IMA Journal of Numerical Analysis (2020) Page 1 of 21
doi:10.1093/imanum/draa016

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, 700 13 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

$$\begin{cases} \partial_t u = \nabla\cdot \boldsymbol{f}(\nabla u) & \text{in } \ \Omega\times(0,T],\\ u=0 & \text{on } \partial\Omega\times(0,T],\\ u(\cdot,0)=u_0 & \text{in } \ \Omega, \end{cases} \tag{1.1}$$

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 $\boldsymbol{f}:\mathbb{R}^d\to\mathbb{R}^d$.

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

$$(\partial_t u,\upsilon)_{L^2(\Omega)} = -E'(u)\upsilon \qquad \text{for all } \upsilon \text{ in a dense and smooth subspace of } H^1_0(\Omega),$$

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

In these applications, the flux function $\boldsymbol{f}$ satisfies the following local ellipticity condition:

$$\nabla_p\boldsymbol{f}(p) \text{ is symmetric and positive definite for every } p\in\mathbb{R}^d. \tag{1.2}$$

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

$$\nabla_p\boldsymbol{f}(p)=\frac{1}{\sqrt{1+|p|^2}}\mathbb{I}_d-\frac{|p|^2}{(1+|p|^2)^{\frac{3}{2}}}\frac{p}{|p|}\otimes\frac{p}{|p|}, \tag{1.3}$$

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

with $\mathbb{I}_d$ the $d \times d$ identity matrix; now, $\nabla_p \boldsymbol{f}(p)$ is symmetric and positive definite for any $p \in \mathbb{R}^d$,

$$\nabla_p \boldsymbol{f}(p)\xi \cdot \xi = \frac{1}{\sqrt{1+|p|^2}}|\xi|^2 - \frac{|p|^2}{(1+|p|^2)^{\frac{3}{2}}}\left(\frac{p}{|p|}\cdot \xi\right)^2 \geqslant \frac{1}{(1+|p|^2)^{\frac{3}{2}}}|\xi|^2 \quad \forall \xi \in \mathbb{R}^d,$$

but the eigenvalues of $\nabla_p \boldsymbol{f}(p)$ 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 $r \geqslant 2$. 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 ℓ 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:

$$u^n_{h,\ell}(x) - u(x,t_n) = [u^n(x) - u(x,t_n)] + [u^n_\ell(x) - u^n(x)] + [u^n_{h,\ell}(x) - u^n_\ell(x)], \tag{1.4}$$

where u^n denotes the time-discrete solution, and u^n_ℓ and $u^n_{h,\ell}$ 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 ℓ 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 $L^q(\Omega)$, 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 Ω is sufficiently regular such that the $W^{2,q}$ elliptic regularity holds for some $d < q \leqslant 6$. In other words, for $g \in L^q(\Omega)$ and $a_{ij} \in W^{1,q}(\Omega)$ such that $a_{ij} = a_{ji}$ and

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

the solution $\upsilon \in H_0^1(\Omega)$ of the boundary value problem for the elliptic equation

$$\begin{cases} \displaystyle\sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}\left(a_{ij}\frac{\partial \upsilon}{\partial x_j}\right) = g & \text{in } \Omega, \\ \upsilon = 0 & \text{on } \partial\Omega, \end{cases} \tag{2.2}$$

satisfies

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

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

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

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

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

Justification of these assumptions can be found in Appendix A. In particular, assumption (a1) holds if Ω is any convex two-dimensional polygon or any three-dimensional polyhedron with interior dihedral angles less than $\frac{3}{4}\pi$ (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 $u_0 \in W^{2,q}(\Omega) \cap W_0^{1,q}(\Omega)$ satisfying the compatibility condition

$$\nabla \cdot \boldsymbol{f}(\nabla u_0) \in W^{2,q}(\Omega) \cap W_0^{1,q}(\Omega), \tag{2.5}$$

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 $u^0 := u_0$, we define a sequence of approximations $u^n \in W^{1,\infty}(\Omega) \cap H_0^1(\Omega)$ to the nodal values $u(t_n) := u(\cdot, t_n)$ of the exact solution, by discretizing (1.1) with the implicit Euler scheme,

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

Let $(S_h)_{0<h<1} \subset H_0^1(\Omega)$ denote a family of finite element spaces of continuous piecewise linear polynomials corresponding to a quasi-uniform triangulation of the domain Ω, with mesh size h. The fully discrete finite element counterpart of the implicit Euler scheme (2.6) is to seek approximations $u_h^n \in S_h$ to $u(t_n)$ such that

$$\left(\frac{u_h^n - u_h^{n-1}}{\tau}, \upsilon_h\right) = -\big(\boldsymbol{f}(\nabla u_h^n), \nabla \upsilon_h\big) \quad \forall \upsilon_h \in S_h, \quad n = 1, \ldots, N, \tag{2.7}$$

where $u_h^0 \in S_h$ is the L^2 projection of the initial value u_0.

The implicit Euler scheme (2.6) and its finite element discretization (2.7) are nonlinear equations that cannot be implemented directly. An efficient way to linearize (2.6), as well as the corresponding fully discrete scheme (2.7), is by Newton's method.

2.2 *Linearization by Newton's method*

Let

$$\mathrm{D}\boldsymbol{f}(p) := \nabla_p \boldsymbol{f}(p), \quad \mathrm{D}^2\boldsymbol{f}(p) := \nabla_p^2 \boldsymbol{f}(p) \quad \text{and} \quad \mathrm{D}^3\boldsymbol{f}(p) := \nabla_p^3 \boldsymbol{f}(p) \quad \text{for } p \in \mathbb{R}^d.$$

Thus $\mathrm{D}^2\boldsymbol{f}(p)$ and $\mathrm{D}^3\boldsymbol{f}(p)$ are 3rd- and 4th-order tensors with components

$$\mathrm{D}^2\boldsymbol{f}(p)_{ijk} = \partial_i\partial_j \boldsymbol{f}_k(p) \quad \text{and} \quad \mathrm{D}^3\boldsymbol{f}(p)_{ijk\ell} = \partial_i\partial_j\partial_k \boldsymbol{f}_\ell(p).$$

Multiplication of the 3rd-order tensor $\mathrm{D}^2\boldsymbol{f}(p)$ with a d-dimensional vector yields a 2nd-order tensor (i.e., matrix) with components

$$(\mathrm{D}^2\boldsymbol{f}(p)\cdot \boldsymbol{\upsilon})_{jk} = \sum_{i=1}^d \upsilon_i \partial_i\partial_j \boldsymbol{f}_k(p).$$

Similarly, multiplication of the 3rd-order tensor $\mathrm{D}^2\boldsymbol{f}(p)$ with a $d\times d$ matrix $M = (M_{ij})$ yields a vector with components

$$(\mathrm{D}^2\boldsymbol{f}(p) : M)_k = \sum_{i,j=1}^d \partial_i\partial_j \boldsymbol{f}_k(p) M_{ij}.$$

Multiplication of the 4rd-order tensor $\mathrm{D}^3\boldsymbol{f}(p)$ with matrices or vectors can be defined similarly.

