Linearization of the finite element method for gradient flows by Newton’s method

GEORGIOS AKRIVIS †
Department of Computer Science and Engineering, University of Ioannina, 451 10 Ioannina, Greece, and Institute of Applied and Computational Mathematics, FORTH, 70013 Heraklion, Crete, Greece.

BUYANG LI ‡
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong.

[Received on 16 June 2019; revised on 13 March 2020; accepted on 13 March 2020]

The implicit Euler scheme for nonlinear partial differential equations of gradient flows is linearized by Newton’s method, discretized in space by the finite element method. With two Newton iterations at each time level, almost optimal order convergence of the numerical solutions is established in both the $L^q(\Omega)$ and $W^{1,q}(\Omega)$ norms. The proof is based on techniques utilizing the resolvent estimate of elliptic operators on $L^q(\Omega)$ and the maximal $L^p$-regularity of fully discrete finite element solutions on $W^{-1,q}(\Omega)$.

Keywords: gradient flow; nonlinear equation; finite element method; linearization; Newton’s iteration; resolvent estimate; maximal $L^p$-regularity.

1. Introduction

We consider the following initial and boundary value problem for a time-dependent nonlinear partial differential equation

\[
\begin{aligned}
\partial_t u &= \nabla \cdot f(\nabla u) \quad \text{in} \quad \Omega \times (0,T], \\
\mathbf{u} &= 0 \quad \text{on} \quad \partial \Omega \times (0,T], \\
u(\cdot,0) &= u_0 \quad \text{in} \quad \Omega,
\end{aligned}
\]  

(1.1)
in a given convex polygon $\Omega \subset \mathbb{R}^2$ or a polyhedron $\Omega \subset \mathbb{R}^3$ with interior edge angles less than $\frac{3}{4} \pi$, up to a given time $T > 0$, with a smooth function $f : \mathbb{R}^d \to \mathbb{R}^d$.

Problems of the form (1.1) occur in many applications, including minimal surface flows (cf. [24, 28]), with $f(p) = p / \sqrt{1 + \|p\|^2}$, regularized models of total variation flows (cf. [8, 9, 21]), with $f(p) = p / \sqrt{\lambda^2 + \|p\|^2}$, and the $L^2(\Omega)$ gradient flow:

\[
(\partial_t u, \mathbf{v})_{L^2(\Omega)} = -E'(u) \mathbf{v} \quad \text{for all} \quad \mathbf{v} \text{ in a dense and smooth subspace of } H^1_0(\Omega),
\]

where $E(u) = \int_{\Omega} F(\nabla u) \, dx$ is an energy functional with a convex function $F : \mathbb{R}^d \to \mathbb{R}$, and $f(p) = \nabla_p F(p)$; see [7, Section 9.6.3].

In these applications, the flux function $f$ satisfies the following local ellipticity condition:

\[
\nabla_p f(p) \text{ is symmetric and positive definite for every } p \in \mathbb{R}^d.
\]  

(1.2)

However, some eigenvalues of $\nabla_p f(p)$ may tend to 0 or $+\infty$ as $|p| \to \infty$. For example, for the flux function $f(p) = p / \sqrt{1 + \|p\|^2}$ appearing in the minimal surface flow problem we have

\[
\nabla_p f(p) = \frac{1}{\sqrt{1 + \|p\|^2}} \mathbb{I}_d - \frac{|p|^2}{(1 + \|p\|^2)^{3/2}} \frac{p \otimes p}{|p|^3},
\]  

(1.3)

†Email: akrivis@cse.uoi.gr
‡Email: bygli@polyu.edu.hk

© The author 2020. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.
with $I_d$ the $d \times d$ identity matrix; now, $\nabla_p f(p)$ is symmetric and positive definite for any $p \in \mathbb{R}^d$,
\[
\nabla_p f(p) \xi \cdot \xi = \frac{1}{\sqrt{1 + |p|^2}} |\xi|^2 - \frac{|p|^2}{(1 + |p|^2)^{\frac{3}{2}}} \left( \frac{p}{|p|} \cdot \xi \right)^2 \geq \frac{1}{(1 + |p|^2)^{\frac{3}{2}}} |\xi|^2 \quad \forall \xi \in \mathbb{R}^d,
\]
but the eigenvalues of $\nabla_p f(p)$ are not uniformly bounded from below by a positive constant as $|p| \to \infty$.
The nonuniform ellipticity leads to some mathematical difficulties in the numerical analysis of this problem. In particular, uniform $W^{1,\infty}$-boundedness of the numerical solutions needs to be proved in the error estimation in order to rule out the possibility of degeneracy.

Optimal order convergence in the discrete $L^\infty(0,T;L^2(\Omega))$ norm for the implicit Euler scheme, combined with finite element spatial discretization, for the regularized total variation flow was first proved in [8, 9] by the energy technique. The $W^{1,\infty}$-boundedness of the numerical solutions was proved by using an inverse inequality of the finite element space, which requires a stepsize restriction $\tau = o(h^2)$.
Under milder conditions, convergence of the numerical solutions was proved by using a compactness argument. Optimal order convergence in $L^\infty(0,T;L^2(\Omega))$ of finite element methods with a linearized semi-implicit Euler scheme was shown in [21]. In order to remove the stepsize restriction $\tau = o(h^2)$, the energy approach used in [21] is limited to two-dimensional problems and finite element methods of polynomial degree $r \geq 2$. In a general $d$-dimensional domain, error analysis of semidiscretization in time was presented in [16] by utilizing the discrete maximal $L^p$-regularity approach. However, since the analysis in [16] is based on estimates in the $L^p(0,T;W^{2,q}(\Omega))$ norm, it cannot be extended to the case of finite element spatial discretization (as the finite element solutions are not in $W^{2,q}(\Omega)$).

This article is concerned with full discretization of (1.1) under the local ellipticity condition (1.2), by using Newton’s iterative method to linearize the nonlinear system obtained by the implicit Euler scheme with the piecewise linear finite element spatial discretization. We assume that the initial and boundary value problem (1.1) admits a sufficiently regular solution, and prove almost optimal order convergence of the numerical solutions with a fixed number of Newton iterations, say $\ell$ iterations at each time level, in two- and three-dimensional domains without the stepsize restriction $\tau = o(h^2)$.

Our idea is to split the error of the Newton iterative finite element solutions into three parts:
\[
u^\ell_{n+1}(x) - u(x,t_n) = [u^n(x) - u(x,t_n)] + [u^n_\ell(x) - u^\ell(x)] + [u^\ell_{n+1}(x) - u^\ell(x)],
\tag{1.4}
\]
where $u^n$ denotes the time-discrete solution, and $u^n_\ell$ and $u^\ell_{n+1}$ denote the Newton iterative solutions of the time-discrete and fully discrete nonlinear systems, respectively. An estimate of the first part on the right-hand side of (1.4) in the $L^p(0,T;W^{2,q}(\Omega)) \cap W^{1,p}(0,T;L^q(\Omega))$ norm was obtained in [16] by using maximal $L^p$-regularity of time discretizations of parabolic equations. This estimate provides a foundation for further analysis of Newton’s iterative method (second part) and the spatial discretization (third part).

We shall prove that the second and third parts in (1.4) are of higher order in time if the number of Newton iterations $\ell$ is at least 2. This further helps to prove the $W^{1,\infty}$-boundedness of the numerical solutions. The technical tools we use are the $L^p(0,T;W^{1,q}(\Omega))$ estimate of discretized parabolic equations, i.e., estimates (2.15)–(2.16), and the best approximation property of finite element approximations to parabolic equations in the discrete $L^p(0,T;L^q(\Omega))$ norm, i.e., estimate (2.18). Both tools are consequences of the discrete maximal $L^p$-regularity theory [4, 10, 13–15, 17, 19, 20, 22], which is a mathematical tool for numerical analysis of nonlinear parabolic equations; see [1, 2, 16, 18, 26]. These articles are mainly concerned with either semidiscretization in time or semilinear parabolic equations; the techniques cannot be applied to the strongly nonlinear problem of gradient flow with fully discrete numerical methods, especially in the case involving linearization by Newton’s iterations.

In Section 2, we introduce the linearized Newton iterative finite element method for (1.1). Then, we present the main theoretical result on the convergence of the numerical solutions. Using resolvent estimates of elliptic operators on $L^2(\Omega)$, in Section 3 we establish error estimates for the time-discrete Newton iterative solutions in $W^{2,q}(\Omega)$ and $W^{1,\infty}(\Omega)$. Then, we view the fully discrete Newton iterative solutions as spatial finite element approximations of the time-discrete Newton iterative solutions, and estimate the difference between the two solutions in Section 4. Under this point of view, we prove...
almost optimal order convergence of the fully discrete Newton iterative finite element solutions in the norms of \( W^{1,q}(\Omega) \) and \( L^q(\Omega) \).

2. Assumptions and main result

In this paper, we work with the following assumptions:

(a1) The domain \( \Omega \) is sufficiently regular such that the \( W^{2,q} \) elliptic regularity holds for some \( d < q \leq 6 \). In other words, for \( g \in L^q(\Omega) \) and \( a_{ij} \in W^{1,q}(\Omega) \) such that \( a_{ij} = a_{ji} \) and

\[
\lambda^{-1} |\xi|^2 \leq \sum_{i,j=1}^d a_{ij}(x) \xi_i \xi_j \leq \lambda |\xi|^2 \quad \forall \xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d \quad \forall x \in \Omega, \tag{2.1}
\]

the solution \( u \in H^1_0(\Omega) \) of the boundary value problem for the elliptic equation

\[
\begin{aligned}
\sum_{i,j=1}^d \frac{\partial}{\partial x_i} \left( a_{ij}(x) \frac{\partial u}{\partial x_j} \right) &= g \quad \text{in } \Omega, \\
\quad u &= 0 \quad \text{on } \partial \Omega,
\end{aligned} \tag{2.2}
\]

satisfies

\[
||u||_{W^{2,p}(\Omega)} \leq c ||g||_{L^p(\Omega)} \quad \forall p \in [2,q], \tag{2.3}
\]

where \( c \) is a positive constant, which may depend on \( \lambda, q, ||a_{ij}||_{W^{1,q}(\Omega)} \) and \( \Omega \).

(a2) The flux function \( f \in C^3(\mathbb{R}^d)^d \) satisfies the local ellipticity condition (1.2).

(a3) The solution of (1.1) is sufficiently regular; more precisely,

\[
u \in C^2 ([0,T];L^q(\Omega)) \cap C([0,T];W^{2,q}(\Omega)) \quad \text{for } q \text{ as in (a1)}.
\tag{2.4}
\]

Justification of these assumptions can be found in Appendix A. In particular, assumption (a1) holds if \( \Omega \) is any convex two-dimensional polygon or any three-dimensional polyhedron with interior dihedral angles less than \( \frac{\pi}{4} \) (such as rectangular parallelepiped); assumption (a2) holds for all examples mentioned in the introduction, including the minimal surface flow, the regularized total variation flow, and the general \( L_2(\Omega) \) gradient flow. Under assumptions (a1)–(a2), for any smooth initial data \( u_0 \in W^{2,q}(\Omega) \cap W^{1,q}_0(\Omega) \) satisfying the compatibility condition

\[
\nabla \cdot f(\nabla u_0) \in W^{2,q}(\Omega) \cap W^{1,q}_0(\Omega), \tag{2.5}
\]

the partial differential equation (PDE) problem (1.1) has a unique solution \( u \) with regularity (2.4) up to some finite time \( T > 0 \) (local existence of smooth solutions), and thus assumption (a3) holds.

