

Error estimates for an Euler implicit, mixed finite element discretization of Richards' equation : equivalence between mixed and conformal approaches

Citation for published version (APA):

Radu, F. A., Pop, I. S., & Knabner, P. (2002). *Error estimates for an Euler implicit, mixed finite element discretization of Richards' equation : equivalence between mixed and conformal approaches*. (RANA : reports on applied and numerical analysis; Vol. 0206). Technische Universiteit Eindhoven.

Document status and date:

Published: 01/01/2002

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

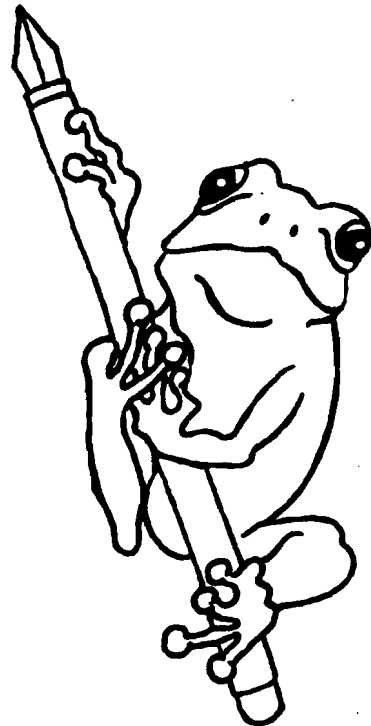
EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computing Science

RANA 02-06
January 2002

Error estimates for an Euler implicit, mixed finite element
discretization of Richards' equation: equivalence
between mixed and conformal approaches

by

F. Radu, I.S. Pop and P. Knabner



Reports on Applied and Numerical Analysis
Department of Mathematics and Computing Science
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven, The Netherlands
ISSN: 0926-4507

ERROR ESTIMATES FOR AN EULER IMPLICIT, MIXED FINITE ELEMENT DISCRETIZATION OF RICHARDS' EQUATION: EQUIVALENCE BETWEEN MIXED AND CONFORMAL APPROACHES

FLORIN RADU[†], IULIU SORIN POP[§], PETER KNABNER[†]

ABSTRACT. We analyse a discretization method for a class of degenerate parabolic problems which includes the Richards' equation. This analysis applies to the pressure-based formulation and includes both variably and fully saturated regimes. As suggested in [2], to overcome the difficulties posed by the lack in regularity, we first integrate the equation in time and then prove equivalence between the conformal formulation and the mixed variational one.

A regularization approach is combined with the Euler implicit scheme to achieve the time discretization. Again, equivalence between the two kinds of formulation is demonstrated for the semi-discrete case. Mixed finite elements are employed for the discretization in space. Error estimates are obtained, showing that the scheme is convergent.

1. INTRODUCTION

A commonly accepted mathematical model of water flow through porous media is the Richards' equation, a nonlinear, possibly degenerate, parabolic differential equation. In the pressure formulation, Richards' equation [4] is expressed as

$$(1.1) \quad \partial_t \Theta(\psi) - \nabla \cdot K(\Theta) \nabla(\psi + z) = 0$$

where ψ is the pressure head, Θ the saturation, K the conductivity and z the height against the gravitational direction. The equation (1.1) models the flow of a wetting fluid (water) in a porous media in the presence of a non-wetting fluid (air) supposed to be at constant pressure, 0. In the saturated region (where only water is present) we have $\psi \geq 0$, while $\psi < 0$ in the unsaturated domain. Different functional dependencies (retention curves) between ψ , K and Θ are proposed in the literature. These are provided essentially by soil particularities and

[†]Institute of Applied Mathematics, University Erlangen-Nürnberg, Martensstr. 3, D-91058 Erlangen, Germany ({raduf, knabner}@am.uni-erlangen.de).

[§]Department of Mathematics and Computer Science, Technical University Eindhoven, P. O. Box 513, 5600 MB Eindhoven, The Netherlands (I.Pop@tue.nl).

allow reducing all the unknowns in the above equation to a single one. Here we are interested in both partially saturated and saturated flow, therefore we retain the pressure ψ as primary unknown.

As suggested in [1], applying the Kirchhoff transformation

$$(1.2) \quad \begin{aligned} \mathcal{K} : \mathbb{R} &\longrightarrow \mathbb{R} \\ \psi &\longmapsto \int_0^\psi K(\Theta(s)) ds \end{aligned}$$

leads to unknowns which are more regular. Since $K(\Theta(s))$ is positive, this transformation can be inverted and equation (1.1) can be rewritten in terms of a new variable, $u := \mathcal{K}(\psi)$. Defining now

$$(1.3) \quad \begin{aligned} b(u) &:= \Theta \circ \mathcal{K}^{-1}(u), \\ k(b(u)) &:= K \circ \Theta \circ \mathcal{K}^{-1}(u), \end{aligned}$$

and letting e_z denote the vertical unit vector, equation (1.1) becomes

$$(1.4) \quad \partial_t b(u) - \nabla \cdot (\nabla u + k(b(u)) e_z) = 0 \quad \text{in } (0, T) \times \Omega.$$

By the above transformation, diffusion becomes linear in equation (1.1). However, the problem may still remain degenerate, leading to solutions lacking regularity. Since this equation models important practical problems, several papers are dealing with analysis and numerical methods for it. Euler methods are often employed for the discretization in time. Adaptive time stepping is studied in [22], [12], or [25]. In case of an implicit discretization, iterative methods are considered (see, for example, [14], [7], whose method was already proposed in [10] and used also in [13], and [12]).

For the spatial discretization, mixed finite elements or finite volumes provide a good approximation of the solution ([15], [3], [5], [9]). The most comprehensive algorithmic approach has been presented in the thesis [22] where hybrid mixed finite elements and an implicit Euler discretization are used. The set of nonlinear equations are solved by a Newton/multigrid method, while and time and space adaptive strategies are constructed on the basis of rigorous error indicators. However, most of the authors are mainly interested in computational aspects and less concerned with rigorous convergence results. With respect to this last aspect we mention recent papers like [26] (for the numerical analysis of a mixed finite element discretization), [9] (where convergence of an implicit finite volume method is proven by compactness arguments), [11] (for a relaxation scheme which applies to this equation too) and [21] (where error estimates are obtained for the unsaturated regime).

Here we consider an increasing and Lipschitz continuous b . Nevertheless, $b'(u)$ may be 0 for some values of u (not necessary isolated). Our

numerical approach employs the lowest order Raviart-Thomas finite elements in space and Euler implicit in time, together with a regularization step. Specifically, with $N > 0$ integer, set $\tau = T/N$ and let \mathcal{T}_h being a decomposition of Ω into closed d -simplices; h stands for the mesh-size. Then the numerical scheme under consideration reads

$$\begin{aligned} b_\epsilon(p_h^n) + \tau \nabla q_h^n &= b_\epsilon(p_h^{n-1}), \\ q_h^n + \nabla p_h^n + k(b(p_h^n))e_z &= 0, \end{aligned}$$

for $n = \overline{1, N}$; p_h^0 approximates u^0 in the finite dimensional approximation space. Here b_ϵ is a regular approximation of b depending on the small parameter $\epsilon > 0$. By p_h^n we denote a piecewise constant approximation of u and q_h^n is a Raviart-Thomas (RT_0) approximation of the flux $-(\nabla u + k(b(u))e_z)$, based on \mathcal{T}_h , both at $t = n\tau$.

As suggested in [2], to overcome the difficulties posed by the lack in regularity, equation (1.4) is first integrated in time. For the resulting problem a mixed variational formulation is stated.

Convergence is shown by obtaining first error estimates for the time discrete scheme, by following the ideas in [19]. Next, using the procedure described in [2], error estimates for the fully discrete scheme are obtained. In this framework applying equivalence between different formulations becomes essential since this allows transferring results obtained for the conforming method to the mixed formulation and viceversa.

The outline of the paper is as follows. First we state the main assumptions and notations used throughout the paper, define the problem to be solved and discuss questions regarding existence and regularity of a solution. In Section 2 the equivalence between a conformal and a mixed variational formulations is proved, for the continuous case as well as for the time discrete one. In Section 3 we investigate the stability of the numerical scheme, while error estimates are derived in Section 4.

1.1. Notations and assumptions. In what follows we let Ω be a domain in \mathbb{R}^d (with $d = 1, 2$ or 3). Let $J = (0, T]$ be a finite time interval. We are interested in solving equation (1.4) endowed with initial and boundary conditions,

$$(1.5) \quad \begin{aligned} \partial_t b(u) - \nabla \cdot (\nabla u + k(b(u))e_z) &= 0 && \text{in } J \times \Omega, \\ u &= u^0 && \text{in } 0 \times \Omega, \\ u &= 0 && \text{on } J \times \Gamma. \end{aligned}$$

Throughout this paper we make use of the following assumptions:

- (A1) $\Omega \subset \mathbb{R}^d$ is bounded with Lipschitz continuous boundary.
- (A2) $b \in C^1$ is non-decreasing and Lipschitz continuous.

- (A3) $k(b(z))$ is continuous and bounded in z and satisfies, for all $z_1, z_2 \in \mathbb{R}$, $|k(b(z_2)) - k(b(z_1))| \leq C_k(b(z_2) - b(z_1))(z_2 - z_1)$.
 (A4) $b(u_0)$ is essentially bounded (by 0 and 1) in Ω and $u_0 \in L^2(\Omega)$.

Remark 1.1. *By (A3), the convection term is bounded. This restriction is not unrealistic since, for Richards' equation, k stands for the conductivity of the medium. This assumption makes our analysis easier, but can be avoided. Moreover, the growth condition on $k(b(\cdot))$ (see also [21]) relaxes the more often assumed Lipschitz continuity of k (see, e. g., [19], [2]). In addition, source terms can also be considered here, provided that they satisfy a similar growth condition as $k(b(u))$.*

Remark 1.2. *In the transformed version, Richards' equation fits in our framework. However, since b is Lipschitz, a vanishing permeability in (1.1) is not allowed, meaning that our analysis is valid in the variably saturated to fully saturated flow regimes, but not in the completely air saturated one.*

Remark 1.3. *For the sake of simplicity, we deal with homogeneous Dirichlet boundary conditions. More general situations can be included in a straightforward manner, with similar results. Here nonlinearities depend only on the unknown u , not on x and t . For more general situations, techniques developed in [2] can be employed.*

Because of the degenerate character, we do not expect smooth solutions for problem (1.5). For defining a solution in a weak sense we let (\cdot, \cdot) stand for the inner product on $L^2(\Omega)$ or the duality pairing between $H_0^1(\Omega)$ and $H^{-1}(\Omega)$, $\|\cdot\|$ for the norm in $L^2(\Omega)$, $\|\cdot\|_1$ and $\|\cdot\|_{-1}$ for the norms in $H^1(\Omega)$, respectively $H^{-1}(\Omega)$. We use analogous notations for the inner product and the corresponding norm on $L^2(0, T; \mathcal{H})$, with \mathcal{H} being either $L^2(\Omega)$, $H^1(\Omega)$, or $H^{-1}(\Omega)$. In addition, we often write u or $u(t)$ instead of $u(t, x)$ and use C to denote a generic positive constant, not depending on the discretization or regularization parameters.