Let ℓ denote the number of Newton iterations at each time level, and set $u_\ell^0 := u_0$. At the nth time level, by choosing the starting value $u_0^n := u_\ell^{n-1}$, Newton's method for the semidiscrete scheme (2.6) seeks $u_m^n \in W^{1,\infty}(\Omega)\cap H_0^1(\Omega), m = 1,\ldots,\ell$, to be the solutions of the following linear equation:

$$\frac{u_m^n - u_\ell^{n-1}}{\tau} = \nabla\cdot \boldsymbol{f}(\nabla u_{m-1}^n) + \nabla\cdot\Big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla(u_m^n - u_{m-1}^n)\Big). \tag{2.8}$$

Similarly, at the nth 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

$$\begin{aligned}&\left(\frac{u_{h,m}^n - u_{h,\ell}^{n-1}}{\tau}, \upsilon_h\right)\\ &= -\big(\boldsymbol{f}(\nabla u_{h,m-1}^n), \nabla\upsilon_h\big) - \big(\mathrm{D}\boldsymbol{f}(\nabla u_{h,m-1}^n)\nabla(u_{h,m}^n - u_{h,m-1}^n), \nabla\upsilon_h\big) \quad \forall \upsilon_h \in S_h,\end{aligned} \tag{2.9}$$

with $u_{h,0}^n := u_{h,\ell}^{n-1}$, and $u_{h,\ell}^0$ 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 $\ell \geqslant 2$. Then, under assumptions (a1)–(a3), there exist positive constants $\tau_\star$ and $h_\star$ such that for $\tau \leqslant \tau_\star$ and $h \leqslant h_\star$ the numerical solutions given by the Newton iterative finite element method (2.9) satisfy the following estimates

$$\max_{1\leqslant n\leqslant N} \|u_{h,\ell}^n - u(t_n)\|_{W^{1,q}(\Omega)} \leqslant c_{\varepsilon,\ell}(\tau + h^{1-\varepsilon}), \tag{2.10}$$

$$\max_{1\leqslant n\leqslant N} \|u_{h,\ell}^n - u(t_n)\|_{L^q(\Omega)} \leqslant c_{\varepsilon,\ell}(\tau + h^{2-\varepsilon}), \tag{2.11}$$

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 τ and n (that may depend on ε, ℓ and T).

The rest of the theoretical part of the article is devoted to the proof of Theorem 2.1. To simplify the notation, for a sequence $(v^n)_{n=0}^k$ with entries in a Banach space X we denote by $\delta_\tau v^n$ the backward difference quotient and by $\|(v^n)_{n=1}^k\|_{L^p(X)}$ the discrete $\ell^p(X)$ norm,

$$\delta_\tau v^n := \frac{v^n - v^{n-1}}{\tau} \qquad \text{and} \qquad \|(v^n)_{n=1}^k\|_{L^p(X)} := \begin{cases} \left(\sum_{n=1}^k \tau \|v^n\|_X^p \right)^{\frac{1}{p}} & \text{if } p \in [1,\infty), \\ \max_{1 \leqslant n \leqslant k} \|v^n\|_X & \text{if } p = \infty. \end{cases}$$

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,

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

and let $\mathscr{A}(x,t)$ denote the symmetric $d \times d$ matrix with entries $a_{ij}(x,t)$. Let $\phi^n \in H_0^1(\Omega)$ and $\phi_h^n \in S_h$, $n = 1,\ldots,N$, be the solutions of

$$\delta_\tau \phi^n - \nabla\cdot(\mathscr{A}(\cdot,t_n)\nabla\phi^n) = g^n \tag{2.12}$$

and

$$\big(\delta_\tau \phi_h^n, v_h\big) + \big(\mathscr{A}(\cdot,t_n)\nabla\phi_h^n, \nabla v_h\big) = (g^n, v_h) \quad \forall\, v_h \in S_h, \tag{2.13}$$

respectively, for some functions $g^n \in L^s(\Omega)$, $s \in (1,\infty)$, $n = 1,\ldots,N$, with starting data $\phi^0 \in H_0^1(\Omega)$ and $\phi_h^0 \in S_h$. Thus

$$\big(\delta_\tau(\phi^n - \phi_h^n), v_h\big) + \big(\mathscr{A}(\cdot,t_n)\nabla(\phi^n - \phi_h^n), \nabla v_h\big) = 0 \quad \forall\, v_h \in S_h. \tag{2.14}$$

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

$$\|(\delta_\tau\phi_h^n)_{n=1}^k\|_{L^p(W^{-1,s}(\Omega))} + \|(\phi_h^n)_{n=1}^k\|_{L^p(W^{1,s}(\Omega))} \leqslant c\|(g^n)_{n=1}^k\|_{L^p(W^{-1,s}(\Omega))} \quad \forall\, s \in (1,\infty), \tag{2.15}$$

$$\|(\delta_\tau\phi^n)_{n=1}^k\|_{L^p(W^{-1,s}(\Omega))} + \|(\phi^n)_{n=1}^k\|_{L^p(W^{1,s}(\Omega))} \leqslant c\|(g^n)_{n=1}^k\|_{L^p(W^{-1,s}(\Omega))} \quad \forall\, s \in (1,\infty), \tag{2.16}$$

$$\|(\delta_\tau\phi^n)_{n=1}^k\|_{L^p(L^s(\Omega))} + \|(\phi^n)_{n=1}^k\|_{L^p(W^{2,s}(\Omega))} \leqslant c\|(g^n)_{n=1}^k\|_{L^p(L^s(\Omega))} \quad \forall\, s \in (1,q], \tag{2.17}$$

and

$$\begin{aligned} \|(\phi^n - \phi_h^n)_{n=1}^k\|_{L^p(L^s(\Omega))} &\leqslant c\|(\phi^n - P_h\phi^n)_{n=1}^k\|_{L^p(L^s(\Omega))} + c\|(\phi^n - R_h^n\phi^n)_{n=1}^k\|_{L^p(L^s(\Omega))} \\ &\leqslant ch^2\|(\phi^n)_{n=1}^k\|_{L^p(W^{2,s}(\Omega))}, \end{aligned} \tag{2.18}$$

where P_h and R_h^n denote the L^2 and Ritz projections onto the finite element space, respectively, with the latter defined by

$$(\mathscr{A}(\cdot,t_n)\nabla(\varphi - R_h^n\varphi), \nabla v_h) = 0 \quad \forall\, v_h \in S_h\ \forall\, \varphi \in H_0^1(\Omega). \tag{2.19}$$

