INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 21, Number 1, Pages 20–64

doi: 10.4208/ijnam2024-1002

NUMERICAL ANALYSIS OF A MIXED FINITE ELEMENT APPROXIMATION OF A COUPLED SYSTEM MODELING BIOFILM GROWTH IN POROUS MEDIA WITH SIMULATIONS

AZHAR ALHAMMALI[1,*], MALGORZATA PESZYNSKA[2], AND CHOAH SHIN[2]

Abstract. In this paper, we consider mixed finite element approximation of a coupled system of nonlinear parabolic advection-diffusion-reaction variational (in)equalities modeling biofilm growth and nutrient utilization in porous media at pore-scale. We study well-posedness of the discrete system and derive an optimal error estimate of first order. Our theoretical estimates extend the work on a scalar degenerate parabolic problem by Arbogast et al, 1997 [4] to a variational inequality; we also apply it to a system. We also verify our theoretical convergence results with simulations of realistic scenarios.

Key words. Parabolic variational inequality, nonlinear coupled system, mixed finite element method, error estimates, biofilm–nutrient model, porous media.

1. Introduction

Biofilms play an important role in a variety of scientific and engineering applications including microbial enhanced oil recovery (MEOR) [26], CO_2 sequestration [29, 17], bioremediation engineering [30], and so on.

Biofilm growth in porous media is affected by the ambient fluid flow and nutrient availability. It is also subject to a volume constraint. In this paper we consider a model proposed in [35] which is a coupled system involving a nonlinear parabolic variational inequality (PVI) equipped with a new nonlinear diffusivity term and subject to Neumann boundary conditions assuming the system is isolated. The model of the biofilm growth is discussed in detail in Sec. 2. We are particularly interested in simulating this model on voxel grids at the pore-scale, i.e., grids corresponding to the x-ray tomography images of porous media at the pore-scale.

We approximate the model with mixed finite element method (MFEM). We believe this choice is better for the problem than the finite element method (FEM) we considered in our earlier work in [1], because of the conservative property of MFEM and its natural way of handling Neumann boundary conditions. (We remark that MFEM works also very well theoretically and computationally when Dirichlet condition is imposed unlike FEM that we succeeded in [1] in deriving an error estimate with Dirichlet conditions only.) Moreover, the implementation of MFEM with the lowest order of Raviart Thomas elements on rectangles and cubes $RT_{[0]}$ as cell centered finite difference method (CCFD) is very easy to implement and to use for voxel grids. We recall that CCFD is equivalent to this mixed FE up to quadrature order of $O(h^2)$ for smooth solutions [34, 40] where the later is more convenient to implement in practice.

MFEM has been studied extensively in literature including the theory developed in [11, 8]. However, most of the work is devoted to either unconstrained problems

Received by the editors on January 15, 2023 and, accepted on September 27, 2023.
2000 *Mathematics Subject Classification.* 35R35, 49J40, 65M60.
*Corresponding author.

such as [13, 4, 23], etc, or constrained stationary problem as in [12]. Semi-discrete mixed finite element approximation for unconstrained parabolic problem was considered in [13] for the linear case and in [23] for the nonlinear case.

There are several challenges in analysing the biofilm–nutrient model considered in this paper. One of challenges is the fact that PVI lacks of regularity. In particular, the second derivative in time of the solution $u_{tt} \notin L^2$ [22, 6]. Johnson in his paper [22] overcomes this challenge by setting some realistic assumptions on the domain and derives the error estimate of the finite element approximation of his problem using summation by parts. We implemented Johnson's approach in our previous work [1] with the finite element approximation of a simple model of biofilm–nutrient dynamic proposed in [32]. However, there are some major differences between the problem in [1] and the problem considered here in this work. In [1] we considered a quasi-linear PVI, where the diffusivity depends only on the spatial variable with no advection term and the boundary conditions are of Dirichlet type. In contrast, the problem in this paper has nonlinear diffusivity and an advection term with Neumann boundary conditions. Johnson's technique used in [22] does not work with MFEM. Therefore, we implement time integration approach used in [4].

Another difficulty is the nonlinearity involved in both the diffusivity and the reaction terms. Woodward and Dawson [41] deal with the nonlinearity using the expanded mixed finite element method which introduces a new variable, and then solves the problem in three unknowns (the primary unknown, its flux, and the new variable). However, as it is described in [33], "the expanded mixed finite element method is not equivalent with the standard mixed finite element method and their results cannot be simply transferred to our method MFEM". Arbogast et al. [4] consider an unconstrained nonlinear parabolic problem, where the nonlinearity is in the reaction term and under time derivative; the diffusivity is nonlinear if the change of variable is used. To derive the error estimate, they use the weighted projection on the approximated space which depends on the diffusivity beside the time integration technique. When we implement this approach to our problem, we need to assume some regularities on the solution which we do not guarantee that they are realistic ones. Therefore, to deal with the nonlinear diffusivity, we first linearize our problem using Kirchhoff transformation as in [33], and then we implement the approach used in [4]. We would like to emphasize here that the works in [4] and [33] are on scalar unconstrained problems whereas our problem is a constrained coupled nonlinear system.
Moreover, there are some studies in literature that regularize the PVI first using Lagrange multipliers then approximate it with finite element method as in [21, 28, 31]. In this paper we keep the PVI formulation in the theoretical analysis, yet use the Lagrange multiplier in computations.

1.1. Outline. Below we set up the notation. In Sec. 2 we provide details of the model. The paper is next broken into two parts: the first deals with the scalar PVI involving nonlinear diffusivity, and the second next deals with the additional challenges due to the coupled nature of the system. In Sec. 3 we provide mathematical details and formulate assumptions on the scalar PVI involving a nonlinear diffusivity. In Sec. 4 we provide details of the discretization and prove well-posedness of the discrete system. In Sec. 5 we prove the result on the convergence of MFEM approximation to this scalar problem. In Sec. 6 we provide the analyses for the full coupled system, and in Sec. 7 we provide examples in $d = 1$ and $d = 2$.

The theoretical results we prove require various assumptions which are specific to the result. In particular, the well-posedness in Sec. 4 and Sec. 6 are derived under the assumption that advection terms are handled explicitly. However, the error analysis in Sec. 5 and Sec. 6 allow for advection terms to be handled implicitly in time. The techniques allow also the study of explicit advection, with small modification, which we do not discuss.

1.2. Notations. Now we set up notation for this paper. Let $\Omega \subset \mathbb{R}^d$, $d = 1, 2$, or 3 be a bounded domain with sufficiently smooth boundary Γ, and let $0 < T < \infty$ and $J = (0, T]$ be the time interval.

We shall use the following functional spaces that are suitable for the MFEM when Neumann boundary conditions are considered.

$$X = H_0(div, \Omega) := \{\mathbf{q} \in H(div; \Omega) : \mathbf{q} \cdot \mathbf{n}|_\Gamma = 0\}, \text{ and } M = L^2(\Omega).$$

Recall that $H(div; \Omega) = \{\mathbf{q} \in (L^2(\Omega))^d; \nabla \cdot \mathbf{q} \in L^2(\Omega)\}$ equipped with the following scalar product and norm:

$$[\mathbf{q}, \boldsymbol{\psi}] = (\mathbf{q}, \boldsymbol{\psi}) + (\nabla \cdot \mathbf{q}, \nabla \cdot \boldsymbol{\psi}), \; \|\mathbf{q}\|_{H(div,\Omega)} = [\mathbf{q}, \mathbf{q}]^{1/2},$$

and the normal trace $\mathbf{q} \cdot \mathbf{n} \in H^{-\frac{1}{2}}(\Gamma)$. We also define the closed subset $K \subset M$ as:

$$K = \{\eta \in M; 0 \leq \eta \leq u^* \text{ a.e on } \Omega\},$$

which includes the box constraint $[0, u^*]$, where $u^* \in \mathbb{R}$.

In the sequel, we use standard notations on $L^\infty(\cdot)$, and Sobolev spaces $H^s(\cdot)$ for some nonnegative integer s; see, e.g., [37] for more details on the Sobolev spaces. Let $\|\cdot\|_0$ denote the norm on $L^2(\cdot)$, $\|\cdot\|_\infty$ the norm on $L^\infty(\cdot)$, and $\|\cdot\|_s$ the norm on $H^s(\cdot)$. For the sake of abbreviation, we write $L^2(L^2)$ to mean $L^2(J; L^2(\Omega))$, and $L^2(H_{div})$ to mean $L^2(J; H(div; \Omega))$, similar abbreviation is applied to other notations on functional spaces. Throughout this paper, let C be a generic constant independent of h or τ, where h is the spatial step size and τ is the time step size. $\mathbf{1}_S$ is the characteristic function on $S \subset \Omega$. We will also use an indicator function I_S defined below. Without abuse of notation, we write the reciprocal of a function g as g^{-1}.

2. Biofilm-nutrient model

Biofilm is a heterogeneous complex structure made of billions of bacterial cells, attached to some wet solid surfaces, and a slimy extracellular polymeric substance (EPS) excreted by the bacteria to protect them from the harsh environmental conditions such as dehydration, ultra-violet radiation, antibacterial chemicals, bacteriophages and phagocytes [42, 3]. The majority of the bacteria exist in biofilm communities in aqueous porous media [25, 14].

Biofilm develops and grows through a reaction process depending on the existence of substrates such as nutrient and oxygen, and also through diffusion and advection process. Initially, the microbes grow in the void space of porous media until their concentration reaches some density $u_* > 0$ indicating that the biofilm forms and becomes mature. Then the biofilm phase continues growing until it reaches a certain density, denoted by $u^* > u_*$, that cannot be exceeded because micro-cells have finite volume, and only so many of them can fit in a particular region. After this density is close to maximal, the majority of the growth occurs through the interface between the biofilm and the ambient fluid [3], which is the free boundary to be modeled. The bulk fluid may penetrate the biofilm region in the permeable and partially permeable zones to transport the substrate so that microbes continue growing in its domain [38, 35].

The particular Biofilm-Nutrient model considered in this paper is proposed in [35] as an enhancement of model in [32] consistent with the ideas which is proposed in [15]. The model is a coupled system of two nonlinear parabolic advection-diffusion-reaction PDEs in a biomass density $u_1(x,t)$, and a nutrient concentration $u_2(x,t)$ required for the microbes to grow [15, 35].

$$\partial_t u_1 + \nabla \cdot (\bar{\mathbf{q}} u_1) - \nabla \cdot (d_1(u_1)\nabla u_1) + \partial I_{[0,u^*]}(u_1) \ni r_1(u_1,u_2) \text{ in } \Omega, \tag{1a}$$

$$\partial_t u_2 + \nabla \cdot (\bar{\mathbf{q}} u_2) - \nabla \cdot (d_2(u_1)\nabla u_2) = r_2(u_1,u_2) \text{ in } \Omega, \tag{1b}$$

for $t > 0$, where Ω is the non-rock domain in the porous medium filled with abundant flowing fluid with some microbes and sufficient nutrient. In fact, Ω can be viewed as the union of the domain of biofilm Ω_b, the domain of fluid Ω_n and the interface between them Γ_{bn}; i.e. $\Omega = \Omega_b \cup \Omega_n \cup \Gamma_{bn}$.

In the model (1) the advective flux $\bar{\mathbf{q}}$ is assumed given, and one can consider different modeling variants in which $\bar{\mathbf{q}}$ is trivial or nontrivial inside the biofilm domain. In particular, $\bar{\mathbf{q}}$ can be obtained by the heterogeneous Brinkman flow model as in [35].

Model (1) can be considered as a free-boundary problem with the free boundary being the interface between the biofilm and the bulk fluid. The free boundary at time t is a set of points x where $u(x,t) = u^*$, i.e., were the biofilm reaches its maximum density.

It is natural to assume that the initial biomass $u_1|_{t=0}$ amount is nonnegative. Even though it is physically justified to expect that u_1 would remain nonnegative at all times, we are not able to prove a maximum principle for the PDE model to show this. Furthermore, discrete mixed methods do not necessarily satisfy maximum principle, thus we impose the non-negativity constraint also on the numerical solutions. In other words, u_1 is subject to the box constraint $0 \le u_1 \le u^*$. The role of the term $\partial I_{[0,u^*]}(u_1)$ in (1a) is to enforce this constraint, and ∂I is the subgradient of the indicator function $I_{[0,u^*]} : L^2(\Omega) \longrightarrow \overline{\mathbb{R}}$ defined to be zero when $0 \le u_1 \le u^*$, and ∞ otherwise:

$$I_{[0,u^*]}(u_1) = \begin{cases} 0, & 0 \le u_1 \le u^*, \\ +\infty & \text{otherwise,} \end{cases} \tag{2}$$

and hence,

$$\partial I_{[0,u^*]}(u_1) = \begin{cases} (-\infty, 0), & u_1 = 0, \\ 0, & 0 < u_1 < u^*, \\ (0,\infty), & u_1 = u^* \end{cases} \tag{3}$$

is a multi-valued function, which explains the inclusion symbol $\ni$ in (1a). We note here that in the distributional sense, ∂I_K can be written as

$$\partial I_K(u) = \{\phi \in L^2(\Omega); (\phi, \eta - u) \le 0, \ \forall \ \eta \in K\}, \ \forall u \in K. \tag{4}$$

In our model it is crucial to define the nonlinear diffusivities d_1, d_2. We postulate that the biofilm diffusivity $d_1 = d_1(u_1)$, and nutrient diffusivity $d_2 = d_2(u_1)$ both depend on the biomass density u_1. In particular, the spreading of biofilm through the interface between the region $0 \le u_1 < u^*$ and $u_1 = u^*$ is modeled by ∂I as well as by fast increasing $d_1(u_1)$ as u_1 approaches u^* (see [15]). $d_1(u_1)$ can be given in the following formula [35]:

$$d_1(u_1) = d_1(\alpha; u_1) = \begin{cases} d_0, & u_1 < 0, \\ d_0 \left[\left(\frac{u_1}{\bar{u^*} - u} \right)^\alpha + 1 \right], & 0 \leq u_1 \leq u^*, \\ d^*, & u_1 > u^*; \end{cases} \tag{5}$$

with $d^* = d_0 \left[\left(\frac{u^*}{\bar{u^*} - u^*} \right)^\alpha + 1 \right]$, where $d_0 > 0$ is the motility coefficient, u^* is the maximum density, and $\bar{u^*} > u^*$ is a barrier parameter. The exponent α is a parameter which expresses the strength of spreading of the biomass u_1 as it approaches u^*. It is important to select α accurately to ensure the growth and spreading without violating the constraint $u_1 \leq u^*$; (see the adaptive model in [35] that computes the optimal α; it also shows that the optimal α has to be greater than or equal to 2). See also similar diffusivity formula in [19].

We explain now the similarities and the differences between our model with the choice of bounded $d_1(u_1)$ and the family of Eberl et al. models [15, 16, 19] where d_1 is a graph at $u_1 = u^*$. First, the formal difference is that the latter model ensures the constraint $u_1 \leq u^*$ by admitting an "infinite diffusivity" close to the constraint. Our model includes this constraint with the operator $\partial I_{[0,u^*]}$. Thus both models achieve the same purpose, even if the actual choice of diffusivity in our model may be less than that in the Eberl et al. models. Second, in numerical implementation, the formally infinite diffusivity requires very small time steps and this is impractical in a fully implicit formulation especially if coupled to the flow. In the end, our approach through variational inequality is more practical computationally.

In turn, the nutrient diffusivity $d_2 = d_2(u_1)$ depends significantly on the density of biofilm phase in the medium. In particular, in the portion of biofilm phase where biofilm is mature with $u_1 \approx u^*$, nutrient diffuses very slowly, but it diffuses somewhat faster in the active layer portion of the biofilm where $u < u^*$ [19]. Following [19, 35], we define $d_2(u_1)$ as:

$$d_2(u_1) = \begin{cases} u^* d_{N,w}, & u_1 < 0, \\ u^* d_{N,w} + u_1(d_{N,b} - d_{N,w}), & 0 \leq u_1 \leq u^*, \\ u^* d_{N,b}, & u_1 > u^*; \end{cases} \tag{6}$$

where $d_{N,w}, d_{N,b}$ are the nutrient diffusivity in the aqueous phase and the biofilm phase, respectively, with $d_{N,b} << d_{N,w}$. Note that d_2 is bounded from below by $u^* d_{N,b}$ and from above by $u^* d_{N,w}$.

The reaction terms r_1 and r_2 are given in terms of Monod function $m(u_2) = \beta \frac{u_2}{u_2 + \gamma}$, where β and γ are fixed constants. β is the specific nutrient uptake rate, and γ is called Monod half-life. $r_1 = \kappa u_1 m(u_2)$ involves the growth rate with a growth constant κ, while $r_2 = -u_1 m(u_2)$ involves the utilization rates.

3. Mathematical details of the nonlinear parabolic variational inequality

In this section we focus on the first component of (1) and study the parabolic variational inequality (PVI): we fix u_2, and rewrite (1a) in terms of $u = u_1$ as follows

$$\partial_t u + \nabla \cdot (\bar{\mathbf{q}} u) - \nabla \cdot (d(u) \nabla u) + \partial I_{[0,u^*]}(u) \ni f(u) \text{ in } \Omega,\ t > 0, \tag{7}$$

where $f(u) = f_0(x) + um(x)$, for some bounded function f_0, and $m(x)$ is the Monod function defined in Sec.2. The diffusivity $d(\cdot) = d_1(\cdot)$ in (5). This function has the following properties which we summarize below.

Lemma 3.1. *Let d be given by (5), and let $\bar{u}^* > u^*$, $d_0 > 0$, $\alpha \geq 2$ be fixed. Then the function $d : [0, \infty) \to [d_0, d^*]$ is Lipschitz continuous with some constant L_d^0. More precisely, we have*

$$\big|d(u) - d(v)\big| \quad \leq \quad L_d^0 |u - v| \,, \; \forall u, v \geq 0, \tag{8a}$$

but also

$$\big|d(u) - d(v)\big| \quad = \quad d^* |u - v| \,, \quad \textit{if } u, v \geq u^*. \tag{8b}$$

The function $1/d(u)$ is well defined, it is bounded from above by d_0^{-1}, below by $(d^*)^{-1}$, and has a global Lipschitz constant L_d^1. Finally, for a fixed u and some $2 \leq \alpha_1 < \alpha_2$, $d(\alpha_1; u) \leq d(\alpha_2; u)$.

Illustrations of the diffusivity $d(\cdot)$ and its reciprocal when $u \geq 0$ are in Figure 1.

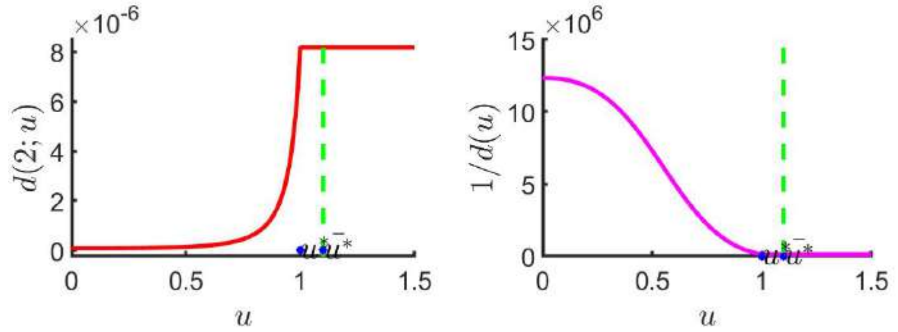

FIGURE 1. Left: biomass diffusivity $d(u) = d(2; u)$. Right: its reciprocal $1/d(u)$.

Now we introduce a new variable $\mathbf{q} = \bar{\mathbf{q}}u - d(u)\nabla u$, and substitute in (7). The model is completed with initial and homogeneous boundary conditions. We rewrite it in the following mixed formulation

$$d^{-1}(u)\mathbf{q} = d^{-1}(u)\bar{\mathbf{q}}u - \nabla u, \text{ in } \Omega, \; t > 0, \tag{9a}$$

$$\partial_t u + \nabla \cdot \mathbf{q} + \partial I_{[0,u^*]}(u) \ni f(u) \text{ in } \Omega, \; t > 0, \tag{9b}$$

$$\mathbf{q}(s, t) \cdot \mathbf{n} = 0 \text{ on } \Gamma, \; t > 0, \tag{9c}$$

$$u(x, 0) = u_{init} \text{ in } \Omega. \tag{9d}$$

Assumption 1. *We make the following assumptions.*

(A) $\bar{\mathbf{q}} = \bar{\mathbf{q}}(x) \in (L^\infty(\Omega))^d$.

(B) *$f = f(x, t; u)$ is a continuous Lipschitz function with respect to u for each $x \in \Omega, t > 0$, with a global Lipschitz constant R .*

(C) *$u^* \in \mathbb{R}$ is given.*

(D) *$u_{init} \in H^1(\Omega)$, and $0 \leq u_{init} \leq u^*$.*

For the sake of numerical analysis, we assume the following regularity of $(\mathbf{q}, u)$.

(E) *$u \in L^2(H^1)$, $u_t \in L^2(H^{-1})$.*

(F) *$\mathbf{q} \in L^2((H^1)^d)$, $\mathbf{q}_t \in L^2((H^{-1})^d)$, $\nabla \cdot \mathbf{q} \in L^2(H^1)$.*

We would like to emphasise here that Assumptions (E) and (F) are not strong regularity assumptions. We set them just for the sake of numerical analysis and they do not depend on some theory. In literature, though, there are quite similar assumptions but for unconstrained problem; see e.g. [4].

Now for each $t > 0$, the solution $(\mathbf{q}(t), u(t)) \in X \times K$ of (9) may be thought as a solution to the mixed formulation of the parabolic variational problem

$$(d^{-1}(u)\mathbf{q}, \boldsymbol{\psi}) - (d^{-1}(u)\bar{\mathbf{q}}u, \boldsymbol{\psi}) - (\nabla \cdot \boldsymbol{\psi}, u) = 0 \quad \forall \boldsymbol{\psi} \in X, \tag{10a}$$

$$(\partial_t u, \eta - u) + (\nabla \cdot \mathbf{q}, \eta - u) \geq (f(u), \eta - u) \quad \forall \eta \in K, \tag{10b}$$

$$u(0) = u_{init}, \tag{10c}$$

To obtain (10a) from (9a), we have multiplied (9a) by some test function $\boldsymbol{\psi} \in X$ and integrated by parts. The variational inequality (10b) is equivalent to (9b) by the definition of ∂I_K in the distributional sense (4).

The well-posedness of the weak formulation of the mixed problem (9) in the case where $\bar{\mathbf{q}} = 0$, and $d = d(x)$ has been shown in [2] where we combined results from [37] and [36]. We also mention the work in [27] where the solvability is shown for similar models but not in mixed formulation.

4. Mixed FE discretization and CCFD implementation

We first explain how the unconstrained problem is discretized in time and in space. Later we incorporate the constraints.

The advection term involving $\bar{\mathbf{q}}$ can be treated explicitly or implicitly in time. In the error analysis we consider the implicit case, which is harder theoretically, and we do not include the analysis of the explicit case. For well-posedness of the discrete system we only discuss the explicit case, which is what we actually use in the implementation.