A weak solution for problem (1.5) is defined as

Definition 1.4. *A function u is called a weak solution for equation (1.5) iff $b(u) \in H^1(0, T; H^{-1}(\Omega))$, $u \in L^2(0, T; H_0^1(\Omega))$, $u(0) = u_0$ (in H^{-1} sense) and for all $\varphi \in L^2(0, T; H_0^1(\Omega))$ it holds*

$$(1.6) \quad \int_0^T (\partial_t b(u(t)), \varphi(t)) + (\nabla u(t) + k(b(u(t)))e_z, \nabla \varphi(t)) dt = 0.$$

Existence, uniqueness and essential bounds for a weak solution of the above problem is studied in several papers (see, for example, [1], [20],

[24] and the references therein). In [1] the following regularity result is obtained

$$(1.7) \quad b(u) \in L^\infty(J; L^1(\Omega)),$$

$$(1.8) \quad \vec{q} := -(\nabla u + k(b(u))e_z) \in L^2(J; (L^2(\Omega))^d).$$

Here $b(u)$ models the water content, hence it is natural to assume that, after scaling, it lies between 0 and 1 for almost every $(t, x) \in J \times \Omega$. For the same reason, in (A4) similar restrictions are imposed to the initial data. Such essential estimates can be shown, for example, if b and k do not depend explicitly on x , or if $k(b(u))$ is constant for $u = 0$ and $u = 1$. Moreover, $u \in L^2(0, T; H_0^1(\Omega))$ yields $b(u) \in L^2(0, T; H_0^1(\Omega))$ due to the Lipschitz continuity of b . Since $b(u) \in H^1(0, T; H^{-1}(\Omega))$ we have $b(u) \in C(0, T; L^2(\Omega))$ (see [18], chapter I), allowing a simplified mixed variational formulation. Following [2] or [26] we integrate (1.5) in time and obtain, for every $t \in J$,

$$(1.9) \quad b(u(t)) + \nabla \cdot \int_0^t \vec{q}(s) ds = b(u^0)$$

in L^2 sense. It follows (see [2] or [22]) that the flux \vec{q} defined in (1.8) satisfies

$$(1.10) \quad \int_0^t \vec{q} d\tau \in H^1(J; (L^2(\Omega))^d) \cap L^2(J; (H^1(\Omega))^d) =: X.$$

2. EQUIVALENT FORMULATIONS

In this section we give the mixed variational formulation and study the equivalence with the conformal one in both continuous and time discrete cases.

2.1. The continuous case. We start with the continuous case. Integrated in time, Problem (1.5) becomes

Problem 1. Find $u \in L^2(J, H_0^1(\Omega))$ such that $b(u) \in L^\infty(J \times \Omega)$, and for all $t \in J$ and $\phi \in H_0^1(\Omega)$ it holds

$$(2.1) \quad (b(u(t)) - b(u^0), \phi) + \int_0^t (\nabla u(s) + k(b(u(s)))e_z, \nabla \phi) ds = 0, .$$

As mentioned in the previous section, this stronger formulation makes sense since $b(u) \in C(J; L^2(\Omega))$.

A mixed formulation for Problem (1.5) reads

Problem 2. Find $(p, \vec{q}) \in L^2(J \times \Omega) \times X$ such that $b(p) \in L^\infty(J \times \Omega)$

and for all $t \in J$ the equations

$$(2.2) \quad (b(p(t)) - b(p^0), w) + (\nabla \tilde{q}(t), w) = 0,$$

$$(2.3) \quad (\tilde{q}(t), v) - \int_0^t (p(s), \nabla v) ds + \int_0^t (k(b(p(s)))e_z, v) ds = 0,$$

hold for all $w \in L^2(\Omega)$ and $v \in H(\operatorname{div}, \Omega)$, with $p^0 = u^0 \in L^2(\Omega)$.

The two problems are equivalent, as stated below. A similar reasoning is already applied in [16].

Proposition 2.1. $u \in L^2(J, H_0^1(\Omega))$ solves Problem 1 iff $(p, \tilde{q}) \in L^2(J \times \Omega) \times X$ defined as

$$(2.4) \quad (p, \tilde{q}) = (u, - \int_0^t (\nabla u(s) + k(b(u(s))))e_z ds)$$

solves Problem 2. Moreover, in this case we have $p \in L^2(J, H_0^1(\Omega))$.

Proof. " \Rightarrow " Let $u \in L^2(J, H_0^1(\Omega))$ be a solution of Problem 1 and (p, \tilde{q}) defined in (2.4). By (1.10) we have $(p, \tilde{q}) \in L^2(J, H_0^1(\Omega)) \times X$. Fixing now $t > 0$, for any $v \in H(\operatorname{div}, \Omega)$, using Green's formula we get

$$\begin{aligned} (\tilde{q}(t), v) &= - \int_0^t (\nabla u(s) + k(b(u(s))))e_z, v ds = \\ &= \int_0^t (p(s), \nabla v) - (k(b(p(s))))e_z, v ds, \end{aligned}$$

so (2.3) is proven.

Next, taking any $\phi \in C_0^\infty(\Omega)$ in equation (2.1) yields

$$\begin{aligned} (b(u(t)) - b(u^0), \phi) &= -(\int_0^t (\nabla u(s) + k(b(u(s))))e_z ds, \nabla \phi) \\ &= (\tilde{q}(t), \nabla \phi) = -(\nabla \tilde{q}(t), \phi). \end{aligned}$$

However, for any $t > 0$, both $b(u(t)) - b(u^0)$ and $\nabla \tilde{q}(t)$ lie in $L^2(\Omega)$, so the above relations still hold for $\phi \in L^2(\Omega)$, implying (2.2).

" \Leftarrow " Let $(p, \tilde{q}) \in L^2(J \times \Omega) \times X$ solving Problem 2 and set $u = p \in L^2(J \times \Omega)$. Taking $v \in (C_0^\infty(\Omega))^d \subset H(\operatorname{div}, \Omega)$ arbitrary, by differentiating (2.3) we get for almost all $t > 0$

$$(2.5) \quad (\partial_t \tilde{q}(t), v) + (k(b(p(t))))e_z, v) = (p(t), \nabla v) = -(\nabla p(t), v),$$

so $\nabla p = -\partial_t \tilde{q} - k(b(p))e_z$ in distributional sense. Since both $\partial_t \tilde{q}$ and $k(b(p))e_z$ are in $L^2(J \times \Omega)$, the same holds for ∇p , so $u = p \in L^2(J, H^1(\Omega))$.

For proving that p vanishes on Γ we use the following ([6], p.91)

Lemma 2.2. *Let $v \in H(\operatorname{div}, \Omega)$ and \vec{n} denote the outer normal to Γ . Then $v \cdot \vec{n}$ is defined in $H^{-1/2}(\Gamma)$ (in the sense of traces) and Green's formula applies for all $p \in H^1(\Omega)$*

$$(2.6) \quad \int_{\Omega} \nabla \cdot v p \, dx + \int_{\Omega} v \cdot \nabla p \, dx = \int_{\Gamma} \vec{n} \cdot v p \, ds.$$

Taking now $v \in H(\operatorname{div}, \Omega)$ in (2.3) gives, for every $t \in J$,

$$- \int_0^t (\nabla p, v) \stackrel{(2.5)}{=} (\tilde{q}(t), v) + \int_0^t (k(b(p))e_z, v) \stackrel{(2.3)}{=} \int_0^t (p, \nabla v).$$

In this way, using (2.6) we get

$$\int_0^t \int_{\Gamma} p v \cdot \vec{n} \, ds = \int_0^t (\nabla p, v) + \int_0^t (p, \nabla v) = 0.$$

Here v was chosen arbitrary, so the trace of p on Γ is zero. Thus $p \in L^2(J, H_0^1(\Omega))$ and the same holds for u .

Moreover, taking any $\phi \in H_0^1(\Omega)$ yields, for all $t > 0$,

$$\begin{aligned} (b(u(t)) - b(u^0), \phi) &\stackrel{(2.2)}{=} -(\nabla \tilde{q}(t), \phi) = (\tilde{q}(t), \nabla \phi) \\ &\stackrel{(2.3)}{=} - \int_0^t (\nabla u(s) + k(b(u(s)))e_z, \nabla \phi) \, ds, \end{aligned}$$

so u solves (2.1). □

2.2. The semi-discrete case. As mentioned in the introduction, for overcoming difficulties due to degeneracy, we first perturb the original equation to obtain a regular parabolic one. Such a technique has been successfully applied in the analysis of degenerate problems, and also allows developing effective numerical schemes (see, e. g. [19]).

In problem (1.5) degeneracy appears due to the vanishing of b' . Therefore we approximate this nonlinearity by b_ϵ , with $\epsilon > 0$ a small perturbation parameter. A possible choice reads

$$(2.7) \quad b_\epsilon(u) = b(u) + \epsilon u.$$

Obviously, b_ϵ is Lipschitz continuous (with the same Lipschitz constant as b , if ϵ is small enough), strictly increasing and its derivative is bounded from below by ϵ . The regularized problem becomes

$$(2.8) \quad \begin{aligned} \partial_t b_\epsilon(u) - \nabla \cdot (\nabla u + k(b(u))e_z) &= 0 && \text{in } (0; T] \times \Omega, \\ u &= u^0 && \text{in } \Omega, \\ u &= 0 && \text{on } J \times \Gamma. \end{aligned}$$

We let $N > 1$ be an integer giving a time step $\tau = T/N$, with $t_n = n\tau$. The regularized semi-discrete conformal problem reads

Problem 3. Let $n = \overline{1, N}$ and u^{n-1} be given. Find $u^n \in H_0^1(\Omega)$ such that, for all $\phi \in H_0^1(\Omega)$,

$$(2.9) \quad (b_\epsilon(u^n) - b_\epsilon(u^{n-1}), \phi) + \tau(\nabla u^n + k(b(u^n))e_z, \nabla \phi) = 0.$$

However, our final aim is a mixed discretization. The time discrete regularized mixed problem becomes

Problem 4. Let $n = \overline{1, N}$ and p^{n-1} given. Find $(p^n, q^n) \in L^2(\Omega) \times H(\text{div}, \Omega)$ such that

$$(2.10) \quad (b_\epsilon(p^n) - b_\epsilon(p^{n-1}), w) + \tau(\nabla q^n, w) = 0,$$

$$(2.11) \quad (q^n, v) - (p^n, \nabla v) + (k(b(p^n))e_z, v) = 0,$$

for all $w \in L^2(\Omega)$, respectively $v \in H(\text{div}, \Omega)$, with $p^0 = u^0 \in L^2(\Omega)$.

As in the continuous case, the two problems above are equivalent.

Proposition 2.3. Let $n = \overline{1, N}$ be fixed and assume $u^{n-1} = p^{n-1}$. Then $u^n \in H_0^1(\Omega)$ solves Problem 3 iff $(p^n, q^n) \in L^2(\Omega) \times H(\text{div}, \Omega)$ defined as

$$(2.12) \quad (p^n, q^n) = (u^n, -(\nabla u^n + k(b(u^n))e_z))$$

solve Problem 4. Moreover, we have $p^n \in H_0^1(\Omega)$.

Proof. " \Rightarrow " Let $u^n \in H_0^1(\Omega)$ be a solution of Problem 3 and (p^n, q^n) be defined in (2.12). For all $v \in H(\text{div}, \Omega)$ we have

$$(q^n, v) = -(\nabla u^n + k(b(u^n))e_z, v) = (p^n, \nabla v) - (k(b(p^n))e_z, v),$$

so (p^n, q^n) verify equation (2.11).

