 (error bound: implicit Euler)**.** *Assume that in* (6) *we have* $\breve{u}\in H^{r+1}(\Omega)$ *and also that* $u\in V_u\cap W_1^1(I;H^{r+1}(\Omega))\cap W_1^2(I;L_2(\Omega))$ *and* $\sigma\in V_\sigma\cap L_\infty(I;H^{r+1}(\Omega)\cap W_\infty^1(\Omega))$ *in* (13) *and* (14) *then, if at least one of the following*

conditions holds,

$$(A)\ \boldsymbol{\nu}=\mathbf{0};\quad (B)\ \|\boldsymbol{\nu}\|_{\mathbb{E}}<\frac{1}{2E}\sqrt{\frac{\check{\gamma}}{\hat{\gamma}}};\quad (C)\ \Gamma_u\cap\Gamma_\sigma=\varnothing;\quad (D)\ \|\boldsymbol{\nu}\|_{\mathbb{E}}<\frac{1}{E}\sqrt{\frac{\lambda}{2\hat{\gamma}}},$$

there is a constant $\hat{k}>0$ *such that whenever* $k<\hat{k}$,

$$\left(k\sum_{i=1}^{j}\left(\frac{1}{\hat{\gamma}}\|\nabla\sigma(t_i)-\nabla\sigma_i^h\|_0^2+\check{\gamma}\|\sigma(t_i)-\sigma_i^h\|_0^2+\|\sigma(t_i)-\sigma_i^h\|_{0,\Gamma_\sigma}^2\right)\right)^{1/2}$$
$$+\left(k\sum_{i=1}^{j}\left(\|\nabla u(t_i)-\nabla u_i^h\|_0^2+\|u(t_i)-u_i^h\|_{0,\Gamma_u}^2\right)\right)^{1/2}$$
$$+\|u(t_j)-u_j^h\|_0\leqslant C_{1,j}^{1/2}h^r+C_{2,j}^{1/2}k.$$

This holds for each $j\in\{1,2,\ldots,N\}$ *and the* $C_{i,j}$ *are, up to a multiplicative constant, those given in Lemma* 7.

Proof. We use Lemma 7 and follow a similar path as for Props. 1 and 2. First, using the estimate,

$$4kE\sum_{i=1}^{j}|(\nabla\zeta_i,\nabla\psi_i)|+2\mu k\sum_{i=1}^{j}|(\zeta_i,\psi_i)|\leqslant\frac{4\epsilon_1\hat{\gamma}E^2k}{\mu}\sum_{i=1}^{j}\|\nabla\psi_i\|_0^2+\frac{2\mu k}{\epsilon_2\check{\gamma}}\sum_{i=1}^{j}\|\psi_i\|_0^2$$
$$+\frac{\mu k}{\epsilon_1}\sum_{i=1}^{j}\|\gamma(u_{i-1}^h)^{-1/2}\nabla\zeta_i\|_0^2+\frac{\epsilon_2\mu k}{2}\sum_{i=1}^{j}\|\gamma(u_{i-1}^h)^{1/2}\zeta_i\|_0^2$$

for all $\epsilon_1,\ \epsilon_2>0$, in Lemma 7 we obtain,

$$\left(1-\frac{2\mu k}{\epsilon_2\check{\gamma}}\right)\|\psi_j\|_0^2+2k\left(1-\frac{2\epsilon_1\hat{\gamma}E^2}{\mu}\right)\sum_{i=1}^{j}\|\nabla\psi_i\|_0^2+2k\lambda\sum_{i=1}^{j}\|\psi_i\|_{0,\Gamma_u}^2$$
$$+\mu k\left(1-\frac{1}{\epsilon_1}\right)\sum_{i=1}^{j}\|\gamma(u_{i-1}^h)^{-1/2}\nabla\zeta_i\|_0^2+\mu k\left(1-\frac{\epsilon_2}{2}\right)\sum_{i=1}^{j}\|\gamma(u_{i-1}^h)^{1/2}\zeta_i\|_0^2$$
$$+\mu k\sum_{i=1}^{j}\|\zeta_i\|_{0,\Gamma_\sigma}^2\leqslant C_{1,j}h^{2r}+C_{2,j}k^2$$
$$+\left(C\|\sigma\|_{L_\infty(0,t_j;W_\infty^1(\Omega))}^2+\frac{2\mu}{\epsilon_2\check{\gamma}}\right)\sum_{i=0}^{j-1}k\|\psi_i\|_0^2+2\mu k\sum_{i=1}^{j}|(\boldsymbol{\nu}\cdot\nabla\psi_i,\zeta_i)|.$$

Now, if condition (A) holds then we choose $\epsilon_1=2$, $\epsilon_2=1$ and any $\mu>4\hat{\gamma}E^2$ to get for some $\hat{k}>0$ that,

$$k\sum_{i=1}^{j}\left(\|\gamma(u_{i-1}^h)^{-1/2}\nabla\zeta_i\|_0^2+\|\gamma(u_{i-1}^h)^{1/2}\zeta_i\|_0^2+\|\zeta_i\|_{0,\Gamma_\sigma}^2\right)$$
$$+\|\psi_j\|_0^2+k\sum_{i=1}^{j}\left(\|\nabla\psi_i\|_0^2+\|\psi_i\|_{0,\Gamma_u}^2\right)\leqslant C_{1,j}h^{2r}+C_{2,j}k^2+Ck\sum_{i=0}^{j-1}\|\psi_i\|_0^2,$$

where the 'generic constant' in the $C_{i,j}$ from Lemma 7 has been adjusted. An application of Grönwall's lemma and a further adjustment of these constants then

produces,

$$k\sum_{i=1}^{j}\left(\|\gamma(u_{i-1}^h)^{-1/2}\nabla\zeta_i\|_0^2+\|\gamma(u_{i-1}^h)^{1/2}\zeta_i\|_0^2+\|\zeta_i\|_{0,\Gamma_\sigma}^2\right) + \|\psi_j\|_0^2+k\sum_{i=1}^{j}\left(\|\nabla\psi_i\|_0^2+\|\psi_i\|_{0,\Gamma_u}^2\right)\leqslant C_{1,j}h^{2r}+C_{2,j}k^2. \tag{35}$$

Now, if $\boldsymbol{\nu}\neq\mathbf{0}$ then we can estimate,

$$2\mu|(\boldsymbol{\nu}\cdot\nabla\psi_i,\zeta_i)|\leqslant\frac{2\mu}{\epsilon_4\check{\gamma}}\|\boldsymbol{\nu}\|_{\mathbb{E}}^2\|\nabla\psi_i\|_0^2+\frac{\epsilon_4\mu}{2}\|\gamma(u_{i-1}^h)^{1/2}\zeta_i\|_0^2,$$

and then get,

$$\left(1-\frac{2\mu k}{\epsilon_2\check{\gamma}}\right)\|\psi_j\|_0^2+2k\left(1-\frac{2\epsilon_1\hat{\gamma}E^2}{\mu}-\frac{\mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{\epsilon_4\check{\gamma}}\right)\sum_{i=1}^{j}\|\nabla\psi_i\|_0^2+2k\lambda\sum_{i=1}^{j}\|\psi_i\|_{0,\Gamma_u}^2$$
$$+\mu k\left(1-\frac{1}{\epsilon_1}\right)\sum_{i=1}^{j}\|\gamma(u_{i-1}^h)^{-1/2}\nabla\zeta_i\|_0^2+\mu k\left(1-\frac{\epsilon_2}{2}-\frac{\epsilon_4}{2}\right)\sum_{i=1}^{j}\|\gamma(u_{i-1}^h)^{1/2}\zeta_i\|_0^2$$
$$+\mu k\sum_{i=1}^{j}\|\zeta_i\|_{0,\Gamma_\sigma}^2\leqslant C_{1,j}h^{2r}+C_{2,j}k^2+\left(C\|\sigma\|_{L_\infty(0,t_j;W_\infty^1(\Omega))}^2+\frac{2\mu}{\epsilon_2\check{\gamma}}\right)\sum_{i=0}^{j-1}k\|\psi_i\|_0^2.$$