2.1 Implicit Euler scheme and finite element method

Let \( N \) be a positive integer and consider a uniform partition \( t_n := n \tau, n = 0, 1, \ldots, N \), of the interval \([0,T]\) with time step \( \tau = T/N \). With the starting value \( u^n := u_0 \), we define a sequence of approximations \( u^n \in W^{1,\infty}(\Omega) \cap H^1_0(\Omega) \) to the nodal values \( u(t_n) := u(\cdot, t_n) \) of the exact solution, by discretizing (1.1) with the implicit Euler scheme,

\[
\frac{u^n - u^{n-1}}{\tau} = \nabla \cdot f(\nabla u^n), \quad n = 1, \ldots, N. \tag{2.6}
\]

Let \( (S_h)_{0<h<1} \subset H^1_0(\Omega) \) denote a family of finite element spaces of continuous piecewise linear polynomials corresponding to a quasi-uniform triangulation of the domain \( \Omega \), with mesh size \( h \). The fully discrete finite element counterpart of the implicit Euler scheme (2.6) is to seek approximations \( u^n_h \in S_h \) to \( u(t_n) \) such that

\[
\left( \frac{u^n_h - u^{n-1}_h}{\tau}, v_h \right) = -\left( f(\nabla u^n_h), \nabla v_h \right) \quad \forall v_h \in S_h, \quad n = 1, \ldots, N, \tag{2.7}
\]
Let us consider the fully discrete scheme (2.7), which cannot be implemented directly. An efficient way to linearize (2.6), as well as the corresponding iterations for (2.8), is by Newton’s method.

2.2 Linearization by Newton’s method

Let
\[ Df(p) := \nabla_p f(p), \quad D^2 f(p) := \nabla_p^2 f(p) \quad \text{and} \quad D^3 f(p) := \nabla_p^3 f(p) \quad \text{for} \quad p \in \mathbb{R}^d. \]

Thus \( D^2 f(p) \) and \( D^3 f(p) \) are 3rd- and 4th-order tensors with components
\[ (D^2 f(p))_{ijkl} = \partial_i \partial_j \partial_k f_k(p) \quad \text{and} \quad (D^3 f(p))_{ijkl} = \partial_i \partial_j \partial_k \partial_l f_l(p). \]

Multiplication of the 3rd-order tensor \( D^2 f(p) \) with a \( d \)-dimensional vector yields a 2nd-order tensor (i.e., matrix) with components
\[ (D^2 f(p) : \mathbf{v})_{jk} = \sum_{i=1}^d v_i \partial_i \partial_j f_k(p). \]

Similarly, multiplication of the 3rd-order tensor \( D^2 f(p) \) with a \( d \times d \) matrix \( M = (M_{ij}) \) yields a vector with components
\[ (D^2 f(p) : M)_k = \sum_{i,j=1}^d \partial_i \partial_j f_k(p) M_{ij}. \]

Multiplication of the 4th-order tensor \( D^3 f(p) \) with matrices or vectors can be defined similarly.

Let \( \ell \) denote the number of Newton iterations at each time level, and set \( u_h^0 := u_0 \). At the \( n \)th time level, by choosing the starting value \( u_h^0 := u_t^{-1} \), Newton’s method for the semidiscrte scheme (2.6) seeks \( u_h^m \in W^{1,\infty}(\Omega) \cap H_0^1(\Omega), m = 1, \ldots, \ell, \) to be the solutions of the following linear equation:
\[ u_h^m - u_h^{m-1} = \nabla \cdot f(\nabla u_h^{m-1}) + \nabla \cdot \left( Df(\nabla u_h^{m-1}) \right). \tag{2.8} \]

Similarly, at the \( n \)th time level, Newton’s method for the fully discrete nonlinear system (2.7) seeks \( u_{h,m}^n \in S_h, m = 1, 2, \ldots, \ell, \) such that
\[ \frac{u_h^n - u_h^{n-1}}{\tau} = -(f(\nabla u_{h,m-1}) \cdot \nabla v_h) - (Df(\nabla u_{h,m-1}) \nabla (u_{h,m-1} - u_{h,m-1}) \cdot \nabla v_h) \quad \forall v_h \in S_h, \tag{2.9} \]

with \( u_h^n \) the \( L^2 \) projection of the initial value \( u_0 \).

Alternatively, (2.9) can also be viewed as the finite element discretization of the semidiscrete Newton iteration scheme (2.8). Based on this point of view, the error of the Newton iterative finite element solutions given by (2.9) can be split into the three parts in (1.4) that can be estimated separately.

The main result of this paper is the following theorem.

**Theorem 2.1** Let \( \ell \geq 2 \). Then, under assumptions (a1)–(a3), there exist positive constants \( \tau_*, \) and \( h_* \) such that for \( \tau \leq \tau_* \) and \( h \leq h_* \) the numerical solutions given by the Newton iterative finite element method (2.9) satisfy the following estimates
\[ \max_{1 \leq n \leq N} \| u_{h,\ell}^n - u(t_n) \|_{W^{1,q}(\Omega)} \leq c_{\varepsilon,\ell}(\tau + h^{1-\varepsilon}), \tag{2.10} \]
\[ \max_{1 \leq n \leq N} \| u_{h,\ell}^n - u(t_n) \|_{L^q(\Omega)} \leq c_{\varepsilon,\ell}(\tau + h^{2-\varepsilon}), \tag{2.11} \]
where \( q \) is the same number as in assumption (a1), \( \varepsilon \in (0, 1) \) can be an arbitrarily small constant, and \( c_{\varepsilon,\ell} \) is a constant independent of \( \tau \) and \( n \) (that may depend on \( \varepsilon, \ell \) and \( T \)).
The main technical tool is the following theorem on discrete maximal $L^p$-regularity of fully discrete finite element solutions of parabolic equations with time-dependent coefficients.

**Theorem 2.2** Let $q > d$ and $a_{ij} = a_{ji} \in C([0,T];W^{1,q}(\Omega))$, $i, j = 1, \ldots, d$, be functions satisfying (2.1) and the Lipschitz condition with respect to $t$, uniformly in $x$,

$$\sup_{x \in \Omega} |a_{ij}(x,t) - a_{ij}(x,s)| \leq c|t-s|, \quad t, s \in [0,T],$$

and let $\mathcal{A}(x,t)$ denote the symmetric $d \times d$ matrix with entries $a_{ij}(x,t)$. Let $\phi^n \in H^1_0(\Omega)$ and $\phi^n_h \in S_h$, $n = 1, \ldots, N$, be the solutions of

$$\delta_t \phi^n - \nabla \cdot (\mathcal{A}(\cdot,t) \nabla \phi^n) = g^n$$

and

$$(\delta_t \phi^n_h, \upsilon_n) + (\mathcal{A}(\cdot,t_n) \nabla \phi^n, \nabla \upsilon_n) = (g^n, \upsilon_n) \quad \forall \upsilon_n \in S_h,$$

respectively, for some functions $g^n \in L^q(\Omega)$, $s \in (1,\infty)$, $n = 1, \ldots, N$, with starting data $\phi^0 \in H^1_0(\Omega)$ and $\phi^0_h \in S_h$. Thus

$$(\delta_t (\phi^n - \phi^n_h), \upsilon_n) + (\mathcal{A}(\cdot,t_n) \nabla (\phi^n - \phi^n_h), \nabla \upsilon_n) = 0 \quad \forall \upsilon_n \in S_h.$$  

Then, under assumption (a1), the following estimates are valid, when $\phi^n_0 = \phi^0 = 0$, for all $1 \leq k \leq N$ and $1 < p < \infty$:

$$\|(\delta_t \phi^n_h)_{n=1}^k\|_{L^p(W^{-1,q}(\Omega))} + \|((\phi^n_h)_{n=1}^k\|_{L^p(W^{1,q}(\Omega))} \leq c\|(g^n)_{n=1}^k\|_{L^p(W^{-1,q}(\Omega))} \quad \forall s \in (1,\infty),$$

$$\|(\delta_t \phi^n)_{n=1}^k\|_{L^p(W^{-1,q}(\Omega))} + \|((\phi^n)_{n=1}^k\|_{L^p(W^{1,q}(\Omega))} \leq c\|(g^n)_{n=1}^k\|_{L^p(W^{-1,q}(\Omega))} \quad \forall s \in (1,\infty),$$

and

$$\|(\phi^n - \phi^n_h)_{n=1}^k\|_{L^p(L^q(\Omega))} \leq c\|(\phi^n - \phi^n_h)_{n=1}^k\|_{L^p(L^q(\Omega))} + c\|(\phi^n - R^\infty h \phi^n)_{n=1}^k\|_{L^p(L^q(\Omega))} \leq c\|h^\ell\|_{L^p(W^{1,q}(\Omega))},$$

where $P_h$ and $R^\infty h$ denote the $L^2$ and Ritz projections onto the finite element space, respectively, with the latter defined by

$$(\mathcal{A}(\cdot,t_n) \nabla (\phi - R^\infty h \phi), \nabla \upsilon_n) = 0 \quad \forall \upsilon_n \in S_h \forall \phi \in H^1_0(\Omega).$$

**Remark 2.1** Estimates (2.15) and (2.16) can be viewed as maximal $L^p$-regularity on the Banach space $W^{-1,q}(\Omega)$ for fully discrete and semidiscrete schemes, respectively.

Under the assumptions of Theorem 2.2, the Ritz projection satisfies the following estimate (cf. [5, (8.5.3)-(8.5.5)])

$$\|\phi - R^\infty h \phi\|_{L^q(\Omega)} + h\|\phi - R^\infty h \phi\|_{W^{1,q}(\Omega)} \leq c\|\phi\|_{W^{\ell,q}(\Omega)} h^{\ell} \quad \forall \phi \in W^{\ell,q}(\Omega), \quad \ell = 1, 2, \quad s \in [2,\infty).$$

Theorem 2.2 is analogous to [23, Theorem 2.1]. The latter is proved in smooth domains with the Neumann boundary condition, while the former is for convex polygonal/polyhedral domains with the Dirichlet boundary condition under assumption (a1). A sketch of the proof for Theorem 2.2 can be found in Appendix B.
3. Newton’s iteration for time discretization

Under assumptions (a1)–(a3), it has been proved in [16, Theorem 2.1] that the semidiscrete solutions $u^n, n = 1, \ldots, N$, are well defined and satisfy the following estimate:

$$
\| (u^n - u(t_n))^N_{n=1} \|_{L^p(W1,\omega)} + \| (u^n - u(t_n))^N_{n=1} \|_{L^{p,q}(W2,q)} \leq c_{p,q} \tau \quad \forall \ p \in (1, \infty),
$$

where $q$ is the number in assumption (a1). This further implies the following regularity estimate (for the time-discrete solution):

$$
\max_{1 \leq n \leq N} ( \| \delta t^n \|_{W1,\omega} + \| u^n \|_{W2,q} ) \leq c,
$$

where $\delta t^n = (u^n - u^{n-1})/\tau$. This regularity estimate plays a crucial role in our analysis of Newton’s method applied to both the implicit Euler scheme (2.6) and its finite element discretization (2.7).

The main result of this section is the following proposition.

**Proposition 3.1** Let $\ell \geq 2$. Then, under assumptions (a1)–(a3), there exists a positive constant $\tau_0$ such that for $\tau \leq \tau_0$ the numerical solutions given by the Newton iterative scheme (2.8) satisfy the following estimates

$$
\max_{1 \leq n \leq N} \| u^n_t - u^n_{t} \|_{H^1(\Omega)} \leq c_{\ell} \tau^{(2\ell+1)/2},
$$

$$
\max_{1 \leq n \leq N} ( \| u^n_t - u^n_{t} \|_{W1,\omega(\Omega)} + \| u^n - u^n_{t} \|_{W2,q(\Omega)} ) \leq c_{\ell} \tau^{(2\ell-1)/2},
$$

$$
\max_{1 \leq n \leq N} ( \tau^{-1} \| u^n_{m-1} - u^n_{m-1} \|_{W1,\omega(\Omega)} + \| u^n_{m} - u^n_{m} \|_{W2,q(\Omega)} ) \leq c_{\ell}, \quad m = 1, \ldots, \ell,
$$

where $c_{\ell}$ is a constant independent of $\tau$ and $n$ (that may depend on $T$ and $\ell$).