4.1. Discrete formulation for the unconstrained problem. We first state the time–discrete mixed formulation for an unconstrained version of (10), where we replace $\eta \in K$ by $\eta \in M$. We apply fully implicit in time formulation, except for the term $\bar{\mathbf{q}}u$ which we discretize explicitly, and diffusivities which can be obtained by time-lagging. For some positive integer N, let $\tau = \frac{T}{N}$, be the time step size, and let $t_n = n\tau$, $J_n = (t_{n-1}, t_n]$ for $n = 1, \dots, N$. At each time step t_n we seek $(\mathbf{q}^n, u^n) \in X \times M$. We use the symbols $n^* = n$ (fully implicit), or $n^* = n - 1$ (time-lagging case). Since the advection term is discretized explicitly, we adapt (9b) and (9a) as follows. We move $-(\nabla \cdot (\bar{\mathbf{q}}u^{n-1}), \eta)$ from the left hand side to the right hand side and absorb it within (f^n, η) in (9b). We use now the symbol $\mathbf{q}$ only for the diffusive flux, i.e., we also modify (9a) to have $d^{-1}(u)\mathbf{q} = -\nabla u$.

The time-discrete form is an identity, since we are not using constraints

$$(d^{-1}(u^{n^*})\mathbf{q}^n, \boldsymbol{\psi}) - (u^n, \nabla \cdot \boldsymbol{\psi}) = 0, \quad \forall \boldsymbol{\psi} \in X, \tag{11a}$$

$$(\nabla \cdot \mathbf{q}^n, \eta) + (\tfrac{u^n - u^{n-1}}{\tau}, \eta) = (f^n, \eta) = (f_0^n + um, \eta), \quad \forall \eta \in M. \tag{11b}$$

4.1.1. Discretization in space. The domain Ω is covered by quadrangulation $\mathcal{T}^h$ which contains rectangular cells denoted by ω_{ij}. We assume that their union covers the domain Ω which is connected. Each cell (ij) has edges $\gamma_{i\pm1/2,j}$ and $\gamma_{i,j\pm1/2}$. It is also convenient to consider some global numbering of the cells $(\omega_{ij})_{ij}$: we assign to each (i,j) some index $1 \leq c \leq N_c$; We also have a global numbering of each of the edges $\gamma_{i\pm1/2,j}, \gamma_{i,j\pm1/2}$ as $\gamma_e, e = 1, \dots Ne$. Each edge e is between the cells in $c(e)$. In 2 dimensions, each edge e is associated with at most two neighboring cells, and each cell c is adjacent to at most four edges.

We shall assume the grid $\mathcal{T}^h$ is quasi-uniform, so that there is a lower h_{min} and an upper h_{max} bound for each $|\omega_{ij}| = h_i \times h_j$, with $h_{min} = \beta_h h_{max}$ and $0 < \beta_h \leq 1$. (A uniform square grid has $\beta_h = 1$.)

On this grid we build the well known spaces $(X_h \times M_h)$ built on $\mathcal{T}^h$ with $X_h = \{\mathbf{q} \in RT_{[0]}; \mathbf{q}\cdot\mathbf{n}|_\Gamma = 0\}$, where $RT_{[0]}$ is the lowest order Raviart-Thomas space on rectangles [9, 4, 40]. That is,

$$\begin{aligned} RT_{[0]}: \quad &= \quad \{\mathbf{q} \in (L^2(\Omega))^2;\ \mathbf{q}|_{w_{ij}} = \begin{pmatrix} ax+b \\ cy+d \end{pmatrix},\ a,b,c,d \in \mathcal{P}_0 \text{ for each } w_{ij} \in \mathcal{T}^h, \\ \text{and} \quad & \quad \mathbf{q}\cdot\mathbf{n}|_e \in \mathcal{P}_0(e) \text{ on each edge } e \in \partial w_{ij}\}. \end{aligned}$$

We choose these lower order approximation spaces because the solution to variational inequalities feature low regularity. The space M_h contains piecewise constants on $\mathcal{T}^h$; i.e. $M_h := \mathcal{M}^{[0]}(\mathcal{T}^h) := \{\eta \in L^2(\Omega); \eta|_{w_{ij}} \in \mathcal{P}_0$ for each $w_{ij} \in \mathcal{T}^h\}$; the basis functions spanning M_h are simply $\mathbf{1}_{\omega_{ij}}$, and $u_h|_{\omega_{ij}} = U_{ij}$ associated with the cell centers x_{ij} of each ω_{ij}.

The vector valued functions in X_h are tensor products of piecewise linears in one coordinate with piecewise constants in the other. In particular, $(\mathbf{q}_h)_1$ is identified by their edge values at the left and right edges $\gamma_{i\pm1/2,j}$ so we have, e.g, $(\mathbf{q}_h)_1|_{\gamma_{i+1/2,j}} = Q_{i+1/2,j}$; analogously $(\mathbf{q}_h)_2$ is identified by values at the bottom and top edges $\gamma_{i,j\pm1/2}$, respectively, $(\mathbf{q}_h)_2|_{\gamma_{i,j-1/2}} = Q_{i,j-1/2}$. The basis functions for the vector valued functions in X_h are $\boldsymbol{\psi}_{i\pm1/2,j}$ for $(\mathbf{q}_h)_1$ and $\boldsymbol{\psi}_{i,j\pm1/2}$ for $(\mathbf{q}_h)_2$. We also have elementwise approximations $d_{ij} \approx d(\alpha;u)|_{\omega_{ij}}$.

In summary, $(Q;U)$ is the vector of the degrees of freedom for $(\mathbf{q}_h, u_h)$ in their bases. In particular, Q is a vector of $Q_{i\pm1/2,j}$ and $Q_{i,j\pm1/2}$ (or of $(Q_e)_{e=1}^{N_e}$), and U is a vector of U_{ij} (or of $(U_c)_{c=1}^{N_c}$).

4.1.2. Fully discrete formulation. Now we discuss how to obtain the approximations to the solutions to (11). These approximations satisfy discrete equations obtained by setting a system similar to (11) but which now must hold in the finite dimensional subspaces of the functional spaces used for (11). For the easiness of implementation we also apply a particular numerical integration rule.

At each time step t_n we seek the approximations $(\mathbf{q}_h^n, u_h^n) \in X_h \times M_h$ to $(\mathbf{q}, u)$ which satisfy a system similar to (11) in which we also apply numerical integration, for $\forall \boldsymbol{\psi} \in X_h, \forall \eta \in M_h$, that

$$(d^{-1}(u_h^{n^*})\mathbf{q}_h^n, \boldsymbol{\psi})_h - (u_h^n, \nabla\cdot\boldsymbol{\psi}) = 0, \tag{12a}$$

$$(\nabla\cdot\mathbf{q}_h^n, \eta) + (\frac{u_h^n - u_h^{n-1}}{\tau}, \eta) = (f^n, \eta) = (f_0^n + u_h^n m(\cdot, t_n), \eta). \tag{12b}$$

The numerical integration $(\cdot,\cdot)_h$ in (12a) to calculate $(d^{-1}\mathbf{q}_h^n, \boldsymbol{\psi})_h$ is a combination of the trapezoidal (T) and midpoint (M) integration rules, respectively, (TM) in x_1 and (MT) in x_2 spatial coordinate, with basis functions associated with the first coordinate and second coordinate. This strategy leads to simplifications and an easy interpretation of the system (12) as a cell-centered finite difference system [34]; see also recent details in [7] recalled now here. Consider $\boldsymbol{\psi} = \boldsymbol{\psi}_{i+1/2,j}$ associated with the edge $\gamma_{i+1/2,j}$ with support on $\omega_{ij}, \omega_{i+1,j}$. Focus on the first component of the inner product $(d^{-1}(u_h^{n^*})\mathbf{q}_h^n, \boldsymbol{\psi})_h$ which is an integral of the product of piecewise linear functions in x_1 on ω_{ij} and $\omega_{i+1,j}$ and of the coefficient d^{-1} which takes piecewise constant values $d_{ij}^{-1} = d(\alpha;U_{ij})^{-1}$, $d_{i+1,j}^{-1} = d(\alpha;U_{i+1,j})^{-1}$, respectively. Once trapezoidal T rule in x_1 direction followed by M in x_2 direction are applied, these yield $\frac{h_j}{2}(d_{ij}^{-1}Q_{i+1/2,j}h_i + d_{i+1,j}^{-1}Q_{i+1/2,j}h_{i+1})$. To simplify, we introduce the notion of edge transmissivities, which are harmonic grid-weighted averages of diffusivities. For example, on $\gamma_{i+1/2,j}$ we have

$$Q_{i+1/2,j} = -D_{i+1/2,j}(U_{i+1,j} - U_{i,j}); \tag{13a}$$

$$D^{-1}_{i+1/2,j} = (h_i d(\alpha; U_{ij})^{-1} + h_{i+1} d(\alpha; U_{i+1,j})^{-1})/2. \tag{13b}$$

This formula defines $D_{i+1/2,j}$ as the harmonic grid-weighted average of d_{ij} and $d_{i+1,j}$. $D_{i,j+1/2}$ is defined similarly, and all $(D_e)_e$ are positive.

Now we are ready to rewrite (12) in the matrix–vector form. The first term of (12a) becomes $\mathcal{D}^{-1}(U)Q^n$, where

$$\mathcal{D}^{-1} = \mathcal{D}^{-1}(U) = \mathtt{diag}(D_e^{-1}(U))_e, \;\; \mathcal{D} = \mathtt{diag}(D_e)_e \tag{14}$$

are diagonal and is made of positive edge transmissivities or their inverses. The second term in (12a) and the first in (12b) are written with the difference matrix $\mathcal{B}$ made of rows which contain -1 and 1, as $\mathcal{B}^T U^n$ and $-\mathcal{B}Q^n$, respectively. Here the non-square matrix $\mathcal{B} : \mathbb{R}^{N_e} \to \mathbb{R}^{N_c}$, with $\mathcal{B}^T : \mathbb{R}^{N_c} \to \mathbb{R}^{N_e}$. We have then

$$\mathcal{D}^{-1}Q^n + \mathcal{B}^T U^n = 0. \tag{15a}$$

The second equation (12b) becomes

$$-\mathcal{B}Q^n + \tfrac{1}{\tau}U^n \;\; = \;\; \mathcal{M}^n U^n + G^n. \tag{15b}$$

Here the term $\mathcal{M}^n U^n$ comes from (mu_h^n, η), where the diagonal matrix $\mathcal{M}^n = \mathtt{diag}(M_c^n)_c$, with the entries $M_c^n = m(x_c, t_n)$. We also have $G^n = F_0^n + \frac{1}{\tau}U^{n-1}$, and where F_0^n are found from integration of (f_0, η) in (12b).

In (15), $\mathcal{D}^{-1}$ is diagonal, and depends on U which we write concisely as $\mathcal{D}^{-1} = \mathcal{D}^{-1}(U^{n^*})$. This means $\mathcal{D}^{-1}$ is computed by time-lagging with $n^* = n$-1, or fully implicitly if $n^* = n$. Whether linear or linearized, the saddle-point structure of the system (15) makes it harder to solve the linear system. However, since $\mathcal{D}^{-1}$ is diagonal, it is easy to transform (15) to a single equation with a positive definite matrix which is easier to solve and precondition. We substitute $Q^n = -\mathcal{D}(U^n)\mathcal{B}^T U^n$ in the second equation multiplied by τ, and rewrite (15) as a single nonlinear system, using $\mathcal{A}(U^{n^*}) = \mathcal{B}\mathcal{D}(U^{n^*})\mathcal{B}^T$. We thus obtain,

$$\mathcal{C}(U^n) = (\tau\mathcal{A}(U^{n^*}) + \mathcal{I} - \tau\mathcal{M}^n)U^n = \tau G^n. \tag{16}$$

We seek the solution of (16) in $W = \mathbb{R}^{N_c}$. If $n^* = n$, the operator is nonlinear, but in the time-lagged case, it is linear. In the time-lagged case, $\mathcal{D} = \mathcal{D}(U^{n-1})$ is known, and this coupled saddle-point system is linear, (without constraints, one can apply Theorem [9](Prop.3.3.1 and Thm 3.6.2)), and we will not handle this case separately.

The more complicated nonlinear case $n^* = n$ will be handled next when we incorporate constraints.

4.2. Constrained problem. Now we incorporate the constraint in (11). In the weak form we have a parabolic variational inequality, and we continue working under the assumption that advection is discretized explicitly in time as in (12b). In the fully discrete case, this means that at every time step t_n, we are seeking the solution $U^n \in K_h \subset M_h$ to a modification of (16)

$$\mathcal{C}(U^n) + \tau\Lambda(U^n) = \tau G^n; \;\; \Lambda \in \partial I_{K_h}(U^n), \tag{17}$$

where $\mathcal{C}, G^n$ are as in (16), and $K_h = M_h \cap K$. Now Λ is a Lagrange multiplier or penalty term which enforces the constraint so that $U^n \in K_h$. Equivalently, the variational inequality reads

$$\langle C(U^n), \eta - U^n \rangle \geq \tau\langle G^n, \eta - U^n \rangle, \;\; \forall \eta \in K_h. \tag{18}$$

4.2.1. Properties of $\mathcal{D}^{-1}$, $\mathcal{D}$, and $\mathcal{M}$. These properties will be needed in the study of well-posedness of (17). We mention that these properties only work for the present finite dimensional case.

Lemma 4.1. *Let $\mathcal{T}^h$ be quasi-uniform, and let $d(\alpha; u)$ be as in Lemma 3.1. Then (i) the entries $\mathcal{D}_e$ are bounded from above by $D_{max} = d^* h_{min}^{-1}$ and from below by $D_{min} = d_0 h_{max}^{-1}$, thus, since $\mathcal{D}$ is diagonal, for any induced matrix norm, $\|\mathcal{D}\| \leq D_{max}$. (ii) We also have the global Lipschitz property for each e, with L_D^e*

$$\left|D_e(U) - D_e(V)\right| \leq L_D^e \max_{c(e)} |U_c - V_c| \,, \ \forall U, V \in M_h. \tag{19a}$$

Thus also with some L_D^0,

$$\left\|\mathcal{D}(U) - \mathcal{D}(V)\right\|_\infty \leq L_D^0 \|U - V\| \,, \ \forall U, V \in M_h. \tag{19b}$$

$$\left\|\mathcal{A}(U)U - \mathcal{A}(V)V\right\| \leq L_D^A \|U - V\| \,, \ \forall U, V \in M_h. \tag{19c}$$

(iii) Finally, there is $C_{\mathcal{J}}$ such that

$$\mathcal{J}(U, V) = \langle \mathcal{A}(U)U - \mathcal{A}(V)V, U - V\rangle \leq C_{\mathcal{J}} \|U - V\|^2; \ \ \forall U, V \in M_h. \tag{19d}$$

Proof. We prove the steps separately.

First, (i) is immediate from (13b).

For part (ii) note that for every edge e, $D_e(U)$ given by (13b) is a differentiable function of each U_c for $c \in c(e)$ (and $D_e(U)$ has globally bounded derivatives), hence, (19a) follows, even if calculation of L_D^e is tedious while it involves h_e. By taking maximum over e, $D_e(U)$ is Lipschitz in $U \in \mathbb{R}^{N_e}$, in any norm, since $\mathbb{R}^{N_e}$ is of finite dimension, thus (19b) holds with some L_D^0. In consequence we get also $\left\|\mathcal{A}(U) - \mathcal{A}(V)\right\| \leq L_A \|U - V\|$ where L_A involves $\|B\|$. To prove (19c), it seems we would need however some handle on $\|U\|, \|V\|$. However, we recall the piecewise definition of $d(u)$ exploited in Lemma 3.1 from which we see that also $d(u)u$ is globally Lipschitz since $d(u)$ is globally bounded and in fact constant for $u \geq u^*$. Thus not only $\mathcal{A}(U)$ but also $\mathcal{A}(U)U$ is globally Lipschitz.

(iii) Now (19d) follows immediately by Cauchy–Schwarz inequality and (19c). □

Next we establish some bounds on the terms $\mathcal{M}^n U^n$. Suppressing n, for every t_n on the right hand side of (15b) we have an obvious result.

Lemma 4.2. *Assume $m(x, t)$ is bounded and is nonnegative for all x, t. Let also $F(U) = \mathcal{M}U + G$ where G is known. Then*

$$\langle F(U) - F(V), U - V\rangle = \langle \mathcal{M}U - \mathcal{M}V, U - V\rangle \geq 0. \tag{20a}$$

We also have, with some constant C_M which depends on $\|m\|_\infty$

$$\left\|F(U) - F(V)\right\| \leq C_m \|U - V\| \,. \tag{20b}$$

4.2.2. Solvability of (17). We will use the well known theorem from [43](Theorem 2.G), and [5](Theorems 5.1.4, 11.2.1, and Ex.11.2.8), and [20](Theorem 11.3.6). We recall these without proof.

Theorem 4.3 (Existence and uniqueness of a nonlinear problem)**.** *Let V be a Hilbert space, and T: $V \to V$ be strongly monotone i.e. $\langle T(u) - T(v), u - v\rangle \geq c_T \|u - v\|^2$, and Lipschitz continuous i.e. $\left\|T(u) - T(v)\right\| \leq L_T \|u - v\|$ for some c_T, L_T. Then (a) for any $b \in V$, there is a unique $u \in V$ which solves*

$$T(u) = b. \tag{21}$$

(b) Furthermore, let $K \subset V$ be non-empty, closed and convex, and T be strongly monotone and Lipschitz on K. Then for any $b \in V$ there exists a unique $u \in K$ such that

$$(T(u), v-u) \geq (b, v-u), \forall v \in K. \tag{22}$$

We use this Theorem now for (17), and apply it to the coupled system (u_1, u_2) system later in Sec. 6.1.

Proposition 1. *Under the assumptions of Lemma 4.1 and Lemma 4.2 the problem* (17) *is uniquely solvable for a sufficient small τ.*

Proof. We apply Theorem 4.3 to (17) with $T = \mathcal{C}$, $V = W_h = \mathbb{R}^{N_c}, K = K_h$. We verify first that $\mathcal{C}$ is Lipschitz on K i.e. $\left\|\mathcal{C}(U) - \mathcal{C}(V)\right\| \leq L_{\mathcal{C}}\|U - V\|$. This follows directly from Lipschitz continuity of $\mathcal{A}(U)U$ established in (19c) which holds on K_h, and from (20b) in Lemma 4.2.
We next show that the operator $\mathcal{C}$ is strongly monotone i.e.,

$$\langle \mathcal{C}(U) - \mathcal{C}(V), U - V\rangle \geq c_{\mathcal{C}}\|U - V\| \tag{23}$$

for some $c_{\mathcal{C}} > 0$ and any $U, V \in K_h$. We first expand

$$\begin{aligned} \langle \mathcal{C}(U) - \mathcal{C}(V), U - V\rangle &= (U - V, U - V) \\ &\quad - \tau(\mathcal{A}(U)U - \mathcal{A}(V)V, V - U) - \tau\langle \mathcal{M}(U - V), U - V\rangle \\ &= \|U - V\|^2 - \tau(\mathcal{A}(U)U - \mathcal{A}(V)V, V - U) - \tau\langle \mathcal{M}(U - V), U - V\rangle. \end{aligned} \tag{24}$$

Now we handle the second and third terms. For the third, since $\mathcal{M}$ has only nonnegative terms, by (20a) it satisfies $-\langle \mathcal{M}(U - V), U - V\rangle \leq 0$, but from (20b) and Cauchy-Schwarz inequality we have $\langle \mathcal{M}(U-V), U-V\rangle \leq C_m\|U - V\|^2$. Thus $-\tau\langle \mathcal{M}(U-V), U-V\rangle \geq -\tau C_m\|U - V\|^2$. For the second term, if $\mathcal{A}$ is independent of U such as in the time-lagging case, it is also a nonnegative definite linear operator, thus we have $\langle \mathcal{A}(U - V), U - V\rangle \geq 0$. In the nonlinear case, we treat the terms involving $\mathcal{A}(U)U$ as a Lipschitz perturbation similar to the $\mathcal{M}$ terms: by (19d) we have

$$\langle \mathcal{A}(U)U - \mathcal{A}(V)V, U - V\rangle \leq C_A\|U - V\|^2. \tag{25}$$

We get the lower bound (23) as long as τ is small enough so that $c_{\mathcal{C}} = 1 - \tau C_{\mathcal{A}} - \tau C_m > 0$.
We conclude by applying Theorem 4.3 to (17). □

Remark 1. *The proof of well-posedness in Proposition 1 applies, as stated, to the case of explicit treatment of advection. For the implicit treatment of advection, the well posedness is proved similarly as long as Assumption 1(A) that the advective flux is bounded holds. Then the corresponding discrete advective operator ΞU is globally Lipschitz in the problem*

$$\widetilde{\mathcal{C}}(U) = (\tau\mathcal{A}(U^{n^*}) + \mathcal{I} + \tau\Xi - \tau\mathcal{M}^n)U^n = \tau G^n. \tag{26}$$

We treat ΞU as an additional Lipschitz perturbation to $\mathcal{C}$ in (16)*, and we see $\widetilde{\mathcal{C}}$ is globally Lipschitz and monotone as long as τ is small enough.*

5. Error estimate for the fully implicit mixed finite approximation of the PVI

Now we proceed to study the error analysis of the fully discrete problem corresponding to the first unknown u_1. Our techniques follows the work in [4] which was formulated with a Kirchhof transformation for a variational equality, i.e., an unconstrained problem. We also use some ideas from [33]. We annotate.

To deal with the nonlinear diffusivity, we transform the PVI (7) using Kirchhoff transformation (as in [33]) defined as:

$$\mathcal{K} : \mathbb{R} \longrightarrow \mathbb{R}, \;\; u \longrightarrow \int_0^u d(s)\, ds. \tag{27}$$

The Kirchhoff transformation can be inverted since $d(s)$ is positive. We denote its inverse by $b(\cdot)$, and take $\xi = \mathcal{K}(u)$, thus, we have $u = b(\xi) = \mathcal{K}^{-1}(\xi)$ and $\nabla\xi = d(u)\nabla u$. Since $0 \le u \le u^*$, we have $0 \le \xi \le \xi^*$, where $\xi^* = \int_0^{u^*} d(s)\, ds$. Thus (7) can be written as

$$\partial_t b(\xi) - \nabla\cdot(\nabla\xi - \bar{\mathbf{q}}b(\xi)) + \partial I_{[0,\xi^*]}(\xi) \ni f(b(\xi)) \text{ in } \Omega, t > 0. \tag{28}$$

Let us next define the closed subset $K^* := \{\eta \in L^2(\Omega); 0 \le \eta \le \xi^* \text{ a.e. on } \Omega\}$, and denote $\mathbf{w} = -\nabla\xi + \bar{\mathbf{q}}b(\xi)$.

We complete the problem with some initial condition and homogeneous boundary condition:

$$\partial_t b(\xi) + \nabla\cdot\mathbf{w} + \partial I_{[0,\xi^*]}(\xi) \ni f(b(\xi)) \text{ in } \Omega, t > 0, \tag{29a}$$

$$\mathbf{w} = -\nabla\xi + \bar{\mathbf{q}}b(\xi) \text{ in } \Omega, t > 0, \tag{29b}$$

$$\mathbf{w}(s,t)\cdot\mathbf{n} = 0 \;\text{ on } \Gamma, t > 0, \tag{29c}$$

$$\xi(x,0) = \xi_{init} \;\text{ in } \Omega, \tag{29d}$$

where $\xi_{init} = \mathcal{K}(u_{init})$.

By properties of $d(\cdot)$ in Lemma 3.1, we have the following result.