If condition (B) holds then we select $\mu=\check{\gamma}\epsilon_4/2\|\boldsymbol{\nu}\|_{\mathbb{E}}^2$ and then we can force,

$$0<1-\frac{2\epsilon_1\hat{\gamma}E^2}{\mu}-\frac{\mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{\epsilon_4\check{\gamma}}=\frac{1}{2}\left(\frac{\check{\gamma}\epsilon_4-8\epsilon_1\hat{\gamma}E^2\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{\check{\gamma}\epsilon_4}\right)$$

with some $\epsilon_1>1$ because condition (B) implies an $\epsilon_4\in(0,2)$ such that $8\hat{\gamma}E^2\|\boldsymbol{\nu}\|_{\mathbb{E}}^2<\epsilon_4\check{\gamma}$. Since we then have $1-1/\epsilon_1>0$ and we can set $\epsilon_2=1-\epsilon_4/2$ we can again determine some $\hat{k}>0$ and arrive at a form of (35) with adjusted constants.

For condition (C) or (D) we begin with,

$$2\mu|(\boldsymbol{\nu}\cdot\nabla\psi_i,\zeta_i)|\leqslant\frac{2\delta\mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{\epsilon_4}\|\psi_i\|_{0,\Gamma_u}^2+\frac{\epsilon_4\delta\mu}{2}\|\zeta_i\|_{0,\Gamma_\sigma}^2 + 2\mu\hat{\gamma}\|\boldsymbol{\nu}\|_{\mathbb{E}}^2\|\psi_i\|_0^2+\frac{\mu}{2}\|\gamma(u_{i-1}^h)^{-1/2}\nabla\zeta_i\|_0^2,$$

where $\delta=0$ for condition (C) and $\delta=1$ for (D). This yields,

$$\left(1-\frac{2\mu k}{\epsilon_2\check{\gamma}}-2\mu k\hat{\gamma}\|\boldsymbol{\nu}\|_{\mathbb{E}}^2\right)\|\psi_j\|_0^2+2k\left(1-\frac{2\epsilon_1\hat{\gamma}E^2}{\mu}\right)\sum_{i=1}^{j}\|\nabla\psi_i\|_0^2$$
$$+2k\left(\lambda-\frac{\delta\mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{\epsilon_4}\right)\sum_{i=1}^{j}\|\psi_i\|_{0,\Gamma_u}^2+\mu k\left(\frac{1}{2}-\frac{1}{\epsilon_1}\right)\sum_{i=1}^{j}\|\gamma(u_{i-1}^h)^{-1/2}\nabla\zeta_i\|_0^2$$
$$+\mu k\left(1-\frac{\epsilon_2}{2}\right)\sum_{i=1}^{j}\|\gamma(u_{i-1}^h)^{1/2}\zeta_i\|_0^2+\mu k\left(1-\frac{\epsilon_4\delta}{2}\right)\sum_{i=1}^{j}\|\zeta_i\|_{0,\Gamma_\sigma}^2$$
$$\leqslant C_{1,j}h^{2r}+C_{2,j}k^2+\left(C\|\sigma\|_{L_\infty(0,t_j;W_\infty^1(\Omega))}^2+\frac{2\mu}{\epsilon_2\check{\gamma}}+2\mu\hat{\gamma}\|\boldsymbol{\nu}\|_{\mathbb{E}}^2\right)\sum_{i=0}^{j-1}k\|\psi_i\|_0^2.$$

For condition (C) we have $\delta=0$ and so we take $\epsilon_2=1$ and some $\mu>4\hat{\gamma}E^2$ so that there exists an ϵ_1 satisfying $2<\epsilon_1<\mu/2\hat{\gamma}E^2$. Once again, for some $\hat{k}>0$ we arrive at a form of (35).

Finally, in this part of the proof, for condition (D) we have $\delta = 1$ and we take ϵ_1, ϵ_2 and μ as for (C). (D) itself implies that we can select μ satisfying $4\hat{\gamma}E^2 < \mu < 2\lambda\|\boldsymbol{\nu}\|_{\mathbb{E}}^{-2}$ which, in turn, implies an ϵ_4 satisfying, $\mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2/\lambda < \epsilon_4 < 2$. Therefore, $1 - \epsilon_4/2 > 0$ and $\lambda - \mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2/\epsilon_4 > 0$ and we have proven that a form of (35) holds under all four conditions with, in each case, some appropriately chosen $\hat{k} > 0$ and a minor scaling adjustment to the 'generic constant' in the $C_{i,j}$.

Lastly, for ξ and ϑ we have,

$$\begin{aligned}\|\xi_j\|_0^2 + k\sum_{i=1}^{j}\Big(\|\nabla\xi_i\|_0^2 + \|\xi_i\|_{0,\Gamma_u}^2\Big)&\\ + k\sum_{i=1}^{j}\Big(\|\gamma(u_{i-1}^h)^{-1/2}\nabla\vartheta_i\|_0^2 + \|\gamma(u_{i-1}^h)^{1/2}\vartheta_i\|_0^2 + \|\vartheta_i\|_{0,\Gamma_\sigma}^2\Big)&\\ \leqslant Ch^{2r}\Big((1+t_j)\|u\|_{L_\infty(0,t_j;H^{r+1}(\Omega))}^2 + t_j\|\sigma\|_{L_\infty(0,t_j;H^{r+1}(\Omega))}^2\Big),&\end{aligned}$$

and the proof is completed by invoking the triangle inequality and the simple fact that $n^{-1/2}(x_1 + \cdots + x_n) \leqslant (x_1^2 + \cdots + x_n^2)^{1/2} \leqslant x_1 + \cdots + x_n$ (each x_n non-negative). □

Finally in this section we need a corollary for the error at the first time step (recall (25)).

Corollary 9 (to Theorem 8)**.** *Assuming further that $u|_{(0,k)} \in V_u \cap H^2(0,k;L_2(\Omega))$ we have for the first step error in the Euler method that,*

$$\|u(k) - u_1^h\|_0 \leqslant C_{\text{ie}}(h^r + k^{3/2}),$$

for a constant, $C_{\text{ie}} > 0$, independent of h and k.