**Proof** Taylor expanding the term on the right-hand side of the implicit Euler scheme (2.6) about $\nabla u^n_{m-1}$, we see that the implicit Euler approximation $u^n_t$ satisfies the relation

$$
\frac{u^n_t - u^n_{t-1}}{\tau} = \nabla \cdot f(\nabla u^n_{m-1}) + \nabla \cdot \left( Df(\nabla u^n_{m-1}) \nabla (u^n - u^n_{m-1}) \right) + \nabla \cdot \left( \int_{0}^{1} D^2f \left( (1-t)\nabla u^n_{m-1} + t\nabla u^n \right) (1-t) d\tau \nabla (u^n - u^n_{m-1}) \cdot \nabla (u^n - u^n_{m-1}) \right).
$$

Letting $e^n_m = u^n_m - u^n$ and subtracting (3.6) from (2.8), we obtain

$$
e^n_m = \nabla \cdot \left( Df(\nabla u^n_{m-1}) \nabla e^n_m \right)
$$

$$
- \nabla \cdot \left( \int_{0}^{1} D^2f \left( (1-t)\nabla u^n_{m-1} + t\nabla u^n \right) (1-t) d\tau \nabla e^n_{m-1} \cdot \nabla e^n_{m-1} \right) + \frac{e^n_{t-1}}{\tau}.
$$

With the second order elliptic partial differential operator

$$
A^n_{m-1} = \nabla \cdot (Df(\nabla u^n_{m-1}) \nabla),
$$

relation (3.7) can be rewritten in the form

$$
A^n_{m-1} e^n_m = -A^n_{m-1} \left( \frac{1}{\tau} - A^n_{m-1} \right)^{-1} \nabla \cdot \left( \int_{0}^{1} D^2f \left( (1-t)\nabla u^n_{m-1} + t\nabla u^n \right) (1-t) d\tau \nabla e^n_{m-1} \cdot \nabla e^n_{m-1} \right)
$$

$$
+ \frac{1}{\tau} \left( \frac{1}{\tau} - A^n_{m-1} \right)^{-1} A^n_{m-1} e^n_{t-1}.
$$

We assume (mathematical induction assumption)

$$
\| e^n_{t-1} \|_{W^{2,q}(\Omega)} \leq \tau, \quad n = 1, \ldots, k,
$$

and prove that

$$
\| e^n_{t} \|_{W^{2,q}(\Omega)} \leq \tau.
$$

To this end, we introduce a second loop of mathematical induction, namely, for $1 \leq n \leq k$ and $1 \leq m \leq i$.
we assume that
\[ \|u_{m-1}^\nu\|_{W^{1-\gamma}(\Omega)} \leq \|u^\nu\|_{W^{1-\gamma}(\Omega)} + 1, \]  
(3.11)
\[ \|u_{m-1}^\nu\|_{W^{2,q}(\Omega)} \leq \|u^\nu\|_{W^{2,q}(\Omega)} + 1. \]  
(3.12)
Then, we prove
\[ \|u_{m}^\nu\|_{W^{1-\gamma}(\Omega)} \leq \|u^\nu\|_{W^{1-\gamma}(\Omega)} + 1, \]  
(3.13)
\[ \|u_{m}^\nu\|_{W^{2,q}(\Omega)} \leq \|u^\nu\|_{W^{2,q}(\Omega)} + 1. \]  
(3.14)
(for the same range of \( n \)) to complete the mathematical induction. Since \( e_0^\nu = e_{k-1}^\nu \) and \( W^{2,q}(\Omega) \hookrightarrow W^{1-\gamma}(\Omega) \), it follows that (3.9) implies (3.11)–(3.12) for \( m = 1 \) (with sufficiently small \( \tau \)). In the sequel, we shall keep the generic constant \( c \) independent of \( n, k, i \) and \( \ell \).

Under the induction assumptions (3.9) and (3.11)–(3.12), the coefficient matrix \( Df(\nabla u_{m-1}^\nu) \) satisfies the following estimates:
\[ \|Df(\nabla u_{m-1}^\nu)\|_{W^{2,q}(\Omega)} \leq c, \]
\[ \lambda^{-1}\|\xi\|^2 \leq Df(\nabla u_{m-1}^\nu)\xi \cdot \xi < \lambda\|\xi\|^2 \quad \forall \xi \in \mathbb{R}^d, \]  
(3.15)
for some positive constants \( c \) and \( \lambda \) depending on \( \|u^\nu\|_{W^{2,q}(\Omega)} \) and \( \|u^\nu\|_{W^{1-\gamma}(\Omega)} \) (implicit Euler approximations of the exact solution of the PDE). Thus, operator \( A_{m-1}^\nu \) generates a bounded analytic semigroup on \( L^q(\Omega) \) (cf. [27, Theorem 3.11]), satisfying the resolvent estimates (cf. [3, Theorem 3.7.11])
\[ \left\| A_{m-1}^\nu \left( \frac{1}{\tau} - A_{m-1}^\nu \right)^{-1} \right\|_{L^q(\Omega) \to L^q(\Omega)} \leq c \quad \text{and} \quad \left\| \frac{1}{\tau} \left( \frac{1}{\tau} - A_{m-1}^\nu \right)^{-1} \right\|_{L^q(\Omega) \to L^q(\Omega)} \leq c. \]
Therefore, (3.8) yields
\[ \|A_{m-1}^\nu e_m^\nu\|_{L^q(\Omega)} \leq c \left\| D f \left( \nabla u_{m-1}^\nu + \tau \nabla u^\nu \right) \left( 1 - \tau \right) \nabla \nabla e_{m-1}^\nu \cdot \nabla e_{m-1}^\nu \right\|_{L^q(\Omega)} + c \|A_{m-1}^\nu e_{k-1}^\nu\|_{L^q(\Omega)} \]
\[ \leq c \left\| D f \left( \nabla u_{m-1}^\nu + \tau \nabla u^\nu \right) \left( 1 - \tau \right) \nabla e_{m-1}^\nu \cdot \nabla e_{m-1}^\nu \right\|_{L^q(\Omega)} \]
\[ + c \left\| D f \left( \nabla u_{m-1}^\nu + \tau \nabla u^\nu \right) \left( 1 - \tau \right) \nabla \nabla e_{m-1}^\nu \cdot \nabla e_{m-1}^\nu \right\|_{L^q(\Omega)} \]
\[ + c \|A_{m-1}^\nu e_{k-1}^\nu\|_{L^q(\Omega)} \]
\[ \leq c \left\| \nabla^2 e_{m-1}^\nu\right\|_{L^q(\Omega)} + c \|\nabla^2 u_{m-1}^\nu\|_{L^q(\Omega)} + c \|\nabla^2 u_{m-1}^\nu\|_{L^q(\Omega)} + c \|\nabla^2 u_{m-1}^\nu\|_{L^q(\Omega)} + c \|A_{m-1}^\nu e_{k-1}^\nu\|_{L^q(\Omega)}, \]
whence (since \( W^{2,q}(\Omega) \hookrightarrow W^{1-\gamma}(\Omega) \) for \( q > d \))
\[ \|A_{m-1}^\nu e_{m}^\nu\|_{L^q(\Omega)} \leq c \|e_{m-1}^\nu\|_{W^{2,q}(\Omega)} + c \|e_{k-1}^\nu\|_{W^{2,q}(\Omega)}. \]  
(3.16)
Under the induction assumptions (3.11)–(3.12), we have the elliptic \( W^{2,q} \) regularity (see assumption (a1))
\[ \|e_{m}^\nu\|_{W^{2,q}(\Omega)} \leq c \|A_{m-1}^\nu e_{m}^\nu\|_{L^q(\Omega)}. \]  
(3.17)
Combining (3.16) and (3.17), we get
\[ \|e_{m}^\nu\|_{W^{2,q}(\Omega)} \leq c \|\nabla e_{m-1}^\nu\|_{W^{2,q}(\Omega)} + c \|e_{k-1}^\nu\|_{W^{2,q}(\Omega)}. \]  
(3.18)
whence, iterating (3.18) with respect to \( m \) to obtain
\[ \|e_{m}^\nu\|_{W^{2,q}(\Omega)} \leq c \|e_{m-1}^\nu\|_{W^{2,q}(\Omega)} + c \|e_{k-1}^\nu\|_{W^{2,q}(\Omega)} \]
\[ \leq c \|e_{m-2}^\nu\|_{W^{2,q}(\Omega)} + c \|e_{k-2}^\nu\|_{W^{2,q}(\Omega)} + c \|e_{k-1}^\nu\|_{W^{2,q}(\Omega)} \]
\[ \leq 2c \|e_{m-2}^\nu\|_{W^{2,q}(\Omega)} + 2c \|e_{k-1}^\nu\|_{W^{2,q}(\Omega)} + c \|e_{k-1}^\nu\|_{W^{2,q}(\Omega)} \]
\[ \leq 2^{i-3}c \|e_{m-3}^\nu\|_{W^{2,q}(\Omega)} + 2^{i-3}c \|e_{k-4}^\nu\|_{W^{2,q}(\Omega)} + 2^{i-3}c \|e_{k-4}^\nu\|_{W^{2,q}(\Omega)} + 2^{i-3}c \|e_{k-4}^\nu\|_{W^{2,q}(\Omega)} + c \|e_{k-1}^\nu\|_{W^{2,q}(\Omega)} \]
where we have used the fact that
\[ \| \nabla u_l^n - \nabla u_l^{n-1} \|_{L^2} \leq c \tau^{1-\varepsilon}, \]
where \( \varepsilon \) can be arbitrarily small, with a constant \( c_\varepsilon \) depending on \( \varepsilon \). In the last inequality we have used
\[ \| u_l^{n-1} - u_l^n \|_{L^2} = \| e_l^n \|_{L^2} \leq \tau, \]
which is due to the induction assumption (3.9), and
\[ \| u_l^{n-1} - u_l^n \|_{L^2} \leq \| u(t_{n-1}) - u(t_n) \|_{L^2} + \| u(t_n) - u_l^n \|_{L^2}, \]
where the last inequality is due to (3.1), in which \( p \) can be arbitrarily large and so \( \| u_l^{n-1} - u_l^n \|_{L^2} \leq c_\varepsilon \tau^{1-\varepsilon}. \) The induction assumption (3.9) implies for sufficiently small \( \tau \) (independent of \( m \) and \( l \))
\[ \| e_m^n \|_{L^2} \leq (2c_\varepsilon \tau^{1-\varepsilon})^{2m} + c\| e_l^{n-1} \|_{L^2} \leq (2c_\varepsilon \tau^{1-\varepsilon})^{2m} + c\| e_l^{n-1} \|_{L^2} \]
and thus
\[ \| e_m^n \|_{L^2} \leq c \tau, \]
which further implies
\[ \| e_m^n \|_{W^{1,\infty}(\Omega)} \leq c \| e_m^n \|_{W^2_0(\Omega)} \leq c \tau. \]

For sufficiently small \( \tau \), the last two estimates imply (3.13)–(3.14), completing the second loop of mathematical induction. Thus (3.13)–(3.14) are valid for all \( 1 \leq m \leq \ell \).