Next, for all $\phi \in C_0^\infty(\Omega)$ (which is dense in $H_0^1(\Omega)$) we get

$$\begin{aligned} (b_\epsilon(p^n) - b_\epsilon(p^{n-1}), \phi) &\stackrel{(2.9)}{=} -\tau(\nabla u^n + k(b(u^n))e_z, \nabla \phi) = \tau(q^n, \nabla \phi) \\ &= -\tau(\nabla q^n, \phi). \end{aligned}$$

But $b_\epsilon(p^n) - b_\epsilon(p^{n-1}) \in L^2(\Omega)$, so $\nabla q^n \in L^2(\Omega)$, implying $q^n \in H(\text{div}, \Omega)$ and equation (2.10) holds by density arguments.

" \Leftarrow " Let $(p^n, q^n) \in L^2(\Omega) \times H(\text{div}, \Omega)$ be a solution of Problem 4 and $u^n = p^n \in L^2(\Omega)$. For any $v \in (C_0^\infty(\Omega))^d \subset H(\text{div}, \Omega)$ we have

$$\begin{aligned} (q^n, v) &\stackrel{(2.11)}{=} (p^n, \nabla v) - (k(b(p^n))e_z, v) \\ &= -(\nabla p^n, v) - (k(b(p^n))e_z, v), \end{aligned}$$

implying

$$\nabla p^n + k(b(p^n))e_z = -q^n$$

in distributional sense. Since both q^n and $k(b(p^n))$ are $L^2(\Omega)$ functions it follows that $p^n \in H^1(\Omega)$. As for the continuous case, using Green's formula (2.6), we get actually $u^n = p^n \in H_0^1(\Omega)$.

Finally, (2.9) results by taking any $\phi \in H_0^1(\Omega)$ in (2.10),

$$\begin{aligned} (b_\epsilon(u^n) - b_\epsilon(u^{n-1}), \phi) &= -\tau(\nabla q^n, \phi) = \tau(q^n, \nabla \phi) \\ &= -\tau((\nabla p^n + k(b(p^n))e_z), \nabla \phi). \end{aligned}$$

□

As resulting from the equivalences proven above, stability and error estimates for the time discrete mixed formulation can be obtained by analyzing the Euler implicit scheme applied to Problem 3. This is the underlying idea in the forthcoming section.

3. STABILITY ESTIMATES

In this section we investigate the stability of our numerical approach. We make use of the lemmas below.

Lemma 3.1. *For any vectors $a_k, b_k \in \mathbb{R}^q$ ($k = \overline{1, N}, q \geq 1$) we have*

$$(3.1) \quad 2 \sum_{n=1}^N a_n \sum_{k=1}^n a_k = \left(\sum_{n=1}^N a_n \right)^2 + \sum_{n=1}^N (a_n)^2,$$

$$(3.2) \quad 2 \sum_{n=1}^N (a_n - a_{n-1}, a_n) = |a_N|^2 - |a_0|^2 + \sum_{n=1}^N |a_n - a_{n-1}|^2,$$

$$(3.3) \quad \sum_{n=1}^N (a_n - a_{n-1}, b_n) = a_N b_N - a_0 b_0 + \sum_{n=1}^N (b_n - b_{n-1}, a_{n-1}).$$

Lemma 3.2. *Under the assumption (A1), for any real sequence x^j , $j = \overline{1, n}$ we have*

$$(3.4) \quad \sum_{j=1}^n (b_\epsilon(x^j) - b_\epsilon(x^{j-1})) x^j \geq -C|x^0|^2 + \frac{\epsilon}{2}|x^n|^2.$$

Proof. Since $b'_\epsilon \geq \epsilon$, one has, for any reals x and y ,

$$((b_\epsilon(x) - b_\epsilon(y))x) \geq \int_y^x sb'_\epsilon ds \quad \text{and} \quad \int_0^x sb'_\epsilon(s) ds \geq \frac{\epsilon}{2}x^2.$$

Furthermore,

$$\begin{aligned} \sum_{j=1}^n (b_\epsilon(x^j) - b_\epsilon(x^{j-1}))x^j &\geq \sum_{j=1}^n \int_{x^{j-1}}^{x^j} sb'_\epsilon(s) ds \\ &= \int_0^{x^n} sb'_\epsilon(s) ds - \int_0^{x^0} sb'_\epsilon(s) ds \geq -C|x^0|^2 + \frac{\epsilon}{2}|x^n|^2, \end{aligned}$$

where the constant C is half of the Lipschitz constant of b . □

3.1. Stability in the time discrete conformal case.

Proposition 3.3. *Assume (A1) - (A4). If u^n solves Problem 3 ($n = \overline{1, N}$), we have*

$$(3.5) \quad \tau \sum_{n=1}^N \|u^n\|_1^2 \leq C.$$

Proof. Taking $\phi = u^n$ in (2.9) and summing up for $n = \overline{1, N}$ gives

$$(3.6) \quad \sum_{n=1}^N (b_\epsilon(u^n) - b_\epsilon(u^{n-1}), u^n) + \sum_{n=1}^N \tau \|\nabla u^n\|^2 \\ + \sum_{n=1}^N \tau (k(b(u^n))e_z, \nabla u^n) = 0.$$

Now we estimate the terms on the left in the above. By (3.4), since $u^0 \in L^2(\Omega)$,

$$\sum_{n=1}^N (b_\epsilon(u^n) - b_\epsilon(u^{n-1}), u^n) \geq -C.$$

The second term needs no further treatment. Finally, since k is bounded, applying the Cauchy-Schwartz inequality, we get

$$\tau \sum_{n=1}^N |(k(b(u^n))e_z, \nabla u^n)| \leq \frac{\tau}{2} \sum_{n=1}^N \|k(b(u^n))e_z\|^2 + \frac{\tau}{2} \sum_{n=1}^N \|\nabla u^n\|^2 \\ \leq C + \frac{\tau}{2} \sum_{n=1}^N \|\nabla u^n\|^2.$$

Inserting the last inequalities into (3.6) and using the inequality of Poincaré gives (3.5). \square

3.2. Stability for the time-discrete mixed formulation. By the equivalence of Problems 3 and 4, Proposition 3.3 provides stability for the time discrete solutions p^n and q^n .

Proposition 3.4. *Assuming (A1) - (A4), if, for any $n = \overline{1, N}$, (p^n, q^n) solve Problem 4, we have*

$$(3.7) \quad \tau \sum_{n=1}^N \|p^n\|_1^2 + \tau \sum_{n=1}^N \|q^n\|^2 \leq C.$$

Proof. The estimate for p^n is a direct consequence of (3.5). Next, taking $w = p^n$ in (2.10) and $v = \tau q^n$ in (2.11) yields

$$(b_\epsilon(p^n) - b_\epsilon(p^{n-1}), p^n) + \tau (\nabla q^n, p^n) = 0, \\ (q^n, \tau q^n) - (p^n, \nabla \tau q^n) + (k(b(p^n))e_z, \tau q^n) = 0.$$

Adding these two equations and summing up for $n = 1$ to N gives

$$\sum_{n=1}^N (b_\epsilon(p^n) - b_\epsilon(p^{n-1}), p^n) + \tau \sum_{n=1}^N \|q^n\|^2 + \tau \sum_{n=1}^N (k(b(p^n))e_z, q^n) = 0,$$

and the rest of the proof follows exactly as in the previous lemma. \square

Other stability estimates can be obtained defining an initial flux $q^0 \in [L^2(\Omega)]^d$. In doing so we take $\rho \in C_0^\infty(B_d(0, 1))$ ($B_d(0, 1)$ being the unit ball in \mathbb{R}^d) so that $\int_{B_d(0,1)} \rho(x)dx = 1$ and consider the mollifier sequence $\left\{ \rho_\mu(x) = \frac{1}{\mu^d} \rho\left(\frac{x}{\mu}\right) \right\}_{1 \gg \mu > 0}$. Defining q^0 as

$$(3.8) \quad q^0 = -\nabla(\rho_\mu * p^0) - k(b(p^0))e_z,$$

with μ to be chosen further and $*$ denoting the convolution operator, for any $v \in H(\text{div}, \Omega)$ we have

$$(3.9) \quad (q^0, v) - (\rho_\mu * p^0, \nabla v) + (k(b(p^0))e_z, v) = 0.$$

A mollifying of p^0 in the above is necessary for having $q^0 \in [L^2(\Omega)]^d$. However, since $p^0 \in L^2(\Omega)$, $\|p^0 - \rho_\mu * p^0\|$ goes to 0 as $\mu \searrow 0$, so $\|q^0\|$ is uniformly bounded with respect to μ . Now the following estimates can be obtained.

Proposition 3.5. *Assuming (A1) - (A4), if, for all $n = \overline{1, N}$, (p^n, q^n) solve Problem 4, for any $k > 0$ we have*

$$(3.10) \quad \sum_{n=1}^k (b_\epsilon(p^n) - b_\epsilon(p^{n-1}), p^n - p^{n-1}) + \tau \|q^k\|^2 + \tau \sum_{n=1}^k \|q^n - q^{n-1}\|^2 \leq C\tau.$$

Proof. First we take $w = p^n - p^{n-1} \in L^2(\Omega)$ in (2.10) and subtract equation (2.11) at time step $n-1$ from the one at time step n . Testing with $v = \tau q^n$ in the resulting equality yields

$$\begin{aligned} (b_\epsilon(p^n) - b_\epsilon(p^{n-1}), p^n - p^{n-1}) + \tau (\nabla q^n, p^n - p^{n-1}) &= 0, \\ \tau (q^n - q^{n-1}, q^n) - \tau (p^n - p^{n-1}, \nabla q^n) \\ + \tau ((k(b(p^n)) - k(b(p^{n-1})))e_z, q^n) &= 0. \end{aligned}$$

For $n = 1$ the second equation above reads

$$\begin{aligned} \tau (q^1 - q^0, q^1) - \tau (p^1 - p^0, \nabla q^1) + \tau ((k(b(p^1)) - k(b(p^0)))e_z, q^1) \\ = \tau (p^0 - \rho_\mu * p^0, \nabla q^1). \end{aligned}$$

Adding the above pairs of equalities and summing the result up for $n = \overline{1, k}$ yields

$$(3.11) \quad \begin{aligned} \sum_{n=1}^k (b_\epsilon(p^n) - b_\epsilon(p^{n-1}), p^n - p^{n-1}) + \tau \sum_{n=1}^k (q^n - q^{n-1}, q^n) \\ + \tau \sum_{n=1}^k ((k(b(p^n)) - k(b(p^{n-1})))e_z, q^n) = \tau (p^0 - \rho_\mu * p^0, \nabla q^1). \end{aligned}$$

Denoting the terms above by T_1, \dots, T_4 , we first notice that T_1 is positive by the monotonicity of b_ϵ . Next, by (3.2),

$$\begin{aligned} T_2 &= \tau \sum_{n=1}^k (q^n - q^{n-1}, q^n) \\ &= \frac{\tau}{2} \|q^k\|^2 - \frac{\tau}{2} \|q^0\|^2 + \frac{\tau}{2} \sum_{n=1}^k \|q^n - q^{n-1}\|^2. \end{aligned}$$