Proof. It is necessary only to note that,

$$\int_0^k \|u_t(t)\|_0^2\,dt \leqslant k\|u_t\|_{L_\infty(0,k;L_2(\Omega))}^2,$$

and,

$$\int_0^k \|u_{tt}(t)\|_0\,dt \leqslant k^{1/2}\|u_{tt}\|_{L_2(0,k;L_2(\Omega))}$$

and then to adjust the constant $C_{2,1}$ in the $j = 1$ case of Theorem 8. □

4.2. Error bound for the Crank-Nicolson method. This material parallels the previous subsection except that we are not so careful about carrying through the dependencies of the constants on u and σ. The first result deals with estimating the error in the nonlinear terms.

Lemma 10 ('nonlinearity error'). *Assuming* (15) *we have for a constant* $\mu > 0$,

$$\mu k \sum_{i=1}^{j} \left| \left(\overline{(\gamma(u)^{-1}\nabla\sigma)}_i - \gamma(\mathscr{E}_i u^h)^{-1}\nabla\bar{\sigma}_i^*, \nabla\bar{\zeta}_i \right) + \left(\overline{(\gamma(u)\sigma)}_i - \gamma(\mathscr{E}_i u^h)\bar{\sigma}_i^*, \bar{\zeta}_i \right) \right|$$

$$\leqslant \mu k \sum_{i=1}^{j} \left(\frac{1}{2\epsilon_A} \|\gamma(\mathscr{E}_i u^h)^{-1/2}\nabla\bar{\zeta}_i\|_0^2 + \frac{1}{2\epsilon_B} \|\gamma(\mathscr{E}_i u^h)^{1/2}\bar{\zeta}_i\|_0^2 \right)$$

$$+(\epsilon_A+\epsilon_B)C \left(\|\sigma\|^2_{L_\infty(0,t_j;W^1_\infty(\Omega))} \sum_{i=1}^{j} k\|u_{i-1/2} - \mathscr{E}_i u^h\|_0^2 + t_j h^{2r} \|\sigma\|^2_{L_\infty(0,t_j;H^{r+1}(\Omega))} \right)$$

$$+ (\epsilon_A + \epsilon_B)Ck^4 \Big(\|\sigma_{tt}\|^2_{L_2(0,t_j;H^1(\Omega))} + \|(\gamma(u)\sigma)_{tt}\|^2_{L_2(0,t_j;L_2(\Omega))}$$

$$+ \|(\gamma(u)^{-1}\nabla\sigma)_{tt}\|^2_{L_2(0,t_j;L_2(\Omega))} \Big)$$

for all $\epsilon_A > 0$ *and* $\epsilon_B > 0$, *where* C *is a constant independent of* h, k, u *and* σ.

Proof. The proof is similar to Lemma 6. First of all we have,

$$\|\overline{(\gamma(u)\sigma)}_i - \gamma(\mathscr{E}_i u^h)\bar{\sigma}_i^*\|_0 \leqslant \|\overline{(\gamma(u)\sigma)}_i - \gamma(u_{i-1/2})\sigma_{i-1/2}\|_0$$
$$+ \|(\gamma(u_{i-1/2}) - \gamma(\mathscr{E}_i u^h))\sigma_{i-1/2}\|_0 + \|\gamma(\mathscr{E}_i u^h)(\sigma_{i-1/2} - \bar{\sigma}_i)\|_0 + \|\gamma(\mathscr{E}_i u^h)(\bar{\sigma}_i - \bar{\sigma}_i^*\|_0,$$
$$\leqslant Ck^{3/2} \Big(\|(\gamma(u)\sigma)_{tt}\|_{L_2(t_{i-1},t_i;L_2(\Omega))} + \hat{\gamma}\|\sigma_{tt}\|_{L_2(t_{i-1},t_i;L_2(\Omega))} \Big)$$
$$+ \hat{\gamma}\|\bar{\vartheta}_i\|_0 + C'_\gamma \|\sigma_{i-1/2}\|_{L_\infty(\Omega)} \|u_{i-1/2} - \mathscr{E}_i u^h\|_0$$

by Lemma 5. Secondly, since

$$\overline{(\gamma(u)^{-1}\nabla\sigma)}_i - \gamma(\mathscr{E}_i u^h)^{-1}\nabla\bar{\sigma}_i^* = \overline{(\gamma(u)^{-1}\nabla\sigma)}_i - \gamma(u_{i-1/2})^{-1}\nabla\sigma_{i-1/2}$$
$$+ \frac{1}{\gamma(u_{i-1/2})\gamma(\mathscr{E}_i u^h)} \Big(\gamma(\mathscr{E}_i u^h)\nabla\sigma_{i-1/2} - \gamma(u_{i-1/2})\nabla\bar{\sigma}_i^* \Big),$$

we have,

$$\|\overline{(\gamma(u)^{-1}\nabla\sigma)}_i - \gamma(\mathscr{E}_i u^h)^{-1}\nabla\bar{\sigma}_i^*\|_0 \leqslant \|\overline{(\gamma(u)^{-1}\nabla\sigma)}_i - \gamma(u_{i-1/2})^{-1}\nabla\sigma_{i-1/2}\|_0$$
$$+ \check{\gamma}^{-2}\|(\gamma(\mathscr{E}_i u^h) - \gamma(u_{i-1/2}))\nabla\sigma_{i-1/2}\|_0 + \check{\gamma}^{-2}\|\gamma(u_{i-1/2})\nabla(\sigma_{i-1/2} - \bar{\sigma}_i)\|_0$$
$$+ \check{\gamma}^{-2}\|\gamma(u_{i-1/2})\nabla(\bar{\sigma}_i - \bar{\sigma}_i^*)\|_0,$$
$$\leqslant Ck^{3/2} \Big(\|(\gamma(u)^{-1}\nabla\sigma)_{tt}\|_{L_2(t_{i-1},t_i;L_2(\Omega))} + \hat{\gamma}\check{\gamma}^{-2}\|\nabla\sigma_{tt}\|_{L_2(t_{i-1},t_i;L_2(\Omega))} \Big)$$
$$+ \hat{\gamma}\check{\gamma}^{-2}\|\nabla\bar{\vartheta}_i\|_0 + C'_\gamma\check{\gamma}^{-2}\|\nabla\sigma_{i-1/2}\|_{L_\infty(\Omega)}\|\mathscr{E}_i u^h - u_{i-1/2}\|_0.$$

To complete the proof we merge these two esimates, use (30), multiply by μk, use Young's inequality and sum over $i = 1, 2, \ldots, j$. □

The next lemma demonstrates how the implicit Euler predictor-corrector at the first step merges with the linear extrapolation at later steps to give the optimal order of k for the Crank-Nicolson method.

Lemma 11 (extrapolation error). *If* $u \in H^2(I; L_2(\Omega)) \cap L_\infty(I; H^{r+1}(\Omega))$ *and* $\breve{u} \in H^{r+1}(\Omega)$ *then,*

$$k \sum_{i=1}^{j} \|u_{i-1/2} - \mathscr{E}_i u^h\|_0^2 \leqslant C(h^{2r} + k^4) + Ck \sum_{i=0}^{j-1} \|\psi_i\|_0^2$$

for $j \in \{1, 2, \ldots\}$ *where* $C > 0$ *is independent of* h *and* k.