By the second mathematical induction, we have shown that \( \| \nabla u_m^n \|_{L^\infty} \) is uniformly bounded with respect to \( \tau \) for all \( 0 \leq m \leq \ell \) (note that \( \| \nabla u_0^n \|_{L^\infty} = \| \nabla u_\ell^{n-1} \|_{L^\infty} \). Now, integrating (3.7) against \( -\nabla \cdot (Df'(\nabla u_{m-1}^n)\nabla e_m^n) \) and using Hölder’s inequality, we obtain, for \( 1 \leq m \leq \ell \),
\[
\frac{1}{2\tau} \left( Df(\nabla u_{m-1}^n)\nabla e_m^n \cdot \nabla e_m^n \right) - \frac{1}{2\tau} \left( Df(\nabla u_{m-1}^n)\nabla e_{l}^{n-1} \cdot \nabla e_{l}^{n-1} \right) + \frac{1}{2} \left\| \nabla \cdot (Df(\nabla u_{m-1}^n)\nabla e_m^n) \right\|^2_{L^2(\Omega)} \leq c \left\| \nabla \cdot (Df'(\nabla u_{m-1}^n)\nabla e_m^n) \right\|^2_{L^2(\Omega)} \]
\[
\leq c \left\| \nabla \cdot \left( \int_0^1 D^2 f \left( (1-t)\nabla u_{m-1}^n + t\nabla u^n \right)(1-t) \, dt \nabla e_{m-1}^n \cdot \nabla e_{m-1}^n \right) \right\|^2_{L^2(\Omega)} \]
\[
\leq c \left\| \int_0^1 D^2 f \left( (1-t)\nabla u_{m-1}^n + t\nabla u^n \right)(1-t) \, dt \cdot \nabla^2 e_{m-1}^n \cdot e_{m-1}^n \right\|^2_{L^2(\Omega)} + c \left\| \int_0^1 D^3 f \left( (1-t)\nabla u_{m-1}^n + t\nabla u^n \right) : \left( (1-t)\nabla^2 u_{m-1}^n + t\nabla^2 u^n \right)(1-t) \, dt \cdot \nabla e_{m-1}^n \cdot \nabla e_{m-1}^n \right\|^2_{L^2(\Omega)} \]
Therefore, by using Hölder’s inequality,

\[
\left\| \int_0^1 D^2 f \left( (1-t)\nabla u_m^{n-1} + t \nabla u^n \right) (1-t) \, dt \cdot \nabla e_m^n , \nabla e_m^{n-1} \right\|_{L^2(\Omega)}^2 \\
+ c \left\| \int_0^1 D^3 f \left( (1-t)\nabla u_m^{n-1} + t \nabla u^n \right) : \left( (1-t)\nabla^2 u_m^{n-1} + t \nabla^2 u^n \right) (1-t) \, dt \cdot \nabla e_m^n , \nabla e_m^{n-1} \right\|_{L^2(\Omega)}^2 \\
\leq c \|\nabla^2 u_m^{n-1}\|_{L^q(\Omega)}^2 \|\nabla e_m^n\|_{L^p(\Omega)}^2 2^{2q \over 2p} L^{2q \over 2p-2}(\Omega) + c \left( \|\nabla^2 u_m^{n-1}\|_{L^q(\Omega)}^2 + \|\nabla^2 u^n\|_{L^q(\Omega)}^2 \right) \|\nabla e_m^n\|_{L^p(\Omega)}^2 2^{2q \over 2p} L^{2q \over 2p-2}(\Omega) \\
\leq c \|e_m^{n-1}\|_{W^{2,q}(\Omega)}^2 \|e_m^n\|_{L^p(\Omega)}^2 2^{2q \over 2p} L^{2q \over 2p-2}(\Omega) + c \|e_m^n\|_{W^{1,q}(\Omega)}^2 \|e_m^{n-1}\|_{W^{1,q}(\Omega)}^2 2^{2q \over 2p} L^{2q \over 2p-2}(\Omega) \\
\leq c \|\nabla u^n_m\|_{L^2(\Omega)}^2 \\
\leq c \|\nabla (D f(\nabla u^n_m) - D f(\nabla u^n_\ell))\|_{L^2(\Omega)}^2 ,
\]  
(3.24)

with \( u^n_\ell := u^n_{m} \) in the case \( m = 1 \), where we have used (3.22) in the second last inequality, and (2.3) in the last inequality. Substituting (3.24) into (3.23) yields

\[
\frac{1}{2\tau} \left\| (D f(\nabla u^n_m) - D f(\nabla u^n_\ell)) \right\|_{L^2(\Omega)}^2 \\
\leq c \tau^2 \left\| \nabla (D f(\nabla u^n_m) - D f(\nabla u^n_\ell)) \right\|_{L^2(\Omega)}^2 + \frac{1}{2\tau} \left\| D f(\nabla u^n_\ell) \nabla e_{\ell}^{n-1}, \nabla e_{\ell}^{n-1} \right\|_{L^2(\Omega)}^2 \\
\leq c \tau^2 \left\| \nabla (D f(\nabla u^n_m) - D f(\nabla u^n_\ell)) \right\|_{L^2(\Omega)}^2 + \frac{1}{2\tau} \left( D f(\nabla u^n_\ell) \nabla e_{\ell}^{n-1}, \nabla e_{\ell}^{n-1} \right) \right).
\]

Again, since \( \|\nabla u^n_\ell\|_{L^2} \) is uniformly bounded for \( 0 \leq m \leq \ell \), it follows that

\[
\left\| D f(\nabla u^n_m) - D f(\nabla u^n_\ell) \right\|_{L^2(\Omega)} \leq C
\]

and thus

\[
\frac{1}{2\tau} \left\| (D f(\nabla u^n_m) - D f(\nabla u^n_\ell)) \right\|_{L^2(\Omega)}^2 \\
\leq c \tau^2 \left\| \nabla (D f(\nabla u^n_m) - D f(\nabla u^n_\ell)) \right\|_{L^2(\Omega)}^2 + \frac{1}{2\tau} \left\| D f(\nabla u^n_\ell) \nabla e_{\ell}^{n-1}, \nabla e_{\ell}^{n-1} \right\|_{L^2(\Omega)}^2 \\
\leq c \tau^2 \left\| \nabla (D f(\nabla u^n_m) - D f(\nabla u^n_\ell)) \right\|_{L^2(\Omega)}^2 + \frac{1}{2\tau} \left( 1 + c \tau \right) \left( D f(\nabla u^n_\ell) \nabla e_{\ell}^{n-1}, \nabla e_{\ell}^{n-1} \right).
\]

Let \( Y_m^n = \left\{ D f(\nabla u^n_m) - D f(\nabla u^n_\ell) \right\} \left\| \nabla (D f(\nabla u^n_m) - D f(\nabla u^n_\ell)) \right\|_{L^2(\Omega)}^2 . \) Then, the last estimate implies

\[
Y_m^n \leq c \tau^2 Y_m^{n-1} + (1 + c \tau) \left\{ D f(\nabla u^n_\ell) \nabla e_{\ell}^{n-1}, \nabla e_{\ell}^{n-1} \right\}.
\]

Iterations of this inequality yield

\[
Y_m^n \leq c \tau^2 Y_m^{n-1} + (1 + c \tau) \left\{ D f(\nabla u^n_\ell) \nabla e_{\ell}^{n-1}, \nabla e_{\ell}^{n-1} \right\} \\
\leq (c \tau^2)^m Y_m^0 + \left( 1 + c \tau^2 \left\{ D f(\nabla u^n_\ell) \nabla e_{\ell}^{n-1}, \nabla e_{\ell}^{n-1} \right\} \right) \\
\leq \cdots \\
\leq (c \tau^2)^m Y_m^0 + \left[ 1 + c \tau^2 + \cdots + (c \tau^2)^{m-1} \right] \left( 1 + c \tau \right) \left\{ D f(\nabla u^n_\ell) \nabla e_{\ell}^{n-1}, \nabla e_{\ell}^{n-1} \right\},
\]

\[
\leq (c \tau^2)^m \left( \left\| \nabla e^n_\ell \right\|_{L^2(\Omega)}^2 + \tau \left\| \nabla e^n_\ell \right\|_{L^2(\Omega)}^2 \right) + \frac{1}{1 - c \tau^2} \left\| \nabla e_{\ell}^{n-1} \right\|_{L^2(\Omega)}^2 .
\]

When \( \tau \) is small enough, the last inequality implies

\[
\left\| \nabla e^n_\ell \right\|_{L^2(\Omega)}^2 \leq (c \tau^2)^m \left( \left\| \nabla e^n_\ell \right\|_{L^2(\Omega)}^2 + \tau \left\| \nabla e^n_\ell \right\|_{L^2(\Omega)}^2 \right) + (1 + c \tau) \left\| \nabla e_{\ell}^{n-1} \right\|_{L^2(\Omega)}^2 .
\]  
(3.25)
In particular, setting $m = \ell$ in the last estimate and using (3.20), we obtain
\[
\|\nabla e^{m}_\ell\|_{L^2(\Omega)}^2 \leq (c\tau^2)\ell^2 + (1 + c\tau)\|\nabla e^{m-1}_\ell\|_{L^2(\Omega)}^2.
\] (3.26)

Iteration of (3.26) gives
\[
\|\nabla e^{m}_\ell\|_{L^2(\Omega)}^2 \leq (c\tau^2)^\ell^2 + (1 + c\tau)\|\nabla e^{m-1}_\ell\|_{L^2(\Omega)}^2 \leq (c\tau^2)^\ell^2 n(1 + c\tau)^n + (1 + c\tau)^n\|\nabla e^0_\ell\|_{L^2(\Omega)}^2 \leq ce^{\ell^2}(c\tau^2)^\ell \tau + ce^{\ell^2}\|\nabla e^0_\ell\|_{L^2(\Omega)}^2.
\]

Therefore, since $e^0_\ell$ vanishes,
\[
\|\nabla e^{m}_\ell\|_{L^2(\Omega)}^2 \leq cc^\ell \tau^{2\ell+1}.
\] (3.27)

Note that $H^1(\Omega) \rightarrow L^6(\Omega)$ for $d \in \{2, 3\}$. From (3.8) we see that, for $q \in (d, 6)$,
\[
\|A^{n}_{m-1} e^m_\ell\|_{L^q(\Omega)} \leq c\|\nabla \int_0^1 D^2 \int (1-t)\nabla u_{m-1} + t\nabla u_\ell)(1-t)dt \nabla e^{m-1}_\ell, \nabla e^m_\ell\|_{L^q(\Omega)} + c\|\nabla e^{m-1}_\ell\|_{L^q(\Omega)} \leq c\|e^{m-1}_\ell\|_{W^{2,q}(\Omega)} + cc\|e^m_\ell\|_{W^{2,q}(\Omega)},
\]
The derivation of the last inequality is analogous to the one of (3.16). Iteration of this estimate yields (in analogy to the derivation of (3.19))
\[
\|e^m_\ell\|_{W^{2,q}(\Omega)} \leq 2^n(c\tau^2)^m + cc\|e^m_\ell\|_{W^{2,q}(\Omega)} (1 + c\sum_{j=1}^{m-1} 2^j (c\tau^2)^{(2\ell-1)/2})^2 \leq (2c\tau^2)^m + cc\|e^m_\ell\|_{W^{2,q}(\Omega)}.
\] (3.28)

For sufficiently small $\tau$, estimate (3.28) implies, for $\ell \geq 2$,
\[
\|e^m_\ell\|_{W^{1,\infty}(\Omega)} + \|e^m_\ell\|_{W^{2,q}(\Omega)} \leq c\tau^\frac{3}{2}, \quad m = 1, \ldots, \ell;
\] (3.29)

therefore, for sufficiently small $\tau$, we have
\[
\|e^m_\ell\|_{W^{2,q}(\Omega)} \leq \tau.
\] (3.30)

This completes the first loop of mathematical induction. Thus, all these estimates hold for $n = 1, \ldots, N$ and $m = 1, \ldots, \ell$. In particular, (3.27)–(3.29) imply (3.3)–(3.5). This completes the proof of Proposition 3.1.

The following result is an immediate consequence of (3.1) and (3.4) by a triangle inequality.