Recalling (A3) and the Cauchy-Schwartz inequality, for T_3 we get

$$\begin{aligned} |T_3| &\leq \frac{\delta_1}{2} \sum_{n=1}^k \|(k(b(p^n)) - k(b(p^{n-1})))e_z\|^2 + \frac{\tau^2}{2\delta_1} \sum_{n=1}^k \|q^n\|^2 \\ &\leq \frac{\delta_1 C_k}{2} \sum_{n=1}^k (b(p^n) - b(p^{n-1}), p^n - p^{n-1}) + \frac{\tau^2}{2\delta_1} \sum_{n=1}^k \|q^n\|^2. \end{aligned}$$

Estimating T_4 follows as before,

$$|T_4| \leq \tau \|p^0 - \rho_\mu * p^0\| \|\nabla q^1\| \leq \delta_2 \|p^0 - \rho_\mu * p^0\|^2 + \frac{\tau^2}{4\delta_2} \|\nabla q^1\|^2.$$

To estimate $\|\nabla q^1\|$ we use (2.10) for $n = 1$, test with $w = \nabla q^1 \in L^2(\Omega)$ and obtain

$$\begin{aligned} \tau \|\nabla q^1\|^2 &\leq \|b_\epsilon(p^1) - b_\epsilon(p^0)\| \|\nabla q^1\| \\ &\leq \frac{C}{2\tau} (b_\epsilon(p^1) - b_\epsilon(p^0), p^1 - p^0) + \frac{\tau}{2} \|\nabla q^1\|^2, \end{aligned}$$

by the Lipschitz continuity of b_ϵ . In this way we get

$$\tau \|\nabla q^1\|^2 \leq \frac{C}{\tau} (b_\epsilon(p^1) - b_\epsilon(p^0), p^1 - p^0).$$

Using these estimates in (3.11) and choosing the δ 's properly gives

$$\begin{aligned} \sum_{n=1}^k (b_\epsilon(p^n) - b_\epsilon(p^{n-1}), p^n - p^{n-1}) + \tau \|q^k\|^2 + \tau \sum_{n=1}^k \|q^n - q^{n-1}\|^2 \\ \leq C_1 \tau + C_2 \|p^0 - \rho_\mu * p^0\|^2 + C_3 \tau^2 \sum_{n=1}^k \|q^n\|^2. \end{aligned}$$

We still have to choose μ in (3.8). Since $\|p^0 - \rho_\mu * p^0\|$ converges to 0, taking μ sufficiently small, the right term in the above becomes

$$C_4 \tau + C_3 \tau^2 \sum_{n=1}^k \|q^n\|^2.$$

Now (3.10) follows by the discrete Gronwall lemma. \square

Remark 3.6. If $p^0 \in H^1(\Omega)$, q^0 can be defined without using a mollifier,

$$(3.12) \quad q^0 = -\nabla p^0 - k(b(p^0))e_z.$$

Then $T_4 = 0$ in (3.11), without changing (3.10).

A direct consequence of the stability estimates above is

Proposition 3.7. *In the setting of Proposition 3.5 we have*

$$(3.13) \quad \sum_{n=1}^N \tau \|\nabla \cdot q^n\|^2 \leq C$$

Proof. Taking $w = \nabla \cdot q^j$ in equation (2.10) and applying the Cauchy-Schwartz inequality one gets

$$\tau \|\nabla \cdot q^j\|^2 \leq \frac{1}{2\tau} \|b_\epsilon(p^j) - b_\epsilon(p^{j-1})\|^2 + \frac{\tau}{2} \|\nabla \cdot q^j\|^2,$$

so

$$\tau \|\nabla \cdot q^j\|^2 \leq \frac{1}{\tau} \|b_\epsilon(p^j) - b_\epsilon(p^{j-1})\|^2.$$

Summing up the above for $j = \overline{1, N}$, using the Lipschitz continuity of b_ϵ and (3.10) leads to (3.13). \square

4. ERROR ESTIMATES

In this section we obtain a priori error estimates for both time discrete scheme, as well as for the fully discrete one.

4.1. Error estimates for the semi-discrete approximation. To obtain error estimates for the time discrete scheme we employ techniques developed in [19] and make use of the Green operator $G : H^{-1}(\Omega) \rightarrow H_0^1(\Omega)$ defined as

$$(4.1) \quad (\nabla(G\psi), \nabla\phi) = (\psi, \phi), \quad \text{for all } \phi \in H_0^1(\Omega).$$

Obviously, G is linear and self adjoint. Moreover, by the Cauchy-Schwartz inequality, using (3.3) yields

Lemma 4.1. *For all $f, f_k \in H^{-1}(\Omega)$ ($k = \overline{1, N}$) and $g \in H^1(\Omega)$ we have*

$$(f, g) \leq \|f\|_{-1} \|\nabla g\|,$$

$$\|\nabla Gf\|^2 = (f, Gf) = \|f\|_{-1}^2,$$

$$2 \sum_{k=1}^N (f_k - f_{k-1}, Gf_k) = \|f_k\|_{-1, \Omega}^2 - \|f_0\|_{-1, \Omega}^2 + \sum_{k=1}^N \|f_k - f_{k-1}\|_{-1, \Omega}^2.$$

Further we use the notations

$$(4.2) \quad \begin{aligned} \bar{u}^n &= \frac{1}{\tau} \int_{t_{n-1}}^{t_n} u(t) dt, \\ u_\Delta(t) &= u^n, \quad \text{for } t \in (t_{n-1}, t_n], \\ e_b(u) &= b(u) - b_\epsilon(u_\Delta), \end{aligned}$$

where $n = \overline{1, N}$ and $\bar{u}^0 = u^0$.

It is worth pointing out here that, by Propositions 2.1 and 2.3, estimates obtained for the conformal discretization can be transferred to the mixed case.

Proposition 4.2. *Assuming (A1) - (A4), if u is the weak solution of Problem 1 and u^n solves, for each $n = \overline{1, N}$, Problem 3, then*

$$(4.3) \quad \begin{aligned} & \max_{n=\overline{1, N}} \left\| \overline{e_b(u)^n} \right\|_{-1}^2 + \|e_b(u)\|_{L^2(J \times \Omega)}^2 + \\ & + \int_0^T (b_\epsilon(u(t)) - b_\epsilon(u_\Delta), u(t) - u_\Delta) dt \leq C(\tau + \epsilon). \end{aligned}$$

Proof. Subtracting equation (2.1) at $t = t_{j-1}$ from the one at $t = t_j$ and then subtracting (2.9) with $n = j$ from the resulting gives

$$\begin{aligned} & (b(u(t_j)) - b(u(t_{j-1})) - b_\epsilon(u^j) + b_\epsilon(u^{j-1}), \phi) \\ & + \tau (\nabla(\overline{u^j} - u^j), \nabla\phi) + \tau \left((\overline{k(b(u))^j} - k(b(u^j)))e_z, \nabla\phi \right) = 0. \end{aligned}$$

Taking $\phi = \overline{Ge_b(u)^j} \in H_0^1(\Omega)$ into above and summing up for $j = \overline{1, n}$ (with $n \leq N$) yields

$$(4.4) \quad \begin{aligned} & \sum_{j=1}^n \left(b(u(t_j)) - b(u(t_{j-1})) - b_\epsilon(u^j) + b_\epsilon(u^{j-1}), \overline{Ge_b(u)^j} \right) \\ & + \sum_{j=1}^n \tau \left(\nabla\overline{u^j} - \nabla u^j, \nabla\overline{Ge_b(u)^j} \right) \\ & + \sum_{j=1}^n \tau \left((\overline{k(b(u))^j} - k(b(u^j)))e_z, \nabla\overline{Ge_b(u)^j} \right) = 0. \end{aligned}$$

We estimate now each of terms in (4.4), denoted by T_1 , T_2 and T_3 .

$$\begin{aligned} T_1 & = \sum_{j=1}^n \left(b(u(t_j)) - \overline{b(u)^j} - b(u(t_{j-1})) + \overline{b(u)^{j-1}}, \overline{Ge_b(u)^j} \right) \\ & + \sum_{j=1}^n \left(\overline{b(u)^j} - b_\epsilon(u^j) - \overline{b(u)^{j-1}} + b_\epsilon(u^{j-1}), \overline{Ge_b(u)^j} \right) \\ & =: T_{11} + T_{12}. \end{aligned}$$

Further, by (3.3) and recalling that $b(u(0)) = \overline{b(u)^0}$ we have

$$\begin{aligned} T_{11} & = \sum_{j=1}^n \left(b(u(t_j)) - \overline{b(u)^j} - b(u(t_{j-1})) + \overline{b(u)^{j-1}}, \overline{Ge_b(u)^j} \right) \\ & = \left(b(u(t_n)) - \overline{b(u)^n}, \overline{Ge_b(u)^n} \right) \\ & \quad - \sum_{j=1}^n \left(b(u(t_{j-1})) - \overline{b(u)^{j-1}}, \overline{Ge_b(u)^j} - \overline{Ge_b(u)^{j-1}} \right) \\ & =: T_{111} - T_{112}. \end{aligned}$$

For T_{111} we make use of Lemma 4.1 and obtain

$$(4.5) \quad \begin{aligned} |T_{111}| & \leq \frac{1}{\tau} \int_{t_{n-1}}^{t_n} \left| \left(b(u(t_n)) - b(u(t)), \overline{Ge_b(u)^n} \right) \right| dt \\ & \leq \frac{1}{\tau} \int_{t_{n-1}}^{t_n} \int_t^{t_n} \left| \left(\partial_s b(u(s)), \overline{Ge_b(u)^n} \right) \right| ds dt \\ & \leq \frac{1}{\tau} \int_{t_{n-1}}^{t_n} \sqrt{\tau} \|\partial_t b(u)\|_{L^2(t_{n-1}, t_n; H^{-1})} \|\overline{e_b(u)^n}\|_{-1} dt \\ & \leq \sqrt{\tau} \|\partial_s b(u)\|_{L^2(t_{n-1}, t_n; H^{-1})} \|\overline{e_b(u)^n}\|_{-1} \\ & \leq \tau \|\partial_s b(u)\|_{L^2(t_{n-1}, t_n; H^{-1})}^2 + \frac{1}{4} \|\overline{e_b(u)^n}\|_{-1}^2. \end{aligned}$$

Proceeding as before, T_{112} can be estimated to

$$(4.6) \quad |T_{112}| \leq \tau \|\partial_t b(u)\|_{L^2(0,t_n;H^{-1})}^2 + \frac{1}{4} \sum_{j=1}^n \|\overline{e_b(u)^j} - \overline{e_b(u)^{j-1}}\|_{-1}^2.$$

Using Lemma 4.1 again, since $\overline{e_b(u)^0} = 0$, T_{12} gives

$$(4.7) \quad \begin{aligned} T_{12} &= \sum_{j=1}^n \left(\overline{b(u)^j} - b_\epsilon(u^j) - \overline{b(u)^{j-1}} + b_\epsilon(u^{j-1}), G\overline{e_b(u)^j} \right) \\ &= \frac{1}{2} \left(\overline{e_b(u)^n}, G\overline{e_b(u)^n} \right) \\ &\quad + \frac{1}{2} \sum_{j=1}^n \left(\overline{e_b(u)^j} - \overline{e_b(u)^{j-1}}, G\overline{e_b(u)^j} - G\overline{e_b(u)^{j-1}} \right) \\ &= \frac{1}{2} \|\overline{e_b(u)^n}\|_{-1}^2 + \frac{1}{2} \sum_{j=1}^n \|\overline{e_b(u)^j} - \overline{e_b(u)^{j-1}}\|_{-1}^2. \end{aligned}$$