REMARK 2.1 Estimates (2.15) and (2.16) can be viewed as maximal L^p-regularity on the Banach space $W^{-1,s}(\Omega)$ 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)]):

$$\|\varphi - R_h^n\varphi\|_{L^s(\Omega)} + h\|\varphi - R_h^n\varphi\|_{W^{1,s}(\Omega)} \leqslant c\|\varphi\|_{W^{\ell,s}(\Omega)}h^\ell \quad \forall\, \varphi \in W^{\ell,s}(\Omega),\ \ell = 1,2,\ s \in [2,\infty). \tag{2.20}$$

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:

$$\|(u^n - u(t_n))_{n=1}^N\|_{L^\infty(W^{1,\infty})} + \|(u^n - u(t_n))_{n=1}^N\|_{L^p(W^{2,q})} \leqslant c_{p,q}\tau \quad \forall p \in (1,\infty), \tag{3.1}$$

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

$$\max_{1\leqslant n\leqslant N} \left(\|\delta_\tau u^n\|_{W^{1,\infty}} + \|u^n\|_{W^{2,q}}\right) \leqslant c \quad \forall q \in (1,\infty), \tag{3.2}$$

where $\delta_\tau u^n = (u^n - u^{n-1})/\tau$. 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 $\ell \geqslant 2$. Then, under assumptions (a1)–(a3), there exists a positive constant τ_0 such that for $\tau \leqslant \tau_0$ the numerical solutions given by the Newton iterative scheme (2.8) satisfy the following estimates

$$\max_{1\leqslant n\leqslant N} \|u_\ell^n - u^n\|_{H^1(\Omega)} \leqslant c_\ell \tau^{(2\ell+1)/2}, \tag{3.3}$$

$$\max_{1\leqslant n\leqslant N} \left(\|u_\ell^n - u^n\|_{W^{1,\infty}(\Omega)} + \|u_\ell^n - u^n\|_{W^{2,q}(\Omega)}\right) \leqslant c_\ell \tau^{(2\ell-1)/2}, \tag{3.4}$$

$$\max_{1\leqslant n\leqslant N} \left(\tau^{-1}\|u_m^n - u_{m-1}^n\|_{W^{1,\infty}(\Omega)} + \|u_m^n\|_{W^{2,q}(\Omega)}\right) \leqslant c_\ell, \quad m = 1, \ldots, \ell, \tag{3.5}$$

where c_ℓ is a constant independent of τ and n (that may depend on T and ℓ).

Proof. Taylor expanding the term on the right-hand side of the implicit Euler scheme (2.6) about ∇u_{m-1}^n, we see that the implicit Euler approximation u^n satisfies the relation

$$\begin{aligned}\frac{u^n - u^{n-1}}{\tau} &= \nabla\cdot \boldsymbol{f}(\nabla u_{m-1}^n) + \nabla\cdot\Big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla(u^n - u_{m-1}^n)\Big)\\ &+ \nabla\cdot\Big(\int_0^1 \mathrm{D}^2\boldsymbol{f}\Big((1-t)\nabla u_{m-1}^n + t\nabla u^n\Big)(1-t)\mathrm{d}t\,\nabla(u^n - u_{m-1}^n)\cdot\nabla(u^n - u_{m-1}^n)\Big).\end{aligned} \tag{3.6}$$

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

$$\begin{aligned}\frac{e_m^n}{\tau} &= \nabla\cdot\Big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla e_m^n\Big)\\ &- \nabla\cdot\Big(\int_0^1 \mathrm{D}^2\boldsymbol{f}\Big((1-t)\nabla u_{m-1}^n + t\nabla u^n\Big)(1-t)\mathrm{d}t\,\nabla e_{m-1}^n\cdot\nabla e_{m-1}^n\Big) + \frac{e_\ell^{n-1}}{\tau}.\end{aligned} \tag{3.7}$$

With the second order elliptic partial differential operator

$$A_{m-1}^n = \nabla\cdot\big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla\big),$$

relation (3.7) can be rewritten in the form

$$\begin{aligned}A_{m-1}^n e_m^n &= -A_{m-1}^n\Big(\frac{1}{\tau} - A_{m-1}^n\Big)^{-1}\nabla\cdot\Big(\int_0^1 \mathrm{D}^2\boldsymbol{f}\Big((1-t)\nabla u_{m-1}^n + t\nabla u^n\Big)(1-t)\mathrm{d}t\,\nabla e_{m-1}^n\cdot\nabla e_{m-1}^n\Big)\\ &+ \frac{1}{\tau}\Big(\frac{1}{\tau} - A_{m-1}^n\Big)^{-1} A_{m-1}^n e_\ell^{n-1}.\end{aligned} \tag{3.8}$$

We assume (mathematical induction assumption)

$$\|e_\ell^{n-1}\|_{W^{2,q}(\Omega)} \leqslant \tau, \quad n = 1, \ldots, k, \tag{3.9}$$

and prove that

$$\|e_\ell^k\|_{W^{2,q}(\Omega)} \leqslant \tau. \tag{3.10}$$

To this end, we introduce a second loop of mathematical induction, namely, for $1 \leqslant n \leqslant k$ and $1 \leqslant m \leqslant i$

we assume that

$$\|u^n_{m-1}\|_{W^{1,\infty}(\Omega)} \leqslant \|u^n\|_{W^{1,\infty}(\Omega)} + 1, \tag{3.11}$$

$$\|u^n_{m-1}\|_{W^{2,q}(\Omega)} \leqslant \|u^n\|_{W^{2,q}(\Omega)} + 1. \tag{3.12}$$

Then, we prove

$$\|u^n_{i}\|_{W^{1,\infty}(\Omega)} \leqslant \|u^n\|_{W^{1,\infty}(\Omega)} + 1, \tag{3.13}$$

$$\|u^n_{i}\|_{W^{2,q}(\Omega)} \leqslant \|u^n\|_{W^{2,q}(\Omega)} + 1 \tag{3.14}$$

(for the same range of n) to complete the mathematical induction. Since $e^n_0 = e^{n-1}_\ell$ and $W^{2,q}(\Omega) \hookrightarrow W^{1,\infty}(\Omega)$, it follows that (3.9) implies (3.11)–(3.12) for $m = 1$ (with sufficiently small τ). In the sequel, we shall keep the generic constant c independent of n, k, i and ℓ.