**Corollary 3.1** Let $\ell \geq 2$ and assume that the solution of the initial and boundary value problem (1.1) satisfies (2.4). Then, there exists a positive constant $\tau_0$ such that for $\tau \leq \tau_0$ the numerical solutions given by Newton’s method (2.8) satisfy the following estimate:
\[
\max_{1 \leq n \leq N} \|u^n_\ell - u(t_n)\|_{W^{1,\infty}(\Omega)} \leq c\tau,
\] (3.31)
where $c$ is a constant independent of $\tau$ and $n$ (that may depend on $T$ and $\ell$).

**Remark 3.1** Assuming uniform ellipticity of the coefficient matrix, error estimates can be established by the energy technique (3.23)–(3.27). This is only a small part of the proof of Proposition 3.1, which mainly consists in deriving $W^{2,q}$ and $W^{1,\infty}$ error estimates in order to rule out the possibility of degeneracy.

4. **Proof of Theorem 2.1**

Theorem 2.1 is a consequence of Proposition 3.1 and the following proposition.

**Proposition 4.1** Let $\ell \geq 2$. Then, under assumptions (a1)–(a3), there exist positive constants $\tau_*$ and...
method (2.9) satisfy the following estimates
\[
\max_{1 \leq n \leq N} \| u_{h,n}^{n} - u_{r}^{n} \|_{W^{1,q} (\Omega)} \leq c_{\epsilon, \ell} h^{1-\epsilon}, \quad (4.1)
\]
\[
\max_{1 \leq n \leq N} \| u_{h,n}^{n} - u_{r}^{n} \|_{L^{q} (\Omega)} \leq c_{\epsilon, \ell} (\tau^{2} + h^{2-\epsilon}), \quad (4.2)
\]
where \( \epsilon \in (0, 1) \) can be an arbitrarily small constant, and \( c_{\epsilon, \ell} \) is a constant independent of \( \tau \) and \( n \) (that may depend on \( \epsilon, \ell \) and \( T \)).

In the rest of this section, we prove Proposition 4.1.

### 4.1 Error equation and mathematical inductions

We recall that \( P_{h} \) denotes the \( L^{2} \) orthogonal projection onto the finite element space \( S_{h} \); it satisfies the following estimates:
\[
\| \phi - P_{h} \phi \|_{W^{k,p} (\Omega)} \leq C h^{m-k} \| \phi \|_{W^{m,p} (\Omega)} \quad \forall \phi \in W^{m,p} (\Omega) \cap H_{0}^{1} (\Omega), \quad (4.3)
\]
for \( k = 0, 1, m = 1, 2, \) and \( 1 \leq p \leq \infty \). Estimate (4.3) is a consequence of [29, Lemma 7.2].

Integrating (2.8) against a test function \( v_{h} \) yields
\[
\left( \frac{u_{m}^{n} - u_{r}^{n-1}}{\tau}, v_{h} \right) = -\left( f(\nabla u_{m-1}^{n}), \nabla v_{h} \right) - (Df(\nabla u_{m-1}^{n}) \nabla (u_{m}^{n} - u_{m-1}^{n}), \nabla v_{h}) \quad \forall v_{h} \in S_{h}. \quad (4.4)
\]

Let
\[ e_{h,m} = u_{h,m} - u_{r}^{n-1}, \quad \theta_{h,m} = u_{h,m} - P_{h} u_{r}^{n} \quad \text{and} \quad \rho_{h,m} = P_{h} u_{r}^{n} - u_{r}^{n}. \]

Obviously, \( e_{h,m} = \theta_{h,m} + \rho_{h,m} \), \( \| \rho_{h,m} \|_{L^{q} (\Omega)} \) and \( \| \theta_{h,m} \|_{W^{1,q} (\Omega)} \) can be easily estimated using (3.5) and (4.3). Therefore, we focus on the estimation of \( \| \theta_{h,m} \|_{L^{q} (\Omega)} \) and \( \| \theta_{h,m} \|_{W^{1,q} (\Omega)} \). By our choice \( u_{h,\ell}^{0} = P_{h} u_{0} = u_{h} \), we have \( \theta_{h,\ell}^{0} = 0 \).

Subtracting (4.4) from (2.9) and noting that \( (e_{h,m}^{n}, v_{h}) = (\theta_{h,m}^{n}, v_{h}) \) for \( v_{h} \in S_{h} \), we obtain
\[
\frac{1}{\tau} (\theta_{h,m}^{n}, v_{h}) = I_{1}^{n} (v_{h}) + I_{2}^{n} (v_{h}) + I_{3}^{n} (v_{h}) = I_{5}^{n} (v_{h}) \quad \forall v_{h} \in S_{h} \quad (4.5)
\]
with
\[
I_{1}^{n} (v_{h}) := -\left( f(\nabla u_{m-1}^{n}), \nabla v_{h} \right),
\]
\[
I_{2}^{n} (v_{h}) := -(Df(\nabla u_{m-1}^{n}) \nabla (e_{h,m}^{n} - e_{h,m-1}^{n}), \nabla v_{h}),
\]
\[
I_{3}^{n} (v_{h}) := -(Df(\nabla u_{m-1}^{n}) - Df(\nabla u_{m-1}^{n})), \nabla v_{h}),
\]
\[
I_{5}^{n} (v_{h}) := \frac{1}{\tau} (e_{h,\ell}^{n-1}, v_{h}).
\]

Now,
\[
I_{1}^{n} (v_{h}) = -\left( \int_{0}^{1} Df((1-s)\nabla u_{m-1}^{n} + s\nabla u_{m-1}^{n} ds \nabla e_{h,m-1}^{n}, \nabla v_{h}) \right.
\]
\[
\quad - (Df(\nabla u_{m-1}^{n}) \nabla e_{h,m-1}^{n}, \nabla v_{h}), \quad (4.6)
\]
whence
\[
I_{1}^{n} (v_{h}) + I_{2}^{n} (v_{h}) = -(Df(\nabla u_{m-1}^{n}) \nabla \theta_{h,m}^{n}, \nabla v_{h}) + I_{5}^{n} (v_{h}) \quad (4.7)
\]
with
\[
I_{6}^{n} (v_{h}) := -(Df(\nabla u_{m-1}^{n}) \nabla \rho_{h,m}^{n}, \nabla v_{h}).
\]
12 of 21

\[ F_\tau^k(v_h) := - \left( \int_0^1 (Df((1-s)\nabla u_{h,m-1}^n + s\nabla u_{m-1}^n) - Df(\nabla u_{m-1}^n))ds \nabla e_{h,m-1}^n, \nabla v_h \right). \]

Thus, (4.5) can be rewritten as

\[ \left( \frac{1}{\tau} - A_{h,m-1}^n \right) \theta_{h,m}^n(v_h) = F_\tau^k(v_h) + F_\tau^0(v_h) + F_\tau^0(v_h) + \| F_\tau^0(v_h) \| \quad \forall v_h \in S_h, \]

with the operator \( A_{h,m-1}^n : S_h \rightarrow S_h \) defined via duality by

\[ (A_{h,m-1}^n w_h, v_h) = - \left( Df(\nabla u_{m-1}^n) \nabla w_h, \nabla v_h \right) \quad \forall w_h, v_h \in S_h. \]

In the following, we estimate \( \| e_{h,m}^{n-1} \|_{W^{1,q}(\Omega)} \) by using the error equation (4.8). To this end, we employ two loops of mathematical induction. The first loop of mathematical induction is as follows. We assume that

\[ \| e_{h,m}^{n-1} \|_{W^{1,q}(\Omega)} \leq h^{\frac{d}{q}+\frac{1}{4}}, \quad n = 1, \ldots, k, \]

and then prove

\[ \| e_{h,m}^k \|_{W^{1,q}(\Omega)} \leq h^{\frac{d}{q}+\frac{1}{4}}. \]

To complete the first loop of mathematical induction, we employ a second loop of mathematical induction: we assume that the following estimates hold for some \( 1 \leq m \leq \ell \)

\[ \| e_{h,m-1}^{n-1} \|_{W^{1,q}(\Omega)} \leq h^{\frac{1}{2} - \frac{d}{q}}, \quad i = 1, \ldots, m, \quad n = 1, \ldots, k, \]

and then prove that

\[ \| e_{h,m}^n \|_{W^{1,q}(\Omega)} \leq h^{\frac{1}{2} - \frac{d}{q}}, \quad n = 1, \ldots, k. \]

Note that (4.10) implies (4.12) with \( i = 1 \) for sufficiently small \( h \). This is a consequence of

\[ \| e_{h,0}^n \|_{W^{1,q}(\Omega)} = \| e_{h,0}^{n-1} \|_{W^{1,q}(\Omega)} \]

\[ \leq \| \theta_{h,0}^{n-1} \|_{W^{1,q}(\Omega)} + \| \rho_{h,0}^{n-1} \|_{W^{1,q}(\Omega)} \]

\[ \leq c h^{\frac{3}{4}} \left( \| e_{h,0}^{n-1} \|_{W^{1,q}(\Omega)} + \| \rho_{h,0}^{n-1} \|_{W^{1,q}(\Omega)} \right) + c h^{1 - \frac{d}{q}} \| u_{h,0}^{n-1} \|_{C^{1,1} - \frac{d}{q}}(\Omega), \] (inverse inequality is used)

\[ \leq c h^{\frac{3}{4}} + c h^{1 - \frac{d}{q}} \| u_{h,0}^{n-1} \|_{W^{2,q}(\Omega)}, \]

where we have used (4.10), (4.3) and \( W^{2,q}(\Omega) \hookrightarrow C^{1,1 - \frac{d}{q}}(\Omega) \) in the last inequality. The boundedness of \( \| u_{h,m}^{n-1} \|_{W^{2,q}(\Omega)} \) is proved in (3.5).

We shall keep the generic constant \( c \) below to be independent of \( n, m \) and \( k \).

### 4.2 Error estimate in \( W^{1,q}(\Omega) \)

In (3.13)–(3.14) we have proved the boundedness of \( \| u_{h,m}^n \|_{W^{1,q}(\Omega)} \) and \( \| u_{h,m}^n \|_{W^{2,q}(\Omega)} \) for all \( 0 \leq m \leq \ell \), which imply the uniform ellipticity (3.15). Under assumption (4.12), we have the following estimates (due to the local Lipschitz continuity of \( f \)):

\[ \| Df(\nabla u_{h,m-1}^n) - Df(\nabla u_{m-1}^n) \|_{W^{-1,q}(\Omega)} \leq c \| e_{h,m-1}^n \|, \quad n = 1, \ldots, k, \]

\[ \| Df((1-s)\nabla u_{h,m}^n + s\nabla u_{m-1}^n) - Df(\nabla u_{m-1}^n) \|_{W^{-1,q}(\Omega)} \leq c \| e_{h,m-1}^n \|, \quad n = 1, \ldots, k. \]

Moreover, we need the following estimates in the subsequent error estimation:

\[ \left\| \frac{1}{\tau}(1 - A_{h,m-1}^n) \right\|_{W^{-1,q}(\Omega)} \leq c \| w_h \|_{W^{-1,q}(\Omega)} \quad \forall w_h \in S_h, \quad \forall 1 < s < \infty, \]

\[ \left\| \frac{1}{\tau}(1 - A_{h,m}^n) \right\|_{W^{-1,q}(\Omega)} \leq c \| w_h \|_{W^{-1,q}(\Omega)} \quad \forall w_h \in S_h, \quad \forall 1 < s < \infty. \]

These estimates are direct consequences of (2.15): indeed, choosing \( k = 1 \) and \( g^n = w_h \) in (2.15) yields

\[ \tau^{-\frac{1}{4}} \| \phi_h^1 / \tau \|_{W^{-1,q}(\Omega)} \leq \tau^{\frac{1}{4}} \| w_h \|_{W^{-1,q}(\Omega)} \]

and

\[ \tau^{-\frac{1}{4}} \| \phi_h^1 \|_{W^{1,q}(\Omega)} \leq c \| w_h \|_{W^{-1,q}(\Omega)}. \]
for $\phi_{h} = \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h}$, which imply (4.16) and (4.17), respectively.