For T_2 we have

$$\begin{aligned} T_2 &= \sum_{j=1}^n \tau \left(\nabla \overline{u^j} - \nabla u^j, \nabla G\overline{e_b(u)^j} \right) \\ &= \tau \sum_{j=1}^n \left(\frac{1}{\tau} \int_{t_{j-1}}^{t_j} (u(t) - u^j) dt, \frac{1}{\tau} \int_{t_{j-1}}^{t_j} (b(u(s)) - b_\epsilon(u^j)) ds \right) \\ &= \sum_{j=1}^n \int_{t_{j-1}}^{t_j} (u(t) - u^j, b(u(t)) - b_\epsilon(u^j)) dt \\ &\quad + \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \left(u(t) - u^j, \frac{1}{\tau} \int_{t_{j-1}}^{t_j} (b(u(s)) - b(u(t))) ds \right) dt \\ &=: T_{21} + T_{22}. \end{aligned}$$

T_{21} can be decomposed as below

$$\begin{aligned} T_{21} &= \sum_{j=1}^n \int_{t_{j-1}}^{t_j} (u(t) - u^j, b(u(t)) - b_\epsilon(u(t))) dt \\ &\quad + \sum_{j=1}^n \int_{t_{j-1}}^{t_j} (u(t) - u^j, b_\epsilon(u(t)) - b_\epsilon(u^j)) dt =: T_{211} + T_{212}. \end{aligned}$$

The definition of b_ϵ in (2.7) gives

$$(4.8) \quad \begin{aligned} |T_{211}| &= \left| \sum_{j=1}^n \int_{t_{j-1}}^{t_j} (u(t) - u^j, \epsilon u(t)) dt \right| \\ &\leq \epsilon \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \|u(t) - u^j\| \|u(t)\| dt \\ &\leq \frac{\epsilon}{4} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \|u(t) - u^j\|^2 dt + \epsilon \|u\|_{L^2(0,t_n;L^2(\Omega))}^2. \end{aligned}$$

Since b_ϵ is monotone, T_{212} is positive; moreover, it holds

$$(4.9) \quad \begin{aligned} T_{212} &\geq \frac{1}{2} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} (u(t) - u^j, b_\epsilon(u(t)) - b_\epsilon(u^j)) dt \\ &\quad + \frac{\epsilon}{2} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \|u(t) - u^j\|^2 dt. \end{aligned}$$

Proceeding as for T_{111} , recalling the apriori estimates in Proposition 3.3, since $b(u) \in H^1(0, T; H^{-1})$ and $u \in L^2(0, T; H^1)$ we obtain

$$\begin{aligned}
(4.10) \quad T_{22} &= \frac{1}{\tau} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t_j} (u(t) - u^j, b(u(s)) - b(u(t))) ds dt \\
&= \frac{1}{\tau} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t_j} \left(\int_t^s (u(t) - u^j, \partial_\tau b(u)) dr \right) ds dt \\
&\leq \frac{1}{\tau} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t_j} \int_t^s \|\nabla(u(t) - u^j)\| \|\partial_\tau b(u)\|_{-1} dr ds dt \\
&\leq \frac{\tau}{2} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \|\nabla(u(t) - u^j)\|^2 + \frac{\tau}{2} \|\partial_\tau b(u)\|_{L^2(0, t_n; H^{-1})}^2 \\
&\leq C \tau.
\end{aligned}$$

For the T_3 we go on as follows

$$\begin{aligned}
|T_3| &\leq \frac{\tau}{4\delta} \sum_{j=1}^n \|\overline{k(b(u))^j} - k(b(u^j))\|^2 + \delta \tau \sum_{j=1}^n \|\overline{e_b(u)^j}\|_{-1}^2 \\
&= T_{31} + \delta \tau \sum_{j=1}^n \|\overline{e_b(u)^j}\|_{-1}^2.
\end{aligned}$$

Applying (A3) and taking $\delta = C_k$ gives

$$\begin{aligned}
(4.11) \quad |T_{31}| &= \frac{\tau}{4\delta} \frac{1}{\tau^2} \sum_{j=1}^n \int_\Omega \left(\int_{t_{j-1}}^{t_j} (k(b(u)) - k(b(u^j))) dt \right)^2 dx \\
&\leq \frac{1}{4\delta\tau} \sum_{j=1}^n \int_\Omega \tau \int_{t_{j-1}}^{t_j} (k(b(u)) - k(b(u^j)))^2 dt dx \\
&\leq \frac{1}{4} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} (b(u) - b(u^j), u - u^j) dt \\
&\leq \frac{1}{4} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} (b_\epsilon(u) - b_\epsilon(u^j), u - u^j) dt.
\end{aligned}$$

Since $b(u) \in H^1(0, T; H^{-1})$ and $u \in L^2(J; H^1(\Omega))$, inserting (4.5) - (4.11) into (4.4) yields

$$\begin{aligned}
&\|\overline{e_b(u)^n}\|_{-1}^2 + \sum_{j=1}^n \|\overline{e_b(u)^j} - \overline{e_b(u)^{j-1}}\|_{-1}^2 + \epsilon \int_{t_{j-1}}^{t_j} \|u(t) - u^j\|^2 dt \\
&+ \sum_{j=1}^n \int_{t_{j-1}}^{t_j} (u(t) - u^j, b_\epsilon(u(t)) - b_\epsilon(u^j)) dt \\
&\leq C(\tau + \epsilon) + 4C_k\tau \sum_{j=1}^n \|\overline{e_b(u)^j}\|_{-1}^2,
\end{aligned}$$

and (4.3) is a direct consequence of the discrete Gronwall lemma. \square

Using the above result an error estimate for the L^2 norm of the time integrated gradient can be obtained. Such an estimate is essential for our analysis because it provides also an error estimate for the time integral of the flux in the mixed formulation.

Proposition 4.3. *Under the assumptions in Proposition 4.2 we have*

$$(4.12) \quad \left\| \int_0^T (u(t) - u_\Delta(t)) dt \right\|_1^2 \leq C(\tau + \epsilon).$$

Proof. Following the ideas in [19], we first add (2.9) for $n = 1$ to N , subtract the resulting from (2.1) at $t = t_N = T$ and end up with

$$\begin{aligned} & (b(u(T)) - b_\epsilon(u^N), \phi) - (b(u(t_0)) - b_\epsilon(u^0), \phi) \\ & + \left(\sum_{j=1}^N \int_{t_{j-1}}^{t_j} \nabla(u(t) - u^j) dt, \nabla \phi \right) \\ & + \left(\sum_{j=1}^N \int_{t_{j-1}}^{t_j} (k(b(u)) - k(b(u^j))) e_z dt, \nabla \phi \right) = 0 \end{aligned}$$

for all $\phi \in H_0^1(\Omega)$. Taking now $\phi = \sum_{j=1}^N \tau (\bar{u}^j - u^j)$ into above gives

$$\begin{aligned} & \left(b(u(T)) - b_\epsilon(u^N), \tau \sum_{j=1}^N (\bar{u}^j - u^j) \right) \\ (4.13) \quad & - \left(\epsilon u^0, \tau \sum_{j=1}^N (\bar{u}^j - u^j) \right) + \left\| \tau \sum_{j=1}^N \nabla (\bar{u}^j - u^j) \right\|^2 \\ & + \left(\sum_{j=1}^N \int_{t_{j-1}}^{t_j} (k(b(u)) - k(b(u^j))) e_z, \nabla \sum_{j=1}^N \tau (\bar{u}^j - u^j) \right) = 0. \end{aligned}$$

Denoting the terms in (4.13) by T_1 , T_2 , T_3 and T_4 we proceed by estimating each of them separately. T_1 yields

$$T_1 = \left(b(u(T)) - \overline{b(u)}^N + \overline{b(u)}^N - b_\epsilon(u^N), \sum_{j=1}^N \tau (\bar{u}^j - u^j) \right) =: T_{11} + T_{12}.$$

As in (4.5), since $\partial_t b(u) \in L^2(0, T; H^{-1})$, T_{11} gives

$$\begin{aligned} (4.14) \quad |T_{11}| & \leq \frac{1}{\tau} \int_{t_{N-1}}^{t_N} \int_t^{t_N} \left| \left(\partial_s b(u), \sum_{j=1}^N \tau (\bar{u}^j - u^j) \right) \right| ds dt \\ & \leq \frac{C_{11}}{2\delta_{11}} \tau + \frac{\delta_{11}}{2} \left\| \sum_{j=1}^N \tau (\bar{u}^j - u^j) \right\|_1^2. \end{aligned}$$

Applying the Cauchy-Schwartz inequality, for T_{12} we obtain

$$(4.15) \quad |T_{12}| \leq \frac{1}{2\delta_{12}} \left\| \overline{e_b(u)}^N \right\|_{-1}^2 + \frac{\delta_{12}}{2} \left\| \sum_{j=1}^N \tau (\bar{u}^j - u^j) \right\|_1^2.$$

Analogous, T_2 gives

$$(4.16) \quad |T_2| \leq \frac{1}{2\delta_2} \epsilon \|u^0\|^2 + \frac{\delta_2}{2} \left\| \sum_{j=1}^N \tau (\bar{u}^j - u^j) \right\|_1^2.$$

For T_3 we recall the inequality of Poincaré

$$(4.17) \quad T_3 = \left\| \sum_{j=1}^N \tau \nabla (\bar{u}^j - u^j) \right\|^2 \geq C \left\| \sum_{j=1}^N \tau (\bar{u}^j - u^j) \right\|_1^2.$$

Analogous, T_4 can be estimated as

$$\begin{aligned} |T_4| & \leq \frac{1}{2\delta_4} \left\| \sum_{j=1}^N \int_{t_{j-1}}^{t_j} (k(b(u)) - k(b(u^j))) e_z dt \right\|^2 \\ & \quad + \frac{\delta_4}{2} \left\| \nabla \sum_{j=1}^N \tau (\bar{u}^j - u^j) \right\|^2. \end{aligned}$$

For the first term above - denoted by T_{41} - we get, by (A3),

$$(4.18) \quad \begin{aligned} T_{41} &\leq N \sum_{j=1}^N \tau \int_{t_{j-1}}^{t_j} \|k(b(u)) - k(b(u^j))\|_0^2 dt \\ &\leq TC_k \sum_{j=1}^N \int_{t_{j-1}}^{t_j} (b(u(t)) - b(u^j), u(t) - u^j). \end{aligned}$$

Inserting (4.14) - (4.18) into (4.13), choosing the δ 's properly and recalling the estimates in Proposition 4.2 proves (4.12). \square