Proof. For $i > 1$ we have,

$$\|u_{i-1/2} - \mathscr{E}_i u^h\|_0 \leqslant \|u_{i-1/2} - \mathscr{E}_i u\|_0 + \|\mathscr{E}_i \xi\|_0 + \|\mathscr{E}_i \psi\|_0,$$

and we just need (33) and Lemma 5. For $i = 1$ we have $\mathscr{E}_1 u^h = \frac{3}{2}u_0^h - \frac{1}{2}u_{-1}^h$ with $u_{-1}^h = 2u_0^h - \hat{u}_1^h$ where $\hat{u}_1^h$ is the first-step solution by the implicit Euler method. Hence,

$$\begin{aligned}\|u_{1/2} - \mathscr{E}_1 u^h\|_0 &\leqslant \|u_{1/2} - \bar{u}_1\|_0 + \frac{1}{2}\|\breve{u} - u_0^h\|_0 + \frac{1}{2}\|u_1 - \hat{u}_1^h\|_0, \\ &\leqslant Ck^{3/2}\|u_{tt}\|_{L_2(0,k;L_2(\Omega))} + Ch^r\|\breve{u}\|_{r+1} + \frac{C_{\text{ie}}}{2}(h^r + k^{3/2}),\end{aligned}$$

by (33), Lemma 5 and Corollary 9. Squaring, merging and summing these then completes the proof. □

We can now state the error bound.

Theorem 12 (Crank-Nicolson: error bound)**.** *Assume that in* (6) *we have* $\breve{u} \in H^{r+1}(\Omega)$ *and also that* $u \in V_u \cap H^1(I; H^{r+1}(\Omega)) \cap H^3(I; L_2(\Omega))$ *and* $\sigma \in V_\sigma \cap H^2(I; H^1(\Omega)) \cap L_\infty(I; H^{r+1}(\Omega) \cap W_\infty^1(\Omega))$ *in* (13) *and* (14)*. Assume also that* $\gamma(u)\sigma \in H^2(I; L_2(\Omega))$ *and* $\gamma(u)^{-1}\nabla\sigma \in H^2(I; L_2(\Omega))$ *then, if at least one of the following conditions holds,*

$$\textit{(A) } \boldsymbol{\nu} = \mathbf{0}; \quad \textit{(B) } \|\boldsymbol{\nu}\|_{\mathbb{E}} < \frac{1}{2E}\sqrt{\frac{\check{\gamma}}{\hat{\gamma}}}; \quad \textit{(C) } \Gamma_u \cap \Gamma_\sigma = \varnothing; \quad \textit{(D) } \|\boldsymbol{\nu}\|_{\mathbb{E}} < \frac{1}{E}\sqrt{\frac{\lambda}{2\hat{\gamma}}},$$

there is a constant $\hat{k} > 0$ *such that whenever* $k < \hat{k}$*,*

$$\begin{aligned}&\left(k\sum_{i=1}^{j}\left(\frac{1}{\hat{\gamma}}\|\nabla\bar{\sigma}_i - \nabla\bar{\sigma}_i^h\|_0^2 + \check{\gamma}\|\bar{\sigma}_i - \bar{\sigma}_i^h\|_0^2 + \|\bar{\sigma}_i - \bar{\sigma}_i^h\|_{0,\Gamma_\sigma}^2\right)\right)^{1/2} \\ &+ \left(k\sum_{i=1}^{j}\left(\|\nabla\bar{u}_i - \nabla\bar{u}_i^h\|_0^2 + \|\bar{u}_i - \bar{u}_i^h\|_{0,\Gamma_u}^2\right)\right)^{1/2} + \|u(t_j) - u_j^h\|_0 \leqslant C(h^r + k^2)\end{aligned}$$

for each $j \in \{1, 2, \ldots, N\}$*. The constant* C *is independent of* h *and* k*.*

Proof. Form the average of (13) at t_i and t_{i-1} and subtract the result from (23). Do the same with (14) and (24) then add the results and rearrange:

$$\begin{aligned}(36)\quad &(\partial_t\psi_i, v) + (\nabla\bar{\psi}_i, \nabla v) + \lambda(\bar{\psi}_i, v)_{\Gamma_u} + (\bar{\zeta}_i, w)_{\Gamma_\sigma} + (\gamma(\mathscr{E}_i u^h)^{-1}\nabla\bar{\zeta}_i, \nabla w) \\ &+ (\gamma(\mathscr{E}_i u^h)\bar{\zeta}_i, w) = (\Delta_i u, v) + (\partial_t\xi_i, v) + (\nabla\bar{\xi}_i, \nabla v) + E(\nabla\bar{\vartheta}_i, \nabla v) \\ &+ \lambda(\bar{\xi}_i, v)_{\Gamma_u} + (\bar{\vartheta}_i, w)_{\Gamma_\sigma} - (\bar{\xi}_i, w) + (\boldsymbol{\nu}\cdot\nabla\bar{\xi}_i, w) \\ &- E(\nabla\bar{\zeta}_i, \nabla v) + (\bar{\psi}_i, w) - (\boldsymbol{\nu}\cdot\nabla\bar{\psi}_i, w) \\ &+ \left((\overline{\gamma(u)^{-1}\nabla\sigma})_i - \gamma(\mathscr{E}_i u^h)^{-1}\nabla\bar{\sigma}_i^*, \nabla w\right) + \left((\overline{\gamma(u)\sigma})_i - \gamma(\mathscr{E}_i u^h)\bar{\sigma}_i^*, w\right)\end{aligned}$$

for all $v \in V_u^h$, for all $w \in V_\sigma^h$ and for each $i \in \{1, 2, \ldots, N\}$.

Now choosing, in (36), $v = 2\bar{\psi}_i \in V_u^h$ and $w = \mu\bar{\zeta}_i \in V_\sigma^h$, for some $\mu > 0$ to be specified later, and noting that $2(\partial_t\psi_i, \bar{\psi}_i) = \partial_t\|\psi_i\|_0^2$, we get,

$$
\begin{aligned}
(37)\quad & \partial_t\|\psi_i\|_0^2 + 2\|\nabla\bar{\psi}_i\|_0^2 + 2\lambda\|\bar{\psi}_i\|_{0,\Gamma_u}^2 + \mu\|\bar{\zeta}_i\|_{0,\Gamma_\sigma}^2 + \mu\|\gamma(\mathscr{E}_i u^h)^{-1/2}\nabla\bar{\zeta}_i\|_0^2 \\
& + \mu\|\gamma(\mathscr{E}_i u^h)^{1/2}\bar{\zeta}_i\|_0^2 = 2(\Delta_i u, \bar{\psi}_i) + 2(\partial_t\xi_i, \bar{\psi}_i) + 2(\nabla\bar{\xi}_i, \nabla\bar{\psi}_i) + 2E(\nabla\bar{\vartheta}_i, \nabla\bar{\psi}_i) \\
& + 2\lambda(\bar{\xi}_i, \bar{\psi}_i)_{\Gamma_u} + \mu(\bar{\vartheta}_i, \bar{\zeta}_i)_{\Gamma_\sigma} - \mu(\bar{\xi}_i, \bar{\zeta}_i) + \mu(\boldsymbol{\nu}\cdot\nabla\bar{\xi}_i, \bar{\zeta}_i) \\
& - 2E(\nabla\bar{\zeta}_i, \nabla\bar{\psi}_i) + \mu(\bar{\psi}_i, \bar{\zeta}_i) - \mu(\boldsymbol{\nu}\cdot\nabla\bar{\psi}_i, \bar{\zeta}_i) \\
& + \mu\big((\overline{\gamma(u)^{-1}\nabla\sigma})_i - \gamma(\mathscr{E}_i u^h)^{-1}\nabla\bar{\sigma}_i^*, \nabla\bar{\zeta}_i\big) + \mu\big((\overline{\gamma(u)\sigma})_i - \gamma(\mathscr{E}_i u^h)\bar{\sigma}_i^*, \bar{\zeta}_i\big).
\end{aligned}
$$