Substituting $u_{h} = \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h}$ into (4.8) yields, for $1 \leq n \leq k$,

$$
(\theta_{h,m}, w_{h}) = \sum_{j=3}^{7} F_{j} \left( \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h} \right) \quad \forall w_{h} \in S_{h},
$$

where

$$
\left| F_{j} \left( \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h} \right) \right|
$$

$$
\leq \left| (Df(\nabla u_{h,m-1}^{n}) - Df(\nabla u_{h,m-1}^{n})) \nabla (\nabla u_{h,m-1}^{n} - \nabla u_{h,m-1}^{n}), \nabla \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h} \right| 
$$

$$
\leq \| Df(\nabla u_{h,m-1}^{n}) - Df(\nabla u_{h,m-1}^{n}) \|_{L^{\infty}(\Omega)} \| \nabla (\nabla u_{h,m-1}^{n} - \nabla u_{h,m-1}^{n}) \|_{L^{2}(\Omega)} \left\| \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h} \right\|_{W^{1,\infty}(\Omega)}
$$

$$
\leq c \| \nabla e_{h,m-1}^{n} \|_{L^{2}(\Omega)} \left\| \nabla e_{h,m-1}^{n} \|_{L^{2}(\Omega)} + \| \nabla e_{h,m-1}^{n} \|_{L^{2}(\Omega)} \right\|_{W^{1,\infty}(\Omega)} \left\| w_{h} \right\|_{W^{1,\infty}(\Omega)} \quad [\text{(4.17) is used}]
$$

$$
\leq c h^{1} \frac{\| \nabla e_{h,m-1}^{n} \|_{L^{2}(\Omega)} + \| \nabla e_{h,m-1}^{n} \|_{L^{2}(\Omega)} \right\|_{W^{1,\infty}(\Omega)} \left\| w_{h} \right\|_{W^{1,\infty}(\Omega)} \tag{4.19}
$$

and, similarly,

$$
\left| F_{j} \left( \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h} \right) \right|
$$

$$
\leq \left| (Df(\nabla u_{h,m-1}^{n}) - Df(\nabla u_{h,m-1}^{n})) \nabla (\nabla u_{h,m-1}^{n} - \nabla u_{h,m-1}^{n}), \nabla \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h} \right| 
$$

$$
\leq c \| \nabla e_{h,m-1}^{n} \|_{L^{2}(\Omega)} \left\| \nabla e_{h,m-1}^{n} \|_{L^{2}(\Omega)} + \| \nabla e_{h,m-1}^{n} \|_{L^{2}(\Omega)} \right\|_{W^{1,\infty}(\Omega)} 
$$

$$
\leq c \| \nabla e_{h,m-1}^{n} \|_{L^{2}(\Omega)} \| \nabla (\nabla u_{m-1}^{n} - \nabla u_{m-1}^{n}) \|_{L^{2}(\Omega)} \| w_{h} \|_{W^{1,\infty}(\Omega)} \quad [\text{(3.5) is used to estimate } \| \nabla (\nabla u_{m-1}^{n} - \nabla u_{m-1}^{n}) \|_{L^{2}(\Omega)}]
\tag{4.20}
$$

and

$$
\left| F_{j} \left( \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h} \right) \right|
$$

$$
\leq \| (Df(\nabla u_{h,m-1}^{n} - \nabla u_{h,m-1}^{n})) \|_{L^{\infty}(\Omega)} \left\| \nabla e_{h,m-1}^{n} \|_{L^{2}(\Omega)} \right\|_{W^{1,\infty}(\Omega)} \left\| \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h} \right\|_{W^{1,\infty}(\Omega)}
$$

$$
\leq c \| \nabla e_{h,m-1}^{n} \|_{L^{2}(\Omega)} \left\| \nabla e_{h,m-1}^{n} \|_{L^{2}(\Omega)} \right\|_{W^{1,\infty}(\Omega)} \left\| w_{h} \right\|_{W^{1,\infty}(\Omega)} \quad [\text{(4.17) is used}]
$$

$$
\leq c \| \nabla e_{h,m-1}^{n} \|_{L^{2}(\Omega)} \left\| \nabla e_{h,m-1}^{n} \|_{L^{2}(\Omega)} \right\|_{W^{1,\infty}(\Omega)} \left\| w_{h} \right\|_{W^{1,\infty}(\Omega)} \quad [\text{(4.12) is used}]
\tag{4.21}
$$

Moreover, we have

$$
\left| F_{j} \left( \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h} \right) \right|
$$

$$
= \left| \nabla e_{h,\ell}^{n-1} \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h} \right|
$$

$$
\leq \| \nabla e_{h,\ell}^{n-1} \|_{W^{1,\infty}(\Omega)} \left\| \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h} \right\|_{W^{1,\infty}(\Omega)}
$$

$$
\leq c \| \nabla e_{h,\ell}^{n-1} \|_{W^{1,\infty}(\Omega)} \| w_{h} \|_{W^{1,\infty}(\Omega)} \quad [\text{(4.16) is used}]
$$

and

$$
\left| F_{j} \left( \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h} \right) \right|
$$

$$
= \left| (Df(\nabla u_{h,m-1}^{n})) \nabla p_{h,m}^{n} \nabla \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h} \right|
$$

$$
\leq c \| p_{h,m}^{n} \|_{W^{1,\infty}(\Omega)} \left\| \left( \frac{1}{r} - A_{h,m-1}^{n} \right)^{-1} w_{h} \right\|_{W^{1,\infty}(\Omega)}
$$

$$
\leq c \| p_{h,m}^{n} \|_{W^{1,\infty}(\Omega)} \| w_{h} \|_{W^{1,\infty}(\Omega)} \quad [\text{(4.17) is used}]
$$

$$
\leq c \| p_{h,m}^{n} \|_{W^{1,\infty}(\Omega)} \| w_{h} \|_{W^{1,\infty}(\Omega)} \quad [\text{(4.3) is used}]
\tag{4.22}
$$
In particular, for an arbitrary function \( w \in W^{-1,q'}(\Omega) \), by choosing \( w_h = P_h w \) and using the stability of the \( L^2 \) projection on \( W^{-1,q'}(\Omega) \), we have

\[
\|(\theta^n_m) - (\theta^n_m, P_h w)\| \leq \|c(\tau + h^{\frac{1}{2} - \frac{d}{2}})\|e^n_{h,m}||_{W^{1,q}(\Omega)} + c\|e^n_{h,m-1}||_{W^{1,q}(\Omega)} + c\|e^n_{h,\ell}||_{W^{1,q}(\Omega)} + c h \|w\|_{W^{-1,q'}(\Omega)},
\]

which implies (via the duality argument)

\[
\theta^n_m||_{W^{1,q}(\Omega)} \leq \|\theta^n_{h,m}||_{W^{1,q}(\Omega)} + \|\theta^n_{h,m-1}||_{W^{1,q}(\Omega)} + \|\theta^n_{h,\ell}||_{W^{1,q}(\Omega)} + c h,
\]

whence

\[
\|e^n_{h,m}||_{W^{1,q}(\Omega)} \leq \|\theta^n_{h,m}||_{W^{1,q}(\Omega)} + \|\theta^n_{h,m-1}||_{W^{1,q}(\Omega)} + \|\theta^n_{h,\ell}||_{W^{1,q}(\Omega)} + c h
\] (4.24)

For sufficiently small \( \tau \) and \( h \), the first term on the right-hand side of (4.24) can be absorbed by the left-hand side. Thus, estimate (4.24) reduces to

\[
\|e^n_{h,m}||_{W^{1,q}(\Omega)} \leq \|\theta^n_{h,m}||_{W^{1,q}(\Omega)} + \|\theta^n_{h,m-1}||_{W^{1,q}(\Omega)} + \|\theta^n_{h,\ell}||_{W^{1,q}(\Omega)} + c h.
\] (4.25)

Iterating estimate (4.25) yields

\[
\|e^n_{h,m}||_{W^{1,q}(\Omega)} \leq \|\theta^n_{h,m}||_{W^{1,q}(\Omega)} + \|\theta^n_{h,m-1}||_{W^{1,q}(\Omega)} + \|\theta^n_{h,\ell}||_{W^{1,q}(\Omega)} + c h
\]

where the last inequality requires \( \tau \) and \( h \) to be sufficiently small (independent of \( n, m, k \) and \( \ell \)), and we have used the identity \( e^n_{h,0} = e^n_{h,\ell} \). The constant \( c \) is independent of \( m \) and \( \ell \). Substituting the induction assumption (4.10) into (4.26) yields

\[
\|e^n_{h,m}||_{W^{1,q}(\Omega)} \leq c h^{\frac{d}{2} + \frac{1}{2}},
\] (4.27)

and the triangle inequality implies

\[
\|\theta^n_{h,m}||_{W^{1,q}(\Omega)} \leq \|\theta^n_{h,m}||_{W^{1,q}(\Omega)} + \|\theta^n_{h,m}||_{W^{1,q}(\Omega)} \leq c h^{\frac{d}{2} + \frac{1}{2}}.
\] (4.28)

The inverse inequality of the finite element space gives

\[
\|e^n_{h,m}||_{W^{1,q}(\Omega)} \leq \|\theta^n_{h,m}||_{W^{1,q}(\Omega)} + \|\theta^n_{h,m}||_{W^{1,q}(\Omega)} \leq c h^{\frac{d}{2} + \frac{1}{2}}
\]

For sufficiently small \( h \), this estimate implies (4.13). This completes the second loop of mathematical induction. Thus, (4.13) holds for \( m = \ell \), i.e.,