Lemma 5.1. *The function $b(\cdot) \in C^1$ is nondecreasing and Lipschitz continuous with Lipschitz constant R_b.*

Remark 2. *By Lemma 5.1, we have*

$$\|b(\phi_1) - b(\phi_2)\|^2 \le R_b(b(\phi_1) - b(\phi_2), \phi_1 - \phi_2), \;\; \forall\phi_1, \phi_2 \in L^2, \tag{30}$$

where R_b is the Lipschitz constant in Lemma 5.1. Furthermore, since $\partial b(\xi)/\partial\xi = b_\xi(\xi) = 1/d(u)$, and by the property of $d(\cdot)$ in Lemma 3.1, there are two constants C_2 and C_3 such that

$$0 < C_2 \le b_\xi(\xi) \le C_3 < \infty \;\; \textit{for all } \xi \in R.$$

Thus,

$$C_2|\eta_1 - \eta_2| \le |b(\eta_1) - b(\eta_2)| \le C_3|\eta_1 - \eta_2|. \tag{31}$$

Assumption 2. *We assume the following for the solution $(\mathbf{w}, \xi)$.*

(A) $\xi \in L^2(H^1)$, $\xi_t \in L^2(H^{-1})$.
(B) $\nabla\cdot\mathbf{w} \in L^2(H^1)$, $\mathbf{w}_t \in L^2((H^{-1})^d)$.
Furthermore, based on the assumptions above, we have
(C) $\int_0^t \mathbf{w}(s)\, ds \in H^1((L^2)^d) \cap L^2(H_{div})$.

To overcome the low regularity of the solution we implement the time integration technique used in [4]. We integrate (29a) in time from 0 to $t \in J$, and use (4) to write problem (29) in the variational formulation. Thus, we have

$$(b(\xi(t)), \eta - \xi(t)) + (\int_0^t \nabla \cdot \mathbf{w}(s)\, ds, \eta - \xi(t))$$

$$\geq (\int_0^t f(b(\xi(s)))\, ds, \eta - \xi(t)) + (b(\xi_{init}), \eta - \xi(t)), \quad \forall \eta \in K^*, \tag{32a}$$

$$(\mathbf{w}(t), \boldsymbol{\psi}) - (\xi(t), \nabla \cdot \boldsymbol{\psi}) - (\bar{\mathbf{q}} b(\xi(t)), \boldsymbol{\psi}) = 0, \ \forall \boldsymbol{\psi} \in X. \tag{32b}$$

Using backward Euler scheme, the fully implicit mixed finite element approximation of (32) is as follows:

For $n \in \{1, \dots, N\}$, we seek a solution $(\mathbf{w}_h^n, \xi_h^n) \in X_h \times K_h^*; K_h^* = K^* \cap M_h$ such that

$$(b(\xi_h^n), \eta_h - \xi_h^n) + \tau(\sum_{j=1}^n \nabla \cdot \mathbf{w}_h^j, \eta_h - \xi_h^n)$$

$$\geq \tau(\sum_{j=1}^n f(b(\xi_h^j)), \eta_h - \xi_h^n) + (b(\xi_{init}^h), \eta_h - \xi_h^n), \quad \forall \eta_h \in K_h^*, \tag{33a}$$

$$(\mathbf{w}_h^n, \boldsymbol{\psi}_h) - (\xi_h^n, \nabla \cdot \boldsymbol{\psi}_h) - (\bar{\mathbf{q}} b(\xi_h^n), \boldsymbol{\psi}_h) = 0, \ \forall \boldsymbol{\psi}_h \in X_h, \tag{33b}$$

where $\xi_{init}^h = \pi_h \xi_{init}$.

To estimate the error between the exact solution and the approximate solution, we need to use approximation properties of finite element spaces. We define the following interpolation operator. (See e.g. ([10], page 150) and ([18], page 217)).

Definition 5.1. *The interpolation operator*

$$\rho_h : (H^1)^d \to RT_{[0]}$$

is defined by

$$\int_e (\boldsymbol{\psi} - \rho_h \boldsymbol{\psi}) \cdot \mathbf{n} = 0 \ \text{ for each edge } e \text{ of the cells in } \mathcal{T}^h, \ \forall \boldsymbol{\psi} \in (H^1)^d.$$

This means that the mean value of the normal component of a given function $\boldsymbol{\psi} \in (H^1)^d$ coincides with the normal component of $\rho_h \boldsymbol{\psi}$ on each edge.

This interpolation operator is related to the orthogonal L^2-projection onto M_h by the following property, for the proof we refer to ([8], Proposition 2.3.2, page 108).

Lemma 5.2 (Minimal Property)**.** *Let $\pi_h : M \to M_h$ be the orthogonal L^2-projection onto M_h, i.e.*

$$(v - \pi_h v, \mu_h) = 0, \ \forall v \in M, \ \forall \mu_h \in M_h.$$

Then

$$\pi_h(\nabla \cdot \boldsymbol{\psi}) = \nabla \cdot (\rho_h \boldsymbol{\psi}), \ \forall \boldsymbol{\psi} \ \in (H^1)^d.$$

The operators ρ_h and π_h defined above satisfy the following properties which we state without proof. We refer to ([8], pages 107-108), ([10], page 151), and ([18], page 217).

Lemma 5.3. *Let ρ_h and π_h be the operators defined in Definition 5.1 and Lemma 5.2, respectively. Then we have the following properties.*

(a) $(\nabla \cdot \rho_h \boldsymbol{\psi}, v_h) = (\nabla \cdot \boldsymbol{\psi}, v_h) \ \forall v_h \in M_h, \ \forall \boldsymbol{\psi} \in X,$

(b) $\|\rho_h \boldsymbol{\psi} - \boldsymbol{\psi}\|_0 \leq Ch|\boldsymbol{\psi}|_1$ *if* $\boldsymbol{\psi} \in (H^1)^d,$

(c) $\|\nabla\cdot\rho_h\boldsymbol{\psi}\|_0 \le C\|\nabla\cdot\boldsymbol{\psi}\|_0\ \forall\boldsymbol{\psi}\in X$,
(d) $\|\nabla\cdot(\boldsymbol{\psi}-\rho_h\boldsymbol{\psi})\|_0 \le Ch|\nabla\cdot\boldsymbol{\psi}|_1$ *if* $\nabla\cdot\boldsymbol{\psi}\in H^1$,
(e) $\|\pi_h v - v\|_0 \le Ch|v|_1$ *if* $v\in H^1$.

We mention that Radu et al. [33] employed Green operator technique to derive the error estimate. Instead, we follow the technique developed in [4] which used a weighted $(L^2(\Omega))^d$ projection on X_h that depends on the diffusivity. However, since the diffusivity in (32) after the Kirchhoff transformaton is the constant 1, we just apply the $(L^2(\Omega))^d$ projection on X_h: $\hat{\mathcal{P}}_h : (L^2(\Omega))^d \longrightarrow X_h$ as for all $\boldsymbol{\psi}\in(L^2(\Omega))^d$, we have

$$(\hat{\mathcal{P}}_h\boldsymbol{\psi}-\boldsymbol{\psi},\boldsymbol{\psi}_h)=0,\ \forall\boldsymbol{\psi}_h\in X_h. \tag{34}$$

It is easy to show that

$$\|\hat{\mathcal{P}}_h\boldsymbol{\psi}-\boldsymbol{\psi}\|_0 \le Ch|\boldsymbol{\psi}|_1,\ \text{if}\ \boldsymbol{\psi}\in(H^1)^d. \tag{35}$$

At the same time, in Sec. 6.2, we make use of the weighted $(L^2(\Omega))^d$ projection on X_h.

In this section, we shall use the following lemma.

Lemma 5.4. *For any vectors* $a_j\in\mathbb{R}^m$*;* $m\ge 1$ *and* $j=1,\dots,n$*, we have*

$$2\sum_{j=1}^{n}(a_j-a_{j-1},a_j)=|a_n|^2-|a_0|^2+\sum_{j=1}^{n}|a_j-a_{j-1}|^2. \tag{36}$$

Theorem 5.5. *Let* $(\mathbf{w}(t),\xi(t))\in X\times K^*$ *be a solution to* (32)*, for each* $t>0$*, that satisfies Assumption 2, and let* $(\mathbf{w}_h^n,\xi_h^n)\in X_h\times K_h^*$ *be a solution to* (33) *for* $n=1,\dots,N$*. Then there exists a constant* $C>0$ *that does not depend on* h *nor* τ *such that*

$$\sum_{j=1}^{n}\|\xi^j-\xi_h^j\|_0^2\tau+\left\|\int_0^{t_n}\mathbf{w}(s)\,ds-\sum_{j=1}^{n}\mathbf{w}_h^j\tau\right\|_0^2\le C\left(h^2+\tau^2\right).$$

Proof. Define $\bar{\mathbf{w}}^n=\frac{1}{\tau}\int_{t_{n-1}}^{t_n}\mathbf{w}(s)\,ds$, $\mathcal{I}_{\mathbf{w}}^n=\tau\sum_{j=1}^n\bar{\mathbf{w}}^j=\int_0^{t_n}\mathbf{w}(s)\,ds$, $\sigma^n=\bar{\mathbf{w}}^n-\mathbf{w}_h^n$, $\hat{\sigma}^n=\tau\sum_{j=1}^n\sigma^j=\mathcal{I}_{\mathbf{w}}^n-\tau\sum_{j=1}^n\mathbf{w}_h^j$, $e^n=\xi^n-\xi_h^n$.

The major difference between the problem (32) and the problem in [4] is that the latter is a system of equalities while (32) involves an inequality. This requires some manipulations to be able to get an estimate for the term $(b(\xi^n)-b(\xi_h^n),\xi^n-\xi_h^n)$. To this end, we take $t=t_n$ and $\eta=\xi_h^n$ in (32a), we obtain

$$\begin{aligned}(b(\xi^n),\xi^n-\xi_h^n)\ &\le\ (\int_0^{t_n}\nabla\cdot\mathbf{w}(s)\,ds,\xi_h^n-\xi^n)\\ &+\ (\int_0^{t_n}f(b(\xi(s)))\,ds,\xi^n-\xi_h^n)+(b(\xi_{init}),\xi^n-\xi_h^n).\end{aligned} \tag{37}$$

Now take $\eta_h=\pi_h\xi^n$ in (33a), we have

$$\begin{aligned}(b(\xi_h^n),\xi_h^n-\pi_h\xi^n)\ &\le\ \tau(\sum_{j=1}^n\nabla\cdot\mathbf{w}_h^j,\pi_h\xi^n-\xi_h^n)\\ &+\ \tau(\sum_{j=1}^n f(b(\xi_h^j)),\xi_h^n-\pi_h\xi^n)+(b(\xi_{init}^h),\xi_h^n-\pi_h\xi^n).\end{aligned} \tag{38}$$

Using the definition of π_h in Lemma 5.2, as $b(\xi_h^n) \in M_h$, the left hand side of (38) can be written as

$$(b(\xi_h^n), \xi_h^n - \pi_h \xi^n) = (b(\xi_h^n), \xi_h^n - \xi^n). \tag{39}$$

Similarly, the third term of the right hand side of (38) can be written as

$$(b(\xi_{init}^h), \xi_h^n - \pi_h \xi^n) = (b(\xi_{init}^h), \xi_h^n - \xi^n). \tag{40}$$

Combining (38)– (40), we obtain

$$\begin{aligned} (b(\xi_h^n), \xi_h^n - \xi^n) \quad &\leq \quad \tau(\sum_{j=1}^{n} \nabla \cdot \mathbf{w}_h^j, \pi_h \xi^n - \xi_h^n) \\ &+ \quad \tau(\sum_{j=1}^{n} f(b(\xi_h^j)), \xi_h^n - \pi_h \xi^n) + (b(\xi_{init}^h), \xi_h^n - \xi^n). \end{aligned} \tag{41}$$

Add (41) to (37), we have

$$\begin{aligned} &(b(\xi^n) - b(\xi_h^n), \xi^n - \xi_h^n) \\ \leq &(\int_0^{t_n} \nabla \cdot \mathbf{w}(s)\, ds, \xi_h^n - \xi^n) + \tau(\sum_{j=1}^{n} \nabla \cdot \mathbf{w}_h^j, \pi_h \xi^n - \xi_h^n) \\ &+ (\int_0^{t_n} f(b(\xi(s)))\, ds, \xi^n - \xi_h^n) - \tau(\sum_{j=1}^{n} f(b(\xi_h^j)), \pi_h \xi^n - \xi_h^n) \\ &+ (b(\xi_{init}) - b(\xi_{init}^h), \xi^n - \xi_h^n). \end{aligned} \tag{42}$$

Take $t = t_n$ in (32b), and subtract (33b) from the obtained equality, we get

$$(\mathbf{w}^n - \mathbf{w}_h^n, \boldsymbol{\psi}_h) = (\xi^n - \xi_h^n, \nabla \cdot \boldsymbol{\psi}_h) + (\bar{\mathbf{q}}[b(\xi^n) - b(\xi_h^n)], \boldsymbol{\psi}_h),\ \forall \boldsymbol{\psi}_h \in X_h.$$

Take $\boldsymbol{\psi}_h = \hat{\mathcal{P}}_h \hat{\sigma}^n$ in the last equality, and notice that by the definition of $\hat{\sigma}^n$, and by adding and subtracting $\rho_h \hat{\sigma}^n$, the function $\hat{\mathcal{P}}_h \hat{\sigma}^n$ can be written as $\hat{\mathcal{P}}_h \hat{\sigma}^n = \rho_h \hat{\sigma}^n + (\hat{\mathcal{P}}_h - \rho_h)\mathcal{I}_{\mathbf{w}}^n$, thus we get

$$\begin{aligned} (\mathbf{w}^n - \mathbf{w}_h^n, \hat{\mathcal{P}}_h \hat{\sigma}^n) \quad &= \quad (\xi^n - \xi_h^n, \nabla \cdot \rho_h \hat{\sigma}^n) \\ &+ \quad (\xi^n - \xi_h^n, \nabla \cdot (\hat{\mathcal{P}}_h - \rho_h)\mathcal{I}_{\mathbf{w}}^n) + (\bar{\mathbf{q}}[b(\xi^n) - b(\xi_h^n)], \hat{\mathcal{P}}_h \hat{\sigma}^n). \end{aligned} \tag{43}$$

Using the definition (34), the left hand side of (43) is:

$$\begin{aligned} (\mathbf{w}^n - \mathbf{w}_h^n, \hat{\mathcal{P}}_h \hat{\sigma}^n) \quad &= \quad (\bar{\mathbf{w}}^n - \mathbf{w}_h^n, \hat{\mathcal{P}}_h \hat{\sigma}^n) + (\mathbf{w}^n - \bar{\mathbf{w}}^n, \hat{\mathcal{P}}_h \hat{\sigma}^n) \\ &= \quad (\sigma^n, \hat{\mathcal{P}}_h \hat{\sigma}^n) + (\mathbf{w}^n - \bar{\mathbf{w}}^n, \hat{\mathcal{P}}_h \hat{\sigma}^n) \\ &= \quad (\hat{\mathcal{P}}_h \sigma^n, \hat{\mathcal{P}}_h \hat{\sigma}^n) + (\mathbf{w}^n - \bar{\mathbf{w}}^n, \hat{\mathcal{P}}_h \hat{\sigma}^n). \end{aligned} \tag{44}$$

Using the properties of π_h and ρ_h in Lemma 5.2 and Lemma 5.3, respectively, the 1st term in the right hand side of (43) can be written as:

$$
\begin{aligned}
&(\nabla\cdot\rho_h\hat{\sigma}^n,\xi^n-\xi_h^n)\\
=&(\nabla\cdot\rho_h\mathcal{I}_{\mathbf{w}}^n,\xi^n-\xi_h^n)\overset{\text{def. of }\hat{\sigma}^n}{-}(\nabla\cdot\tau\sum_{j=1}^n\mathbf{w}_h^j,\xi^n-\xi_h^n)\\
=&(\nabla\cdot\mathcal{I}_{\mathbf{w}}^n,\xi^n-\xi_h^n)+(\nabla\cdot(\rho_h\mathcal{I}_{\mathbf{w}}^n-\mathcal{I}_{\mathbf{w}}^n),\xi^n-\xi_h^n)-(\nabla\cdot\tau\sum_{j=1}^n\mathbf{w}_h^j,\xi^n-\xi_h^n)\hat{\sigma}^n\\
\overset{\text{Lemma 5.3 (a)}}{=}&(\nabla\cdot\mathcal{I}_{\mathbf{w}}^n,\xi^n-\xi_h^n)+(\nabla\cdot(\rho_h\mathcal{I}_{\mathbf{w}}^n-\mathcal{I}_{\mathbf{w}}^n),\xi^n-\pi_h\xi^n)\\
&\overset{\text{Lemma 5.2}}{-}(\nabla\cdot\tau\sum_{j=1}^n\mathbf{w}_h^j,\pi_h\xi^n-\xi_h^n). \qquad (45)
\end{aligned}
$$

Combining (43)–(45), we have

$$
\begin{aligned}
(\hat{\mathcal{P}}_h\sigma^n,\hat{\mathcal{P}}_h\hat{\sigma}^n)=&(\bar{\mathbf{w}}^n-\mathbf{w}^n,\hat{\mathcal{P}}_h\hat{\sigma}^n)+(\nabla\cdot\mathcal{I}_{\mathbf{w}}^n,\xi^n-\xi_h^n)\\
&+(\nabla\cdot(\rho_h\mathcal{I}_{\mathbf{w}}^n-\mathcal{I}_{\mathbf{w}}^n),\xi^n-\pi_h\xi^n)-(\nabla\cdot\tau\sum_{j=1}^n\mathbf{w}_h^j,\pi_h\xi^n-\xi_h^n)\\
&+(\nabla\cdot(\hat{\mathcal{P}}_h-\rho_h)\mathcal{I}_{\mathbf{w}}^n,\xi^n-\xi_h^n)+(\bar{\mathbf{q}}[b(\xi^n)-b(\xi_h^n)],\hat{\mathcal{P}}_h\hat{\sigma}^n). \qquad (46)
\end{aligned}
$$

Add (46) to (42), we get

$$
\begin{aligned}
&(b(\xi^n)-b(\xi_h^n),\xi^n-\xi_h^n)+(\hat{\mathcal{P}}_h\sigma^n,\hat{\mathcal{P}}_h\hat{\sigma}^n)\\
\le&(\bar{\mathbf{w}}^n-\mathbf{w}^n,\hat{\mathcal{P}}_h\hat{\sigma}^n)+(\nabla\cdot(\rho_h\mathcal{I}_{\mathbf{w}}^n-\mathcal{I}_{\mathbf{w}}^n),\xi^n-\pi_h\xi^n)\\
&+(\int_0^{t_n}f(b(\xi(s)))\,ds,\xi^n-\xi_h^n)-\tau(\sum_{j=1}^n f(b(\xi_h^j)),\pi_h\xi^n-\xi_h^n)\\
&+(\nabla\cdot(\hat{\mathcal{P}}_h-\rho_h)\mathcal{I}_{\mathbf{w}}^n,\xi^n-\xi_h^n)+(\bar{\mathbf{q}}[b(\xi^n)-b(\xi_h^n)],\hat{\mathcal{P}}_h\hat{\sigma}^n)\\
&+(b(\xi_{init})-b(\xi_{init}^h),\xi^n-\xi_h^n). \qquad (47)
\end{aligned}
$$

Since $\hat{\sigma}^n\in X$; $\forall n$, we have by (35)

$$(\hat{\mathcal{P}}_h\hat{\sigma}^n,\boldsymbol{\psi}_h)=(\hat{\sigma}^n,\boldsymbol{\psi}_h),\ \forall\boldsymbol{\psi}_h\in X_h,$$

and

$$(\hat{\mathcal{P}}_h\hat{\sigma}^{n-1},\boldsymbol{\psi}_h)=(\hat{\sigma}^{n-1},\boldsymbol{\psi}_h),\forall\boldsymbol{\psi}_h\in X_h.$$

Subtract the last two equations and note that $\hat{\sigma}^n=\hat{\sigma}^{n-1}+\tau\sigma^n$, so we have

$$(\hat{\mathcal{P}}_h\hat{\sigma}^n-\hat{\mathcal{P}}_h\hat{\sigma}^{n-1},\boldsymbol{\psi}_h)=(\sigma^n,\boldsymbol{\psi}_h)\tau,\ \forall\boldsymbol{\psi}_h\in X_h.$$

Take $\boldsymbol{\psi}_h=\hat{\mathcal{P}}_h\hat{\sigma}^n$ in the last equation, we obtain

$$(\sigma^n,\hat{\mathcal{P}}_h\hat{\sigma}^n)\tau=(\hat{\mathcal{P}}_h\hat{\sigma}^n-\hat{\mathcal{P}}_h\hat{\sigma}^{n-1},\hat{\mathcal{P}}_h\hat{\sigma}^n). \qquad (48)$$

Now, replace n by j in (47) and multiply by τ and take the sum from 1 through n, we have

$$\begin{aligned}
&\sum_{j=1}^{n}(b(\xi^j)-b(\xi_h^j),\xi^j-\xi_h^j)\tau+\sum_{j=1}^{n}(\hat{\mathcal{P}}_h\sigma^j,\hat{\mathcal{P}}_h\hat{\sigma}^j)\tau\\
\leq&\sum_{j=1}^{n}(\bar{\mathbf{w}}^j-\mathbf{w}^j,\hat{\mathcal{P}}_h\hat{\sigma}^j)\tau+\sum_{j=1}^{n}(\nabla\cdot(\rho_h\mathcal{I}_{\mathbf{w}}^j-\mathcal{I}_{\mathbf{w}}^j),\xi^j-\pi_h\xi^j)\tau\\
&+\sum_{j=1}^{n}\tau(\int_0^{t_j}f(b(\xi(s)))\,ds,\xi^j-\xi_h^j)-\tau^2\sum_{j=1}^{n}(\sum_{k=1}^{j}f(b(\xi_h^k)),\pi_h\xi^j-\xi_h^j)\\
&+\sum_{j=1}^{n}(\nabla\cdot(\hat{\mathcal{P}}_h-\rho_h)\mathcal{I}_{\mathbf{w}}^j,\xi^j-\xi_h^j)\tau+\sum_{j=1}^{n}(\bar{\mathbf{q}}[b(\xi^j)-b(\xi_h^j)],\hat{\mathcal{P}}_h\hat{\sigma}^j)\tau\\
&+\sum_{j=1}^{n}(b(\xi_{init})-b(\xi_{init}^h),\xi^j-\xi_h^j)\tau. \qquad (49)
\end{aligned}$$

By (48) and Lemma 5.4, we have

$$\begin{aligned}
\sum_{j=1}^{n}(\sigma^j,\hat{\mathcal{P}}_h\hat{\sigma}^j)\tau=&\sum_{j=1}^{n}(\hat{\mathcal{P}}_h\hat{\sigma}^j-\hat{\mathcal{P}}_h\hat{\sigma}^{j-1},\hat{\mathcal{P}}_h\hat{\sigma}^j)\\
=&\frac{1}{2}\|\hat{\mathcal{P}}_h\hat{\sigma}^n\|_0^2-\frac{1}{2}\|\hat{\mathcal{P}}_h\hat{\sigma}^0\|_0^2+\frac{1}{2}\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h\hat{\sigma}^j-\hat{\mathcal{P}}_h\hat{\sigma}^{j-1}\|_0^2. \qquad (50)
\end{aligned}$$