The next step is to estimate the first eight terms on the right. Labelling them I, II, …, $VIII$ we have $III + V = 0$ because of (32) and,

$$
\begin{aligned}
\Big|I + II + IV + VI + VII + VIII\Big| \leqslant{}& \frac{1}{\epsilon_1}\|\Delta_i u\|_0^2 + \epsilon_1\|\bar{\psi}_i\|_0^2 + \frac{1}{\epsilon_2}\|\partial_t\xi_i\|_0^2 + \epsilon_2\|\bar{\psi}_i\|_0^2 \\
& + \frac{E^2}{\epsilon_4}\|\nabla\bar{\vartheta}_i\|_0^2 + \epsilon_4\|\nabla\bar{\psi}_i\|_0^2 + \frac{\mu}{2\epsilon_6}\|\bar{\vartheta}_i\|_{0,\Gamma_\sigma}^2 + \frac{\epsilon_6\mu}{2}\|\bar{\zeta}_i\|_{0,\Gamma_\sigma}^2 \\
& + \frac{\mu}{2\epsilon_7\check{\gamma}}\|\bar{\xi}_i\|_0^2 + \frac{\epsilon_7\mu}{2}\|\gamma(\mathscr{E}_i u^h)^{1/2}\bar{\zeta}_i\|_0^2 + \frac{\mu\|\boldsymbol{\nu}\|_{\mathbb{E}}^2}{2\epsilon_8\check{\gamma}}\|\nabla\bar{\xi}_i\|_0^2 + \frac{\epsilon_8\mu}{2}\|\gamma(\mathscr{E}_i u^h)^{1/2}\bar{\zeta}_i\|_0^2.
\end{aligned}
$$

Now, use these, multiply by k, sum over $i = 1, 2, \ldots, j$, note that $\|\psi_0\|_0^2 = (\xi_0, \psi_0)$, use Lemmas 4, 5, 10, 11, with the estimates (30) and (33) to obtain,

$$
\begin{aligned}
& \left(1 - \left(\frac{\epsilon_1 + \epsilon_2}{2}\right)\hat{k}\right)\|\psi_j\|_0^2 + (2 - \epsilon_4)k\sum_{i=1}^{j}\|\nabla\bar{\psi}_i\|_0^2 + 2\lambda k\sum_{i=1}^{j}\|\bar{\psi}_i\|_{0,\Gamma_u}^2 \\
& \quad + \left(1 - \frac{\epsilon_6}{2}\right)\mu k\sum_{i=1}^{j}\|\bar{\zeta}_i\|_{0,\Gamma_\sigma}^2 + \left(1 - \frac{1}{2\epsilon_A}\right)\mu k\sum_{i=1}^{j}\|\gamma(\mathscr{E}_i u^h)^{-1/2}\nabla\bar{\zeta}_i\|_0^2 \\
& \quad + \left(1 - \frac{\epsilon_7}{2} - \frac{\epsilon_8}{2} - \frac{1}{2\epsilon_B}\right)\mu k\sum_{i=1}^{j}\|\gamma(\mathscr{E}_i u^h)^{1/2}\bar{\zeta}_i\|_0^2 \\
& \leqslant \left(1 + \frac{1}{\epsilon_2} + \frac{1}{\epsilon_4} + \frac{1}{\epsilon_6} + \frac{1}{\epsilon_7} + \frac{1}{\epsilon_8} + \epsilon_A + \epsilon_B\right)Ch^{2r} + \left(\frac{1}{\epsilon_1} + \epsilon_A + \epsilon_B\right)Ck^4 \\
& \quad + (\epsilon_1 + \epsilon_2 + \epsilon_A + \epsilon_B)Ck\sum_{i=0}^{j-1}\|\psi_i\|_0^2 \\
& \quad + k\sum_{i=1}^{j}\left(2E|(\nabla\bar{\zeta}_i, \nabla\bar{\psi}_i)| + \mu|(\bar{\psi}_i, \bar{\zeta}_i)| + \mu|(\boldsymbol{\nu}\cdot\nabla\bar{\psi}_i, \bar{\zeta}_i)|\right).
\end{aligned}
$$

Choosing $\epsilon_1 = \epsilon_2 = 1/2\hat{k}$, $\epsilon_4 = \epsilon_6 = \epsilon_A = 1$, $\epsilon_7 = \epsilon_8 = 1/4$, $\epsilon_B = 2$ and multiplying through by two gives,

$$
\begin{aligned}
& \mu k\sum_{i=1}^{j}\left(\|\gamma(\mathscr{E}_i u^h)^{-1/2}\nabla\bar{\zeta}_i\|_0^2 + \|\gamma(\mathscr{E}_i u^h)^{1/2}\bar{\zeta}_i\|_0^2 + \|\bar{\zeta}_i\|_{0,\Gamma_\sigma}^2\right) \\
& \quad + \|\psi_j\|_0^2 + 2k\sum_{i=1}^{j}\left(\|\nabla\bar{\psi}_i\|_0^2 + \lambda\|\bar{\psi}_i\|_{0,\Gamma_u}^2\right) \leqslant C(h^{2r} + k^4) + Ck\sum_{i=0}^{j-1}\|\psi_i\|_0^2 \\
& \quad + 4kE\sum_{i=1}^{j}|(\nabla\bar{\zeta}_i, \nabla\bar{\psi}_i)| + 2\mu k\sum_{i=1}^{j}|(\bar{\psi}_i, \bar{\zeta}_i)| + 2\mu k\sum_{i=1}^{j}|(\boldsymbol{\nu}\cdot\nabla\bar{\psi}_i, \bar{\zeta}_i)|,
\end{aligned}
$$

TABLE 1. Table of errors for the implicit Euler scheme illustrating the convergence rate predicted by Theorem 8 for $j = N$.

	M				
N	8	16	32	64	128
8	$2.041_{(0.843)}$	$1.124_{(0.886)}$	$0.7036_{(0.76)}$	$0.5477_{(0.558)}$	$0.5011_{(0.43)}$
16	$1.9_{(0.942)}$	$1.002_{(1.03)}$	$0.5569_{(1.01)}$	$0.3661_{(0.942)}$	$0.2997_{(0.87)}$
32	$1.892_{(0.849)}$	$0.9782_{(0.957)}$	$0.5078_{(0.98)}$	$0.2834_{(0.974)}$	$0.1892_{(0.953)}$
64	$1.891_{(0.85)}$	$0.972_{(0.961)}$	$0.4932_{(0.988)}$	$0.2552_{(0.993)}$	$0.1429_{(0.988)}$
128	$1.891_{(0.85)}$	$0.9706_{(0.962)}$	$0.4894_{(0.99)}$	$0.2472_{(0.996)}$	$0.1279_{(0.997)}$

which, apart from the first term on the right, is a Crank-Nicolson analogue of Lemma 7. The remainder of this proof can, therefore, be completed in the same way as for the proof of Theorem 8. □

The next section gives some numerical demonstrations of these results.