\[
\|e^n_{h,\ell}||_{W^{1,q}(\Omega)} \leq h^{\frac{d}{2} + \frac{1}{2}}, \quad n = 1, \ldots, k.
\] (4.29)
Substituting $m = \ell$ into (4.8), we rewrite the resulting equation as
\[
 (\delta_{t,\ell} \theta_{h,\ell}^n - A_{h,\ell}^n \theta_{h,\ell}^n, \upsilon_h) = I^n_j (\upsilon_h) + I^n_j (\upsilon_h) + I^n_j (\upsilon_h) \quad \forall \upsilon_h \in S_h,
\]
(4.30)
where $I^n_j (\upsilon_h)$ from (4.8) has been absorbed by the left-hand side of (4.30). Note that $I^n_j, j = 3, 4, 6, 7$, can be identified with linear functionals on $W_0^{1,q} (\Omega)$. Then
\[
 \| I^n_j \|_{W^{-1,q} (\Omega)} \leq \| Df (\nabla u_{h,\ell-1}) - Df (\nabla u_{h,\ell}) \|_{L^2 (\Omega)} \| \nabla (e^n_{h,\ell} - e^n_{h,\ell-1}) \|_{L^2 (\Omega)}
\]
\[
 \leq c \| \nabla e^n_{h,\ell-1} \|_{L^2 (\Omega)} (\| \nabla e^n_{h,\ell} \|_{L^q (\Omega)} + \| \nabla e^n_{h,\ell-1} \|_{L^q (\Omega)})
\]
\[
 \leq c h^{\frac{1}{2} - \frac{d}{2q}} (\| \nabla e^n_{h,\ell} \|_{L^q (\Omega)} + \| \nabla e^n_{h,\ell-1} \|_{L^q (\Omega)})
\]
\[(4.12) \text{ is used}.
\]
This estimate agrees with the estimate of $I^n_j$ in (4.19). Similarly, from the estimates of $I^n_j, I^n_j$ and $I^n_j$ in (4.20), (4.23) and (4.21) we see that
\[
 \| I^n_j \|_{W^{-1,q} (\Omega)} + \| I^n_j \|_{W^{-1,q} (\Omega)} + \| I^n_j \|_{W^{-1,q} (\Omega)} \leq c (\tau + h^{\frac{1}{2} - \frac{d}{2q}}) \| \nabla e^n_{h,\ell-1} \|_{L^q (\Omega)} + ch.
\]
By denoting $\alpha (\cdot, t) = Df (\nabla u_{h,\ell-1})$ and constructing $\alpha (\cdot, t)$ as the piecewise linear interpolant of $\alpha (\cdot, t_n), n = 0, 1, \ldots, N$, the resulting matrix $\alpha (\cdot, t)$ satisfies the conditions of Theorem 2.2 in view of (3.5). Since $\theta_{h,\ell}^n = 0$, applying (2.15) of Theorem 2.2 to (4.30) yields, for $1 < p < \infty$,
\[
 \| (\delta_{t,\ell} \theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))} + \| (\theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))}
\]
\[
 \leq c \| (I^n_j)^k \|_{L^p (W^{1,q} (\Omega))} + \| (I^n_j)^k \|_{L^p (W^{1,q} (\Omega))} + c \| (I^n_j)^k \|_{L^p (W^{1,q} (\Omega))}
\]
\[
 \leq c (\tau + h^{\frac{1}{2} - \frac{d}{2q}}) \| (e^n_{h,\ell-1})^k \|_{L^p (W^{1,q} (\Omega))} + \| (e^n_{h,\ell-1})^k \|_{L^p (W^{1,q} (\Omega))} + ch
\]
\[
 \leq c (\tau + h^{\frac{1}{2} - \frac{d}{2q}}) \| (\theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))} + \| (\theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))} + ch,
\]
(4.31)
where the last inequality is due to
\[
 \| (e^n_{h,\ell-1})^k \|_{L^p (W^{1,q} (\Omega))} \leq \| (\theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))} + \| (\theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))}
\]
\[
 \leq \| (\theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))} + ch.
\]
For sufficiently small $\tau$ and $h$, estimate (4.31) reduces to
\[
 \| (\delta_{t,\ell} \theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))} + \| (\theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))}
\]
\[
 \leq c (\tau + h^{\frac{1}{2} - \frac{d}{2q}}) \| (\theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))} + ch.
\]
(4.32)
In view of the relation $\theta_{h,m} = e_{h,m} - \rho_{h,m}$, setting $m = \ell - 1$ in (4.26) we have
\[
 \| \theta_{h,\ell-1}^n \|_{W^{1,q} (\Omega)} \leq c \| \theta_{h,\ell-1}^n \|_{W^{1,q} (\Omega)} + ch.
\]
Substituting this estimate into (4.32), we obtain
\[
 \| (\delta_{t,\ell} \theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))} + \| (\theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))}
\]
\[
 \leq c (\tau + h^{\frac{1}{2} - \frac{d}{2q}}) \| (\theta_{h,\ell-1}^n)^k \|_{L^p (W^{1,q} (\Omega))} + ch.
\]
(4.33)
For sufficiently small $\tau$ and $h$, the first term on the right-hand side of (4.33) can be absorbed by the left-hand side (since $\rho_{h,\ell}^n = 0$). This yields
\[
 \| (\delta_{t,\ell} \theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))} + \| (\theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))} \leq ch \quad \forall 1 < p < \infty.
\]
(4.34)
The inverse inequality of the finite element space gives
\[
 \| (\delta_{t,\ell} \theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))} \leq ch^{-1} \| (\theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))} \leq ch^{-1}.
\]
(4.35)
Then, the Sobolev interpolation inequality implies
\[
 \| (\theta_{h,\ell}^n)^k \|_{L^p (W^{1,q} (\Omega))}
16 of 21

\[ \leq c \left( \| \theta_{h,k}^n \|_{L^2(W^{1,q}(\Omega))} \right)^{1-\frac{1}{p}} \left( \| \delta \theta_{h,k}^n \|_{L^2(W^{1,q}(\Omega))} \right)^{\frac{1}{p}} \] (see Appendix C)

\[ \leq (ch)^{1-\frac{1}{p}} \left( ch^{-1} \right)^{\frac{1}{p}} \leq ch^{1-\frac{1}{p}} \quad \forall 1 < p < \infty. \] (4.36)

Thus (by the triangle inequality)

\[ \left\| (e_{h,k}^n)^k \right\|_{L^2(W^{1,q}(\Omega))} \leq \left\| (\theta_{h,k}^n)^k \right\|_{L^2(W^{1,q}(\Omega))} + \left\| (\rho_{h,k}^n)^k \right\|_{L^2(W^{1,q}(\Omega))} \]

\[ \leq ch^{1-\frac{1}{p}} \quad \forall 1 < p < \infty. \] (4.37)

For sufficiently small \( \tau \) and \( h \), estimate (4.37) implies

\[ \left\| (e_{h,k}^n)^k \right\|_{L^2(W^{1,q}(\Omega))} \leq h^{\frac{1}{2} + \delta}. \] (4.38)

This proves (4.11) and thus completes the first loop of mathematical induction.

To conclude, (4.37) implies (4.1).

### 4.3 Error estimate in \( L^2(\Omega) \)

Note that (4.30) can be rewritten as

\[ (\delta \theta_{h,k}^n - A_{h,k} e_{h,k}^n, \nu_b) = F_h^n (\nu_b) + F_h^n (\nu_b) + F_h^n (\nu_b) \quad \forall \nu_b \in S_h, \] (4.39)

where \( F_h^n (\nu_b) \) in (4.30) has been absorbed into the left-hand side. Let \( \eta_b^n \in S_h \) be the solution of the finite element equation

\[ (\delta \eta_b^n - A_{h,k} e_{h,k}^n, \nu_b) = F_h^n (\nu_b) + F_h^n (\nu_b) + F_h^n (\nu_b) \quad \forall \nu_b \in S_h, \] (4.40)

with zero initial value \( \eta_b^0 = 0 \). Then \( \eta_b^n \) satisfies the following estimate (cf. (2.15) of Theorem 2.2)

\[ \left\| (\delta \eta_b^n)^N \right\|_{L^2(W^{1,q/2}(\Omega))} + \left\| (\eta_b^n)^N \right\|_{L^2(W^{1,q/2}(\Omega))} \]

\[ \leq c \left( \| F_h^n \| + \| F_h^n \| \right) \left\| (\eta_b^n)^N \right\|_{L^2(W^{1,q/2}(\Omega))} \]

\[ \leq c \left( \| \nabla e_{h,k}^n \|_{L^2(\Omega)} + \left\| (e_{h,k}^n - \theta_{h,k}^n)^N \right\|_{L^2(L^q(\Omega))} \right) \]

\[ + c \left( \| e_{h,k}^n \|_{L^2(\Omega)} + \| \nabla (u_d^n - u_t^{n-1}) \|_{L^2(L^q(\Omega))} \right) \]

\[ + c \left( \| e_{h,k}^n \|_{L^2(\Omega)} + \| \nabla (u_d^n - u_t^{n-1}) \|_{L^2(\Omega)} \right) \]

\[ \leq c \left( \| \nabla e_{h,k}^n \|_{L^2(\Omega)} \right)^2 \| e_{h,k}^n \|_{L^2(\Omega)} + c \left( \| \nabla e_{h,k}^n \|_{L^2(\Omega)} \right)^2 \| e_{h,k}^n \|_{L^2(\Omega)} \]

where we have used the estimate \( \| \nabla (u_d^n - u_t^{n-1}) \|_{L^2(\Omega)} \leq c \tau \), which is a consequence of (3.5). Setting \( m = \ell - 1 \) in (4.26) yields

\[ \left\| e_{h,k}^{\ell-1} \right\|_{W^{1,q}(\Omega)} \leq c \left\| e_{h,k}^{\ell-1} \right\|_{W^{1,q}(\Omega)} + ch. \]

The last two estimates together imply

\[ \left\| (\delta \eta_b^n)^N \right\|_{L^2(W^{1,q/2}(\Omega))} + \left\| (\eta_b^n)^N \right\|_{L^2(W^{1,q/2}(\Omega))} \]

\[ \leq c \left( \| \nabla e_{h,k}^n \|_{L^2(\Omega)} + \| \nabla (u_d^n - u_t^{n-1}) \|_{L^2(\Omega)} \right) \]

\[ \leq c \left( \tau + h^{1-\frac{1}{p}} \right) \quad \forall 1 < p < \infty, \]

where the last inequality is due to (4.37). The inverse inequality of the finite element space yields

\[ \left\| (\delta \eta_b^n)^N \right\|_{L^2(W^{1,q/2}(\Omega))} \leq c \left( \tau + h^{1-\frac{1}{p}} \right) \]

Since \( \eta_b^0 = 0 \), the Sobolev interpolation inequality implies

\[ \left\| (\eta_b^n)^N \right\|_{L^2(W^{1,q/2}(\Omega))} \leq c \left( \| \theta_{h,k}^n \|_{L^2(W^{1,q/2}(\Omega))} \right)^{1-\frac{1}{p}} \left( \| \delta \theta_{h,k}^n \|_{L^2(W^{1,q/2}(\Omega))} \right)^{\frac{1}{p}} \]

\[ \leq ch^{1-\frac{1}{p}} \left( \tau + h^{1-\frac{1}{p}} \right) \quad \forall 1 < p < \infty. \] (4.41)
Since $W^{1,q/2}(\Omega) \hookrightarrow L^q(\Omega)$ for $q > d$, it follows that
\[
\|(\eta_h^n)_{n=1}^{N}\|_{L^q(\Omega)} \leq c \|h^{1-\frac{1}{q}}(\tau + h^{1-\frac{1}{q}}) \leq c(\tau^2 + h^{2-\frac{3}{q}}) \quad \forall 1 < p < \infty.
\] (4.42)

Subtracting (4.40) from (4.39), we have
\[
\{(\delta_t(u_h^n - u_h^0), \nu_h) + (Df(\nabla u_h^n - \nabla u_h^0), \nabla \nu_h) = 0 \quad \forall \nu_h \in S_h,
\]
whence $u_h^n - u_h^0$ and $u_h^0$ play the same roles as $\phi_h^n$ and $\phi$ in Theorem 2.2, respectively. Thus, (2.18) implies
\[
\|(u_h^n - u_h^0)_{n=1}^{N}\|_{L^p(\Omega)} \leq c h^2.
\]
and therefore
\[
\|(u_h^n - u_h^0 - P_h u_h^0)_{n=1}^{N}\|_{L^p(\Omega)} \leq \|(u_h^n - u_h^0)_{n=1}^{N}\|_{L^p(\Omega)} + \|(u_h^n - P_h u_h^0)_{n=1}^{N}\|_{L^p(\Omega)} \leq c h^2.
\] (4.44)

Then (4.43) implies
\[
\|(\delta_t(u_h^n - u_h^0 - P_h u_h^0), \nu_h)\| \leq \|(Df(\nabla u_h^n - \nabla u_h^0 - \nabla u_h^0), \nabla \nu_h)\| + \|(Df(\nabla u_h^n - \nabla u_h^0), \nabla \nu_h)\|
\leq c h^{-2} \|u_h^n - u_h^0 - P_h u_h^0\|_{L^p(\Omega)} \|\nu_h\|_{L^q(\Omega)} + \|u_h^n - P_h u_h^0\|_{L^p(\Omega)} \|\nu_h\|_{L^q(\Omega)}
\leq c h^{-2} \|u_h^n - u_h^0 - P_h u_h^0\|_{L^p(\Omega)} \|\nu_h\|_{L^q(\Omega)} + \|u_h^n - P_h u_h^0\|_{L^p(\Omega)} + c \|\nu_h\|_{L^q(\Omega)},
\]
where we have used (3.5) with $m = \ell$. By using the duality argument and (4.44), we obtain
\[
\|(\delta_t(u_h^n - u_h^0 - P_h u_h^0))_{n=1}^{N}\|_{L^p(\Omega)} \leq c h^{-2} \|(u_h^n - u_h^0 - P_h u_h^0)_{n=1}^{N}\|_{L^p(\Omega)} + c \leq c.
\] (4.45)

Since $u_h^0 - \eta_h^0 - P_h u_h^0 = 0$, the Sobolev interpolation inequality gives
\[
\|(u_h^n - \eta_h^n - P_h u_h^n)_{n=1}^{N}\|_{L^p(\Omega)} \leq c \|(u_h^n - \eta_h^n - P_h u_h^n)_{n=1}^{N}\|_{L^p(\Omega)} \|(\delta_t(u_h^n - \eta_h^n - P_h u_h^n))_{n=1}^{N}\|_{L^p(\Omega)} \leq c \|h^{1-\frac{1}{q}}\| c \|h^{2-\frac{3}{q}}\| \quad \forall 1 < p < \infty.
\] (4.46)

Estimates (4.42) and (4.46) yield
\[
\|(u_h^n - u_h^0)_{n=1}^{N}\|_{L^p(\Omega)} \leq c \|h^{1-\frac{1}{q}}\| c \|h^{2-\frac{3}{q}}\| \quad \forall 1 < p < \infty,
\] (4.47)
which implies (4.2). This completes the proof of Proposition 4.1.