Combining (49)–(50) and using (30),(31) and (34), we get

$$\begin{aligned}
&\frac{C_2}{R_b}\sum_{j=1}^{n}\|e^j\|_0^2\tau+\frac{1}{2}\|\hat{\mathcal{P}}_h\hat{\sigma}^n\|_0^2\\
\leq&\sum_{j=1}^{n}(b(\xi^j)-b(\xi_h^j),\xi^j-\xi_h^j)\tau+\frac{1}{2}\|\hat{\mathcal{P}}_h\hat{\sigma}^n\|_0^2+\frac{1}{2}\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h\hat{\sigma}^j-\hat{\mathcal{P}}_h\hat{\sigma}^{j-1}\|_0^2\\
\leq&\sum_{l=1}^{6}T_l; \qquad (51)
\end{aligned}$$

where

$$T_1=\sum_{j=1}^{n}(\bar{\mathbf{w}}^j-\mathbf{w}^j,\hat{\mathcal{P}}_h\hat{\sigma}^j)\tau, \qquad (52)$$

$$T_2=\sum_{j=1}^{n}(\nabla\cdot(\rho_h\mathcal{I}_{\mathbf{w}}^j-\mathcal{I}_{\mathbf{w}}^j),\xi^j-\pi_h\xi^j)\tau, \qquad (53)$$

$$T_3=\sum_{j=1}^{n}\tau(\int_0^{t_j}f(b(\xi(s)))\,ds,\xi^j-\xi_h^j)-\tau^2\sum_{j=1}^{n}(\sum_{k=1}^{j}f(b(\xi_h^k)),\pi_h\xi^j-\xi_h^j), \qquad (54)$$

$$T_4=\sum_{j=1}^{n}(\nabla\cdot(\hat{\mathcal{P}}_h-\rho_h)\mathcal{I}_{\mathbf{w}}^j,\xi^j-\xi_h^j)\tau, \qquad (55)$$

$$T_5 = \sum_{j=1}^{n} (\bar{\mathbf{q}}[b(\xi^j) - b(\xi_h^j)], \hat{\mathcal{P}}_h \hat{\sigma}^j)\tau, \tag{56}$$

$$T_6 = \sum_{j=1}^{n} (b(\xi_{init}) - b(\xi_{init}^h), \xi^j - \xi_h^j)\tau + \frac{1}{2}\|\hat{\mathcal{P}}_h \hat{\sigma}^0\|_0^2. \tag{57}$$

Now we estimate each T_l; $l = 1, \ldots, 6$, and use the properties in Lemma 5.2, Lemma 5.3, and (35), we have

$$\begin{aligned}
T_1 &= \sum_{j=1}^{n} \int_\Omega \hat{\mathcal{P}}_h \sigma^j \int_{t_{j-1}}^{t_j} (\mathbf{w}(s) - \mathbf{w}^j)\, ds\, dx \\
&= \sum_{j=1}^{n} \int_\Omega \hat{\mathcal{P}}_h \sigma^j \int_{t_{j-1}}^{t_j} \int_{t_j}^{s} \mathbf{w}_t(t)\, dt\, ds\, dx \\
&\leq \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \int_{s}^{t_j} \int_\Omega |\hat{\mathcal{P}}_h \sigma^j||\mathbf{w}_t(t)|\, dx\, dt\, ds \\
&\leq \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \int_{s}^{t_j} \|\hat{\mathcal{P}}_h \sigma^j\|_0 \|\mathbf{w}_t(t)\|_0\, dt\, ds \\
&\leq \tau^{3/2} \sum_{j=1}^{n} \|\hat{\mathcal{P}}_h \sigma^j\|_0 \|\mathbf{w}_t\|_{L^2(J_j;L^2)} \\
&\leq \frac{\varepsilon_1}{2} \sum_{j=1}^{n} \|\hat{\mathcal{P}}_h \sigma^j\|_0^2 \tau + \frac{1}{2\varepsilon_1} \tau^2 \sum_{j=1}^{n} \|\mathbf{w}_t\|_{L^2(J_j;L^2)}^2 \\
&\leq \frac{\varepsilon_1}{2} \sum_{j=1}^{n} \|\hat{\mathcal{P}}_h \sigma^j\|_0^2 \tau + \frac{1}{2\varepsilon_1} \tau^2 \|\mathbf{w}_t\|_{L^2(L^2)}^2. \qquad (58)
\end{aligned}$$

$$\begin{aligned}
T_2 &= \sum_{j=1}^{n} (\nabla \cdot (\rho_h \mathcal{I}_{\mathbf{w}}^j - \mathcal{I}_{\mathbf{w}}^j), \xi^j - \pi_h \xi^j)\tau \\
&\leq \frac{1}{2} \sum_{j=1}^{n} \|\nabla \cdot (\rho_h \mathcal{I}_{\mathbf{w}}^j - \mathcal{I}_{\mathbf{w}}^j)\|_0^2 \tau + \frac{1}{2} \sum_{j=1}^{n} \|\xi^j - \pi_h \xi^j\|_0^2 \tau \\
&\leq \frac{C}{2} h^2 \sum_{j=1}^{n} |\nabla \cdot \mathcal{I}_{\mathbf{w}}^j|_1^2 \tau + \frac{C}{2} h^2 \sum_{j=1}^{n} |\xi^j|_1^2 \tau. \qquad (59)
\end{aligned}$$

$$\begin{aligned}
T_3 &= \sum_{j=1}^{n} \tau \left(\int_0^{t_j} f(b(\xi(s)))\, ds - \tau \sum_{k=1}^{j} f(b(\xi_h^k)), e^j \right) \\
&= \sum_{j=1}^{n} \tau \left(\int_0^{t_j} f(b(\xi(s)))\, ds - \tau \sum_{k=1}^{j} f(b(\xi^k)), e^j \right) \\
&+ \sum_{j=1}^{n} \tau^2 \left(\sum_{k=1}^{j} [f(b(\xi^k)) - f(b(\xi_h^k))], e^j \right) \\
&= T_{31} + T_{32}, \qquad (60)
\end{aligned}$$

with obvious notations of T_{31} and T_{32}.

Using the Lipschitz continuity of f and b in Assumption 1: (B), and Lemma 5.1, respectively, we have

$$\begin{aligned}
T_{31} &= \tau \sum_{j=1}^{n} \left(\int_0^{t_j} f(b(\xi(s)))\, ds - \tau \sum_{k=1}^{j} f(b(\xi^k)), e^j \right) \\
&= \tau \sum_{j=1}^{n} \left(\sum_{k=1}^{j} \int_{t_{k-1}}^{t_k} [f(b(\xi(s))) - f(b(\xi^k))]\, ds, e^j \right) \\
&\leq RR_b \tau \sum_{j=1}^{n} \left(\sum_{k=1}^{j} \int_{t_{k-1}}^{t_k} |\xi(s) - \xi^k|\, ds, e^j \right) \\
&= RR_b \tau \sum_{j=1}^{n} \left(\sum_{k=1}^{j} \int_{t_{k-1}}^{t_k} \int_s^{t_k} |\xi_t(t)|\, dt\, ds, e^j \right) \\
&\leq RR_b \tau \sum_{j=1}^{n} \|e^j\|_0 \sum_{k=1}^{j} \int_{t_{k-1}}^{t_k} \int_s^{t_k} \|\xi_t(t)\|_0\, dt\, ds \\
&\leq RR_b \sum_{j=1}^{n} \sum_{k=1}^{j} (\|e^j\|_0 \tau \|\xi_t(\cdot)\|_{L^2(J_k;L^2)} \tau^{3/2}) \\
&\leq \frac{\varepsilon_2}{2} \sum_{j=1}^{n} \sum_{k=1}^{j} (\|e^j\|_0^2 \tau) \tau + \frac{R^2 R_b^2}{2\varepsilon_2} \sum_{j=1}^{n} \sum_{k=1}^{j} \tau^3 \|\xi_t(\cdot)\|^2_{L^2(J_k;L^2)} \\
&\leq \frac{T\varepsilon_2}{2} \sum_{j=1}^{n} \|e^j\|_0^2 \tau + \tau^2 \left(\frac{R^2 R_b^2 T}{2\varepsilon_2} \|\xi_t(\cdot)\|^2_{L^2(L^2)}\right). \qquad (61)
\end{aligned}$$

$$\begin{aligned}
T_{32} &= \sum_{j=1}^{n} \tau^2 \left(\sum_{k=1}^{j} [f(b(\xi^k)) - f(b(\xi_h^k))], e^j \right) \\
&\leq \tau^2 RR_b \sum_{j=1}^{n} \sum_{k=1}^{j} \|e^k\|_0 \|e^j\|_0 \\
&\leq \frac{\varepsilon_3}{2} \sum_{j=1}^{n} \left(\sum_{k=1}^{j} \|e^k\|_0^2 \tau \right) \tau + \frac{R^2 R_b^2 T}{2\varepsilon_3} \sum_{j=1}^{n} \|e^j\|_0^2 \tau. \qquad (62)
\end{aligned}$$

$$\begin{aligned}
T_4 &= \sum_{j=1}^{n} (\nabla \cdot (\hat{\mathcal{P}}_h - \rho_h) \mathcal{I}_{\mathbf{w}}^j, \xi^j - \xi_h^j) \tau \\
&\leq \frac{1}{2\varepsilon_4} \sum_{j=1}^{n} \|(\nabla \cdot (\hat{\mathcal{P}}_h - \rho_h) \mathcal{I}_{\mathbf{w}}^j\|_0^2 \tau + \frac{\varepsilon_4}{2} \sum_{j=1}^{n} \|e^j\|_0^2 \tau \\
&\leq \frac{1}{2\varepsilon_4} \sum_{j=1}^{n} \|(\nabla \cdot (\hat{\mathcal{P}}_h \mathcal{I}_{\mathbf{w}}^j - \mathcal{I}_{\mathbf{w}}^j)\|_0^2 \tau \\
&\quad + \frac{1}{2\varepsilon_4} \sum_{j=1}^{n} \|(\nabla \cdot (\mathcal{I}_{\mathbf{w}}^j - \rho_h \mathcal{I}_{\mathbf{w}}^j)\|_0^2 \tau + \frac{\varepsilon_4}{2} \sum_{j=1}^{n} \|e^j\|_0^2 \tau \\
&\leq h^2 \frac{C}{\varepsilon_4} \sum_{j=1}^{n} |\nabla \cdot \mathcal{I}_{\mathbf{w}}^j|_1^2 \tau + \frac{\varepsilon_4}{2} \sum_{j=1}^{n} \|e^j\|_0^2 \tau. \qquad (63)
\end{aligned}$$

$$
\begin{aligned}
T_5 &= \sum_{j=1}^{n}(\bar{\mathbf{q}}[b(\xi^j)-b(\xi_h^j)],\hat{\mathcal{P}}_h\hat{\sigma}^j)\tau \\
&\le \|\bar{\mathbf{q}}\|_\infty R_b\tau\sum_{j=1}^{n}\|e^j\|_0\|\hat{\mathcal{P}}_h\hat{\sigma}^j\|_0 \\
&\le \frac{\|\bar{\mathbf{q}}\|_\infty^2 R_b^2\varepsilon_5}{2}\sum_{j=1}^{n}\|e^j\|_0^2\tau+\frac{1}{2\varepsilon_5}\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h\hat{\sigma}^j\|_0^2\tau. \qquad (64)
\end{aligned}
$$

$$
\begin{aligned}
T_6 &= \sum_{j=1}^{n}(b(\xi_{init})-b(\xi_{init}^h),\xi^j-\xi_h^j)\tau+\frac{1}{2}\|\hat{\mathcal{P}}_h\hat{\sigma}^0\|_0^2 \\
&\le \sum_{j=1}^{n}R_b^2\frac{\varepsilon_6}{2}\|\xi_{init}-\xi_{init}^h\|_0^2\tau+\frac{1}{2\varepsilon_6}\sum_{j=1}^{n}\|e^j\|_0^2\tau+\frac{1}{2}\|\hat{\mathcal{P}}_h\hat{\sigma}^0\|_0^2 \\
&\le \frac{\varepsilon_6 TR_b^2}{2}h^2+\frac{1}{2\varepsilon_6}\sum_{j=1}^{n}\|e^j\|_0^2\tau+\frac{1}{2}\|\hat{\mathcal{P}}_h\hat{\sigma}^0\|_0^2. \qquad (65)
\end{aligned}
$$

Combining (51)–(65), and by appropriate choices of ε_i's , we have

$$
\begin{aligned}
\sum_{j=1}^{n}\|e^j\|_0^2\tau &+ \|\hat{\mathcal{P}}_h\hat{\sigma}^n\|_0^2 \\
&\le C\left[h^2+\tau^2+\|\hat{\mathcal{P}}_h\hat{\sigma}^0\|_0^2+\sum_{j=1}^{n}\left(\sum_{k=1}^{j}\|e^k\|_0^2\tau\right)\tau+\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h\hat{\sigma}^j\|_0^2\tau\right]. \qquad (66)
\end{aligned}
$$

Using Gronwall's Lemma, the proof is complete. □

6. Analysis of the nonlinear coupled system

Now consider the coupled biofilm-nutrient system which extends (17), and which is solved for $U = (U_1; U_2)$ with U_1 denoting the biofilm concentration and U_2 denoting the nutrient. We extend the inner $\langle\cdot,\cdot\rangle$ product on M to the inner product on $M\times M$, with $\|U\|_{L^2\times L^2}^2 = \|U_1\|_0^2+\|U_2\|_0^2$, and we work with $K_h \times M_h$.

We seek $(U_1; U_2) \in K_h \times M_h$ which satisfies $C(U) = (\mathcal{C}_1(U); \mathcal{C}_2(U)) = \tau G = \tau(G_1; G_2)$ or the fully discrete system, a counterpart of (1). The unconstrained system reads

$$\mathcal{C}_1(U_1;U_2) = (\tau\mathcal{A}_1(U_1^n)+\mathcal{I}-\tau\mathcal{M}_1(U_2^n))U_1^n = \tau G_1^n, \qquad (67a)$$

$$\mathcal{C}_2(U_1;U_2) = \tau\mathcal{M}_2({U_2}^n)U_1^n+(\tau\mathcal{A}_2(U_1^n)+\mathcal{I})U_2^n = \tau G_2^n. \qquad (67b)$$

The coupling between the equations is in the diffusivity $\mathcal{A}_k = \mathcal{A}_k(U_1)$ for each $k = 1, 2$ in each equation which depends on U_1. In addition, each equation has coupling terms representing the growth $U_1\mathcal{M}_1(U_2)$ and decay $= -U_1\mathcal{M}_2(U_2)$, where each diagonal nonnegative matrix $\mathcal{M}_k(U_2) = \texttt{diag}((m_k(U_c))_c$ is associated with a nonnegative, bounded, and Lipschitz continuous functions $m_k(u_2)$.

Under the constraints, we have a system extending (17)

$$\mathcal{C}_1(U_1;U_2)+\tau\Lambda_1(U_1^n) = \tau G_1^n, \qquad (68a)$$

$$\mathcal{C}_2(U_1;U_2) = \tau G_2^n. \qquad (68b)$$

Below we first prove well-posedness of the coupled system. Next we derive error estimates for this system.

6.1. Well-posedness for the coupled system.

Proposition 2. *Assume that each function $m_k : [0, \infty) \to \mathbb{R}$ for $k = 1, 2$ is a Lipschitz continuous nonnegative and nondecreasing function bounded from above. We will also use assumptions on u^*, and the diffusivity as above. Then the system* (68) *is solvable on the set $K_h \times M_h \subset M_h \times M_h$ as long as τ is small enough.*

Proof. We will use Theorem 4.3 again following similar steps as in the proof of Proposition 1. We first note that each of $\mathcal{C}_1(U_1, U_2), \mathcal{C}_2(U_1, U_2)$ is globally Lipschitz with respect to each U_1, U_2, since the individual operators $\mathcal{A}_k, \mathcal{M}_k$, $k = 1, 2$ are either constant in each of the variables U_j or depend nonlinearly on a particular U_j, with the nonlinearity globally Lipschitz and bounded as in Proposition 1. Moreover, $\mathcal{M}_k$ are globally bounded due to the properties of each m_k.
Writing each $\mathcal{C}_k(U) = U_k + \tau\Psi_k(U_1, U_2)$ where $\Psi_1(U) = \mathcal{A}_1(U_1)U_1 - \mathcal{M}_1(U_2)U_1$ and $\Psi_2(U) = \mathcal{A}_2(U_1)U_2 + \mathcal{M}_2(U_2)U_1$ we have that each Ψ_k is globally Lipschitz in each U_j thus also in $U = (U_1, U_2)$ in any product norm, thus we conclude that $\mathcal{C} = (\mathcal{C}_1; \mathcal{C}_2)$ is also globally Lipschitz in U.
It remains to check that $\mathcal{C}$ is strongly monotone. However, each $\mathcal{C}_k$ is made of $\mathcal{I}$ plus a globally Lipschitz term multiplied by τ. With τ small enough, we obtain strong monotonicity of each $\mathcal{C}_k$ and in turn of $\mathcal{C}$, since now

$$
\begin{aligned}
(\mathcal{C}(U) - \mathcal{C}(V), U - V) = &\|U_1 - V_1\|^2 + \|U_2 - V_2\|^2 \\
&+ \tau(\Psi_1(U) - \Psi_1(V), U_1 - V_1) + \tau(\Psi_2(U) - \Psi_2(V), U_2 - V_2).
\end{aligned} \tag{69}
$$

The proof is complete after we apply Theorem 4.3. □

Remark 3. *The well-posedness in Proposition 2 holds regardless whether advection is treated explicitly or implicitly, as explained in Remark 1.*

6.2. Error Estimate for the coupled system. Now we consider the coupled system (1); we follow the same technique we used with the scalar PVI in Sec. (5).
Let $\xi = \mathcal{K}(u_1)$; where

$$\mathcal{K}(u_1) = \int_0^{u_1} d_1(s)\, ds$$

is the Kirchhoff transformation, and let $b(\cdot)$ be its inverse. Using the following notations

$$
\begin{aligned}
\xi^* &:= \int_0^{u^*} d_1(s)\, ds, \\
\mathbf{w} &:= -\nabla\xi + \bar{\mathbf{q}}b(\xi), \\
\mu &:= u_2, \\
\mathbf{z} &:= -d_2(b(\xi))\nabla\mu + \bar{\mathbf{q}}\mu,
\end{aligned}
$$

system (1) can be written equivalently in the following mixed formulation:

For each $t > 0$, we seek a solution $[(\mathbf{w}(t), \xi(t)), (\mathbf{z}(t), \mu(t))] \in (X \times K^*) \times (X \times M)$ such that

$$(b(\xi(t)), \eta - \xi(t)) + (\int_0^t \nabla \cdot \mathbf{w}(s)\, ds, \eta - \xi(t))$$

$$\geq (\int_0^t r_1(b(\xi(s)), \mu(s))\, ds, \eta - \xi(t)) + (b(\xi_{init}), \eta - \xi(t)), \quad \forall \eta \in K^*, \tag{70a}$$

$$(\mathbf{w}(t), \boldsymbol{\psi}) - (\xi(t), \nabla \cdot \boldsymbol{\psi}) - (\bar{\mathbf{q}} b(\xi(t)), \boldsymbol{\psi}) = 0, \ \forall \boldsymbol{\psi} \in X, \tag{70b}$$

$$(\mu(t), \gamma) + (\int_0^t \nabla \cdot \mathbf{z}(s)\, ds, \gamma) = (\int_0^t r_2(b(\xi(s)), \mu(s))\, ds, \gamma)$$

$$+ (\mu_{init}, \gamma), \quad \forall \gamma \in M, \tag{70c}$$

$$(d_2^{-1}(b(\xi))\mathbf{z}(t), \boldsymbol{\zeta}) - (\mu(t), \nabla \cdot \boldsymbol{\zeta}) - (d_2^{-1}(b(\xi))\bar{\mathbf{q}}\mu, \boldsymbol{\zeta}) = 0, \ \forall \boldsymbol{\zeta} \in X, \tag{70d}$$

where K^* is defined analogously to that in Sec. 5.

The fully implicit MFE approximation of (70) is:

For $n \in \{1, \ldots, N\}$, we seek a solution $[(\mathbf{w}_h^n, \xi_h^n), (\mathbf{z}_h^n, \mu_h^n)] \in (X_h \times K_h^*) \times (X_h \times M_h); K_h^* = K^* \cap M_h$ such that

$$(b(\xi_h^n), \eta_h - \xi_h^n) + \tau(\sum_{j=1}^n \nabla \cdot \mathbf{w}_h^j, \eta_h - \xi_h^n)$$

$$\geq \tau(\sum_{j=1}^n r_1(b(\xi_h^j), \mu_h^j), \eta_h - \xi_h^n) + (b(\xi_{init}^h), \eta_h - \xi_h^n), \quad \forall \eta_h \in K_h^*, \tag{71a}$$

$$(\mathbf{w}_h^n, \boldsymbol{\psi}_h) - (\xi_h^n, \nabla \cdot \boldsymbol{\psi}_h) - (\bar{\mathbf{q}} b(\xi_h^n), \boldsymbol{\psi}_h) = 0, \ \forall \boldsymbol{\psi}_h \in X_h, \tag{71b}$$

$$(\mu_h^n, \gamma_h) + \tau(\sum_{j=1}^n \nabla \cdot \mathbf{z}_h^j, \gamma_h) = \tau(\sum_{j=1}^n r_2(b(\xi_h^j), \mu_h^j), \gamma_h)$$

$$+ (\mu_{init}^h, \gamma_h), \quad \forall \gamma_h \in M_h, \tag{71c}$$

$$(d_2^{-1}(b(\xi_h^n))\mathbf{z}_h^n, \boldsymbol{\zeta}_h) - (\mu_h^n, \nabla \cdot \boldsymbol{\zeta}_h) - (d_2^{-1}(b(\xi_h^n))\bar{\mathbf{q}}\mu_h^n, \boldsymbol{\zeta}_h) = 0, \ \forall \boldsymbol{\zeta}_h \in X_h, \tag{71d}$$

where $\xi_{init}^h = \pi_h \xi_{init}$ and $\mu_{init}^h = \pi_h \mu_{init}^h$.

Assumption 3. *In this section we make use of the following assumptions*

(A) *d_1 satisfies the same conditions $d(\cdot)$ satisfies in Lemma 3.1.*

(B) *By formula* (6), *$d_2^{-1}(\cdot)$ is a smooth function, and there are constants ν_1 and ν_2 such that*

$$0 < \nu_1 \leq d_2^{-1}(s) \leq \nu_2 \textit{ for } \forall s \in \mathbb{R}.$$

Moreover, $d_2^{-1}(\cdot)$ is a continuous Lipschitz function with a Lipschitz constant L_d^2.

(C) $\bar{\mathbf{q}} = \bar{\mathbf{q}}(\mathbf{x}) \in (L^\infty(\Omega))^d$.

(D) *$r_1(\cdot,\cdot)$ and $r_2(\cdot,\cdot)$ are smooth functions on $\mathbb{R}^2$ with a global Lipschitz constant R, we also assume that r_1 and r_2 are uniformly bounded on $\mathbb{R}^+ = [0, \infty)$.*

We also assume the following regularities:

(E) $\xi, \mu \in L^2(H^1)$, $\xi_t, \mu_t \in L^2(H^{-1})$, $\xi_t \in L^\infty(L^\infty)$.

(F) $\nabla \cdot \mathbf{w}, \nabla \cdot \mathbf{z} \in L^2(H^1)$, $\mathbf{w}_t, \mathbf{z}_t \in L^2((H^{-1})^d)$.

Furthermore, based on the assumptions above, we have

(G) $\int_0^t \mathbf{w}(s)\, ds, \int_0^t \mathbf{z}(s)\, ds \in H^1((L^2)^d) \cap L^2(H_{div})$.

To deal with the coupling, we also assume the following for $(\mu, \mathbf{z})$

(H) $\mu \in L^\infty(L^\infty)$ *and* $\mathbf{z} \in L^\infty((L^\infty)^d)$.

To derive an error estimate between the solutions of (70) and (71), we shall use the properties of the operators defined in Lemma 5.3, and the property of the projection $\hat{\mathcal{P}}_h$ in (35).