Propositions 4.2 and 4.3 can be summarized in the following

Theorem 4.4. *If u is the solution of Problem 1 and u^n solves Problem 3 ($n = \overline{1, N}$), we have*

$$(4.19) \quad \begin{aligned} &\max_{n=\overline{1, N}} \|\overline{e_b(u)^n}\|_{-1}^2 + \|e_b(u)\|_{L^2(J \times \Omega)}^2 + \left\| \int_0^T (u(t) - u_\Delta(t)) dt \right\|_1^2 \\ &\quad + \int_0^T (b_\epsilon(u(t)) - b_\epsilon(u_\Delta(t)), u(t) - u_\Delta(t)) dt \leq C(\tau + \epsilon). \end{aligned}$$

Remark 4.5. *The estimates above do not change if we replace the last term on the left by $\int_0^T (b(u(t)) - b(u_\Delta(t)), u(t) - u_\Delta(t)) dt$.*

Since Problems 3 and 4 are equivalent we immediately obtain

Theorem 4.6. *In the setting of Theorem 4.4, if (p^n, q^n) solve Problem 4 ($n = \overline{1, N}$), we get*

$$(4.20) \quad \begin{aligned} &\sum_{n=1}^N \int_{t_{n-1}}^{t_n} (b_\epsilon(u(t)) - b_\epsilon(p^n), u(t) - p^n) dt \\ &+ \left\| \sum_{n=1}^N \int_{t_{n-1}}^{t_n} (u(t) - p^n) dt \right\|_1^2 + \|\tilde{q}(T) - \tau \sum_{n=1}^N q^n\|^2 \\ &\leq C(\tau + \epsilon). \end{aligned}$$

Remark 4.7. *As in Remark 4.5, we can replace the scalar product in (4.20) by $\int_0^T (b(u(t)) - b(u_\Delta(t)), u(t) - u_\Delta(t)) dt$. This immediately implies an error estimate for the saturation,*

$$\sum_{n=1}^N \int_{t_{n-1}}^{t_n} \|b(u(t)) - b(p^n)\|^2 dt \leq C(\tau + \epsilon).$$

4.2. Error estimates for the fully discrete mixed discretization.

The next step in our analysis is proving error estimates for the fully discrete approximation. We first estimate the error for the flux variable and then proceed with estimates for the p unknowns.

In doing so we denote by W and V the spaces $L^2(\Omega)$ and $H(\text{div}, \Omega)$. Let \mathcal{T}_h be a regular decomposition of $\Omega \subset \mathbb{R}^d$ into closed d -simplices; h stands for the mesh-size (see [8]). Here we assume $\overline{\Omega} = \cup_{T \in \mathcal{T}_h} T$, hence Ω is polygonal. Thus we neglect the errors caused by an approximation of a non-polygonal domain, avoiding so an excess of technicalities (a complete analysis in this sense can be found in [19]).

The discrete subspaces $W_h \times V_h \subset W \times V$ are defined as

$$(4.21) \quad \begin{aligned} W_h &:= \{p \in W \mid p \text{ is constant on each element } T \in \mathcal{T}_h\}, \\ V_h &:= \{\vec{q} \in V \mid \vec{q}|_T = \vec{a} + b \vec{x} \text{ for all } T \in \mathcal{T}_h\}. \end{aligned}$$

So W_h denotes the space of piecewise constant functions, while V_h is the RT_0 space (see [6]). Further we make use of the usual L^2 projector

$$(4.22) \quad P_h : L^2(\Omega) \rightarrow W_h, \quad ((P_h w - w), w_h) = 0 \quad \forall w_h \in W_h.$$

Taking an \tilde{V} slightly better than V (for example, $V \cap (L^s(\Omega))^d$ with an $s > 2$), a projector Π_h can be defined as (see [6], p.131)

$$(4.23) \quad \Pi_h : \tilde{V} \rightarrow V_h, \quad (\nabla \cdot (\Pi_h v - v), w_h) = 0$$

for all $w_h \in W_h$. With $r \geq 0$, for the operators defined above we have

$$(4.24) \quad \begin{aligned} \|w - P_h w\| &\leq Ch^r \|w\|_r, \\ \|v - \Pi_h v\| &\leq Ch^r \|v\|_r, \end{aligned}$$

for any $w \in H^r(\Omega)$ and $v \in (H^r(\Omega))^d$.

The following technical lemma is proven in [22]

Lemma 4.8. *Assuming (A1), taking $f_h \in W_h$, a $v_h \in V_h$ exists so that*

$$\begin{aligned} \nabla \cdot v_h &= f_h, \\ \|v_h\| &\leq C \|\nabla \cdot v_h\|, \end{aligned}$$

$C > 0$ being a generic constant not depending on h , f_h or v_h .

Before proceeding with the fully discrete approximation scheme we rewrite Problem 4 (continuous in space) as

Problem 5. Let $n = \overline{1, N}$. Find $(p^n, q^n) \in W \times V$ such that

$$(4.25) \quad (b_\epsilon(p^n), w) - (b_\epsilon(p^0), w) + \tau \left(\sum_{j=1}^n \nabla q^j, w \right) = 0,$$

$$(4.26) \quad (q^n, v) - (p^n, \nabla v) + (k(b(p^n))e_z, v) = 0$$

for all $w \in W$ and $v \in V$, with $p^0 = u^0$.

The fully discrete mixed finite element approximation reads

Problem 6. Let $n = \overline{1, N}$. Find $(p_h^n, q_h^n) \in W_h \times V_h$ such that

$$(4.27) \quad (b_\epsilon(p_h^n), w_h) + \tau \left(\sum_{j=1}^n \nabla q_h^j, w_h \right) = (b_\epsilon(p_h^0), w_h),$$

$$(4.28) \quad (q_h^n, v_h) - (p_h^n, \nabla v_h) + (k(b(p_h^n))e_z, v_h) = 0$$

for all $w_h \in W_h$ and $v_h \in V_h$.

Initially we take $p_h^0 = b_\epsilon^{-1}(P_h b_\epsilon(u^0))$. Since $P_h b_\epsilon(u^0)$ is constant on any $T \in \mathcal{T}_h$, the same holds for $b_\epsilon^{-1}(P_h b_\epsilon(u^0))$, so $p_h^0 \in W_h$. Moreover, with this choice, for all $w_h \in W_h$, we obtain

$$(b_\epsilon(p_h^0), w_h) = (b_\epsilon(u^0), w_h) = (b_\epsilon(p^0), w_h).$$

We start with apriori estimates for the fully discrete case.

Proposition 4.9. *Assuming (A1) - (A4), if (p_h^n, q_h^n) solve Problem 6 ($n = \overline{1, N}$), we have*

$$(4.29) \quad \begin{aligned} \|p_h^n\|^2 + \|q_h^n\|^2 &\leq C, \\ (b_\epsilon(p_h^n) - b_\epsilon(p_h^{n-1}), p_h^n - p_h^{n-1}) &\leq C\tau. \end{aligned}$$

Proof. We subtract equation (4.27) written at two consecutive time steps and obtain, for all $w_h \in W_h$,

$$(4.30) \quad (b_\epsilon(p_h^n) - b_\epsilon(p_h^{n-1}), w_h) + \tau(\nabla q_h^n, w_h) = 0.$$

Taking $w_h = p_h^n - p_h^{n-1} \in W_h$ into above, respectively $v_h = \tau q_h^n \in V_h$ into (4.28), and adding the resulting gives

$$(b_\epsilon(p_h^n) - b_\epsilon(p_h^{n-1}), p_h^n - p_h^{n-1}) + \tau\|q_h^n\|^2 + \tau(k(b(p_h^n))e_z, q_h^n) = 0.$$

Since, by (A3), k is bounded, the last term above can be estimated as

$$\tau |(k(b(p_h^n))e_z, q_h^n)| \leq C\tau + \frac{\tau}{2}\|q_h^n\|^2.$$

Since b_ϵ is monotone we immediately get

$$(4.31) \quad 0 \leq (b_\epsilon(p_h^n) - b_\epsilon(p_h^{n-1}), p_h^n - p_h^{n-1}) + \tau\|q_h^n\|^2 \leq C\tau.$$

By Lemma 4.8, for any $n = \overline{1, N}$, a $v_h^n \in V_h$ exists so that $\nabla v_h^n = p_h^n$ and $\|v_h^n\| \leq C\|\nabla \cdot v_h^n\| = C\|p_h^n\|$. Testing in (4.28) with v_h^n given above yields

$$(q_h^n, v_h^n) - \|p_h^n\|^2 + (k(b(p_h^n))e_z, v_h^n) = 0.$$

As above, using now (4.31) we obtain

$$|(q_h^n, v_h^n)| + |(k(b(p_h^n))e_z, v_h^n)| \leq (C_1 + C_2)\|v_h^n\| \leq C\|p_h^n\|.$$

These two equations imply

$$(4.32) \quad \|p_h^n\| \leq C,$$

which, together with (4.31) proves the result. \square

Applying now techniques developed in [2] we estimate the errors induced by the spatial discretization.

Proposition 4.10. *Let $n = \overline{1, N}$. If $(p^n, q^n) \in W \times V$, $(p_h^n, q_h^n) \in W_h \times V_h$ solve Problem 5, respectively 6, assuming (A1) - (A4) yields*

$$(4.33) \quad \sum_{n=1}^N \{(b_\epsilon(p^n) - b_\epsilon(p_h^n), p^n - p_h^n) + \tau \|\Pi_h q^n - q_h^n\|^2\} + \tau \left\| \sum_{n=1}^N (\Pi_h q^n - q_h^n) \right\|^2 \\ \leq C \sum_{n=1}^N \{\|q^n - \Pi_h q^n\|^2 + \|P_h p^n - p^n\|^2\}.$$

Proof. Subtracting (4.27) from (4.25) and (4.28) from (4.26) gives

$$(b_\epsilon(p^n) - b_\epsilon(p_h^n), w_h) + \tau \left(\sum_{j=1}^n \nabla(q^j - q_h^j), w_h \right) = 0,$$

$$(q^n - q_h^n, v_h) - (p^n - p_h^n, \nabla v_h) + ((k(b(p^n)) - k(b(p_h^n)))e_z, v_h) = 0.$$

Taking $w_h = P_h p^n - p_h^n \in W_h$ and $v_h = \tau \sum_{j=1}^n (\Pi_h q^j - q_h^j) \in V_h$ into above leads to

$$(b_\epsilon(p^n) - b_\epsilon(p_h^n), P_h p^n - p_h^n) + \tau \left(\sum_{j=1}^n \nabla(\Pi_h q^j - q_h^j), P_h p^n - p_h^n \right) = 0, \\ \tau (q^n - q_h^n, \sum_{j=1}^n (\Pi_h q^j - q_h^j)) - \tau (P_h p^n - p_h^n, \nabla \sum_{j=1}^n (\Pi_h q^j - q_h^j)) \\ + \tau ((k(b(p^n)) - k(b(p_h^n)))e_z, \sum_{j=1}^n (\Pi_h q^j - q_h^j)) = 0.$$

Adding these equalities and summing the result up from 1 to N yields

$$\sum_{n=1}^N (b_\epsilon(p^n) - b_\epsilon(p_h^n), P_h p^n - p_h^n) + \sum_{n=1}^N (q^n - q_h^n, \sum_{j=1}^n \tau (\Pi_h q^j - q_h^j)) \\ (4.34) \quad + \sum_{n=1}^N ((k(b(p^n)) - k(b(p_h^n)))e_z, \sum_{j=1}^n \tau (\Pi_h q^j - q_h^j)) = 0.$$

We estimate now each of the terms above, denoted by T_1 , T_2 and T_3 .