Under the induction assumptions (3.9) and (3.11)–(3.12), the coefficient matrix $\mathrm{D}\boldsymbol{f}(\nabla u^n_{m-1})$ satisfies the following estimates:

$$\begin{aligned}
&\|\mathrm{D}\boldsymbol{f}(\nabla u^n_{m-1})\|_{W^{1,q}(\Omega)} \leqslant c,\\
&\lambda^{-1}|\xi|^2 \leqslant \mathrm{D}\boldsymbol{f}(\nabla u^n_{m-1})\xi\cdot\xi \leqslant \lambda|\xi|^2 \quad \forall \xi \in \mathbb{R}^d,
\end{aligned} \tag{3.15}$$

for some positive constants c and λ depending on $\|u^n\|_{W^{2,q}(\Omega)}$ and $\|u^n\|_{W^{1,\infty}(\Omega)}$ (implicit Euler approximations of the exact solution of the PDE). Thus, operator A^n_{m-1} generates a bounded analytic semigroup on $L^q(\Omega)$ ([27, Theorem 3.1]), satisfying the resolvent estimates (cf. [3, Theorem 3.7.11])

$$\left\|A^n_{m-1}\left(\frac{1}{\tau} - A^n_{m-1}\right)^{-1}\right\|_{L^q(\Omega)\to L^q(\Omega)} \leqslant c \quad \text{and} \quad \left\|\frac{1}{\tau}\left(\frac{1}{\tau} - A^n_{m-1}\right)^{-1}\right\|_{L^q(\Omega)\to L^q(\Omega)} \leqslant c.$$

Therefore, (3.8) yields

$$\begin{aligned}
&\|A^n_{m-1}e^n_m\|_{L^q(\Omega)}\\
&\leqslant c\left\|\nabla\cdot\Big(\int_0^1 \mathrm{D}^2\boldsymbol{f}\Big((1-t)\nabla u^n_{m-1} + t\nabla u^n\Big)(1-t)\mathrm{d}t\,\nabla e^n_{m-1}\cdot\nabla e^n_{m-1}\Big)\right\|_{L^q(\Omega)} + c\|A^n_{m-1}e^{n-1}_\ell\|_{L^q(\Omega)}\\
&\leqslant c\left\|\int_0^1 \mathrm{D}^2\boldsymbol{f}\Big((1-t)\nabla u^n_{m-1} + t\nabla u^n\Big)(1-t)\mathrm{d}t : \nabla^2 e^n_{m-1}\cdot\nabla e^n_{m-1}\right\|_{L^q(\Omega)}\\
&\quad + c\left\|\int_0^1 \mathrm{D}^3\boldsymbol{f}\Big((1-t)\nabla u^n_{m-1} + t\nabla u^n\Big) : \Big((1-t)\nabla^2 u^n_{m-1} + t\nabla^2 u^n\Big)(1-t)\mathrm{d}t\cdot\nabla e^n_{m-1}\cdot\nabla e^n_{m-1}\right\|_{L^q(\Omega)}\\
&\quad + c\|A^n_{m-1}e^{n-1}_\ell\|_{L^q(\Omega)}\\
&\leqslant c\|\nabla^2 e^n_{m-1}\|_{L^q(\Omega)}\|\nabla e^n_{m-1}\|_{L^\infty(\Omega)} + (c\|\nabla^2 u^n_{m-1}\|_{L^q(\Omega)} + c)\|\nabla e^n_{m-1}\|^2_{L^\infty(\Omega)} + c\|A^n_{m-1}e^{n-1}_\ell\|_{L^q(\Omega)},
\end{aligned}$$

whence (since $W^{2,q}(\Omega) \hookrightarrow W^{1,\infty}(\Omega)$ for $q > d$)

$$\|A^n_{m-1}e^n_m\|_{L^q(\Omega)} \leqslant c\|e^n_{m-1}\|^2_{W^{2,q}(\Omega)} + c\|e^{n-1}_\ell\|_{W^{2,q}(\Omega)}. \tag{3.16}$$

Under the induction assumptions (3.11)–(3.12), we have the elliptic $W^{2,q}$ regularity (see assumption (a1))

$$\|e^n_m\|_{W^{2,q}(\Omega)} \leqslant c\|A^n_{m-1}e^n_m\|_{L^q(\Omega)}. \tag{3.17}$$

Combining (3.16) and (3.17), we get

$$\|e^n_m\|_{W^{2,q}(\Omega)} \leqslant c\|e^n_{m-1}\|^2_{W^{2,q}(\Omega)} + c\|e^{n-1}_\ell\|_{W^{2,q}(\Omega)}, \tag{3.18}$$

whence, iterating (3.18) with respect to m we obtain

$$\begin{aligned}
\|e^n_m\|_{W^{2,q}(\Omega)} &\leqslant c\|e^n_{m-1}\|^2_{W^{2,q}(\Omega)} + c\|e^{n-1}_\ell\|_{W^{2,q}(\Omega)}\\
&\leqslant c(c\|e^n_{m-2}\|^2_{W^{2,q}(\Omega)} + c\|e^{n-1}_\ell\|_{W^{2,q}(\Omega)})^2 + c\|e^{n-1}_\ell\|_{W^{2,q}(\Omega)}\\
&\leqslant 2c^3\|e^n_{m-2}\|^4_{W^{2,q}(\Omega)} + 2c^3\|e^{n-1}_\ell\|^2_{W^{2,q}(\Omega)} + c\|e^{n-1}_\ell\|_{W^{2,q}(\Omega)}\\
&\leqslant 2^{1+3}c^{1+2+4}\|e^n_{m-3}\|^8_{W^{2,q}(\Omega)} + 2^{1+3}c^{1+2+4}\|e^{n-1}_\ell\|^4_{W^{2,q}(\Omega)} + 2c^3\|e^{n-1}_\ell\|^2_{W^{2,q}(\Omega)} + c\|e^{n-1}_\ell\|_{W^{2,q}(\Omega)}
\end{aligned}$$