We also define the following weighted projections $\hat{\mathcal{P}}_h^{d_2}$ which depends on the diffusivity $d_2(\cdot)$ as:

$$\hat{\mathcal{P}}_h^{d_2} : (L^2(\Omega))^d \longrightarrow X_h$$

such that

$$(d_2^{-1}(b(\xi(t))(\hat{\mathcal{P}}_h^{d_2}\mathbf{z} - \mathbf{z}), \boldsymbol{\zeta}_h) = 0, \ \ \forall \boldsymbol{\zeta}_h \in X_h. \tag{72}$$

The weighted projector $\hat{\mathcal{P}}_h^{d_2}$ satisfies the following property.

Lemma 6.1. *Let* $\hat{\mathcal{P}}_h^{d_2}$ *be defined as in* (72)*, then if* $\boldsymbol{\psi} \in (H^1)^d$*, we have*

$$\|\hat{\mathcal{P}}_h^{d_2}\boldsymbol{\psi} - \boldsymbol{\psi}\|_0 \le h|\boldsymbol{\psi}|_1. \tag{73}$$

Proof. Using the properties of $d_2^{-1}(\cdot)$ in Assumption 3 (B), and the definition of $\hat{\mathcal{P}}_h^{d_2}$ in (72), we have for all $\boldsymbol{\psi} \in (H^1)^d$

$$\begin{aligned}
\nu_1\|\hat{\mathcal{P}}_h^{d_2}\boldsymbol{\psi} - \boldsymbol{\psi}\|_0^2 &\le (d_2^{-1}(u)(\hat{\mathcal{P}}_h^{d_2}\boldsymbol{\psi} - \boldsymbol{\psi}), \hat{\mathcal{P}}_h\boldsymbol{\psi} - \boldsymbol{\psi}) \\
&= (d_2^{-1}(u)(\hat{\mathcal{P}}_h^{d_2}\boldsymbol{\psi} - \boldsymbol{\psi}), \hat{\mathcal{P}}_h\boldsymbol{\psi}) - (d_2^{-1}(u)(\hat{\mathcal{P}}_h^{d_2}\boldsymbol{\psi} - \boldsymbol{\psi}), \boldsymbol{\psi}) \\
&= (d_2^{-1}(u)(\hat{\mathcal{P}}_h^{d_2}\boldsymbol{\psi} - \boldsymbol{\psi}), \rho_h\boldsymbol{\psi} - \boldsymbol{\psi}) - (d_2^{-1}(u)(\hat{\mathcal{P}}_h^{d_2}\boldsymbol{\psi} - \boldsymbol{\psi}), \rho_h\boldsymbol{\psi}) \\
&\le \nu_2\|\hat{\mathcal{P}}_h^{d_2}\boldsymbol{\psi} - \boldsymbol{\psi}\|_0\|\rho_h\boldsymbol{\psi} - \boldsymbol{\psi}\|_0.
\end{aligned}$$

Thus by the property (b) in Lemma 5.3, we have

$$\|\hat{\mathcal{P}}_h^{d_2}\boldsymbol{\psi} - \boldsymbol{\psi}\|_0 \ \le \ \frac{\nu_2}{\nu_1}\|\rho_h\boldsymbol{\psi} - \boldsymbol{\psi}\|_0 \le Ch|\boldsymbol{\psi}|_1.$$

□

Theorem 6.2. *Let* $[(\mathbf{w}(t), \xi(t)), (\mathbf{z}(t), \mu(t))] \in (X \times K^*) \times (X \times M)$ *be a solution to* (70)*, for each* $t > 0$*, that satisfies Assumption 3, and let* $[(\mathbf{w}_h^n, \xi_h^n), (\mathbf{z}_h^n, \mu_h^n)] \in (X_h \times K_h^*) \times (X_h \times M_h)$ *be a solution to* (71) *for* $n = 1, \dots, N$*. Then there exists a constant* $C > 0$ *that does not depend on* h *nor* τ *such that*

$$\sum_{j=1}^{n}\left(\|\xi^j - \xi_h^j\|_0^2 + \|\mu^j - \mu_h^j\|_0^2\right)\tau + \left\|\int_0^{t_n}\mathbf{w}(s)\,ds - \sum_{j=1}^{n}\mathbf{w}_h^j\tau\right\|_0^2 + \left\|\int_0^{t_n}\mathbf{z}(s)\,ds - \sum_{j=1}^{n}\mathbf{z}_h^j\tau\right\|_0^2 \le C\left(h^2 + \tau^2\right).$$

Proof. As in Sec. 5, we define $\bar{\mathbf{w}}^n = \frac{1}{\tau}\int_{t_{n-1}}^{t_n}\mathbf{w}(s)\,ds$, $\mathcal{I}_\mathbf{w}^n = \tau\sum_{j=1}^n \bar{\mathbf{w}}^j = \int_0^{t_n}\mathbf{w}(s)\,ds$. Analogous notations for $\bar{\mathbf{z}}^n$, and $\mathcal{I}_\mathbf{z}^n$. We also define $\sigma_\mathbf{w}^n = \bar{\mathbf{w}}^n - \mathbf{w}_h^n$, $\hat{\sigma}_\mathbf{w}^n = \tau\sum_{j=1}^n\sigma_\mathbf{w}^j = \mathcal{I}_\mathbf{w}^n - \tau\sum_{j=1}^n\mathbf{w}_h^j$. Analogous notations for $\sigma_\mathbf{z}^n$ and $\hat{\sigma}_\mathbf{z}^n$. Also define $e_\xi^n = \xi^n - \xi_h^n$ and $e_\mu^n = \mu^n - \mu_h^n$.

Take $t = t_n$ and $\eta = \xi_h^n$ in (70a), and take $\eta_h = \pi_h \xi^n$ in (71a). Using the definition of π_h, and add the result together, we obtain

$$\begin{aligned}
& (b(\xi^n) - b(\xi_h^n), \xi^n - \xi_h^n) \\
\leq \ & (\int_0^{t_n} \nabla \cdot \mathbf{w}(s)\, ds, \xi_h^n - \xi^n) - \tau(\sum_{j=1}^{n} \nabla \cdot \mathbf{w}_h^j, \xi_h^n - \pi_h \xi^n) \\
+ \ & (\int_0^{t_n} r_1(b(\xi(s)), \mu(s))\, ds, \xi^n - \xi_h^n) - \tau(\sum_{j=1}^{n} r_1(b(\xi_h^j), \mu_h^j), \xi^n - \xi_h^n) \\
+ \ & (b(\xi_{init}) - b(\xi_{init}^h), \xi^n - \xi_h^n).
\end{aligned} \tag{74}$$

Now take $t = t_n$ and $\gamma = \mu^n - \mu_h^n$ in (70c), and take $\gamma_h = \mu_h^n - \pi_h \mu^n$ in (71c), we get

$$\begin{aligned}
\|e_\mu^n\|_0^2 \ = \ & (\int_0^{t_n} \nabla \cdot \mathbf{z}(s)\, ds, \mu_h^n - \mu^n) - \tau(\sum_{j=1}^{n} \nabla \cdot \mathbf{z}_h^j, \mu_h^n - \pi_h \mu^n) \\
+ \ & (\int_0^{t_n} r_2(b(\xi(s)), \mu(s))\, ds, \mu^n - \mu_h^n) - \tau(\sum_{j=1}^{n} r_2(b(\xi_h^j), \mu_h^j), \mu^n - \mu_h^n) \\
+ \ & (\mu_{init} - \mu_{init}^h, \mu^n - \mu_h^n).
\end{aligned} \tag{75}$$

Take $t = t_n$ in (70b), then subtract (71b) from the obtained equality, and take $\boldsymbol{\psi}_h = \hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^n = \rho_h \hat{\sigma}_{\mathbf{w}}^n + (\hat{\mathcal{P}}_h - \rho_h)\mathcal{I}_{\mathbf{w}}^n$, we obtain

$$\begin{aligned}
(\hat{\mathcal{P}}_h \sigma_{\mathbf{w}}^n, \hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^n) \ = \ & (\bar{\mathbf{w}}^n - \mathbf{w}^n, \hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^n) + (\nabla \cdot \mathcal{I}_{\mathbf{w}}^n, \xi^n - \xi_h^n) \\
+ \ & (\nabla \cdot (\rho_h \mathcal{I}_{\mathbf{w}}^n - \mathcal{I}_{\mathbf{w}}^n), \xi^n - \pi_h \xi^n) - (\nabla \cdot \tau \sum_{j=1}^{n} \mathbf{w}_h^j, \pi_h \xi^n - \xi_h^n) \\
+ \ & (\nabla \cdot (\hat{\mathcal{P}}_h - \rho_h)\mathcal{I}_{\mathbf{w}}^n, \xi^n - \xi_h^n) + (\bar{\mathbf{q}}[b(\xi^n) - b(\xi_h^n)], \hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^n).
\end{aligned} \tag{76}$$

Now take $t = t_n$ in (70d), and subtract (71d) from the obtained equality, we get

$$\begin{aligned}
& (d_2^{-1}(b(\xi^n))\mathbf{z}^n, \boldsymbol{\zeta}_h) - (d_2^{-1}(b(\xi_h^n))\mathbf{z}_h^n, \boldsymbol{\zeta}_h) \\
- \ & ([d_2^{-1}(b(\xi^n)) - d_2^{-1}(b(\xi_h^n))]\bar{\mathbf{q}}\mu^n, \boldsymbol{\zeta}_h) - (d_2^{-1}(b(\xi_h^n))\bar{\mathbf{q}}(\mu^n - \mu_h^n), \boldsymbol{\zeta}_h) \\
= \ & (\nabla \cdot \boldsymbol{\zeta}_h, \mu^n - \mu_h^n), \quad \forall \boldsymbol{\zeta}_h \in X_h.
\end{aligned} \tag{77}$$

The first two terms in the left hand side of (77) can be written as

$$\begin{aligned}
& (d_2^{-1}(b(\xi^n))\mathbf{z}^n, \boldsymbol{\zeta}_h) - (d_2^{-1}(b(\xi_h^n))\mathbf{z}_h^n, \boldsymbol{\zeta}_h) \\
= \ & ([d_2^{-1}(b(\xi^n)) - d_2^{-1}(b(\xi_h^n))]\mathbf{z}^n, \boldsymbol{\zeta}_h) \\
+ \ & (d_2^{-1}(b(\xi_h^n))(\mathbf{z}^n - \bar{\mathbf{z}}^n), \boldsymbol{\zeta}_h) + (d_2^{-1}(b(\xi_h^n))\sigma_{\mathbf{z}}^n, \boldsymbol{\zeta}_h), \quad \forall \boldsymbol{\zeta}_h \in X_h.
\end{aligned} \tag{78}$$

Combine (77) and (78), and take $\boldsymbol{\zeta}_h = \hat{\mathcal{P}}_h^{d_2} \hat{\sigma}_{\mathbf{z}}^n = \rho_h \hat{\sigma}_{\mathbf{z}}^n + (\hat{\mathcal{P}}_h^{d_2} - \rho_h)\mathcal{I}_{\mathbf{z}}^n$ in the obtained result, and use the definition of $\hat{\mathcal{P}}_h^{d_2}$, π_h, and the property of ρ_h in (72), Lemma 5.2, and Lemma 5.3 (a), respectively, we get

$$\begin{aligned}
& (d_2^{-1}(b(\xi_h^n))\hat{\mathcal{P}}_h^{d_2}\sigma_{\mathbf{z}}^n, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n) \\
= \; & (d_2^{-1}(b(\xi_h^n))(\bar{\mathbf{z}}^n - \mathbf{z}^n), \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n) \\
- \; & ([d_2^{-1}(b(\xi^n)) - d_2^{-1}(b(\xi_h^n))]\mathbf{z}^n, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n) \\
+ \; & ([d_2^{-1}(b(\xi^n)) - d_2^{-1}(b(\xi_h^n))]\bar{\mathbf{q}}\mu^n, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n) + (d_2^{-1}(b(\xi_h^n))\bar{\mathbf{q}}e_\mu^n, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n) \\
+ \; & (\nabla \cdot (\rho_h \mathcal{I}_{\mathbf{z}}^n - \mathcal{I}_{\mathbf{z}}^n), \mu^n - \pi_h \mu^n) + (\nabla \cdot \mathcal{I}_{\mathbf{z}}^n, \mu^n - \mu_h^n) \\
- \; & \tau(\nabla \cdot \sum_{j=1}^{n} \mathbf{z}_h^j, \pi_h\mu^n - \mu_h^n) + (\nabla \cdot (\hat{\mathcal{P}}_h^{d_2} - \rho_h)\mathcal{I}_{\mathbf{z}}^n, \mu^n - \mu_h^n).
\end{aligned} \tag{79}$$

Combining the ine(qualities) (74)–(76), and (79), and using Assumption 3 (B), we have

$$\begin{aligned}
& (b(\xi^n) - b(\xi_h^n), \xi^n - \xi_h^n) + \|e_\mu^n\|_0^2 \\
+ \; & (\hat{\mathcal{P}}_h\sigma_{\mathbf{w}}^n, \hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^n) + (d_2^{-1}(b(\xi^n))\hat{\mathcal{P}}_h^{d_2}\sigma_{\mathbf{z}}^n, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n) \\
\leq \; & (\int_0^{t_n} r_1(b(\xi(s)), \mu(s))\, ds, \xi^n - \xi_h^n) - \tau(\sum_{j=1}^{n} r_1(b(\xi_h^j), \mu_h^j), \xi^n - \xi_h^n) \\
+ \; & (\int_0^{t_n} r_2(b(\xi(s)), \mu(s))\, ds, \mu^n - \mu_h^n) - \tau(\sum_{j=1}^{n} r_2(b(\xi_h^j), \mu_h^j), \mu^n - \mu_h^n) \\
+ \; & (\bar{\mathbf{w}}^n - \mathbf{w}^n, \hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^n) + (d_2^{-1}(b(\xi_h^n))(\bar{\mathbf{z}}^n - \mathbf{z}^n), \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n) \\
+ \; & (\nabla \cdot (\rho_h \mathcal{I}_{\mathbf{w}}^n - \mathcal{I}_{\mathbf{w}}^n), \xi^n - \pi_h\xi^n) + (\nabla \cdot (\rho_h \mathcal{I}_{\mathbf{z}}^n - \mathcal{I}_{\mathbf{z}}^n), \mu^n - \pi_h\mu^n) \\
+ \; & (\nabla \cdot (\hat{\mathcal{P}}_h - \rho_h)\mathcal{I}_{\mathbf{w}}^n, \xi^n - \xi_h^n) + (\nabla \cdot (\hat{\mathcal{P}}_h^{d_2} - \rho_h)\mathcal{I}_{\mathbf{z}}^n, \mu^n - \mu_h^n) \\
+ \; & (\bar{\mathbf{q}}[b(\xi^n) - b(\xi_h^n)], \hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^n) \\
- \; & ([d_2^{-1}(b(\xi^n)) - d_2^{-1}(b(\xi_h^n))]\mathbf{z}^n, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n) \\
+ \; & ([d_2^{-1}(b(\xi^n)) - d_2^{-1}(b(\xi_h^n))]\bar{\mathbf{q}}\mu^n, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n) + (d_2^{-1}(b(\xi_h^n))\bar{\mathbf{q}}e_\mu^n, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n) \\
+ \; & (b(\xi_{init}) - b(\xi_{init}^h), \xi^n - \xi_h^n) + (\mu_{init} - \mu_{init}^h, \mu^n - \mu_h^n).
\end{aligned} \tag{80}$$

Note that

$$\hat{\sigma}_{\mathbf{w}}^{n-1} = \hat{\sigma}_{\mathbf{w}}^n - \sigma_{\mathbf{w}}^n \tau, \ \forall n.$$

Similarly,

$$\hat{\sigma}_{\mathbf{z}}^{n-1} = \hat{\sigma}_{\mathbf{z}}^n - \sigma_{\mathbf{z}}^n \tau, \ \forall n.$$

Since $\hat{\sigma}_{\mathbf{w}}^n$ and $\hat{\sigma}_{\mathbf{w}}^{n-1} \in X, \ \forall n$, we have by the definition of the projection $\hat{\mathcal{P}}_h$ on X_h,

$$(\hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^n - \hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^{n-1}, \boldsymbol{\psi}_h) = (\sigma_{\mathbf{w}}^n, \boldsymbol{\psi}_h)\tau, \ \forall \boldsymbol{\psi}_h \in X_h.$$

Take $\boldsymbol{\psi}_h = \hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^n$ in the last equation, we obtain

$$(\sigma_{\mathbf{w}}^n, \hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^n)\tau = (\hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^n - \hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^{n-1}, \hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^n). \tag{81}$$

Now, by the definition of $\hat{\mathcal{P}}_h^{d_2}$ in (72) and since $\hat{\sigma}_{\mathbf{z}}^n$ and $\hat{\sigma}_{\mathbf{z}}^{n-1} \in X$, we have

$$(d_2^{-1}(b(\xi^n))(\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n - \hat{\sigma}_{\mathbf{z}}^n), \boldsymbol{\zeta}_h) = 0; \forall \boldsymbol{\zeta}_h \in X_h,$$

and

$$(d_2^{-1}(b(\xi^{n-1}))(\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{n-1} - \hat{\sigma}_{\mathbf{z}}^{n-1}), \boldsymbol{\zeta}_h) = 0; \forall \boldsymbol{\zeta}_h \in X_h.$$

Subtract the last two qualities, we get

$$\begin{aligned}
& (d_2^{-1}(b(\xi^n))\sigma_{\mathbf{z}}^n, \boldsymbol{\zeta}_h)\tau \\
= \; & (d_2^{-1}(b(\xi^n))(\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n - \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{n-1}), \boldsymbol{\zeta}_h) \\
+ \; & ([d_2^{-1}(b(\xi^n)) - d_2^{-1}(b(\xi^{n-1}))](\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{n-1} - \hat{\sigma}_{\mathbf{z}}^{n-1}), \boldsymbol{\zeta}_h); \forall \boldsymbol{\zeta}_h \in X_h.
\end{aligned}$$

Take $\boldsymbol{\zeta}_h = \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n$ in the last equality, we get

$$\begin{aligned}
& (d_2^{-1}(b(\xi^n))\sigma_{\mathbf{z}}^n, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n)\tau \\
= \; & (d_2^{-1}(b(\xi^n))(\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n - \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{n-1}), \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n) \\
+ \; & ([d_2^{-1}(b(\xi^n)) - d_2^{-1}(b(\xi^{n-1}))](\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{n-1} - \hat{\sigma}_{\mathbf{z}}^{n-1}), \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n).
\end{aligned}$$

That is,

$$\begin{aligned}
& (d_2^{-1}(b(\xi^n))\sigma_{\mathbf{z}}^n, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n)\tau \\
= \; & \frac{1}{2}\Big[(d_2^{-1}(b(\xi^n))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n) - (d_2^{-1}(b(\xi^{n-1}))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{n-1}, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{n-1})\Big] \\
+ \; & \frac{1}{2}\Big[(d_2^{-1}(b(\xi^n))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n) + (d_2^{-1}(b(\xi^{n-1}))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{n-1}, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{n-1})\Big] \\
+ \; & ([d_2^{-1}(b(\xi^n)) - d_2^{-1}(b(\xi^{n-1}))](\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{n-1} - \hat{\sigma}_{\mathbf{z}}^{n-1}), \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n) \\
- \; & (d_2^{-1}(b(\xi^n))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{n-1}, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n). \qquad (82)
\end{aligned}$$

Replace n by j in (80) and multiply by τ and take the sum from 1 through n, we obtain

$$\begin{aligned}
& \sum_{j=1}^n (b(\xi^j) - b(\xi_h^j), \xi^j - \xi_h^j)\tau + \sum_{j=1}^n \|e_\mu^j\|_0^2\tau \\
+ \; & \sum_{j=1}^n (\hat{\mathcal{P}}_h\sigma_{\mathbf{w}}^j, \hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^j)\tau + \sum_{j=1}^n (d_2^{-1}(b(\xi^j))\hat{\mathcal{P}}_h^{d_2}\sigma_{\mathbf{z}}^j, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j)\tau \\
\le \; & \sum_{j=1}^n \Big(\int_0^{t_j} r_1(b(\xi(s)), \mu(s))\, ds, \xi^j - \xi_h^j\Big)\tau - \tau^2 \sum_{j=1}^n \Big(\sum_{k=1}^j r_1(b(\xi_h^k), \mu_h^k), \xi^j - \xi_h^j\Big) \\
+ \; & \sum_{j=1}^n \Big(\int_0^{t_j} r_2(b(\xi(s)), \mu(s))\, ds, \mu^j - \mu_h^j\Big)\tau - \tau^2 \sum_{j=1}^n \Big(\sum_{k=1}^j r_2(b(\xi_h^k), \mu_h^k), \mu^j - \mu_h^j\Big) \\
+ \; & \sum_{j=1}^n (\bar{\mathbf{w}}^j - \mathbf{w}^j, \hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^j)\tau + \sum_{j=1}^n (d_2^{-1}(b(\xi_h^j))(\bar{\mathbf{z}}^j - \mathbf{z}^j), \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j)\tau \\
+ \; & \sum_{j=1}^n (\nabla\cdot(\rho_h\mathcal{I}_{\mathbf{w}}^j - \mathcal{I}_{\mathbf{w}}^j), \xi^j - \pi_h\xi^j)\tau + \sum_{j=1}^n (\nabla\cdot(\rho_h\mathcal{I}_{\mathbf{z}}^j - \mathcal{I}_{\mathbf{z}}^j), \mu^j - \pi_h\mu^j)\tau \\
+ \; & \sum_{j=1}^n (\nabla\cdot(\hat{\mathcal{P}}_h - \rho_h)\mathcal{I}_{\mathbf{w}}^j, \xi^j - \xi_h^j)\tau + \sum_{j=1}^n (\nabla\cdot(\hat{\mathcal{P}}_h^{d_2} - \rho_h)\mathcal{I}_{\mathbf{z}}^j, \mu^j - \mu_h^j)\tau
\end{aligned}$$

$$\begin{aligned}
&+ \sum_{j=1}^{n}(\bar{\mathbf{q}}[b(\xi^j) - b(\xi_h^j)], \hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^j)\tau \\
&- \sum_{j=1}^{n}([d_2^{-1}(b(\xi^j)) - d_2^{-1}(b(\xi_h^j))]\mathbf{z}^j, \hat{\mathcal{P}}_h^{d_2} \hat{\sigma}_{\mathbf{z}}^j)\tau \\
&+ \sum_{j=1}^{n}([d_2^{-1}(b(\xi^j)) - d_2^{-1}(b(\xi_h^j))]\bar{\mathbf{q}}\mu^j, \hat{\mathcal{P}}_h^{d_2} \hat{\sigma}_{\mathbf{z}}^j)\tau + \sum_{j=1}^{n}(d_2^{-1}(b(\xi_h^j))\bar{\mathbf{q}} e_\mu^j, \hat{\mathcal{P}}_h^{d_2} \hat{\sigma}_{\mathbf{z}}^j)\tau \\
&+ \sum_{j=1}^{n}(b(\xi_{init}) - b(\xi_{init}^h), \xi^j - \xi_h^j)\tau + \sum_{j=1}^{n}(\mu_{init} - \mu_{init}^h, \mu^j - \mu_h^j)\tau. \qquad (83)
\end{aligned}$$

By (81) and Lemma 5.4, we have

$$\begin{aligned}
\sum_{j=1}^{n}(\sigma_{\mathbf{w}}^j, \hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^j)\tau &= \sum_{j=1}^{n}(\hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^j - \hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^{j-1}, \hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^j) \\
&= \frac{1}{2}\|\hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^n\|_0^2 - \frac{1}{2}\|\hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^0\|_0^2 + \frac{1}{2}\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^j - \hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^{j-1}\|_0^2. \qquad (84)
\end{aligned}$$

By (82), we have

$$\begin{aligned}
&\sum_{j=1}^{n}(d_2^{-1}(b(\xi^j))\sigma_{\mathbf{z}}^j, \hat{\mathcal{P}}_h^{d_2} \hat{\sigma}_{\mathbf{z}}^j)\tau \\
&= \frac{1}{2}\sum_{j=1}^{n}\Big[(d_2^{-1}(b(\xi^j))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j) - (d_2^{-1}(b(\xi^{j-1}))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1})\Big] \\
&+ \frac{1}{2}\sum_{j=1}^{n}\Big[(d_2^{-1}(b(\xi^j))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j) + (d_2^{-1}(b(\xi^{j-1}))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1})\Big] \\
&+ \sum_{j=1}^{n}([d_2^{-1}(b(\xi^j)) - d_2^{-1}(b(\xi^{j-1}))](\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1} - \hat{\sigma}_{\mathbf{z}}^{j-1}), \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j) \\
&- \sum_{j=1}^{n}(d_2^{-1}(b(\xi^j))(\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j). \qquad (85)
\end{aligned}$$

Note that

$$\begin{aligned}
&\frac{1}{2}\sum_{j=1}^{n}\Big[(d_2^{-1}(b(\xi^j))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j) - (d_2^{-1}(b(\xi^{j-1}))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1})\Big] \\
&= \frac{1}{2}(d_2^{-1}(b(\xi^n))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^n) \\
&- \frac{1}{2}(d_2^{-1}(b(\xi^0))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^0, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^0). \qquad (86)
\end{aligned}$$

Combining (83)–(86), we have

$$\begin{aligned}
&\sum_{j=1}^{n}(b(\xi^j) - b(\xi_h^j), \xi^j - \xi_h^j)\tau + \sum_{j=1}^{n}\|e_\mu^j\|_0^2\tau + \frac{1}{2}\|\hat{\mathcal{P}}_h \sigma_{\mathbf{w}}^n\|_0^2 \\
&\qquad + \frac{1}{2}\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^j - \hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^{j-1}\|_0^2 + \frac{\nu_1}{2}\|\hat{\mathcal{P}}_h^{d_2}\sigma_{\mathbf{z}}^n\|_0^2 \le \sum_{l=1}^{14} T_l, \qquad (87)
\end{aligned}$$

with obvious notations of T_l's.