5. Numerical test

To support our theoretical analysis, we present a numerical example by solving the initial and boundary value problem
\[
\begin{cases}
\partial_t u = \nabla \cdot \left( \frac{\nabla u}{\sqrt{1 + |\nabla u|^2}} \right) + g \quad \text{in} \quad \Omega \times (0,T], \\
u = 0 \quad \text{on} \quad \partial \Omega \times (0,T], \\
u(\cdot,0) = u_0 \quad \text{in} \quad \Omega,
\end{cases}
\]
\[
\text{with the proposed method (2.9) in the domain } \Omega = [0,1] \times [0,1] \text{ up to time } T = 1, \text{ where the function } g \text{ and initial data } u_0 \text{ are chosen corresponding to the exact solution}
\]
\[
u(x,y,t) = e^t \sin(\pi x) \sin(\pi y).
\] (5.2)

Inclusion of such a source term $g$ in the equation does not affect the error analysis in this paper. The computations are performed by the software FreeFEM++; see [12]. We perform two Newton iterations at every time level, $\ell = 2$.

The rectangular domain is partitioned into regular right triangles with $M + 1$ uniformly distributed
points on each side. In order to test the convergence with respect to the spatial mesh size, a sufficiently small time stepsize $\tau = 10^{-3}$ is used so that the time discretization error is negligible. The $L^2$ and $H^1$ errors of the numerical solutions are presented in Table 1, where $h := 1/M$. The convergence rates are calculated based on the numerical results of the two finest meshes.

To test the convergence with respect to the time stepsize, a sufficiently small mesh size $h = 2^{-9}$ is used so that the spatial discretization error is negligible. The $L^2$ and $H^1$ errors of the numerical solutions are presented in Table 2.

From Tables 1–2, we see that the numerical results (order of convergence) are consistent with the theoretical results proved in Theorem 2.1.

### Table 1. Errors of the numerical solutions with $\tau = 10^{-3}$.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$|u_h^N - u(t_N)|_{L^2(\Omega)}$</th>
<th>$|u_h^N - u(t_N)|_{H^1(\Omega)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/8</td>
<td>6.7246E-02</td>
<td>1.1894E-01</td>
</tr>
<tr>
<td>1/16</td>
<td>1.7987E-02</td>
<td>5.9414E-01</td>
</tr>
<tr>
<td>1/32</td>
<td>4.2704E-03</td>
<td>2.9663E-01</td>
</tr>
</tbody>
</table>

convergence rate $O(h^{2/3})$  $O(h^{1.01})$

### Table 2. Errors of the numerical solutions with $h = 2^{-9}$.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$|u_h^N - u(t_N)|_{L^2(\Omega)}$</th>
<th>$|u_h^N - u(t_N)|_{H^1(\Omega)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/8</td>
<td>5.0080E-02</td>
<td>2.2621E-01</td>
</tr>
<tr>
<td>1/16</td>
<td>2.4763E-02</td>
<td>1.1302E-01</td>
</tr>
<tr>
<td>1/32</td>
<td>1.2308E-02</td>
<td>5.8460E-02</td>
</tr>
</tbody>
</table>

convergence rate $O(\tau^{1/3})$  $O(\tau^{0.95})$  

### Appendix

A. Justification of assumptions (a1)–(a3)

**Assumption (a1).** We first assume that $a_{ij} = a_{ji}$ are constants satisfying (2.1). In this case, if $\Omega$ is a convex polygon in $\mathbb{R}^2$, then (2.3) holds for some $q > 2$ (cf. [11, Theorem 4.4.3.7], in which the set of $m$ such that $-\frac{2}{q} < \lambda_{j,m} < 0$ is empty for every $j$, provided $q > 2$ is sufficiently close to 2). If $\Omega$ is a polyhedron with all the interior angles of the edges less than $\frac{3}{4}\pi$, then [6, Corollary 3.9 and Section 4.c] implies (2.3) for some $q > 3$. For example, a rectangular parallelepiped in $\mathbb{R}^3$ satisfies the condition (cf. [6, Corollary 3.14]).

In general, if $a_{ij} \in W^{1,q}(\Omega)$ with $q > d$, then the perturbation argument in [6, Section 5] can be applied (details omitted), which reduces the case of variable coefficients to the case of constant coefficients.

**Assumption (a2).** Expression (1.3) for $\nabla_p f(p)$ shows that assumption (a2) is valid for the minimal surface flow problem. Similarly, it also holds for regularized total variation flows. For a general $L^2(\Omega)$ gradient flow, we have $\nabla_p f(p) = \nabla_p^2 F(p)$ for $p \in \mathbb{R}^d$, which is symmetric and positive definite due to the convexity of the function $F : \mathbb{R}^d \to \mathbb{R}$. Thus, all examples mentioned in the introduction section satisfy assumption (a2).

**Assumption (a3).** Suppose that $u_0 \in W^{2,q}(\Omega) \cap W^{1,q}_0(\Omega)$ satisfies the compatibility condition (2.5). Let $D = W^{2,q}(\Omega)$ and $X = L^q(\Omega)$. Then, the conditions of [25, Theorem 8.1.1 (1)] are satisfied, which implies the existence of a local solution $u \in C^1([0,T];D)$ for (1.1). Differentiation of (1.1) with respect...
to \( t \) yields

\[
\begin{cases}
\partial_t (\partial_t u) = \nabla \cdot (Df(\nabla u) \partial_t u) & \text{in } \Omega \times (0,T], \\
\partial_t u = 0 & \text{on } \partial \Omega \times (0,T], \\
\partial_t u(\cdot,0) = \nabla \cdot f(\nabla u_0) \in D & \text{in } \Omega.
\end{cases}
\] (A.1)

Then, [25, Theorem 8.1.1 (2)] implies \( u_t \in C^1([0,T];X) \cap C([0,T];D) \). This shows that \( u \) has the regularity (2.4).

**B. Sketch of proof of Theorem 2.2**

*Proof of (2.16) and (2.17).* Let \( k \) be fixed and define the operator \( A(t_k) : W_0^{1,s}(\Omega) \to W^{-1,s}(\Omega) \) by

\[
A(t_k) \phi^a := \nabla \cdot (\alpha(\cdot,t_k) \nabla \phi^a).
\]

Then, [15, Theorem 3.1] implies the discrete maximal \( L^p \)-regularity (cf. [22, Lemma 2.1]), i.e., the solution of the initial value problem

\[
\begin{cases}
\partial_t \psi - A(t_k) \psi = \chi, & t > 0, \\
\psi|_{t=0} = 0
\end{cases}
\] (B.1)

with the coefficient frozen at \( t = t_k \), satisfies

\[
\begin{align*}
||\partial_t \psi||_{L^p(\Omega;W^{-1,s}(\Omega))} + ||\psi||_{L^p(\Omega;W^{1,s}(\Omega))} & \leq c||\chi||_{L^p(\Omega;W^{1,s}(\Omega))} \quad \forall p, s \in (1, \infty), \\
||\partial_t \psi||_{L^p(\Omega;L^1(\Omega))} + ||A(t_k) \psi||_{L^p(\Omega;L^1(\Omega))} & \leq c||\chi||_{L^p(\Omega;L^1(\Omega))} \quad \forall p, s \in (1, \infty).
\end{align*}
\] (B.2)

(B.3)

Then, [15, Theorem 3.1] implies the discrete maximal \( L^p \)-regularity in \( W^{-1,q}(\Omega) \), i.e., the solution of the autonomous equation

\[
\partial_t \phi^a - A(t_k) \phi^a = g^n
\] (B.4)

with \( \phi^0 = 0 \), satisfies (2.16) and (2.17). Now, we rewrite the nonautonomous equation (2.12) as

\[
\partial_t \phi^a - A(t_k) \phi^a = g^n - (A(t_k) - A(t_n)) \phi^n,
\] (B.5)

and note that the coefficient on the left-hand side is frozen at \( t = t_k \). Let

\[
B_0 = 0 \quad \text{and} \quad B_n = ||(\phi^a)^n||_{L^p(\Omega)} - \sum_{m=1}^n ||\phi^m||_{W^{1,s}(\Omega)} \quad \text{for } 1 \leq n \leq N
\].

Then \( ||\phi^n||_{W^{1,s}(\Omega)} = B_n - B_{n-1} \) and therefore, applying (2.16) yields (raised to power \( p \))

\[
\begin{align*}
&||\partial_t \phi^a||_{L^p(\Omega)}^k + ||A(t_k) \phi^a||_{L^p(\Omega)}^k \\
&\leq c||g^n||_{L^p(\Omega)}^k + c\left|\|A(t_k) - A(t_n)\|_{L^p(\Omega)}^k\right| + c\left|\|\partial_t \phi^a||_{L^p(\Omega)}^k\right|
\end{align*}
\]

\[
= c||g^n||_{L^p(\Omega)}^k + c\sum_{n=1}^k |t_k - t_n|^p ||\phi^n||_{W^{1,s}(\Omega)}^p
\]

\[
= c||g^n||_{L^p(\Omega)}^k + c\sum_{n=1}^k |t_k - t_n|^p (B_n - B_{n-1})
\]

\[
= c||g^n||_{L^p(\Omega)}^k + c\sum_{n=1}^k (|t_k - t_n|^p - |t_k - t_{n+1}|^p)B_n \quad \text{(summation by parts is used)}
\]

\[
= c||g^n||_{L^p(\Omega)}^k + c\sum_{n=1}^k (|t_k - t_n|^p - |t_k - t_{n+1}|^p) ||(\phi^n)^m||_{L^p(\Omega)}^p
\]
\[ \theta(t) = \frac{t_n - t}{\tau} \theta_{n-1} + \frac{t - t_n-1}{\tau} \theta_n \quad \forall t \in [t_n-1, t_n], \]

for \( t \in [0, T], \) of \( \theta_{0, \ell}, \ldots, \theta_{n, \ell} \), which satisfies \( \theta(0) = 0 \). We have

\[ \max_{t \in [0, T] } \| \theta(t) \|_{L^{1,q}(\Omega)} \leq c \| \theta \|_{L^p(0, T; W^{1,q}(\Omega))} \| \partial_t \theta \|_{L^p(0, T; W^{1,q}(\Omega))}, \]

which furthermore implies (C.1) because

\[ \| \theta \|_{L^p(0, T; W^{1,q}(\Omega))} \leq c \| (\theta_n^k)_{n=1} \|_{L^p(0, T; W^{1,q}(\Omega))} \]

and

\[ \| \partial_t \theta \|_{L^p(0, T; W^{1,q}(\Omega))} \leq c \| (\delta_t \theta_n^k)_{n=1} \|_{L^p(0, T; W^{1,q}(\Omega))}. \]
REFERENCES