$$\leqslant \cdots \leqslant (2c)^{2^m}\|e_0^n\|_{W^{2,q}(\Omega)}^{2^m} + \|e_\ell^{n-1}\|_{W^{2,q}(\Omega)}\left(1+\sum_{j=1}^{m-1}(2c)^{2^j}\|e_\ell^{n-1}\|_{W^{2,q}(\Omega)}^{2^j-1}\right),$$

and thus

$$\|e_m^n\|_{W^{2,q}(\Omega)} \leqslant (2c_\varepsilon\tau^{1-\varepsilon})^{2^m} + \|e_\ell^{n-1}\|_{W^{2,q}(\Omega)}\left(1+c\sum_{j=1}^{m-1}(2c)^{2^j-1}\|e_\ell^{n-1}\|_{W^{2,q}(\Omega)}^{2^j-1}\right), \tag{3.19}$$

where we have used the fact that

$$\begin{aligned}\|e_0^n\|_{W^{2,q}(\Omega)} &= \|u_0^n - u^n\|_{W^{2,q}(\Omega)}\\ &= \|u_\ell^{n-1} - u^n\|_{W^{2,q}(\Omega)}\\ &\leqslant \|u_\ell^{n-1} - u^{n-1}\|_{W^{2,q}(\Omega)} + \|u^{n-1} - u^n\|_{W^{2,q}(\Omega)}\\ &\leqslant c_\varepsilon\tau^{1-\varepsilon},\end{aligned} \tag{3.20}$$

where ε can be arbitrarily small, with a constant c_ε depending on ε. In the last inequality we have used $\|u_\ell^{n-1} - u^{n-1}\|_{W^{2,q}(\Omega)} = \|e_\ell^{n-1}\|_{W^{2,q}(\Omega)} \leqslant \tau$, which is due to the induction assumption (3.9), and

$$\begin{aligned}\|u^{n-1} - u^n\|_{W^{2,q}(\Omega)} &\leqslant \|u^{n-1} - u(t_{n-1})\|_{W^{2,q}(\Omega)} + \|u(t_{n-1}) - u(t_n)\|_{W^{2,q}(\Omega)} + \|u(t_n) - u^n\|_{W^{2,q}(\Omega)}\\ &\leqslant c\tau^{1-\frac{1}{p}} + c\tau + c\tau^{1-\frac{1}{p}} \leqslant c\tau^{1-\frac{1}{p}},\end{aligned}$$

where the last inequality is due to (3.1), in which p can be arbitrarily large and so $\|u^{n-1} - u^n\|_{W^{2,q}(\Omega)} \leqslant c_\varepsilon\tau^{1-\varepsilon}$. The induction assumption (3.9) implies for sufficiently small τ (independent of m and ℓ)

$$\|e_m^n\|_{W^{2,q}(\Omega)} \leqslant (2c_\varepsilon\tau^{1-\varepsilon})^{2^m} + c\|e_\ell^{n-1}\|_{W^{2,q}(\Omega)} \leqslant (2c_\varepsilon\tau^{1-\varepsilon})^{2^m} + c\|e_\ell^{n-1}\|_{W^{2,q}(\Omega)}$$

and thus

$$\|e_m^n\|_{W^{2,q}(\Omega)} \leqslant c\tau, \tag{3.21}$$

which further implies

$$\|e_m^n\|_{W^{1,\infty}(\Omega)} \leqslant c\|e_m^n\|_{W^{2,q}(\Omega)} \leqslant c\tau. \tag{3.22}$$

For sufficiently small τ, 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 $1 \leqslant m \leqslant \ell$.

By the second mathematical induction, we have shown that $\|\nabla u_m^n\|_{L^\infty}$ is uniformly bounded with respect to τ for all $0 \leqslant m \leqslant \ell$ (note that $\|\nabla u_0^n\|_{L^\infty} = \|\nabla u_\ell^{n-1}\|_{L^\infty}$). Now, integrating (3.7) against $-\nabla\cdot\big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla e_m^n\big)$ and using Hölder's inequality, we obtain, for $1 \leqslant m \leqslant \ell$,

$$\begin{aligned}&\frac{1}{2\tau}\big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla e_m^n, \nabla e_m^n\big) - \frac{1}{2\tau}\big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla e_\ell^{n-1}, \nabla e_\ell^{n-1}\big)\\ &\quad + \frac{1}{2}\big\|\nabla\cdot\big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla e_m^n\big)\big\|_{L^2(\Omega)}^2\\ &\leqslant c\left\|\nabla\cdot\left(\int_0^1 \mathrm{D}^2\boldsymbol{f}\Big((1-t)\nabla u_{m-1}^n + t\nabla u^n\Big)(1-t)\mathrm{d}t\,\nabla e_{m-1}^n\cdot\nabla e_{m-1}^n\right)\right\|_{L^2(\Omega)}^2\\ &\leqslant c\left\|\int_0^1 \mathrm{D}^2\boldsymbol{f}\Big((1-t)\nabla u_{m-1}^n + t\nabla u^n\Big)(1-t)\mathrm{d}t\,:\nabla^2 e_{m-1}^n\cdot\nabla e_{m-1}^n\right\|_{L^2(\Omega)}^2\\ &\quad + c\left\|\int_0^1 \mathrm{D}^3\boldsymbol{f}\Big((1-t)\nabla u_{m-1}^n + t\nabla u^n\Big):\Big((1-t)\nabla^2 u_{m-1}^n + t\nabla^2 u^n\Big)(1-t)\mathrm{d}t\;\cdot\nabla e_{m-1}^n\cdot\nabla e_{m-1}^n\right\|_{L^2(\Omega)}^2.\end{aligned} \tag{3.23}$$

Since $\|\nabla u_m^n\|_{L^\infty}$ is uniformly bounded, it follows that

$$\Big|\mathrm{D}^2\boldsymbol{f}\Big((1-t)\nabla u_{m-1}^n + t\nabla u^n\Big)\Big| + \Big|\mathrm{D}^3\boldsymbol{f}\Big((1-t)\nabla u_{m-1}^n + t\nabla u^n\Big)\Big| \leqslant C.$$