Now we estimate each T_l, $l = 1, \ldots, 14$.

$$
\begin{aligned}
T_1 &= \sum_{j=1}^{n}\left(\left[\int_0^{t_j} r_1(b(\xi(s)),\mu(s))\ ds - \sum_{k=1}^{j} r_1(b(\xi_h^k),\mu_h^k)\tau\right], e_\xi^j\right)\tau \\
&= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_1(b(\xi(s)),\mu(s)) - r_1(b(\xi_h^k),\mu_h^k)]\ ds, e_\xi^j)\tau \\
&= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_1(b(\xi(s)),\mu(s)) - r_1(b(\xi(s)),\mu_h^k)]\ ds, e_\xi^j)\tau \\
&+ \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_1(b(\xi(s)),\mu_h^k) - r_1(b(\xi_h^k),\mu_h^k)]\ ds, e_\xi^j)\tau \\
&= k_{11} + k_{12}, \qquad (88)
\end{aligned}
$$

with obvious notations of k_{11} and k_{12}.

$$
\begin{aligned}
k_{11} &= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_1(b(\xi(s)),\mu(s)) - r_1(b(\xi(s)),\mu_h^k)]\ ds, e_\xi^j)\tau \\
&= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_1(b(\xi(s)),\mu(s)) - r_1(b(\xi(s)),\mu^k)]\ ds, e_\xi^j)\tau \\
&+ \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_1(b(\xi(s)),\mu^k) - r_1(b(\xi(s)),\mu_h^k)]\ ds, e_\xi^j)\tau \\
&= r_{111} + r_{112}. \qquad (89)
\end{aligned}
$$

$$
\begin{aligned}
r_{111} &= \sum_{j=1}^{n}\tau\sum_{k=1}^{j}\int_\Omega\int_{t_{k-1}}^{t_k}[r_1(b(\xi(s)),\mu(s)) - r_1(b(\xi(s)),\mu^k)]\ e_\xi^j\ ds\ dx \\
&\le R\sum_{j=1}^{n}\tau\sum_{k=1}^{j}\int_\Omega\int_{t_{k-1}}^{t_k}|\mu(s)-\mu^k|\ e_\xi^j\ ds\ dx \\
&\le R\sum_{j=1}^{n}\tau\sum_{k=1}^{j}\int_\Omega\int_{t_{k-1}}^{t_k}\int_s^{t_k}|\mu_t(t)|\ e_\xi^j\ dt\ ds\ dx \\
&\le R\sum_{j=1}^{n}\tau\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}\int_s^{t_k}\|\mu_t(t)\|_0\|e_\xi^j\|_0\ dt\ ds \\
&\le R\sum_{j=1}^{n}\tau\|e_\xi^j\|_0\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}\tau^{1/2}\|\mu_t(t)\|_{L^2(J_k;L^2)}\ ds \\
&= R\sum_{j=1}^{n}\tau^2\|e_\xi^j\|_0\sum_{k=1}^{j}\tau^{1/2}\|\mu_t(t)\|_{L^2(J_k;L^2)}. \qquad (90)
\end{aligned}
$$

By Cauchy-Schwarz inequality, we have

$$\begin{aligned}\sum_{k=1}^{j}\tau^{1/2}\|\mu_t(t)\|_{L^2(J_k;L^2)} &\leq \frac{1}{2}(\sum_{k=1}^{j}\tau+\sum_{k=1}^{j}\|\mu_t(t)\|^2_{L^2(J_k;L^2)})\\ &\leq \frac{1}{2}(T+\|\mu_t(t)\|^2_{L^2(J;L^2)}).\end{aligned} \tag{91}$$

Insert (91) in (90), we get

$$\begin{aligned} r_{111} &\leq \frac{R}{2}(T+\|\mu_t(t)\|^2_{L^2(L^2)})\sum_{j=1}^{n}\tau^2\|e_\xi^j\|_0\\ &\leq \frac{\varepsilon_1}{2}(\frac{R}{2}(T+\|\mu_t(t)\|^2_{L^2(L^2)})^2T\tau^2+\frac{1}{2\varepsilon_1}\sum_{j=1}^{n}\tau\|e_\xi^j\|_0^2.\\ &= C\tau^2+\frac{1}{2\varepsilon_1}\sum_{j=1}^{n}\tau\|e_\xi^j\|_0^2.\end{aligned} \tag{92}$$

$$\begin{aligned} r_{112} &= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_1(b(\xi(s)),\mu^k)-r_1(b(\xi(s)),\mu_h^k)]\ ds), e_\xi^j)\tau\\ &\leq \sum_{j=1}^{n}\tau\sum_{k=1}^{j}\int_\Omega\int_{t_{k-1}}^{t_k}|r_1(b(\xi(s)),\mu^k)-r_1(b(\xi(s)),\mu_h^k)|\ e_\xi^j\ ds\ dx\\ &\leq R\sum_{j=1}^{n}\tau\sum_{k=1}^{j}\int_\Omega\int_{t_{k-1}}^{t_k}|\mu^k-\mu_h^k|\ e_\xi^j\ ds\ dx\\ &= R\sum_{j=1}^{n}\tau^2\sum_{k=1}^{j}\int_\Omega|e_\mu^k||e_\xi^j|\ dx\\ &\leq R\sum_{j=1}^{n}\sum_{k=1}^{j}\tau^2\|e_\mu^k\|_0\|e_\xi^j\|_0\\ &\leq R\frac{\varepsilon_2}{2}\sum_{j=1}^{n}\left(\sum_{k=1}^{j}\|e_\mu^k\|_0^2\tau\right)\tau+\frac{T}{2\varepsilon_2}\sum_{j=1}^{n}\|e_\xi^j\|_0^2\tau.\end{aligned} \tag{93}$$

$$\begin{aligned} k_{12} &= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_1(b(\xi(s)),\mu_h^k)-r_1(b(\xi_h^k),\mu_h^k)]\ ds, e_\xi^j)\tau\\ &= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_1(b(\xi(s)),\mu_h^k)-r_1(b(\xi^k),\mu_h^k)]\ ds, e_\xi^j)\tau\\ &+ \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_1(b(\xi^k),\mu_h^k)-r_1(b(\xi_h^k),\mu_h^k)]\ ds, e_\xi^j)\tau\\ &= r_{121}+r_{122}.\end{aligned} \tag{94}$$

$$
\begin{aligned}
r_{121} &= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_1(b(\xi(s)),\mu_h^k)-r_1(b(\xi^k),\mu_h^k)]\ \ ds, e_\xi^j)\tau \\
&\le RR_b\sum_{j=1}^{n}\tau\sum_{k=1}^{j}\int_\Omega\int_{t_{k-1}}^{t_k}|\xi(s)-\xi^k|\ \ e_\xi^j\ ds\ dx \\
&\le RR_b\sum_{j=1}^{n}\tau\sum_{k=1}^{j}\int_\Omega\int_{t_{k-1}}^{t_k}\left(\int_s^{t_k}|\xi_t(t)|dt\right)\ \ e_\xi^j\ ds\ dx \\
&\le RR_b\sum_{j=1}^{n}\tau\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}\int_s^{t_k}\|\xi_t(t)\|_0\|e_\xi^j\|_0\ dt\ ds \\
&\le RR_b\sum_{j=1}^{n}\tau^2\sum_{k=1}^{j}\|e_\xi^j\|_0\tau^{1/2}\|\xi_t(t)\|_{L^2(J_k;L^2)} \\
&\le \frac{RR_b}{2}\sum_{j=1}^{n}\tau^2\|e_\xi^j\|_0\left[\sum_{k=1}^{j}\tau+\sum_{k=1}^{j}\|\xi_t(t)\|_{L^2(J_k;L^2)}^2\right] \\
&\le \frac{RR_b}{2}\sum_{j=1}^{n}\tau^2\|e_\xi^j\|_0\left[T+\|\xi_t(t)\|_{L^2(L^2)}^2\right] \\
&\le \frac{R^2R_b^2T\varepsilon_3}{8}\left[T+\|\xi_t(t)\|_{L^2(L^2)}^2\right]^2\tau^2+\frac{1}{2\varepsilon_3}\sum_{j=1}^{n}\|e_\xi^j\|_0^2\tau \\
&\le C\tau^2+\frac{1}{2\varepsilon_3}\sum_{j=1}^{n}\|e_\xi^j\|_0^2\tau. \qquad (95)
\end{aligned}
$$

$$
\begin{aligned}
r_{122} &= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_1(b(\xi^k),\mu_h^k)-r_1(b(\xi_h^k),\mu_h^k)]\ \ ds, e_\xi^j)\tau \\
&\le RR_b\sum_{j=1}^{n}\tau^2\sum_{k=1}^{j}\int_\Omega|\xi^k-\xi_h^k|\ \ e_\xi^j\ \ dx \\
&\le \sum_{j=1}^{n}RR_b\tau^2\sum_{k=1}^{j}\|e_\xi^k\|_0\|e_\xi^j\|_0 \\
&\le \frac{TR^2R_b^2\varepsilon_4}{2}\sum_{j=1}^{n}\|e_\xi^j\|_0^2\tau+\frac{1}{2\varepsilon_4}\sum_{j=1}^{n}\left(\sum_{k=1}^{j}\|e_\xi^k\|_0^2\tau\right)\tau. \qquad (96)
\end{aligned}
$$

$$
\begin{aligned}
T_2 &= \sum_{j=1}^{n}\left(\left[\int_0^{t_j}r_2(b(\xi(s)),\mu(s))\ \ ds-\sum_{k=1}^{j}r_2(b(\xi_h^k),\mu_h^k)\tau\right], e_\mu^j\right)\tau \\
&= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_2(b(\xi(s)),\mu(s))-r_2(b(\xi_h^k),\mu_h^k)]\ \ ds, e_\mu^j)\tau \\
&= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_2(b(\xi(s)),\mu(s))-r_2(b(\xi(s)),\mu_h^k)]\ \ ds, e_\mu^j)\tau
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_2(b(\xi(s)),\mu_h^k)-r_2(b(\xi_h^k),\mu_h^k)]\ \ ds, e_\mu^j)\tau \\
&= \quad k_{21}+k_{22} \qquad (97)
\end{aligned}
$$

$$
\begin{aligned}
k_{21} &= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_2(b(\xi(s)),\mu(s))-r_2(b(\xi(s)),\mu_h^k)]\ \ ds, e_\mu^j)\tau \\
&= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_2(b(\xi(s)),\mu(s))-r_2(b(\xi(s)),\mu^k)]\ \ ds, e_\mu^j)\tau \\
&+ \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_2(b(\xi(s)),\mu^k)-r_2(b(\xi(s)),\mu_h^k)]\ \ ds, e_\mu^j)\tau \\
&= r_{211}+r_{212} \qquad (98)
\end{aligned}
$$

$$
\begin{aligned}
r_{211} &= \sum_{j=1}^{n}\tau\sum_{k=1}^{j}\int_\Omega\int_{t_{k-1}}^{t_k}[r_2(b(\xi(s)),\mu(s))-r_2(b(\xi(s)),\mu^k)]\ e_\mu^j\ \ ds\ \ dx \\
&\leq R\sum_{j=1}^{n}\tau\sum_{k=1}^{j}\int_\Omega\int_{t_{k-1}}^{t_k}|\mu(s)-\mu^k|\ e_\mu^j\ \ ds\ \ dx \\
&\leq R\sum_{j=1}^{n}\tau\sum_{k=1}^{j}\int_\Omega\int_{t_{k-1}}^{t_k}\int_s^{t_k}|\mu_t(t)|\ e_\mu^j\ \ dt\ \ ds\ \ dx \\
&\leq R\sum_{j=1}^{n}\tau\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}\int_s^{t_k}\|\mu_t(t)\|_0\|e_\mu^j\|_0\ \ dt\ \ ds \\
&\leq R\sum_{j=1}^{n}\tau^2\|e_\mu^j\|_0\sum_{k=1}^{j}\tau^{1/2}\|\mu_t(t)\|_{L^2(J_k;L^2)} \\
&\leq \frac{R}{2}(T+\|\mu_t(t)\|^2_{L^2(L^2)})\sum_{j=1}^{n}\tau^2\|e_\mu^j\|_0 \\
&\leq \frac{\varepsilon_5}{2}(\frac{R}{2}(T+\|\mu_t(t)\|^2_{L^2(L^2)})^2T\tau^2+\frac{1}{2\varepsilon_5}\sum_{j=1}^{n}\tau\|e_\mu^j\|_0^2 \\
&= C\tau^2+\frac{1}{2\varepsilon_5}\sum_{j=1}^{n}\|e_\mu^j\|_0^2\tau. \qquad (99)
\end{aligned}
$$

$$
\begin{aligned}
r_{212} &= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_2(b(\xi(s)),\mu^k)-r_2(b(\xi(s)),\mu_h^k)]\ \ ds), e_\mu^j)\tau \\
&\leq R\sum_{j=1}^{n}\tau\sum_{k=1}^{j}\int_\Omega\int_{t_{k-1}}^{t_k}|\mu^k-\mu_h^k|\ \ e_\mu^j\ \ ds\ \ dx \\
&\leq R\sum_{j=1}^{n}\sum_{k=1}^{j}\tau^2\|e_\mu^k\|_0\|e_\mu^j\|_0 \leq R\frac{\varepsilon_6}{2}\sum_{j=1}^{n}\left(\sum_{k=1}^{j}\|e_\mu^k\|_0^2\tau\right)\tau+\frac{T}{2\varepsilon_6}\sum_{j=1}^{n}\|e_\mu^j\|_0^2\tau. \qquad (100)
\end{aligned}
$$

$$
\begin{aligned}
k_{22} &= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_2(b(\xi(s)),\mu_h^k)-r_2(b(\xi_h^k),\mu_h^k)]\ \ ds, e_\mu^j)\tau \\
&= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_2(b(\xi(s)),\mu_h^k)-r_2(b(\xi^k),\mu_h^k)]\ \ ds, e_\mu^j)\tau \\
&+ \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_2(b(\xi^k),\mu_h^k)-r_2(b(\xi_h^k),\mu_h^k)]\ \ ds, e_\mu^j)\tau \\
&= r_{221}+r_{222}. \qquad (101)
\end{aligned}
$$

$$
\begin{aligned}
r_{221} &= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_2(b(\xi(s)),\mu_h^k)-r_2(b(\xi^k),\mu_h^k)]\ \ ds, e_\mu^j)\tau \\
&\le RR_b\tau\sum_{j=1}^{n}\sum_{k=1}^{j}\int_\Omega\left(\int_{t_{k-1}}^{t_k}|\xi(s)-\xi^k|\ \ ds\right)e_\mu^j\ dx \\
&\le RR_b\tau\sum_{j=1}^{n}\sum_{k=1}^{j}\int_\Omega e_\mu^j\int_{t_{k-1}}^{t_k}\left(\int_s^{t_k}|\xi_t(t)|dt\right)ds\ \ dx \\
&\le RR_b\tau\sum_{j=1}^{n}\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}\int_s^{t_k}\int_\Omega|e_\mu^j||\xi_t(t)|dx\ dt\ ds \\
&\le RR_b\sum_{j=1}^{n}\tau^2\sum_{k=1}^{j}\|e_\mu^j\|_0\tau^{1/2}\|\xi_t(t)\|_{L^2(J_k;L^2)} \\
&\le \frac{RR_b}{2}\sum_{j=1}^{n}\tau^2\|e_\mu^j\|_0\left[\sum_{k=1}^{j}\tau+\sum_{k=1}^{j}\|\xi_t(t)\|^2_{L^2(J_k;L^2)}\right] \\
&\le \frac{RR_b}{2}\sum_{j=1}^{n}\tau^2\|e_\mu^j\|_0\left[T+\|\xi_t(t)\|^2_{L^2(L^2)}\right] \\
&\le \frac{R^2R_b^2T\varepsilon_7}{8}\left[T+\|\xi_t(t)\|^2_{L^2(L^2)}\right]^2\tau^2+\frac{1}{2\varepsilon_7}\sum_{j=1}^{n}\|e_\mu^j\|_0^2\tau \\
&\le C\tau^2+\frac{1}{2\varepsilon_7}\sum_{j=1}^{n}\|e_\mu^j\|_0^2\tau. \qquad (102)
\end{aligned}
$$

$$
\begin{aligned}
r_{222} &= \sum_{j=1}^{n}(\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}[r_2(b(\xi^k),\mu_h^k)-r_2(b(\xi_h^k),\mu_h^k)]\ \ ds, e_\mu^j)\tau \\
&\le RR_b\sum_{j=1}^{n}\tau\sum_{k=1}^{j}\int_\Omega e_\mu^j\int_{t_{k-1}}^{t_k}|\xi^k-\xi_h^k|\ ds\ dx \\
&\le RR_b\sum_{j=1}^{n}\tau^2\sum_{k=1}^{j}\|e_\xi^k\|_0\|e_\mu^j\|_0 \\
&\le \frac{R^2R_b^2T\varepsilon_8}{2}\sum_{j=1}^{n}\|e_\mu^j\|_0^2\tau+\frac{1}{2\varepsilon_8}\sum_{j=1}^{n}\left(\sum_{k=1}^{j}\|e_\xi^k\|_0^2\tau\right)\tau. \qquad (103)
\end{aligned}
$$

$$\begin{aligned}
T_3 &= \sum_{j=1}^{n}(\bar{\mathbf{w}}^j - \mathbf{w}^j, \hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^j)\tau \\
&= \sum_{j=1}^{n}\tau\int_\Omega\left(\hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^j\frac{1}{\tau}\int_{t_{j-1}}^{t_j}\big(\mathbf{w}(s)-\mathbf{w}(t_j)\big)\,ds\,dx\right) \\
&\leq \sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\int_\Omega\hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^j\left(\int_s^{t_j}|\mathbf{w}_t|\,dt\right)\,dx\,ds \\
&\leq \sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\int_s^{t_j}\int_\Omega|\hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^j||\mathbf{w}_t(t)|\,dx\,dt\,ds \\
&\leq \sum_{j=1}^{n}\|\hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^j\|_0\int_{t_{j-1}}^{t_j}\int_s^{t_j}\|\mathbf{w}_t(t)\|_0\,dt\,ds \\
&\leq \sum_{j=1}^{n}\tau^{3/2}\|\hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^j\|_0\|\mathbf{w}_t\|_{L^2(J_j;L^2)} \\
&\leq \frac{\varepsilon_9}{2}\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^j\|_0^2\tau + \tau^2\frac{1}{2\varepsilon_9}\|\mathbf{w}_t\|_{L^2(L^2)}. \qquad (104)
\end{aligned}$$

$$\begin{aligned}
T_4 &= \sum_{j=1}^{n}(d_2^{-1}(b(\xi_h^j))(\bar{\mathbf{z}}^j - \mathbf{z}^j), \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j)\tau \\
&= \sum_{j=1}^{n}\tau\left(\int_\Omega d_2^{-1}(b(\xi_h^j))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j\frac{1}{\tau}\int_{t_{j-1}}^{t_j}(\mathbf{z}(s)-\mathbf{z}(t_j))\,ds\,dx\right) \\
&\leq \sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\int_\Omega d_2^{-1}(b(\xi_h^j))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j\left(\int_s^{t_j}|\mathbf{z}_t|\,dt\right)\,dx\,ds \\
&\leq \nu_2\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\int_s^{t_j}\|\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j\|_0\|\mathbf{z}_t(t)\|_0\,dt\,ds \\
&\leq \sum_{j=1}^{n}\tau^{3/2}\nu_2\|\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j\|_0\|\mathbf{z}_t\|_{L^2(J_j;L^2)} \\
&\leq \frac{\varepsilon_{10}\nu_2^2}{2}\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j\|_0^2\tau + \tau^2\frac{1}{2\varepsilon_{10}}\|\mathbf{z}_t\|_{L^2(L^2)}^2. \qquad (105)
\end{aligned}$$

$$\begin{aligned}
T_5 &= \sum_{j=1}^{n}(\nabla\cdot(\rho_h\mathcal{I}_{\mathbf{w}}^j - \mathcal{I}_{\mathbf{w}}^j), \xi^j - \pi_h\xi^j)\tau \\
&\leq \frac{1}{2}\sum_{j=1}^{n}\|\nabla\cdot(\rho_h\mathcal{I}_{\mathbf{w}}^j - \mathcal{I}_{\mathbf{w}}^j)\|_0^2\tau + \frac{1}{2}\sum_{j=1}^{n}\|\xi^j - \pi_h\xi^j\|_0^2\tau \\
&\leq \frac{C}{2}h^2\sum_{j=1}^{n}|\nabla\cdot\mathcal{I}_{\mathbf{w}}^j|_1^2\tau + \frac{C}{2}h^2\sum_{j=1}^{n}|\xi^j|_1^2\tau. \qquad (106)
\end{aligned}$$