5. Numerical results

In this section we use an artificial exact solution in order to demonstrate the convergence rates claimed by Theorems 8 and 12, and then we go on to show the results of some numerical experiments under more demanding conditions. We will see that although the Crank-Nicolson method is theoretically superior to the implicit Euler method (and that this shows through in the convergence tests) it may not be so adequate in dealing with data that generate steep travelling fronts, such as those observed in [12, 11].

All of the computations detailed here were carried out using version 2.24 of *Freefem++* (see `www.freefem.org/ff++`) in *Windows XP* and *SuSE Gnu/Linux 10.3*, and the graphics were generated by *MATLAB R2007b*. In all cases we used the unit square, $\Omega = (0,1)^2$, as the spatial domain and created a uniform mesh of triangles by forming an M by M array of the building block $\boxtimes$. We also took $\Gamma_u = \Gamma_\sigma = \partial\Omega$ for all the examples that follow and, unless explicitly mentioned otherwise, we used linear finite elements for the implicit Euler calculations and quadratic elements for the Crank-Nicolson ones.

In order to demonstrate the theoretically predicted convergence rates we add functions $f^\natural$ and $f^\sharp$ to the right hand sides of (4) and (5) and a function $\sigma^\flat$ to the right hand side of (9). These, along with $u^\flat$ and $\breve{u}$ are then chosen so that the exact solutions are given by $u(x,y,t) = \cos(2\pi x)\cos(3\pi y)\cos(6\pi t)$ and $\sigma(x,y,t) = \cos(\pi x)\cos(2\pi y)\cos(8\pi t)$.

The remaining data are chosen as $T = 1$, $\gamma_G = \hat{\gamma} = 1$, $\gamma_R = \check{\gamma} = 0.1$, $\Delta = 2$, $u_c = 8$, $E = 20$, $\boldsymbol{\nu} = (0.01, 0.03)^T$, $\lambda = 3$ and the uniform time step is given by $k = T/N$. For these data it is readily checked that $\|\boldsymbol{\nu}\|_{\mathbb{E}} \approx 0.032$ while $(1/2E)\sqrt{(\check{\gamma}/\hat{\gamma})} \approx 0.008$ and $(1/E)\sqrt{(\lambda/2\hat{\gamma})} \approx 0.061$, and so the conditions of Theorems 8 and 12 are met.

Table 1 shows the errors resulting from the implicit Euler scheme while Table 2 shows those for the Crank-Nicolson scheme. The errors are measured in the norm bounded in the relevant theorem. In each of these tables the subscript shows the order of convergence estimated by that error as compared to the error immediately north-west (the first columns and rows are using error not shown here). The first-order convergence of the implicit Euler scheme is evident as is the second-order convergence of the Crank-Nicolson scheme.

TABLE 2. Table of errors for the Crank-Nicolson scheme illustrating the convergence rate predicted by Theorem 12 for $j = N$.

	M				
N	8	16	32	64	128
8	$0.2413_{(2.12)}$	$0.2087_{(1.48)}$	$0.2063_{(1.35)}$	$0.2061_{(1.34)}$	$0.2061_{(1.34)}$
16	$0.2952_{(0.717)}$	$0.09699_{(1.32)}$	$0.06356_{(1.72)}$	$0.06079_{(1.76)}$	$0.06061_{(1.77)}$
32	$0.3353_{(1.58)}$	$0.08933_{(1.72)}$	$0.02768_{(1.81)}$	$0.01739_{(1.87)}$	$0.01653_{(1.88)}$
64	$0.3468_{(1.75)}$	$0.09101_{(1.88)}$	$0.02341_{(1.93)}$	$0.00716_{(1.95)}$	$0.004472_{(1.96)}$
128	$0.3498_{(1.78)}$	$0.0917_{(1.92)}$	$0.02325_{(1.97)}$	$0.005922_{(1.98)}$	$0.001806_{(1.99)}$

FIGURE 1. Snapshots for the Implicit Euler scheme.

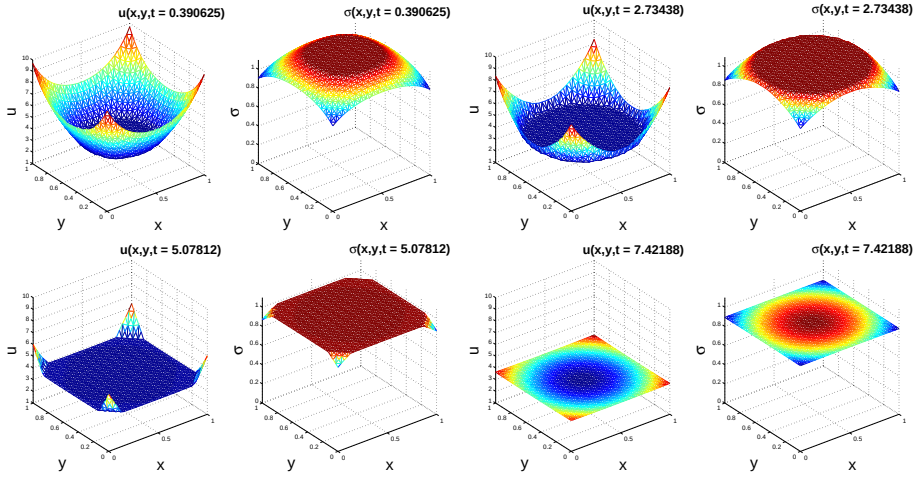


FIGURE 2. Snapshots for the Crank-Nicolson scheme.

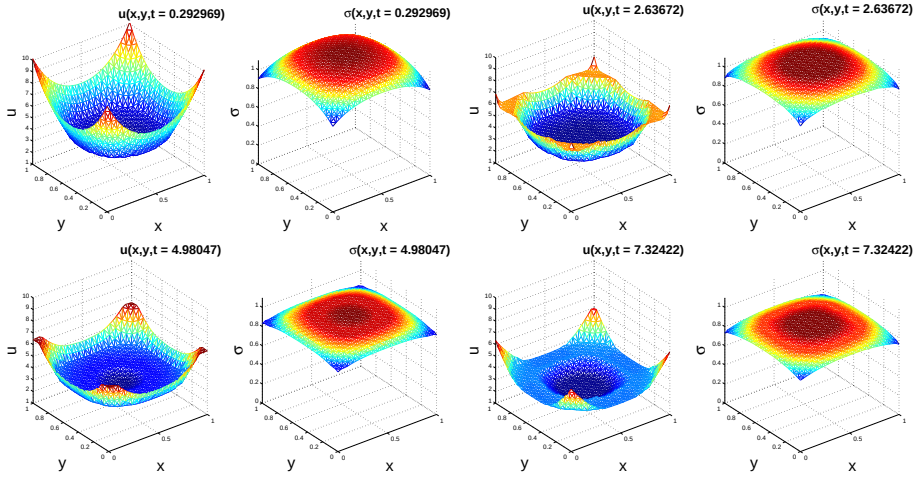
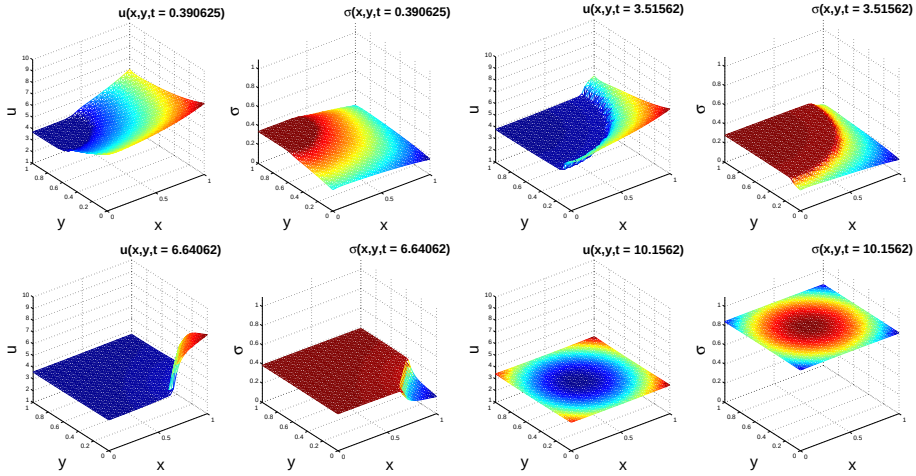