Therefore, by using Hölder's inequality,

$$
\begin{aligned}
&\Big\| \int_0^1 \mathrm{D}^2 \boldsymbol{f}\Big((1-t)\nabla u_{m-1}^n + t\nabla u^n\Big)(1-t)\mathrm{d}t \,:\, \nabla^2 e_{m-1}^n \cdot \nabla e_{m-1}^n \Big\|_{L^2(\Omega)}^2 \\
&\quad + c\Big\| \int_0^1 \mathrm{D}^3 \boldsymbol{f}\Big((1-t)\nabla u_{m-1}^n + t\nabla u^n\Big) : \Big((1-t)\nabla^2 u_{m-1}^n + t\nabla^2 u^n\Big)(1-t)\mathrm{d}t \;\cdot \nabla e_{m-1}^n \cdot \nabla e_{m-1}^n \Big\|_{L^2(\Omega)}^2 \\
&\leqslant c\|\nabla^2 e_{m-1}^n\|_{L^q(\Omega)}^2 \|\nabla e_{m-1}^n\|_{L^{\frac{2q}{q-2}}(\Omega)}^2 + c(\|\nabla^2 u_{m-1}^n\|_{L^q(\Omega)}^2 + \|\nabla^2 u^n\|_{L^q(\Omega)}^2)\|\nabla e_{m-1}^n\|_{L^{\frac{2q}{q-2}}(\Omega)}^2 \|\nabla e_{m-1}^n\|_{L^\infty(\Omega)}^2 \\
&\leqslant c\|e_{m-1}^n\|_{W^{2,q}(\Omega)}^2 \|e_{m-1}^n\|_{H^2(\Omega)}^2 + c\|e_{m-1}^n\|_{H^2(\Omega)}^2 \|e_{m-1}^n\|_{W^{1,\infty}(\Omega)}^2 \quad (\text{since } H^2(\Omega) \hookrightarrow W^{1,\frac{2q}{q-2}}(\Omega)) \\
&\leqslant c\tau^2 \|e_{m-1}^n\|_{H^2(\Omega)}^2 \\
&\leqslant c\tau^2 \big\|\nabla \cdot \big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-2}^n)\nabla e_{m-1}^n\big)\big\|_{L^2(\Omega)}^2,
\end{aligned}
\tag{3.24}
$$

with $u_{-1}^n := u_0^n$ 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

$$
\begin{aligned}
&\frac{1}{2\tau}\big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla e_m^n, \nabla e_m^n\big) + \frac{1}{2}\big\|\nabla \cdot \big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla e_m^n\big)\big\|_{L^2(\Omega)}^2 \\
&\leqslant c\tau^2 \big\|\nabla \cdot \big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-2}^n)\nabla e_{m-1}^n\big)\big\|_{L^2(\Omega)}^2 + \frac{1}{2\tau}\big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla e_\ell^{n-1}, \nabla e_\ell^{n-1}\big) \\
&\leqslant c\tau^2 \big\|\nabla \cdot \big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-2}^n)\nabla e_{m-1}^n\big)\big\|_{L^2(\Omega)}^2 + \frac{1}{2\tau}\big((\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n) - \mathrm{D}\boldsymbol{f}(\nabla u_\ell^n))\nabla e_\ell^{n-1}, \nabla e_\ell^{n-1}\big) \\
&\quad + \frac{1}{2\tau}\big(\mathrm{D}\boldsymbol{f}(\nabla u_\ell^n)\nabla e_\ell^{n-1}, \nabla e_\ell^{n-1}\big).
\end{aligned}
$$

Again, since $\|\nabla u_m^n\|_{L^\infty}$ is uniformly bounded for $0 \leqslant m \leqslant \ell$, it follows that

$$
|\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n) - \mathrm{D}\boldsymbol{f}(\nabla u_\ell^n)| \leqslant C
$$

and thus

$$
\begin{aligned}
&\frac{1}{2\tau}\big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla e_m^n, \nabla e_m^n\big) + \frac{1}{2}\big\|\nabla \cdot \big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla e_m^n\big)\big\|_{L^2(\Omega)}^2 \\
&\leqslant c\tau^2 \big\|\nabla \cdot \big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-2}^n)\nabla e_{m-1}^n\big)\big\|_{L^2(\Omega)}^2 + c\|\nabla e_\ell^{n-1}\|_{L^2(\Omega)}^2 + \frac{1}{2\tau}\big(\mathrm{D}\boldsymbol{f}(\nabla u_\ell^n)\nabla e_\ell^{n-1}, \nabla e_\ell^{n-1}\big) \\
&\leqslant c\tau^2 \big\|\nabla \cdot \big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-2}^n)\nabla e_{m-1}^n\big)\big\|_{L^2(\Omega)}^2 + \frac{1}{2\tau}(1 + c\tau)\big(\mathrm{D}\boldsymbol{f}(\nabla u_\ell^n)\nabla e_\ell^{n-1}, \nabla e_\ell^{n-1}\big).
\end{aligned}
$$

Let $Y_m^n = \big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla e_m^n, \nabla e_m^n\big) + \tau\big\|\nabla \cdot \big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla e_m^n\big)\big\|_{L^2(\Omega)}^2$. Then, the last estimate implies

$$
Y_m^n \leqslant c\tau^2 Y_{m-1}^n + (1 + c\tau)\big(\mathrm{D}\boldsymbol{f}(\nabla u_\ell^n)\nabla e_\ell^{n-1}, \nabla e_\ell^{n-1}\big).
$$

Iterations of this inequality yield

$$
\begin{aligned}
Y_m^n &\leqslant c\tau^2 Y_{m-1}^n + (1 + c\tau)\big(\mathrm{D}\boldsymbol{f}(\nabla u_\ell^n)\nabla e_\ell^{n-1}, \nabla e_\ell^{n-1}\big) \\
&\leqslant (c\tau^2)^2 Y_{m-2}^n + (1 + c\tau^2)(1 + c\tau)\big(\mathrm{D}\boldsymbol{f}(\nabla u_\ell^n)\nabla e_\ell^{n-1}, \nabla e_\ell^{n-1}\big) \\
&\leqslant \cdots \\
&\leqslant (c\tau^2)^m Y_0^n + \big[1 + c\tau^2 + \cdots + (c\tau^2)^{m-1}\big](1 + c\tau)\big(\mathrm{D}\boldsymbol{f}(\nabla u_\ell^n)\nabla e_\ell^{n-1}, \nabla e_\ell^{n-1}\big) \\
&\leqslant (c\tau^2)^m(\|\nabla e_0^n\|_{L^2(\Omega)}^2 + \tau\|e_0^n\|_{H^2(\Omega)}^2) + \big[1 + c\tau^2 + \cdots + (c\tau^2)^{m-1}\big](1 + c\tau)\|\nabla e_\ell^{n-1}\|_{L^2(\Omega)}^2 \\
&\leqslant (c\tau^2)^m(\|\nabla e_0^n\|_{L^2(\Omega)}^2 + \tau\|e_0^n\|_{H^2(\Omega)}^2) + \frac{1 + c\tau}{1 - c\tau^2}\|\nabla e_\ell^{n-1}\|_{L^2(\Omega)}^2.
\end{aligned}
$$