Similarly,

$$\begin{aligned} T_6 &= \sum_{j=1}^{n}(\nabla\cdot(\rho_h\mathcal{I}_{\mathbf{z}}^j-\mathcal{I}_{\mathbf{z}}^j),\mu^j-\pi_h\mu^j)\tau \\ &\le \frac{C}{2}h^2\sum_{j=1}^{n}|\nabla\cdot\mathcal{I}_{\mathbf{z}}^j|_1^2\tau+\frac{C}{2}h^2\sum_{j=1}^{n}|\mu^j|_1^2\tau. \end{aligned} \tag{107}$$

$$\begin{aligned} T_7 &= \sum_{j=1}^{n}(\nabla\cdot(\hat{\mathcal{P}}_h-\rho_h)\mathcal{I}_{\mathbf{w}}^j,\xi^j-\xi_h^j)\tau \\ &\le \frac{1}{2\varepsilon_{11}}\sum_{j=1}^{n}\|\nabla\cdot(\hat{\mathcal{P}}_h-\rho_h)\mathcal{I}_{\mathbf{w}}^j\|_0^2\tau+\frac{\varepsilon_{11}}{2}\sum_{j=1}^{n}\|e_\xi^j\|_0^2\tau \\ &\le \frac{1}{2\varepsilon_{11}}\sum_{j=1}^{n}\|\nabla\cdot(\hat{\mathcal{P}}_h\mathcal{I}_{\mathbf{w}}^j-\mathcal{I}_{\mathbf{w}}^j)\|_0^2\tau \\ &+ \frac{1}{2\varepsilon_{11}}\sum_{j=1}^{n}\|\nabla\cdot(\mathcal{I}_{\mathbf{w}}^j-\rho_h\mathcal{I}_{\mathbf{w}}^j)\|_0^2\tau+\frac{\varepsilon_{11}}{2}\sum_{j=1}^{n}\|e_\xi^j\|_0^2\tau \\ &\le h^2\frac{C}{\varepsilon_{11}}\sum_{j=1}^{n}|\nabla\cdot\mathcal{I}_{\mathbf{w}}^j|_1^2\tau+\frac{\varepsilon_{11}}{2}\sum_{j=1}^{n}\|e_\xi^j\|_0^2\tau. \end{aligned} \tag{108}$$

Similarly,

$$\begin{aligned} T_8 &= \sum_{j=1}^{n}(\nabla\cdot(\hat{\mathcal{P}}_h^{d_2}-\rho_h)\mathcal{I}_{\mathbf{z}}^j,\mu^j-\mu_h^j)\tau \\ &\le h^2\frac{C}{\varepsilon_{12}}\sum_{j=1}^{n}|\nabla\cdot\mathcal{I}_{\mathbf{z}}^j|_1^2\tau+\frac{\varepsilon_{12}}{2}\sum_{j=1}^{n}\|e_\mu^j\|_0^2\tau. \end{aligned} \tag{109}$$

$$\begin{aligned} T_9 &= \sum_{j=1}^{n}(\bar{\mathbf{q}}[b(\xi^j)-b(\xi_h^j)],\hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^j)\tau \\ &\le \|\bar{\mathbf{q}}\|_\infty R_b\tau\sum_{j=1}^{n}\|e_\xi^j\|_0\|\hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^j\|_0 \\ &\le \frac{\|\bar{\mathbf{q}}\|_\infty^2R_b^2\varepsilon_{13}}{2}\sum_{j=1}^{n}\|e_\xi^j\|_0^2\tau+\frac{1}{2\varepsilon_{13}}\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^j\|_0^2\tau. \end{aligned} \tag{110}$$

$$\begin{aligned} T_{10} &= -\sum_{j=1}^{n}([d_2^{-1}(b(\xi^j))-d_2^{-1}(b(\xi_h^j))]\mathbf{z}^j,\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j)\tau \\ &\le \sum_{j=1}^{n}\tau\int_\Omega|d_2^{-1}(b(\xi^j))-d_2^{-1}(b(\xi_h^j))||\mathbf{z}^j||\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j|\,dx \\ &\le L_d^2\|\mathbf{z}^j\|_\infty\sum_{j=1}^{n}\tau\int_\Omega|e_\xi^j||\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j|\,dx \\ &\le (L_d^2)^2\|\mathbf{z}\|_{L^\infty(L^\infty)}^2\frac{\varepsilon_{14}}{2}\sum_{j=1}^{n}\|e_\xi^j\|_0^2\tau+\frac{1}{2\varepsilon_{14}}\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j\|^2\tau. \end{aligned} \tag{111}$$

$$
\begin{aligned}
T_{11} &= \sum_{j=1}^{n}([d_2^{-1}(b(\xi^j)) - d_2^{-1}(b(\xi_h^j))]\bar{\mathbf{q}}\mu^j, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j)\tau \\
&\le \sum_{j=1}^{n}\tau\int_\Omega |d_2^{-1}(b(\xi^j)) - d_2^{-1}(b(\xi_h^j))||\bar{\mathbf{q}}||\mu^j||\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j|\,dx \\
&\le L_d^2\|\bar{\mathbf{q}}\|_\infty\|\mu^j\|_\infty\sum_{j=1}^{n}\tau\int_\Omega |e_\xi^j||\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j|\,dx \\
&\le (L_d^2)^2\|\bar{\mathbf{q}}\|_\infty^2\|\mu\|_{L^\infty(L^\infty)}^2\frac{\varepsilon_{15}}{2}\sum_{j=1}^{n}\|e_\xi^j\|_0^2\tau + \frac{1}{2\varepsilon_{15}}\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j\|^2\tau. \qquad (112)
\end{aligned}
$$

$$
\begin{aligned}
T_{12} &= \sum_{j=1}^{n}(d_2^{-1}(b(\xi_h^j))\bar{\mathbf{q}}e_\mu^j, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j)\tau \\
&\le \sum_{j=1}^{n}\nu_2\|\bar{\mathbf{q}}\|_\infty\|e_\mu^j\|_0\|\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j\|_0\tau \\
&\le \frac{\varepsilon_{16}}{2}\nu_2^2\|\bar{\mathbf{q}}\|_\infty^2\sum_{j=1}^{n}\|e_\mu^j\|_0^2\tau + \frac{1}{2\varepsilon_{16}}\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_\mu^j\|_0^2\tau. \qquad (113)
\end{aligned}
$$

$$
\begin{aligned}
T_{13} &= \sum_{j=1}^{n}\Big\{-\frac{1}{2}\Big[(d_2^{-1}(b(\xi^j))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j) \\
&+ (d_2^{-1}(b(\xi^{j-1}))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1})\Big] \\
&+ (d_2^{-1}(b(\xi^j))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j) \\
&- ([d_2^{-1}(b(\xi^j)) - d_2^{-1}(b(\xi^{j-1}))](\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1} - \hat{\sigma}_{\mathbf{z}}^{j-1}), \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j)\Big\} \\
&= -\frac{1}{2}\sum_{j=1}^{n}\Big[(d_2^{-1}(b(\xi^j))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j) \\
&+ (d_2^{-1}(b(\xi^j))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}) \\
&- 2(d_2^{-1}(b(\xi^j))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_\xi^j)\Big] \\
&+ \frac{1}{2}\sum_{j=1}^{n}([d_2^{-1}(b(\xi^j)) - d_2^{-1}(b(\xi^{j-1}))]\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}) \\
&- \sum_{j=1}^{n}([d_2^{-1}(b(\xi^j)) - d_2^{-1}(b(\xi^{j-1}))](\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1} - \hat{\sigma}_{\mathbf{z}}^{j-1}), \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j) \\
&= -\frac{1}{2}\sum_{j=1}^{n}\Big(d_2^{-1}(b(\xi^j))(\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j - \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}), \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j - \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}\Big) \\
&+ \frac{1}{2}\sum_{j=1}^{n}([d_2^{-1}(b(\xi^j)) - d_2^{-1}(b(\xi^{j-1}))]\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}) \\
&- \sum_{j=1}^{n}([d_2^{-1}(b(\xi^j)) - d_2^{-1}(b(\xi^{j-1}))](\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1} - \hat{\sigma}_{\mathbf{z}}^{j-1}), \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j) \\
&= k_{131} + k_{132} + k_{133}. \qquad (114)
\end{aligned}
$$

$$
\begin{aligned}
(115) \qquad k_{131} &= -\frac{1}{2}\sum_{j=1}^{n}\|d_2^{(-1/2)}(b(\xi^j))(\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j - \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1})\|_0^2. \\
k_{132} &= \frac{1}{2}\sum_{j=1}^{n}([d_2^{-1}(b(\xi^j)) - d_2^{-1}(b(\xi^{j-1}))]\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}, \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}) \\
&\leq \frac{1}{2}\sum_{j=1}^{n}\int_\Omega |d_2^{-1}(b(\xi^j)) - d_2^{-1}(b(\xi^{j-1}))||\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}|^2\, dx \\
&\leq \frac{L_d^2 R_b}{2}\sum_{j=1}^{n}\int_\Omega |\xi^j - \xi^{j-1}||\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}|^2\, dx \\
&\leq \frac{L_d^2 R_b}{2}\sum_{j=1}^{n}\int_\Omega \left(\int_{t_{j-1}}^{t_j} \xi_t(t)\, dt\right)|\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}|^2\, dx \\
(116) \qquad &\leq \frac{L_d^2 R_b}{2}\|\xi_t\|_{L^\infty(L^\infty)}\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1}\|^2\tau.
\end{aligned}
$$

Using the definition of $\hat{\sigma}_{\mathbf{z}}^{j-1}$, we have

$$
\begin{aligned}
k_{133} &= -\sum_{j=1}^{n}([d_2^{-1}(b(\xi^j)) - d_2^{-1}(b(\xi^{j-1}))](\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^{j-1} - \hat{\sigma}_{\mathbf{z}}^{j-1}), \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j) \\
&= -\sum_{j=1}^{n}([d_2^{-1}(b(\xi^j)) - d_2^{-1}(b(\xi^{j-1}))](\hat{\mathcal{P}}_h^{d_2}\mathcal{I}_{\mathbf{z}}^{j-1} - \mathcal{I}_{\mathbf{z}}^{j-1}), \hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j) \\
&\leq \sum_{j=1}^{n}\int_\Omega |d_2^{-1}(b(\xi^j)) - d_2^{-1}(b(\xi^{j-1}))||(\hat{\mathcal{P}}_h^{d_2}\mathcal{I}_{\mathbf{z}}^{j-1} - \mathcal{I}_{\mathbf{z}}^{j-1})||\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j|\, dx \\
&\leq L_d^2 R_b\|\xi_t\|_{L^\infty(L^\infty)}\sum_{j=1}^{n}\tau\int_\Omega |(\hat{\mathcal{P}}_h^{d_2}\mathcal{I}_{\mathbf{z}}^{j-1} - \mathcal{I}_{\mathbf{z}}^{j-1})||\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j|\, dx \\
&\leq L_d^2 R_b\|\xi_t\|_{L^\infty(L^\infty)}\sum_{j=1}^{n}\tau\|\hat{\mathcal{P}}_h^{d_2}\mathcal{I}_{\mathbf{z}}^{j-1} - \mathcal{I}_{\mathbf{z}}^{j-1}\|_0\|\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j\|_0 \\
&\leq L_d^2 R_b\|\xi_t\|_{L^\infty(L^\infty)}\frac{\varepsilon_{17}}{2}\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h^{d_2}\mathcal{I}_{\mathbf{z}}^{j-1} - \mathcal{I}_{\mathbf{z}}^{j-1}\|_0^2\tau + \frac{1}{2\varepsilon_{17}}\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j\|_0^2\tau \\
(117) \qquad &\leq L_d^2 R_b\|\xi_t\|_{L^\infty(L^\infty)}\frac{\varepsilon_{17}}{2}h^2\sum_{j=1}^{n}|\nabla\cdot\mathcal{I}_{\mathbf{z}}^{j-1}|_1^2\tau + \frac{1}{2\varepsilon_{17}}\sum_{j=1}^{n}\|\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^j\|_1^2\tau.
\end{aligned}
$$

$$
\begin{aligned}
T_{14} &= \sum_{j=1}^{n}(b(\xi_{init}) - b(\xi_{init}^h), \xi^j - \xi_h^j)\tau + \sum_{j=1}^{n}(\mu_{init} - \mu_{init}^h, \mu^j - \mu_h^j)\tau \\
&+ \frac{1}{2}\|\hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^0\|_0^2 + \frac{1}{2}\|d_2^{(-1/2)}(b(\xi^0))\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^0\|_0^2 \\
&\leq \sum_{j=1}^{n}R_b\frac{\varepsilon_{18}}{2}\|\xi_{init} - \xi_{init}^h\|_0^2\tau + \frac{1}{2\varepsilon_{18}}\sum_{j=1}^{n}\|e_\xi^j\|_0^2\tau + \sum_{j=1}^{n}\frac{\varepsilon_{19}}{2}\|\mu_{init} - \mu_{init}^h\|_0^2\tau \\
&+ \frac{1}{2\varepsilon_{19}}\sum_{j=1}^{n}\|e_\mu^j\|_0^2\tau + \frac{1}{2}\|\hat{\mathcal{P}}_h\hat{\sigma}_{\mathbf{w}}^0\|_0^2 + \frac{\nu_2}{2}\|\hat{\mathcal{P}}_h^{d_2}\hat{\sigma}_{\mathbf{z}}^0\|_0^2
\end{aligned}
$$

$$\leq \quad Ch^2 + \frac{1}{2\varepsilon_{18}} \sum_{j=1}^{n} \|e_\xi^j\|_0^2 \tau + \frac{1}{2\varepsilon_{19}} \sum_{j=1}^{n} \|e_\mu^j\|_0^2 \tau$$

$$+ \quad \frac{1}{2} \|\hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^0\|_0^2 + \frac{\nu_2}{2} \|\hat{\mathcal{P}}_h^{d_2} \hat{\sigma}_{\mathbf{z}}^0\|_0^2. \tag{118}$$

Combining (87)–(118), using the property of $b(\cdot)$ in (30),(31), and Assumption 3(B), and taking appropriate choices of ε_i's, we have

$$\sum_{j=1}^{n} \|e_\xi^j\|_0^2 \tau \quad + \quad \sum_{j=1}^{n} \|e_\mu^j\|_0^2 \tau + \|\hat{\mathcal{P}}_h \sigma_{\mathbf{w}}^n\|_0^2 + \|\hat{\mathcal{P}}_h^{d_2} \sigma_{\mathbf{z}}^n\|_0^2$$

$$\leq \quad C \left[h^2 + \tau^2 + \sum_{j=1}^{n} \left(\sum_{k=1}^{j} \|e_\xi^k\|_0^2 \tau \right) + \sum_{j=1}^{n} \left(\sum_{k=1}^{j} \|e_\mu^k\|_0^2 \tau \right) \right.$$

$$\left. + \quad \sum_{j=1}^{n} \|\hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^j\|_0^2 \tau + \sum_{j=1}^{n} \|\hat{\mathcal{P}}_h^{d_2} \hat{\sigma}_{\mathbf{z}}^j\|_0^2 \tau + \|\hat{\mathcal{P}}_h \hat{\sigma}_{\mathbf{w}}^0\|_0^2 + \|\hat{\mathcal{P}}_h^{d_2} \hat{\sigma}_{\mathbf{z}}^0\|_0^2 \right]. \tag{119}$$

Using Gronwall's Lemma, the proof is complete. □

7. Numerical Examples

In this section, we conduct two experiments, one in 1D, where we compute the error, and the other in 2D, where we study the behavior of biofilm U_1 and nutrient U_2 in a complex porous medium with realistic data.

We consider system (68). We recall, from Sec. 4.1.2, equation (15b), that

$$G_1^n \quad = \quad -\nabla_h \cdot (\bar{\mathbf{q}} U_1^{n-1}) + \frac{1}{\tau} U_1^{n-1}, \tag{120a}$$

$$G_2^n \quad = \quad -\nabla_h \cdot (\bar{\mathbf{q}} U_2^{n-1}) + \frac{1}{\tau} U_2^{n-1}, \tag{120b}$$

where ∇_h denotes explicit upwind flux. At each time step n, we implement the operator splitting method [24]. That is, we first find the solution $(\tilde{U}_1^n, \tilde{U}_2^n)$ explicitly for the advection part

$$\tilde{U}_1^n = \tau G_1^n, \tag{121a}$$

$$\tilde{U}_2^n = \tau G_2^n. \tag{121b}$$

Then we find the solution $(U_1^n, U_2^n, \Lambda^n)$ for the diffusion–reaction system

$$(\tau \mathcal{A}_1(\tilde{U}_1^n) + \mathcal{I} - \tau \mathcal{M}_1(U_2^n)) U_1^n + \tau \Lambda^n = \tilde{U}_1^n, \tag{122a}$$

$$\tau \mathcal{M}_2(U_2^n) U_1^n + (\tau \mathcal{A}_2(\tilde{U}_1^n) + \mathcal{I}) U_2^n = \tilde{U}_2^n, \tag{122b}$$

which is solved along side with the equation $U_1^n - P_{[0,u^*]}(U_1^n - \Lambda^n) = 0$, where $P_{[0,u^*]}(U) = \max\{0, \min(U, u^*)\}$. Semi-smooth Newton method [39] is used to solve the last nonlinear system.

Example 7.1. *In this example, we consider a tube in 1D of length* 1*, with boundaries at* $x = 0$ *and* $x = 1$*. We assume that the medium is rich of nutrient with an initial nutrient* $u_{init}^2 = 1$*. The nutrient is also constantly injected at the wall* $x = 1$*. The initial biomass is equal to* $u_{init}^1 = 0.8$*, which is not mature yet. We assume that it becomes mature at* $u_* = 0.9$*, and its maximum density* $u^* = 1$*. The fluid in the medium flows at rate equal to* 5×10^{-5}*. Table 1 provides the data we use in this experiment.*

TABLE 1. Parameters used in Example 7.1.

d_0	α	$d_{N,w}$	β	γ	κ	$\bar{\mathbf{q}}$	u^1_{init}	$u^2_{init} = u^2_D$
10^{-4}	2	6	1	1.18×10^{-3}	0.44	5×10^{-5}	$0.8u^*$	1

Figure 2 shows that the mature biofilm forms around the time $t = 0.25$. Then it continues growing until it reaches its maximum density $u^* = 1$ at time $t = 2.5$ when Λ changes from being 0 to being a positive number to prevent the biofilm from exceeding its maximum density. After that, biofilm stops growing upward, but it continues growing forward exploiting the availability of nutrient that is transported by the flow in the biofilm domain.

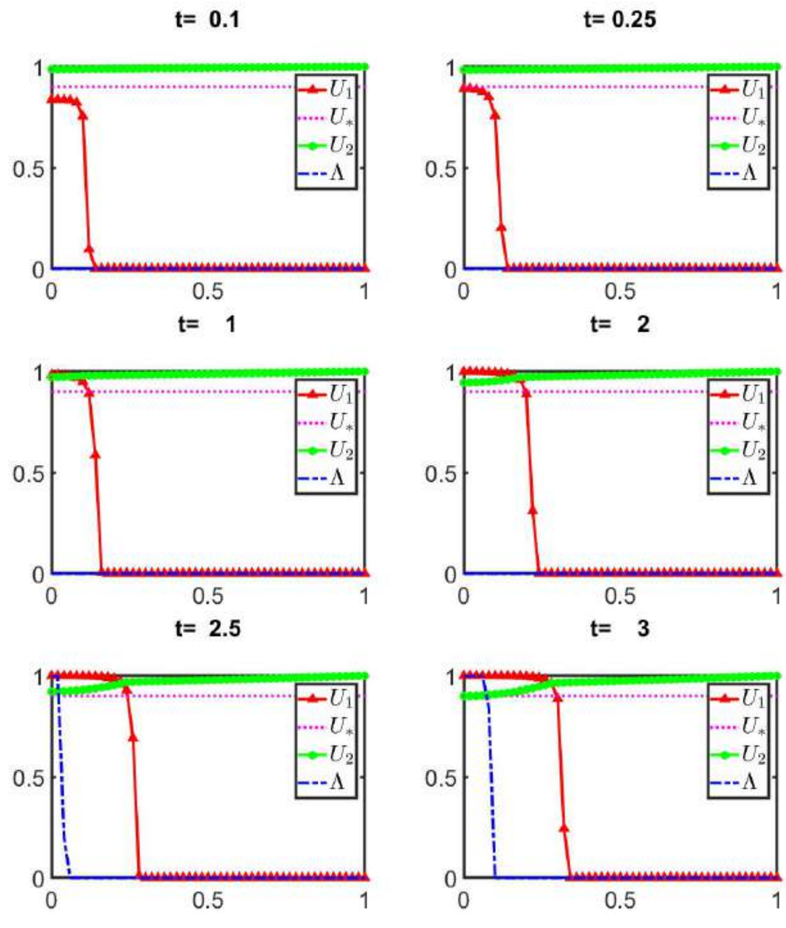

FIGURE 2. Evolution of the solution (U_1, U_2) of Example 7.1, $h = 0.2, \tau = 10^{-3}$. The solution U_1 satisfies the constraint $U_1 \leq u^* = 1$, and the Lagrange multiplier Λ becomes active in the region where $U_1 = u^*$. The diffusivity of nutrient is lower whenever $U_1 \geq u_* = 0.9$ which is visible in the profile of U_2.

Example 7.2. *The aim of this example is to test the errors* $\sqrt{\sum_n \|e_1^n\|_0^2 \tau}$ *and* $\sqrt{\sum_n \|e_2^n\|_0^2 \tau}$, *where* $e_1^n = u_1^n - U_1^n$, $e_2^n = u_2^n - U_2^n$. *We consider the same data in*

Example 7.1 in 1D which is given in Table 1 except some slight changes, so U_1 hits its maximum $u^* = 1$ so quickly faster than it takes in Example 7.1 which allows us to compute the error when the constraint takes place. We start at initial biomass equal to $u^1_{init} = 1$, the uptake rate $\beta = 2$, and flow velocity equal to 0.01.

We compute the numerical solution (U_1, U_2) at different values of h and τ shown in Table 2. Since the analytical solution of the system (1) *is difficult to obtain, we compare the numerical solutions with a solution (U_1^{fine}, U_2^{fine}) computed at a fine spatial step size $h_{fine} = 0.001$ and a fine temporal step size $\tau_{fine} = 10^{-4}$.*

Table 2 shows the order of convergence of $L^2(Err_1) = \sqrt{\sum_n \|e_1^n\|_0^2 \tau}$ and $L^2(Err_2) = \sqrt{\sum_n \|e_2^n\|_0^2 \tau}$ for Example 7.2. We would like to note that since it is hard to compute these errors at each time step, we rather compute the error at some steps n, with $t_n \in \{0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.85\}$. As it is shown in the table, the error converges of the first order as we expected. We also compute the error in the max norm in time for the same example in Table 3, i.e., we compute $\max(Err_1) = \max_n \|e_1^n\|_0$ and $\max(Err_2) = \max_n \|e_2^n\|_0$ at the same steps as before. As we notice, the order of convergence of $Ł^2(Err_1)$, which we theoretically analysed, exceeds that of $\max_n(Err_1)$. Figure 3 illustrates the order of convergence of the errors in these two norms. See also the solutions at $t_{0.55}, t_{0.65}, t_{0.75}$ in Figure 4.