FIGURE 3. Snapshots for the Implicit Euler scheme.



On the basis of these results it seems that the Crank-Nicolson scheme is the superior of the two but this may not be the case.

To see this we now select the data $\hat{\gamma} = 1$, $\check{\gamma} = 0.01$, $\Delta = 0.01$, $u_c = 4$, $E = 20$, $\boldsymbol{\nu} = (0, 0)^T$, $\lambda = 0.01$, $\breve{u} = 5$, $u^\flat = 1$, $T = 25$ and take $M = 32$ and $N = 128$. We have no exact solution here and so no $f^\natural$, $f^\sharp$ or $\sigma^\flat$ are used for these calculations.

Figure 1 shows snapshots of u and σ at various times for the implicit Euler method while Figure 2 shows the calculations for the Crank-Nicolson method (at a closest mid-time). We can see that the Euler method predicts 'umbrella shaped' surfaces morphing smoothly into approximately flat surfaces. The Crank-Nicolson method on the other hand predicts a grossly similar behaviour but suggests that more detail on the surface for u actually exists. In fact, an independently-coded calculation using the proprietary *COMSOL Multiphysics* finite element package produces results that are in agreement with the Euler method and casts doubt on the Crank-Nicolson calculation. We will return to this claim in the conclusions section.

As a last example in this section we alter the data just given so that $\boldsymbol{\nu} = (1, -1)^T$ (which, given the other data, violates the conditions on $\|\boldsymbol{\nu}\|_{\mathbb{E}}$ required by the error estimates) and plot the results in Figure 3 for the Euler method. We can see that the Euler method predicts a steep front travelling smoothly across the domain and settling to an approximately flat surface whereas the Crank-Nicolson method gives surfaces (not shown here) that, seemingly, are again spurious. *COMSOL Multiphysics* again agrees with the Euler method for this case.

6. Concluding remarks

Although the error bounds and convergence tests suggest that the extrapolated Crank-Nicolson method is superior to the linearised implicit Euler method the illustrations in Figures 1 and 2 suggest that the Euler scheme is in fact the more robust of the two.

To test whether or not it is the extrapolation that is the cause of this inadequacy in the Crank-Nicolson method we implemented two other schemes. A Newton-Crank-Nicolson (NCN) method and a predictor-corrector Crank-Nicolson (PCCN) method. The NCN method was based upon using the Euler method to obtain a

TABLE 3. Table of errors for the Newton-Crank-Nicolson scheme for $j = N$.

	M				
N	4	8	16	32	64
4	1.051	0.5802	0.5231	0.519	0.5187
8	0.4851	$0.2413_{(2.12)}$	$0.2087_{(1.48)}$	$0.2062_{(1.34)}$	$0.2061_{(1.33)}$
16	1.004	$0.2948_{(0.719)}$	$0.09576_{(1.33)}$	$0.06166_{(1.76)}$	$0.0588_{(1.81)}$
32	1.163	$0.3353_{(1.58)}$	$0.08911_{(1.73)}$	$0.02696_{(1.83)}$	$0.01623_{(1.93)}$
64	1.204	$0.3468_{(1.75)}$	$0.09099_{(1.88)}$	$0.02334_{(1.93)}$	$0.00694_{(1.96)}$

TABLE 4. Table of errors for the predictor-corrector Crank-Nicolson scheme for $j = N$.

	M				
N	4	8	16	32	64
4	1.051	0.5799	0.5228	0.5186	0.5184
8	0.4851	$0.2413_{(2.12)}$	$0.2087_{(1.47)}$	$0.2062_{(1.34)}$	$0.2061_{(1.33)}$
16	1.004	$0.2948_{(0.719)}$	$0.09575_{(1.33)}$	$0.06164_{(1.76)}$	$0.05877_{(1.81)}$
32	1.163	$0.3353_{(1.58)}$	$0.08911_{(1.73)}$	$0.02695_{(1.83)}$	$0.01622_{(1.93)}$
64	1.204	$0.3468_{(1.75)}$	$0.09099_{(1.88)}$	$0.02334_{(1.93)}$	$0.006941_{(1.96)}$

FIGURE 4. Snapshot for the NCN scheme.

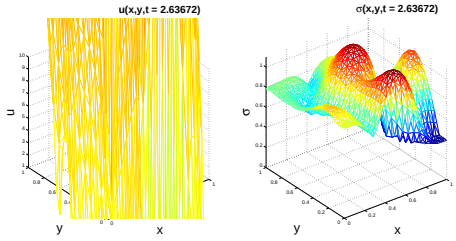


FIGURE 5. Snapshot for the PCCN scheme.

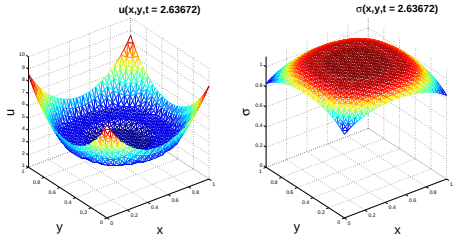


FIGURE 6. Snapshot for the 'hi-fi' scheme.

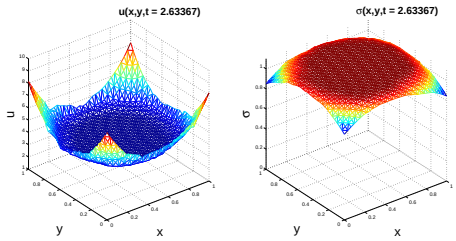
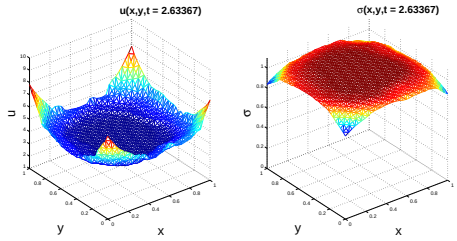


FIGURE 7. Snapshot for the 'hi-fi' P_1 scheme problem 3.



starting value and, thereafter, the solution from the previous time step was used as a starting value for the current time step. At each time step the equations were iterated until the $L_2(\Omega)$ norm of the difference in iterates was below a tolerance of 10^{-4} or until thirty iterations had been made.