When τ is small enough, the last inequality implies

$$
\|\nabla e_m^n\|_{L^2(\Omega)}^2 \leqslant (c\tau^2)^m(\|\nabla e_0^n\|_{L^2(\Omega)}^2 + \tau\|e_0^n\|_{H^2(\Omega)}^2) + (1 + c\tau)\|\nabla e_\ell^{n-1}\|_{L^2(\Omega)}^2.
\tag{3.25}
$$

In particular, setting $m=\ell$ in the last estimate and using (3.20), we obtain

$$\|\nabla e_\ell^n\|_{L^2(\Omega)}^2 \leqslant (c\tau^2)^\ell \tau^2 + (1+c\tau)\|\nabla e_\ell^{n-1}\|_{L^2(\Omega)}^2. \tag{3.26}$$

Iteration of (3.26) gives

$$\begin{aligned}\|\nabla e_\ell^n\|_{L^2(\Omega)}^2 &\leqslant (c\tau^2)^\ell \tau^2 + (1+c\tau)\|\nabla e_\ell^{n-1}\|_{L^2(\Omega)}^2\\ &\leqslant (c\tau^2)^\ell \tau^2\big[1+(1+c\tau)+\cdots+(1+c\tau)^{n-1}\big] + (1+c\tau)^n\|\nabla e_\ell^0\|_{L^2(\Omega)}^2\\ &\leqslant (c\tau^2)^\ell \tau^2 n(1+c\tau)^n + (1+c\tau)^n\|\nabla e_\ell^0\|_{L^2(\Omega)}^2\\ &\leqslant ce^{cT}(c\tau^2)^\ell \tau + ce^{cT}\|\nabla e_\ell^0\|_{L^2(\Omega)}^2;\end{aligned}$$

therefore, since e_ℓ^0 vanishes,

$$\|\nabla e_\ell^n\|_{L^2(\Omega)}^2 \leqslant cc^\ell \tau^{2\ell+1}. \tag{3.27}$$

Note that $H^1(\Omega)\hookrightarrow L^6(\Omega)$ for $d\in\{2,3\}$. From (3.8) we see that, for $q\in(d,6]$,

$$\begin{aligned}\|A_{m-1}^n e_m^n\|_{L^q(\Omega)} &\leqslant c\left\|\nabla\cdot\Big(\int_0^1 \mathrm{D}^2\boldsymbol{f}\Big((1-t)\nabla u_{m-1}^n + t\nabla u^n\Big)(1-t)\mathrm{d}t\,\nabla e_{m-1}^n\cdot\nabla e_{m-1}^n\Big)\right\|_{L^q(\Omega)} + c\left\|\frac{1}{\tau}e_\ell^{n-1}\right\|_{L^q(\Omega)}\\ &\leqslant c\|e_{m-1}^n\|_{W^{2,q}(\Omega)}^2 + cc^\ell\tau^{(2\ell-1)/2};\end{aligned}$$

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))

$$\begin{aligned}\|e_m^n\|_{W^{2,q}(\Omega)} &\leqslant 2^m(c_\varepsilon\tau^{1-\varepsilon})^{2^m} + cc^\ell\tau^{(2\ell-1)/2}\left(1+c\sum_{j=1}^{m-1}2^j\big(cc^\ell\tau^{(2\ell-1)/2}\big)^{2^j-1}\right)\\ &\leqslant (2c_\varepsilon\tau^{1-\varepsilon})^{2^m} + cc^\ell\tau^{(2\ell-1)/2}.\end{aligned} \tag{3.28}$$

For sufficiently small τ, estimate (3.28) implies, for $\ell\geqslant 2$,

$$\|e_m^n\|_{W^{1,\infty}(\Omega)} + \|e_m^n\|_{W^{2,q}(\Omega)} \leqslant c\tau^{\frac{3}{2}}, \quad m=1,\dots,\ell; \tag{3.29}$$

therefore, for sufficiently small τ, we have

$$\|e_\ell^n\|_{W^{2,q}(\Omega)} \leqslant \tau. \tag{3.30}$$

This completes the first loop of mathematical induction. Thus, all these estimates hold for $n=1,\dots,N$ and $m=1,\dots,\ell$. 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 $\ell\geqslant 2$ and assume that the solution of the initial and boundary value problem (1.1) satisfies (2.4). Then, there exists a positive constant τ_0 such that for $\tau\leqslant\tau_0$ the numerical solutions given by Newton's method (2.8) satisfy the following estimate:

$$\max_{1\leqslant n\leqslant N}\|u_\ell^n - u(t_n)\|_{W^{1,\infty}(\Omega)} \leqslant c\tau, \tag{3.31}$$

where c is a constant independent of τ and n (that may depend on T and ℓ).

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 $\ell\geqslant 2$. Then, under assumptions (a1)–(a3), there exist positive constants $\tau_\star$ and

$h_\star$ such that for $\tau \leqslant \tau_\star$ and $h \leqslant h_\star$ the numerical solutions given by the Newton iterative finite element method (2.9) satisfy the following estimates

$$\max_{1\leqslant n\leqslant N} \|u_{h,\ell}^n - u_\ell^n\|_{W^{1,q}(\Omega)} \leqslant c_{\varepsilon,\ell} h^{1-\varepsilon}, \tag{4.1}$$

$$\max_{1\leqslant n\leqslant N} \|u_{h,\ell}^n - u_\ell^n\|_{L^q(\Omega)} \leqslant c_{\varepsilon,\ell} (\tau^2 + h^{2-\varepsilon}), \tag{4.2}$$

where $\varepsilon \in (0,1)$ can be an arbitrarily small constant, and $c_{\varepsilon,\ell}$ is a constant independent of τ and n (that may depend on ε, ℓ 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:

$$\|\varphi - P_h\varphi\|_{W^{k,p}(\Omega)} \leqslant Ch^{m-k}\|\varphi\|_{W^{m,p}(\Omega)} \quad \forall \varphi \in W^{m,p}(\Omega) \cap H_0^1(\Omega), \tag{4.3}$$

for $k = 0,1$, $m = 1,2$, and $1 \leqslant p \leqslant \infty$. Estimate (4.3) is a consequence of [29, Lemma 7.2]. Similarly, the Lagrangian interpolation operator $\Pi_h : C(\overline{\Omega}) \to S_h$ satisfies

$$\|\varphi - \Pi_h\varphi\|_{L^p(\Omega)} + h\|\nabla(\varphi - \Pi_h\varphi)\|_{L^p(\Omega)} \leqslant ch^2\|\varphi\|_{W^{2,p}(\Omega)} \quad \forall \varphi \in W^{2,p}(\Omega). \tag{4.4}$$