TABLE 2. Order of convergence for Example 7.2; $L^2(Err_1) = \sqrt{\sum_n \|e_1^n\|_0^2 \tau}$, $L^2(Err_2) = \sqrt{\sum_n \|e_2^n\|_0^2 \tau}$.

h	τ	$L^2(Err_1)$	$L^2(Err_2)$	$L^2(Err_1)$ order	$L^2(Err_2)$ order
0.02	0.02	0.042379	0.0010275	-	-
0.01	0.01	0.0151	0.00030577	1.4888	1.7487
0.005	0.005	0.0074604	8.9094e-05	1.0172	1.779
0.0025	0.0025	0.0020002	2.1826e-05	1.8991	2.0293

TABLE 3. Order of convergence for Example 7.2; $\max(Err_1) = \max_n \|e_1^n\|_0$, $\max(Err_2) = \max_n \|e_2^n\|_0$.

h	τ	$\max(Err_1)$	$\max(Err_2)$	$\max(Err_1)$ order	$\max(Err_2)$ order
0.02	0.02	0.151	0.0031494	-	-
0.01	0.01	0.071414	0.0013377	1.0802	1.2353
0.005	0.005	0.054772	0.00055935	0.38277	1.258
0.0025	0.0025	0.017866	0.00018686	1.6163	1.5818

Next, we study the effect of the flow rate $\bar{\mathbf{q}}$ on the growth. This flow rate is assumed known in this paper, but it is important to study whether $\bar{\mathbf{q}}$ is trivial or nontrivial in Ω_b, in other words, respectively, whether Ω_b is considered impermeable or permeable to the flow. In addition, we want to see whether the character of the flow in Ω_b plays a role (such as Stokes-like or Darcy-like). In our computational experiments $\bar{\mathbf{q}}$ is determined by a coupled heterogeneous Brinkman flow model $-\mu\Delta\bar{\mathbf{q}} + k_b\chi_{\Omega_b}\bar{\mathbf{q}} + \nabla p = 0; \nabla \cdot \bar{\mathbf{q}} = 0$, in which this bio-gel permeability in Ω_b is denoted by k_b. To study the aforementioned scenarios we set $k_b = 0$ for the impermeable case, k_b moderate for Darcy-like flow in Ω_b, and use a large $k_b \uparrow \infty$ for the case when the flow in Ω_b is Stokes-like. We use realistic data form [35] listed in Table 4.

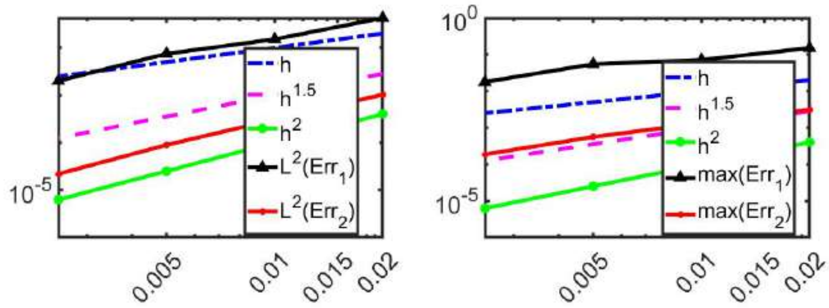

FIGURE 3. Order of convergence of $L^2(Err_1) = \sqrt{\sum_n \|e_1^n\|_0^2 \tau}$, $L^2(Err_2) = \sqrt{\sum_n \|e_2^n\|_0^2 \tau}$ in the left, and of $\max(Err_1) = \max_n \|e_1^n\|_0$, $\max(Err_2) = \max_n \|e_2^n\|_0$ in the right for Example 7.2.

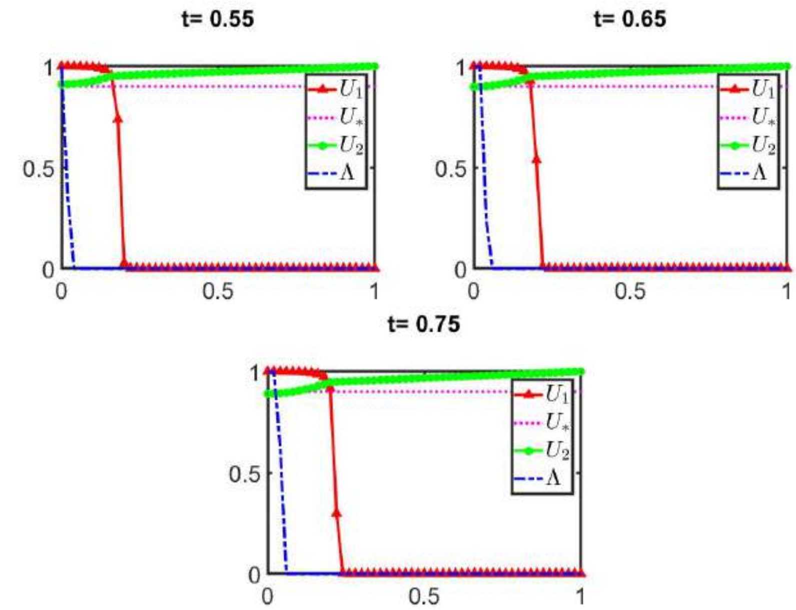

FIGURE 4. Numerical solution (U_1, U_2) of Example 7.2 at $t = 0.55$ in the left, $t = 0.65$ in the middle, and $t = 0.75$ in the right. The front keeps moving in time due to the high availability of nutrient. The meaning of U_*, Λ is as explained in Figure (2).

Example 7.3. *We consider a single-pore medium* $\Omega = (0,1)^2[\mathrm{mm}]^2$ *shown in Figure 5. Initially, ten percent of the non-rock region is filled with biomass* ($u^1_{init} = 0.6$) *with no nutrient. Then the nutrient is injected through the left boundary of* Ω*. We set Neumann no-nutrient flux conditions on the rest of the boundaries. We also assume that there is no biomass flux on all of the boundaries. The ambient fluid flows from left to right at initial rate of* $\bar{\mathbf{q}}_{init} = 3.6 \times 10^{-3}[\mathrm{mm/h}]$*. See the parameters used in this example in Table 4, which is realistic data obtained form* [35] *with slight changes.*

TABLE 4. Parameters used in Example 7.3.

d_0	α	$d_{N,w}$	β	γ	κ	$\bar{\mathbf{q}}_{init}$	u^1_{init}	u^2_{init}	u^2_D
10^{-4}	2	6	1	1.18×10^{-3}	0.5	3.6×10^{-3}	$0.6u^*$	0	1

FIGURE 5. A sample of a porous medium $\Omega = \Omega_n \cup \Omega_r$, where Ω_n in white, and Ω_r in black.

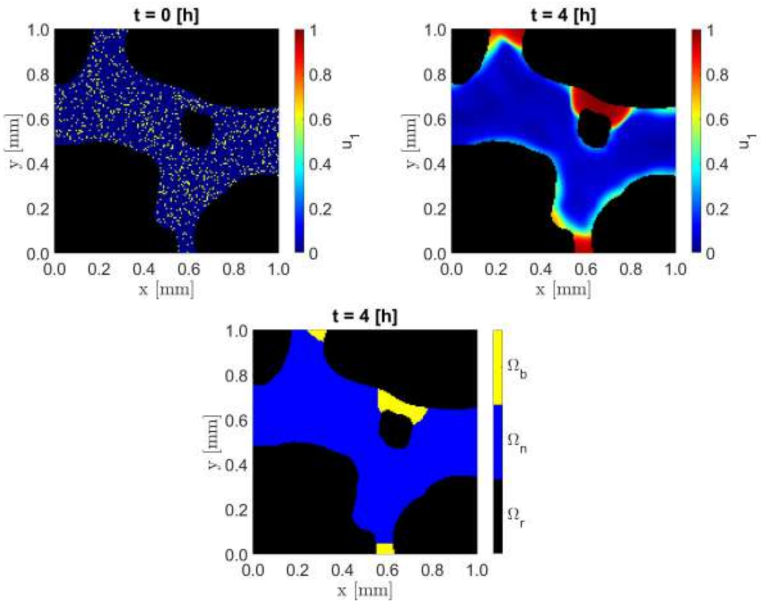

FIGURE 6. Accumulation of biomass near the rock surface . Left: initial biomass. Middle and right: biomass concentration and biofilm domain, respectively, after 4 hours of providing nutrient. Ω_b: the biofilm domain, Ω_n : the fluid domain, and Ω_r : the rock domain.

Once the flow starts, biomass accumulates in the region near the rock surface as it is shown in Figure 6. At high flow rate, the immature biomass is driven away, therefore, most growth occurs at low flow rate as it is shown in Figure 7. As biofilm grows, the flow region decreases, and hence its velocity increases.

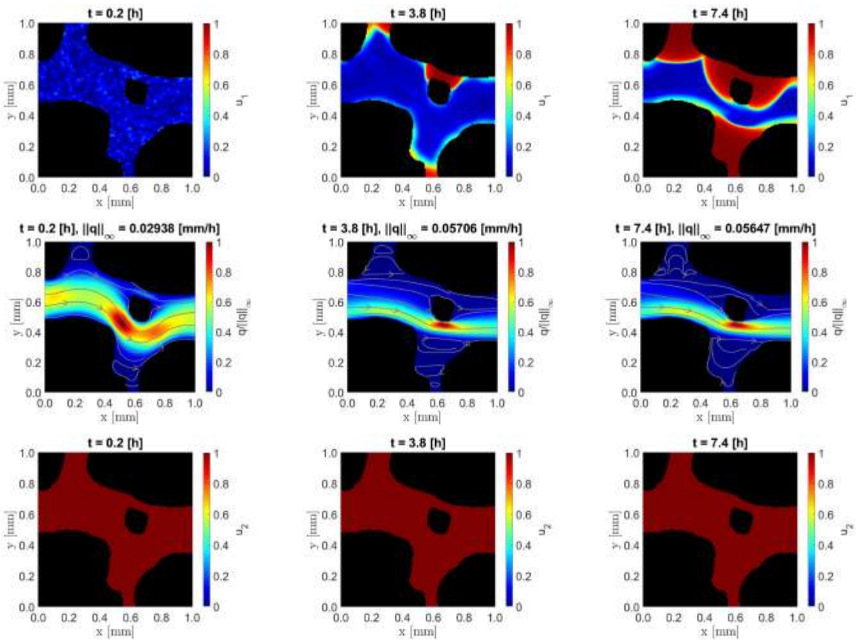

FIGURE 7. The effect of flow on biofilm growth when Ω_b is assumed permeable to the flow and the flux $\bar{\mathbf{q}}$ is obtained with $k_b = 10^{-5}$. Top: the biofilm evolution. Middle: the flow velocity. Bottom: the nutrient.

Next, we consider various scenarios of the character of the flow $\bar{\mathbf{q}}$ which gives us availability of nutrient inside Ω_b. To this aim, we consider different values of bio-gel permeability $k_b \uparrow \infty$ when Ω_b is permeable, $k_b = 10^{-5}$ when Ω_b is partially permeable, and $k_b = 0$ when Ω_b is impermeable. As it is illustrated in Figure 8, as k_b increases, the biofilm grows faster and fills up the pore more quickly. We also see that the flow velocity is affected by the permeability of bio-gel. Table 5 shows the time needed for the biofilm of different k_b to clog and fill up the pore.

TABLE 5. Time taken for the biofilm to clog and fill up the pore with different permeability k_b of Example 7.3.

	Permeable	Partially Permeable	Impermeable
Biofilm permeability k_b	∞	$10^{-5}[mm^2]$	0
Time when clogged	$9.1[h]$	$9.18[h]$	$9.2[h]$
Time when filled up completely	$12.75[h]$	$12.83[h]$	$12.91[h]$

8. Summary and future work

In this paper we carried out rigorous analysis of a mixed finite element approximation (MFEM) of lowest order for a nonlinear constrained parabolic system modeling biofilm growth, with advection. We also illustrated the results with numerical experiments. We believe our results are first for such a system, and that they also extend known theory for MFEM for scalar parabolic variational inequalities to the case with nonlinear diffusivity and advection.

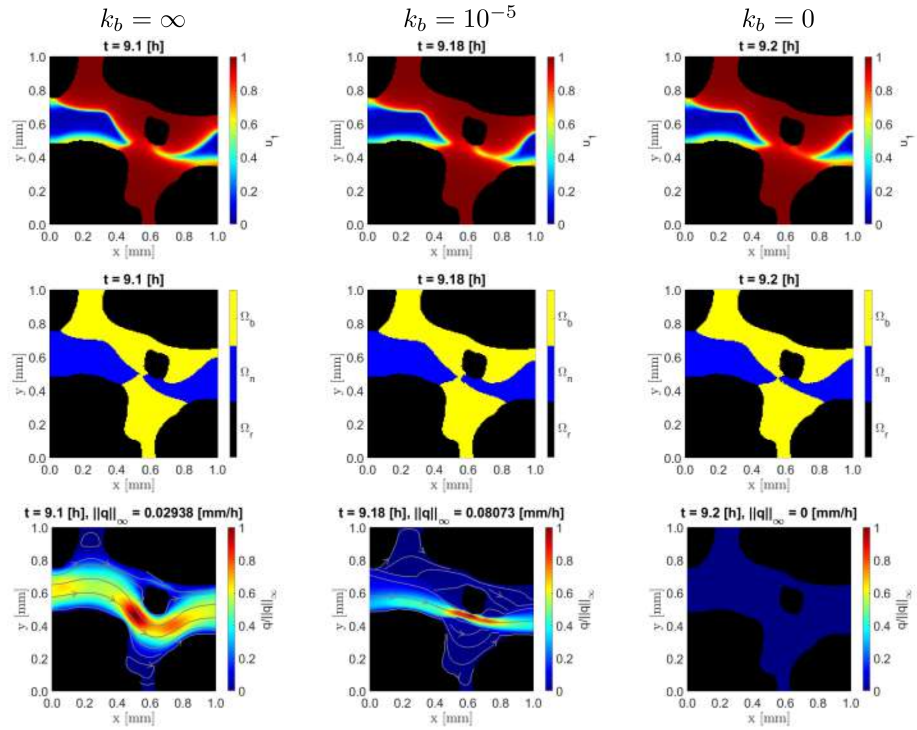

FIGURE 8. The effect of assumption on the accessibility of nutrient by advection associated with the permeability of biofilm domain k_b on the biofilm growth. Top: the biomass concentration. Middle: the biofilm domain. Bottom: the flow velocity. Recall the case $k_b = \infty$ (left column) allows Stokest-type flow on the domain, $k_b = 0$ (right column) makes $\bar{\mathbf{q}}|_{\Omega_b} = 0$, and the intermediate k_b (middle column) allows Darcy-type flow in Ω_b.

In future work we aim to relax the somewhat stringent assumptions on the regularity of the solutions required for convergence. We also plan to study the error in the coupled flow problem, and consider other modeling and theoretical extensions.

Acknowledgements

The authors would like to thank the editors and the anonymous reviewers for their comments which helped to improve the exposition. Some portion of the theoretical work in this paper was completed when the first author was a PhD student at Oregon State University supported by the scholarship from Imam Abdulrahman Bin Faisal University; see [2] Chapter 5. Some computational models were completed when the third author was a PhD student at Oregon State University.
This research was also partially supported by the grant NSF DMS-1912938 "Modeling with Constraints and Phase Transitions in Porous Media", the NSF IRD plan 2019-21, and NSF DMS-1522734 "Phase transitions in porous media across multiple scales", PI: Malgorzata Peszynska.

References

[1] A. Alhammali and M. Peszynska. Numerical analysis of a parabolic variational inequality system modeling biofilm growth at the porescale. Numerical Methods for PDEs, 36:941–971, 2020.

[2] Azhar Alhammali. Numerical Analysis of a System of Parabolic Variational Inequalities with Application to Biofilm Growth. PhD thesis, Oregon State University, 2019.

[3] Stergiani I Arampatzi, Georgios Giannoglou, and Eudoxia Diza. Biofilm formation: A complicated microbiological process. Aristotle University Medical Journal, 38(2):21–29, 2011.

[4] Todd Arbogast, Mary F. Wheeler, and Nai-Ying Zhang. A nonlinear mixed finite element method for a degenerate parabolic equation arising in flow in porous media. SIAM J. Numer. Anal., 33(4):1669–1687, 1996.

[5] Kendall Atkinson and Weimin Han. Theoretical numerical analysis, volume 39 of Texts in Applied Mathematics. Springer, New York, second edition, 2005. A functional analysis framework.

[6] Claudio Baiocchi. Discretization of evolution variational inequalities. Partial Differential Equations and the Calculus of Variations, pages 59–92, 1989.

[7] Lisa Bigler, Malgorzata Peszynska, and Naren Vohra. Heterogeneous Stefan problem and permafrost models with P0-P0 finite elements and fully implicit monolithic solver. Electron. Res. Arch., 30(4):1477–1531, 2022.

[8] Daniele Boffi, Franco Brezzi, and Michel Fortin. Mixed Finite Element Methods and Applications. Springer, Berlin Heidlberg, 2013.

[9] Daniele Boffi, Michel Fortin, and F. Brezzi. Mixed finite element methods and applications. Springer series in computational mathematics; 44. Springer, 2013.

[10] Dietrich Braess. Finite elements. Cambridge Univ. Press, Cambridge, 3 edition, 2007.

[11] Franco Brezzi and Michel Fortin. Mixed and hybrid finite element methods, volume 15 of Springer Series in Computational Mathematics. Springer-Verlag, New York, 1991.

[12] Franco Brezzi, William W. Hager, and P. A. Raviart. Error estimates for the finite element solution of variational inequalities. part ii. mixed methods. Numerische Mathematik, 31(1):1–16, 1978.

[13] Vidar Thomee Claes Johnson. Error estimates for some mixed finite element methods for parabolic type problems. RAIRO. Analyse numérique, 15(1):41–78, 1981.

[14] Knut Drescher, Yi Shen, Bonnie L. Bassler, and Howard A. Stone. Biofilm streamers cause catastrophic disruption of flow with consequences for environmental and medical systems. Proceedings of the National Academy of Sciences, 110(11):4345–4350, 2013.

[15] Herman J Eberl, David F Parker, and Mark CM Vanloosdrecht. A new deterministic spatio-temporal continuum model for biofilm development. Computational and Mathematical Methods in Medicine, 3(3):161–175, 2001.

[16] Hermann J. Eberl and Laurent Demaret. A finite difference scheme for a degenerated diffusion equation arising in microbial ecology. Electronic journal of differential equations, Conference(15):77–95, 2007.

[17] Anozie Ebigbo, Rainer Helmig, Alfred B. Cunningham, Holger Class, and Robin Gerlach. Modelling biofilm growth in the presence of carbon dioxide and water flow in the subsurface. ADVANCES IN WATER RESOURCES, 33(7):762–781, 2010.

[18] Alexandre Ern and Jean-Luc Guermond. Theory and practice of finite elements. Springer, New York, 2004.

[19] Mallory R Frederick, Christina Kuttler, Burkhard A Hense, and Hermann J Eberl. A mathematical model of quorum sensing regulated eps production in biofilm communities. Theoretical Biology and Medical Modelling, 8(1):1–29, 2011.

[20] Weimin Han and Kendall E. Atkinson. Theoretical Numerical Analysis: A Functional Analysis Framework, volume 39. Springer New York, New York, NY, 3 edition, 2009.

[21] Kazufumi Ito and Karl Kunisch. Parabolic variational inequalities: The lagrange multiplier approach. Journal de mathématiques pures et appliquées, 85(3):415–449, 2006.

[22] Claes Johnson. A convergence estimate for an approximation of a parabolic variational inequality. SIAM Journal on Numerical Analysis, 13(4):599606, 1976.

[23] M.-Y. Kim, F.A. Milner, and E.-J. Park. Some observations on mixed methods for fully nonlinear parabolic problems in divergence form. Applied Mathematics Letters, 9(1):75–81, 1996.

[24] Randall J. LeVeque. Finite difference methods for ordinary and partial differential equations. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2007. Steady-state and time-dependent problems.

[25] Iram Liaqat. An overview of biofilm formation, properties and control. International journal of medical and biological frontiers, 17(4/5):415, 2011.
[26] F. A. MacLeod, H. M. Lappin-Scott, and J. W. Costerton. Plugging of a model rock system by using starved bacteria. American society for microbiology, 54(6):1365–1372, 1988.
[27] Marek Niezgódka Maria Gokieli, Nobuyuki Kenmochi. Variational inequalities of navierstokes type with time dependent constraints. Journal of Mathematical Analysis and Applications, 449(2):1229–1247, 2017.
[28] SAJID Memon. Finite element method for american option pricing: a penalty approach. International Journal of Numerical Analysis and Modelling, Series B: Computing and Information, 3(3):345–370, 2012.
[29] Andrew C. Mitchell, Adrienne J. Phillips, Randy Hiebert, Robin Gerlach, Lee H. Spangler, and Alfred B. Cunningham. Biofilm enhanced geologic sequestration of supercritical co_2. International Journal of Greenhouse Gas Control, 3(1):90–99, 2009.
[30] Manisha Mukherjee, Yidan Hu, Chuan H. Tan, Scott A. Rice, and Bin Cao. Engineering a light-responsive, quorum quenching biofilm to mitigate biofouling on water purification membranes. Science advances, 4(12):eaau1459–eaau1459, 2018.
[31] Pradeepa Nair and Amiya K Pani. Finite element methods for parabolic variational inequalities with a volterra term. Numerical Functional Analysis and Optimization, 24(1-2):107–127, 2003.
[32] Malgorzata Peszynska, Anna Trykozko, Gabriel Iltis, Steffen Schlueter, and Dorthe Wildenschild. Biofilm growth in porous media: Experiments, computational modeling at the porescale, and upscaling. Advances in Water Resources, 95:288–301, 2016.
[33] Florin Radu, Iuliu Sorin Pop, and Peter Knabner. Order of convergence estimates for an euler implicit, mixed finite element discretization of richards' equation. SIAM Journal on Numerical Analysis, 42(4):1452–1478, 2004.
[34] T. F. Russell and M. F. Wheeler. Finite element and finite difference methods for continuous flows in porous media. In R. E. Ewing, editor, The Mathematics of Reservoir Simulation, pages 35–106. SIAM, Philadelphia, 1983.
[35] Choah Shin, Azhar Alhammali, Lisa Bigler, Naren Vohra, and Malgorzata Peszynska. Coupled flow and biomass-nutrient growth at pore-scale with permeable biofilm, adaptive singularity and multiple species. Mathematical Biosciences and Engineering, 18(3), 2021.
[36] R. E Showalter. Nonlinear degenerate evolution equations in mixed formulation. SIAM journal on mathematical analysis, 42(5):2114–2131, 2010.
[37] Ralph E Showalter. Monotone operators in Banach space and nonlinear partial differential equations. American Math. Soc., Providence, RI, 1997.
[38] Martin Thullner, Josef Zeyer, and Wolfgang Kinzelbach. Influence of microbial growth on hydraulic properties of pore networks. Transport in Porous Media, 49(1):99–122, 2002.
[39] Michael Ulbrich. Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces, volume 11 of MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011.
[40] Alan Weiser and Mary Fanett Wheeler. On convergence of block-centered finite differences for elliptic problems. SIAM J. Numer. Anal., 25(2):351–375, 1988.
[41] Carol S Woodward and Clint N Dawson. Analysis of expanded mixed finite element methods for a nonlinear parabolic equation modeling flow into variably saturated porous media. SIAM Journal on Numerical Analysis, 37(3):701–724, 2000.
[42] Wen Yin, Yiting Wang, Lu Liu, and Jin He. Biofilms: The microbial "protective clothing" in extreme environments. International journal of molecular sciences, 20(14):3423, 2019.
[43] Eberhard Zeidler. Applied functional analysis, volume 108 of Applied Mathematical Sciences. Springer-Verlag, New York, 1995. Applications to mathematical physics.

[1] Department of Mathematics, College of Science, Imam Abdulrahman Bin Faisal University, Ar Rayyan, Dammam 32256, Saudi Arabia and [2] Department of Mathematics, Oregon State University, Corvallis, OR 97331, USA

E-mail: aalhammali@iau.edu.sa and mpesz@math.oregonstate.edu and shinc@oregonstate.edu