The PCCN method used at each time step the Euler method to predict a value for u_i^h as $\hat{u}_i^h$ and then this was used to form $\bar{u}_i^h$ in the $\gamma(\cdot)$ terms in the usual Crank-Nicolson scheme.

To illustrate the convergence behaviour of the NCN and PCCN methods we choose the same data as for the previous convergence tests and give the errors in Table 3 for the NCN method and in Table 4 for the PCCN method. The second-order convergence in space and time is evident and suggests that the methods are implemented correctly. (In fact Tables 2, 3 and 4 are almost indistinguishable.)

However, when we try to reproduce the solution in Figure 1 we obtain (at or near the time $t = 2.73\ldots$) Figure 4 for the NCN method and Figure 5 for the PCCN method. Clearly Figure 4 suggests that the NCN scheme is hopeless.

To investigate this recall that Figure 4 is based on $T = 25$ and $N = 128$ (so $k = 0.1953125$). Reducing this time step by 10 (i.e. $N = 1280$) still gives very poor results beyond the first time step and in fact only when we take $k = 0.001953125$ do we get results of the correct appearance for the first seven time steps. Unfortunately, after t_7 the solution quality degraded rapidly.

Two further investigations were also carried out. In Figure 6 we show the results of a 'hi-fidelity' calculation in time where the only difference in data, from that used to produce Figure 2, is that $N = 512$.

Also, as a matter of completeness, in Figure 7 we show the result of using this same 'hi-fi' calculation but with linear finite elements instead of quadratics.

Obviously, Figures 1, 2, 4, 5, 6 and 7 should, at least in the 'eye norm', be the same and yet they are not. Moreover, Figures 2, 4, 5, 6 and 7 are not comparable whereas Figure 1 agrees with an independent calculation using alternative software.

In conclusion, these numerical experiments suggest that while the Crank-Nicolson scheme is a theoretically more accurate approximation it can produce solution surfaces which appear to exhibit gross oscillations and smaller travelling fronts. These features appear to be spurious and it seems as though impractically fine discretizations would be needed to eliminate them. Overall, it seems likely that the Euler method is the more robust of the two schemes under consideration.

Of course, we should bear in mind that the Crank-Nicolson method demands higher temporal regularity but the numerical solutions do not suggest that this is unreasonable in the cases presented. A further investigation could be based on so-called 'overkill' computations in an effort to judge whether or not the norms required by the error bounds do in fact exist.

To close we recall that our model problem was drawn from that in [6] and, in particular, the term νu_x there gave rise to the term $\boldsymbol{\nu}\cdot\nabla u$ in (5) and (14). In a later study, [7], Edwards replaced that term with νu_x^2 in order to eliminate any 'preferred directions'. It seems that the foregoing material could be adapted to deal with an analogue of that replacement in the following way. If in (5) and (14) we replace $\boldsymbol{\nu}\cdot\nabla u$ with $\chi\|\nabla u\|_{\mathbb{E}}$, for any p norm, $\|\cdot\|_{\mathbb{E}}$, on $\mathbb{R}^d$ and some real χ, then we would have $\boldsymbol{\nu}\cdot\nabla u_i^h$ (resp. $\boldsymbol{\nu}\cdot\nabla\bar{u}_i^h$) replaced with $\chi\|\nabla u_i^h\|_{\mathbb{E}}$ (resp. $\chi\|\nabla\bar{u}_i^h\|_{\mathbb{E}}$) in (22) (resp. (24)).

Focussing on the implicit Euler scheme and examining the effect on the proof of (the resulting analogue of) Lemma 7 we see that (with $\chi=1$ for simplicity) we would have to deal with the difference $(\|\nabla u_i\|_{\mathbb{E}},w)-(\|\nabla u_i^h\|_{\mathbb{E}},w)$. But with the inverse triangle inequality (i.e. $\big|\,\|\boldsymbol{a}\|_{\mathbb{E}}-\|\boldsymbol{b}\|_{\mathbb{E}}\,\big|\leqslant\|\boldsymbol{a}-\boldsymbol{b}\|_{\mathbb{E}}$ for all $\boldsymbol{a}$ and $\boldsymbol{b}$) this can estimated as $\big|(\|\nabla u_i\|_{\mathbb{E}}-\|\nabla u_i^h\|_{\mathbb{E}},w)\big|\leqslant\|\nabla\xi_i\|_0\|w\|_0+\|\nabla\psi_i\|_0\|w\|_0$ and we can just pick up the proof as presented earlier.

On the other hand, if we follow [7] more literally and work instead with a quadratic term, $\chi\|\nabla u\|_{\mathbb{E}}^2$ say, then we can estimate with,

$$\big|(\|\nabla u_i\|_{\mathbb{E}}^2-\|\nabla u_i^h\|_{\mathbb{E}}^2,w)\big|\leqslant\big(\|\nabla u_i\|_{L_\infty(\Omega)}+\|\nabla u_i^h\|_{L_\infty(\Omega)}\big)\big(\|\nabla\xi_i\|_0+\|\nabla\psi_i\|_0\big)\|w\|_0$$

and we would need an *a priori* $W^1_\infty(\Omega)$ bound on each u_i^h in order to proceed. We do not pursue these models any further here, but note that the implementation for the implicit Euler method could be simplified considerably by lagging this nonlinear gradient term (i.e. evaluating it at the previous time step).

References

[1] Kendall Atkinson and Weimin Han. *Theoretical numerical analysis: a functional analysis framework*. Springer-Verlag (TAM39), 2001.

[2] Norbert Bauermeister and Simon Shaw. Finite-element approximation of non-Fickian polymer diffusion. *IMA J. Numer. Anal.*, 30:702—730, 2010. doi: 10.1093/imanum/drn071; BURA: http://hdl.handle.net/2438/3113 (BICOM Tech. Rep. 08/1, see `www.brunel.ac.uk/bicom`).

[3] Susanne C. Brenner and L. Ridgway Scott. *The mathematical theory of finite element methods*. Springer-Verlag New York, Inc, 1994.

[4] 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.

[5] D. A. Edwards and D. S. Cohen. An unusual moving boundary condition arising in anomalous diffusion problems. *SIAM J. Appl. Math.*, 55:662—676, 1995.

[6] David A. Edwards. A spatially nonlocal model for polymer-penetrant diffusion. *Z. angew. Math. Phys.*, 52:254—288, 2001.

[7] David A. Edwards. A spatially nonlocal model for polymer desorption. *Journal of Engineering Mathematics*, 53:221—238, 2005.

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

[9] Robert B. Lowrie. A comparison of implicit time integration methods for nonlinear relaxation and diffusion. *J. Comp. Phys.*, 196:566—590, 2004.

[10] 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.

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

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

[13] 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.

Brunel Institute of Computational Mathematics, Brunel University UB8 3PH, England
E-mail: `simon.shaw@brunel.ac.uk`
URL: `people.brunel.ac.uk/~icsrsss`