$$(4.35) \quad T_1 = \sum_{n=1}^N (b_\epsilon(p^n) - b_\epsilon(p_h^n), p^n - p_h^n) \\ + \sum_{n=1}^N (b_\epsilon(p^n) - b_\epsilon(p_h^n), P_h p^n - p_h^n) =: T_{11} + T_{12}.$$

With $C > 0$ independent on τ or h , for T_{11} we obtain

$$(4.36) \quad T_{11} \geq \frac{1}{2} \sum_{n=1}^N (b_\epsilon(p^n) - b_\epsilon(p_h^n), p^n - p_h^n) \\ + C \left(\sum_{n=1}^N \|b_\epsilon(p^n) - b_\epsilon(p_h^n)\|^2 + \epsilon \sum_{n=1}^N \|p^n - p_h^n\|^2 \right).$$

Applying the inequality of Cauchy T_{12} yields

$$(4.37) \quad |T_{12}| \leq \frac{1}{2} \sum_{n=1}^N \|b_\epsilon(p^n) - b_\epsilon(p_h^n)\|^2 + \frac{1}{2} \sum_{n=1}^N \|P_h p^n - p^n\|^2.$$

Rewriting T_2 as

$$(4.38) \quad T_2 = \sum_{n=1}^N (q^n - \Pi_h q^n, \sum_{j=1}^n \tau (\Pi_h q^j - q_h^j)) \\ + \sum_{n=1}^N (\Pi_h q^n - q_h^n, \sum_{j=1}^n \tau (\Pi_h q^j - q_h^j)) =: T_{21} + T_{22},$$

we estimate T_{21} and T_{22} . For T_{21} we get

$$(4.39) \quad |T_{21}| \leq \frac{1}{2} \sum_{n=1}^N \|q^n - \Pi_h q^n\|^2 + \frac{\tau^2}{2} \sum_{n=1}^N \left\| \sum_{j=1}^n (\Pi_h q^j - q_h^j) \right\|^2,$$

while for T_{22} we use (3.1) to obtain

$$(4.40) \quad T_{22} = \frac{\tau}{2} \left\| \sum_{n=1}^N (\Pi_h q^n - q_h^n) \right\|^2 + \frac{\tau}{2} \sum_{n=1}^N \|\Pi_h q^n - q_h^n\|^2.$$

Using (A3), T_3 gets

$$(4.41) \quad |T_3| \leq \frac{\delta}{2} \sum_{n=1}^N \|k(b(p^n)) - k(b(p_h^n))\|^2 + \frac{\tau^2}{2\delta} \sum_{n=1}^N \left\| \sum_{j=1}^n (\Pi_h q^j - q_h^j) \right\|^2 \\ \leq \frac{C_k \delta}{2} \sum_{n=1}^N (b(p^n) - b(p_h^n), p^n - p_h^n) + \frac{\tau^2}{2\delta} \sum_{n=1}^N \left\| \sum_{j=1}^n (\Pi_h q^j - q_h^j) \right\|^2.$$

Inserting (4.35) - (4.41) into (4.34) and choosing δ properly gives

$$\sum_{n=1}^N \{ (b_\epsilon(p^n) - b_\epsilon(p_h^n), p^n - p_h^n) + \tau \|\Pi_h q^n - q_h^n\|^2 \} + \tau \left\| \sum_{n=1}^N (\Pi_h q^n - q_h^n) \right\|^2 \\ \leq C \sum_{n=1}^N \left\{ \|q^n - \Pi_h q^n\|^2 + \|P_h p^n - p^n\|^2 + \tau^2 \left\| \sum_{j=1}^n (\Pi_h q^j - q_h^j) \right\|^2 \right\}.$$

Finally (4.33) follows applying the discrete Gronwall lemma. \square

Remark 4.11. *By the equivalence proven in Proposition 2.3, $p^n \in H^1(\Omega)$ for all n . Using now (4.24) and (3.7) we get*

$$\sum_{n=1}^N \|P_h p^n - p^n\|^2 \leq C h^2 \sum_{n=1}^N \|p^n\|_1^2 \leq C \frac{h^2}{\tau},$$

and the estimates (4.33) can be modified accordingly.

Similar estimates can be obtained for the p -unknowns.

Proposition 4.12. *Under the assumptions of Proposition 4.10 we have*

$$(4.42) \quad \tau \left\| \sum_{n=1}^N (p^n - p_h^n) \right\|^2 \leq C \left\{ \sum_{n=1}^N (b_\epsilon(p^n) - b_\epsilon(p_h^n), p^n - p_h^n) + \tau \left\| \sum_{n=1}^N (\Pi_h q^n - q_h^n) \right\|^2 + \sum_{n=1}^N \|q^n - \Pi_h q^n\|^2 \right\}.$$

Proof. Subtracting (4.28) from (4.26), recalling the definition of P_h and summing up for $n = 1$ to N yields

$$(4.43) \quad \left(\sum_{n=1}^N (q^n - q_h^n), v_h \right) - \left(\sum_{n=1}^N (P_h p^n - p_h^n), \nabla v_h \right) + \left(\sum_{n=1}^N \tau (k(b(p^n)) - k(b(p_h^n))) e_z, v_h \right) = 0,$$

for any $v_h \in V_h$. Using now Lemma (4.8), a $v_h \in V_h$ exists such that

$$(4.44) \quad \nabla v_h = \sum_{n=1}^N \tau (P_h p^n - p_h^n)$$

and $\|v_h\| < C \|\tau \sum_{n=1}^N (P_h p^n - p_h^n)\|$. In this case (4.43) gives

$$(4.45) \quad \tau \left\| \sum_{n=1}^N (P_h p^n - p_h^n) \right\|^2 = \left(\sum_{n=1}^N (q^n - q_h^n), v_h \right) + \left(\sum_{n=1}^N (k(b(p^n)) - k(b(p_h^n))) e_z, v_h \right)$$

Denoting by T_1 and T_2 the terms on the right into above, applying the inequality of Cauchy and recalling the estimates on $\|v_h\|$ leads to

$$(4.46) \quad \begin{aligned} |T_1| &\leq \frac{\tau}{2\delta_1} \left\| \sum_{n=1}^N (q_h^n - q^n) \right\|^2 + \frac{\delta_1}{2\tau} \|v_h\|^2 \\ &\leq \frac{\tau}{2\delta_1} \left\| \sum_{n=1}^N (q_h^n - q^n) \right\|^2 + \frac{C\tau\delta_1}{2} \left\| \sum_{n=1}^N (p_h^n - P_h p_h^n) \right\|^2. \end{aligned}$$

Similarly, by (A3) we obtain

$$(4.47) \quad \begin{aligned} |T_2| &\leq \frac{\tau}{2\delta_2} \left\| \sum_{n=1}^N (k(b(p_h^n)) - k(b(p^n))) \right\|^2 + \frac{\delta_2}{2\tau} \|v_h\|^2 \\ &\leq \frac{C}{2\delta_2} \sum_{n=1}^N (b(p_h^n) - b(p^n), p_h^n - p^n) \\ &\quad + \frac{C\tau\delta_2}{2} \left\| \sum_{n=1}^N (p_h^n - P_h p_h^n) \right\|^2. \end{aligned}$$

Choosing δ_1 and δ_2 properly, (4.45) - (4.47) gives

$$\begin{aligned} &\tau \left\| \sum_{n=1}^N (p^n - p_h^n) \right\|^2 \\ &\leq C \left\{ \sum_{n=1}^N (b_\epsilon(p^n) - b_\epsilon(p_h^n), p^n - p_h^n) + \tau \left\| \sum_{n=1}^N (q^n - q_h^n) \right\|^2 \right\}. \end{aligned}$$

The last term above can be rewritten as

$$\begin{aligned} \tau \left\| \sum_{n=1}^N (q^n - q_h^n) \right\|^2 &\leq \tau \left\| \sum_{n=1}^N (\Pi_h q^n - q_h^n) \right\|^2 + \tau \left\| \sum_{n=1}^N (q^n - \Pi_h q^n) \right\|^2 \\ &\leq \tau \left\| \sum_{n=1}^N (\Pi_h q^n - q_h^n) \right\|^2 + T \sum_{n=1}^N \|q^n - \Pi_h q^n\|^2, \end{aligned}$$

which completes the proof. \square

The following is a direct consequence of Propositions 4.10 and 4.12.

Theorem 4.13. *Assuming (A1)-(A4), if $(p^n, q^n) \in W \times V$, $(p_h^n, q_h^n) \in W_h \times V_h$ solve, for $n = \overline{1, N}$, Problems 5 and 6, we obtain*

$$\begin{aligned} (4.48) \quad &\sum_{n=1}^N (b_\epsilon(p^n) - b_\epsilon(p_h^n), p^n - p_h^n) + \tau \sum_{n=1}^N \|\Pi_h q^n - q_h^n\|^2 \\ &+ \tau \left\| \sum_{n=1}^N (q^n - q_h^n) \right\|^2 + \tau \left\| \sum_{n=1}^N (p^n - p_h^n) \right\|^2 \\ &\leq C \left(\sum_{n=1}^N \|q^n - \Pi_h q^n\|^2 + \sum_{n=1}^N \|P_h p^n - p^n\|^2 \right). \end{aligned}$$

Combining the estimates in Theorems 4.6 and 4.13 and recalling Remark 4.11 we get, for the fully discrete scheme

Theorem 4.14. *Assuming (A1)-(A4), we get*

$$\begin{aligned} (4.49) \quad &\left\| \sum_{n=1}^N \int_{t_{n-1}}^{t_n} (u(t) - p_h^n) dt \right\|^2 + \|\tilde{q}(T) - \tau \sum_{n=1}^N q_h^n\|^2 \\ &\leq C(\tau + \epsilon + h^2 + \tau \sum_{n=1}^N \|q^n - \Pi_h q^n\|^2). \end{aligned}$$

Proof. Let T_1 and T_2 denote the terms on the left in (4.49). For T_1 , by the properties of norms we have

$$T_1 \leq 2 \left\| \sum_{n=1}^N \int_{t_{n-1}}^{t_n} (u(t) - p^n) dt \right\|^2 + 2\tau^2 \left\| \sum_{n=1}^N (p^n - p_h^n) \right\|^2.$$

Estimates (4.20) in Theorem 4.4 give

$$\left\| \sum_{n=1}^N \int_{t_{n-1}}^{t_n} (u(t) - p^n) dt \right\|^2 \leq C(\tau + \epsilon),$$

which, together with (4.48) imply

$$(4.50) \quad T_1 \leq C(\tau + \epsilon).$$

Analogously, for T_2 we obtain

$$\begin{aligned} (4.51) \quad T_2 &\leq 2 \|\tilde{q}(T) - \tau \sum_{n=1}^N q_h^n\|^2 + 2\tau^2 \left\| \sum_{n=1}^N (q^n - q_h^n) \right\|^2 \\ &\leq C_1(\tau + \epsilon) + C_2(\tau + \epsilon + h^2 + \tau \sum_{n=1}^N \|q^n - \Pi_h q^n\|^2), \end{aligned}$$

by the arguments above. Now (4.49) follows from (4.50) and (4.51). \square

Corollary 4.15. *As in Theorem 4.14, for the scalar product we have*

$$\begin{aligned} (4.52) \quad &\sum_{n=1}^N \int_{t_{n-1}}^{t_n} (b_\epsilon(u(t)) - b_\epsilon(p_h^n), u(t) - p_h^n) dt \\ &\leq C \left(\tau^{\frac{1}{2}} + \epsilon^{\frac{1}{2}} + h^2/\tau^{\frac{1}{2}} + \tau^{\frac{1}{2}} \sum_{n=1}^N \|q^n - \Pi_h q^n\|^2 \right). \end{aligned}$$

Proof. We decompose the scalar product as follows

$$\begin{aligned} & \sum_{n=1}^N \int_{t_{n-1}}^{t_n} (b_\epsilon(u(t)) - b_\epsilon(p_h^n), u(t) - p_h^n) dt \\ &= \sum_{n=1}^N \int_{t_{n-1}}^{t_n} (b_\epsilon(u(t)) - b_\epsilon(p^n), u(t) - p_h^n) dt \\ &+ \sum_{n=1}^N \int_{t_{n-1}}^{t_n} (b_\epsilon(p^n) - b_\epsilon(p_h^n), u(t) - p_h^n) dt =: T_1 + T_2. \end{aligned}$$

Applying the inequality of Cauchy, T_1 yields

$$\begin{aligned} |T_1| &\leq \sum_{n=1}^N \int_{t_{n-1}}^{t_n} \|b_\epsilon(u(t)) - b_\epsilon(p^n)\| \|u(t) - p_h^n\| dt \\ &\leq \frac{1}{4\delta_1} \sum_{n=1}^N \int_{t_{n-1}}^{t_n} \|b_\epsilon(u(t)) - b_\epsilon(p^n)\|^2 dt + \delta_1 \sum_{n=1}^N \int_{t_{n-1}}^{t_n} \|u(t) - p_h^n\|^2 dt. \end{aligned}$$

Since b_ϵ is Lipschitz, using (4.20), the first sum gives

$$\begin{aligned} & \sum_{n=1}^N \int_{t_{n-1}}^{t_n} \|b_\epsilon(u(t)) - b_\epsilon(p^n)\|^2 dt \\ &\leq C \sum_{n=1}^N \int_{t_{n-1}}^{t_n} (b_\epsilon(u(t)) - b_\epsilon(p^n), u(t) - p^n) dt \\ &\leq C(\tau + \epsilon). \end{aligned}$$

Having $u \in L^2(J \times \Omega)$, by (4.29) we get

$$\sum_{n=1}^N \int_{t_{n-1}}^{t_n} \|u(t) - p_h^n\|^2 dt \leq C.$$

In this way, choosing $\delta_1 = (\tau + \epsilon)^{\frac{1}{2}}$ yields

$$(4.53) \quad |T_1| \leq C(\tau + \epsilon)^{\frac{1}{2}} \leq C(\tau^{\frac{1}{2}} + \epsilon^{\frac{1}{2}}).$$

For T_2 we obtain

$$\begin{aligned} |T_2| &\leq \sum_{n=1}^N \int_{t_{n-1}}^{t_n} \|b_\epsilon(p^n) - b_\epsilon(p_h^n)\| \|u(t) - p_h^n\| dt \\ &\leq \frac{\tau^{\frac{1}{2}}}{4} \sum_{n=1}^N \|b_\epsilon(p^n) - b_\epsilon(p_h^n)\|^2 + \tau^{\frac{1}{2}} \sum_{n=1}^N \int_{t_{n-1}}^{t_n} \|u(t) - p_h^n\|^2 dt. \end{aligned}$$

As before, the second sum above is uniformly bounded, while for the first one we write

$$\sum_{n=1}^N \|b_\epsilon(p^n) - b_\epsilon(p_h^n)\|^2 \leq C \sum_{n=1}^N (b_\epsilon(p^n) - b_\epsilon(p_h^n), p^n - p_h^n).$$

Using (4.48) gives

$$\sum_{n=1}^N \|b_\epsilon(p^n) - b_\epsilon(p_h^n)\|^2 \leq C(h^2/\tau + \sum_{n=1}^N \|q^n - \Pi_h q^n\|^2),$$

so T_2 is bounded by

$$(4.54) \quad |T_2| \leq C(\tau^{\frac{1}{2}} + h^2/\tau^{\frac{1}{2}} + \tau^{\frac{1}{2}} \sum_{n=1}^N \|q^n - \Pi_h q^n\|^2).$$

The result follows now from (4.53) and (4.54). \square

Assuming additionally

(A5) $q^n \in H^1(\Omega)^d$ for all $n = \overline{1, N}$,

recalling (4.24), the estimates in Theorem 4.14 and Corollary 4.15 become

$$\begin{aligned} & \left\| \sum_{n=1}^N \int_{t_{n-1}}^{t_n} (u(t) - p_h^n) dt \right\|^2 + \|\tilde{q}(T) - \tau \sum_{n=1}^N q_h^n\|^2 \leq C(\tau + \epsilon + h^2), \\ & \sum_{n=1}^N \int_{t_{n-1}}^{t_n} (b_\epsilon(u(t)) - b_\epsilon(p_h^n), u(t) - p_h^n) dt \leq C \left(\tau^{\frac{1}{2}} + \epsilon^{\frac{1}{2}} + h^2/\tau^{\frac{1}{2}} \right). \end{aligned}$$

Remark 4.16. Obviously (A5) is fulfilled if Ω is one-dimensional, since $H(\text{div}, \Omega)$ and $H^1(\Omega)$ coincide. Assumption (A5) holds also in the multi-dimensional case provided $\partial\Omega$ is smooth enough and k is derivable. Then, using (A2) and (A3), $k(b(\cdot)) \in C^1(0, 1)$ and we have

$$|\partial_u k(b(u))| \leq \lim_{\delta \rightarrow 0} \left| \frac{k(b(u + \delta)) - k(b(u))}{\delta} \right| \leq \sqrt{C_k \frac{b(u + \delta) - b(u)}{\delta}} \leq C.$$

Following [17] (Chapter 4, Theorems 5.1 and 5.2), for any $n = \overline{1, N}$, u^n solving Problem 3 is in $H^2(\Omega)$ and the corresponding norm is uniformly bounded in n by a constant which, nevertheless, may depend on τ . Therefore $q^n \in H^1(\Omega)$ for all $n \geq 1$ and $\|q^n\|_1 \leq C(\tau)$.

Acknowledgements. We would like to thank Prof. C. J. van Duijn and Dr. E. F. Kaasschieter for useful discussions and suggestions. The work of the second author was supported by the Netherlands Organization for Scientific Research (NWO), through Project 809.62.010 of Earth and Life Sciences.

REFERENCES

- [1] H. W. Alt, S. Luckhaus, Quasilinear elliptic-parabolic equations, Math. Z., 183(1983), 311 - 341.
- [2] T. Arbogast, M. F. Wheeler, N. Y. Zhang, A nonlinear mixed finite element method for a degenerate parabolic equation arising in flow in porous media, SIAM J. Numer. Anal. 33(1996), 1669 - 1687.
- [3] R. G. Baca, J. N. Chung, D. J. Mulla, Mixed transform finite element method for solving the non-linear equation for flow in variably saturated porous media, Int. J. Numer. Meth. Fluids, 24(1997), 441 - 455.
- [4] J. Bear, Y. Bachmat, Introduction to Modelling of Transport Phenomena in Porous Media, Kluwer Academic, Dordrecht, 1991.

- [5] L. Berganaschi, M. Putti, Mixed finite elements and Newton-type linearizations for the solution of Richards' equation, *Int. J. Numer. Meth. Engineering*, 45(1999), 1025 - 1046.
- [6] F. Brezzi, M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer Verlag, New York, 1991.
- [7] M. A. Celia, E. T. Bouloutas, R. L. Zarba, A general mass-conservative numerical solution for the unsaturated flow equation, *Water Resour. Res.*, 26(1990), 1483 - 1496.
- [8] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978.
- [9] R. Eymard, M. Gutnic, D. Hillhorst, The finite volume method for Richards equation, *Comput. Geosci.*, 3(1999), 259 - 294.
- [10] U. Hornung, W. Messing, *Poröse Medien - Methoden und Simulation* (in German), Verlag Beiträge zur Hydrologie, Kirchzarten, 1984.
- [11] W. Jäger, J. Kačur, Solution of doubly nonlinear and degenerate parabolic problems by relaxation schemes, *M²AN (Math. Model. Numer. Anal.)*, 29(1995), 605 - 627.
- [12] D. Kavetski, P. Binning & S. W. Sloan, Adaptive time stepping and error control in a mass conservative numerical solution of the mixed form of Richards equation, *Adv. Water Res.* 24(2001), 595 - 605.
- [13] P. Knabner, Finite element simulation of saturated-unsaturated flow through porous media, in *Large Scale Scientific Computing* (P. Deuhlhard et al., eds.), *Progress in Scientific Computing* 7(1987), Boston, 83-93.
- [14] P. Knabner, E. Schneid, Adaptive hybrid mixed finite element discretization of instationary variably saturated flow in porous media, in *High Performance Scientific and Engineering Computing* (M. Breuer et al., eds.), Springer, Berlin 2002.
- [15] P. Knabner, E. Schneid, Numerical solution of unsteady saturated/unsaturated flow through porous media, in *Numerical Modelling in Continuum Mechanics, Part II* (M. Feistauer et al., eds.), Prague, 1997, 337-343.
- [16] P. Knabner, G. Summ, Efficient realization of the mixed finite element discretization for nonlinear problems, submitted to *Mathematics of Computation*.
- [17] O. A. Ladyzhenskaya, N. N. Ural'tseva, *Linear and Quasilinear Elliptic Equations*, Academic Press, London, 1968.
- [18] J. L. Lions, E. Magenes, *Non Homogenous Boundary Value Problems and Applications, Vol. I*, Springer Verlag, Berlin, 1972.
- [19] R. H. Nochetto, C. Verdi, Approximation of degenerate parabolic problems using numerical integration, *SIAM J. Numer. Anal.* 25(1988), 784 - 814.
- [20] F. Otto, L^1 -contraction and uniqueness for quasilinear elliptic-parabolic equations, *J. Differ. Equations*, 131(1996), 20 - 38.
- [21] I. S. Pop, Error estimates for a time discretization method for the Richards' equation, RANA Preprint 16-01(2001), Eindhoven University of Technology.
- [22] E. Schneid, *Hybrid-Gemischte Finite-Elemente-Diskretisierung der Richards-Gleichung* (in German), PhD-thesis, University of Erlangen Nürnberg, 2000, also available at http://www.am.uni-erlangen.de/am1/publications/dipl_phd_thesis/dipl_phd_thesis.html.
- [23] C. Wagner, Numerical methods for diffusion-reaction-transport processes in unsaturated porous media, *Computing and Visualisation in Science*, 1(1998), 97 - 105.

- [24] M. Watanabe, An approach by difference to the porous medium equation with convection, *Hiroshima Math. J.*, 25(1995), 623 - 645.
- [25] G. A. Williams, C. T. Miller, An evaluation of temporally adaptive transformation approaches for solving Richards' equation, *Adv. Water Res.* 22(1999), 831 - 840.
- [26] C. Woodward, C. Dawson, Analysis of expanded mixed finite element methods for a nonlinear parabolic equation modeling flow into variably saturated porous media, *SIAM J. Numer. Anal.* 37(2000), 701 - 724.