Integrating (2.8) against a test function υ_h yields

$$\left(\frac{u_m^n - u_\ell^{n-1}}{\tau}, \upsilon_h\right) = -\big(\boldsymbol{f}(\nabla u_{m-1}^n), \nabla\upsilon_h\big) - \big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla(u_m^n - u_{m-1}^n), \nabla\upsilon_h\big) \quad \forall \upsilon_h \in S_h. \tag{4.5}$$

Let

$$e_{h,m}^n = u_{h,m}^n - u_m^n, \quad \theta_{h,m}^n = u_{h,m}^n - P_h u_m^n \quad \text{and} \quad \rho_{h,m}^n = P_h u_m^n - u_m^n.$$

Obviously, $e_{h,m}^n = \theta_{h,m}^n + \rho_{h,m}^n$; $\|\rho_{h,m}^n\|_{L^q(\Omega)}$ and $\|\rho_{h,m}^n\|_{W^{1,q}(\Omega)}$ can be easily estimated using (3.5) and (4.3). Therefore, we focus on the estimation of $\|\theta_{h,m}^n\|_{L^q(\Omega)}$ and $\|\theta_{h,m}^n\|_{W^{1,q}(\Omega)}$. By our choice $u_{h,\ell}^0 = P_h u_0 = P_h u_\ell^0$, we have $\theta_{h,\ell}^0 = 0$.

Subtracting (4.5) from (2.9) and noting that $(e_{h,m}^n, \upsilon_h) = (\theta_{h,m}^n, \upsilon_h)$ for $\upsilon_h \in S_h$, we obtain

$$\frac{1}{\tau}(\theta_{h,m}^n, \upsilon_h) = I_1^n(\upsilon_h) + I_2^n(\upsilon_h) + I_3^n(\upsilon_h) + I_4^n(\upsilon_h) + I_5^n(\upsilon_h) \quad \forall \upsilon_h \in S_h \tag{4.6}$$

with

$$\begin{aligned}
I_1^n(\upsilon_h) &:= -\big(\boldsymbol{f}(\nabla u_{h,m-1}^n) - \boldsymbol{f}(\nabla u_{m-1}^n), \nabla\upsilon_h\big),\\
I_2^n(\upsilon_h) &:= -\big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla(e_{h,m}^n - e_{h,m-1}^n), \nabla\upsilon_h\big),\\
I_3^n(\upsilon_h) &:= -\big((\mathrm{D}\boldsymbol{f}(\nabla u_{h,m-1}^n) - \mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n))\nabla(e_{h,m}^n - e_{h,m-1}^n), \nabla\upsilon_h\big),\\
I_4^n(\upsilon_h) &:= -\big((\mathrm{D}\boldsymbol{f}(\nabla u_{h,m-1}^n) - \mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n))\nabla(u_m^n - u_{m-1}^n), \nabla\upsilon_h\big),\\
I_5^n(\upsilon_h) &:= \frac{1}{\tau}(e_{h,\ell}^{n-1}, \upsilon_h).
\end{aligned}$$

Now,

$$\begin{aligned}
I_1^n(\upsilon_h) &= -\left(\int_0^1 \mathrm{D}\boldsymbol{f}((1-s)\nabla u_{h,m-1}^n + s\nabla u_{m-1}^n)\mathrm{d}s\, \nabla e_{h,m-1}^n, \nabla\upsilon_h\right)\\
&= -\big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla e_{h,m-1}^n, \nabla\upsilon_h\big)\\
&\quad - \left(\int_0^1 (\mathrm{D}\boldsymbol{f}((1-s)\nabla u_{h,m-1}^n + s\nabla u_{m-1}^n) - \mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n))\mathrm{d}s\, \nabla e_{h,m-1}^n, \nabla\upsilon_h\right),
\end{aligned} \tag{4.7}$$

whence

$$I_1^n(\upsilon_h) + I_2^n(\upsilon_h) = -\big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla\theta_{h,m}^n, \nabla\upsilon_h\big) + I_6^n(\upsilon_h) + I_7^n(\upsilon_h) \tag{4.8}$$

with

$$I_6^n(v_h) := -\big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla\rho_{h,m}^n, \nabla v_h\big),$$
$$I_7^n(v_h) := -\left(\int_0^1 (\mathrm{D}\boldsymbol{f}((1-s)\nabla u_{h,m-1}^n + s\nabla u_{m-1}^n) - \mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n))\mathrm{d}s\,\nabla e_{h,m-1}^n, \nabla v_h\right).$$

Thus, (4.6) can be rewritten as

$$\left(\left(\frac{1}{\tau} - A_{h,m-1}^n\right)\theta_{h,m}^n, v_h\right) = I_3^n(v_h) + I_4^n(v_h) + I_5^n(v_h) + I_6^n(v_h) + I_7^n(v_h) \quad \forall v_h \in S_h, \tag{4.9}$$

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

$$(A_{h,m-1}^n w_h, v_h) = -\big(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla w_h, \nabla v_h\big) \quad \forall w_h, v_h \in S_h. \tag{4.10}$$

In the following, we estimate $\|e_{h,m}^n\|_{W^{1,q}(\Omega)}$ by using the error equation (4.9). To this end, we employ two loops of mathematical induction. The first loop of mathematical induction is as follows. We assume that

$$\|e_{h,\ell}^{n-1}\|_{W^{1,q}(\Omega)} \leqslant h^{\frac{d}{2q}+\frac{1}{2}}, \quad n = 1,\ldots,k, \tag{4.11}$$

and then prove

$$\|e_{h,\ell}^{k}\|_{W^{1,q}(\Omega)} \leqslant h^{\frac{d}{2q}+\frac{1}{2}}. \tag{4.12}$$

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 $1 \leqslant m \leqslant \ell$

$$\|e_{h,i-1}^n\|_{W^{1,\infty}(\Omega)} \leqslant h^{\frac{1}{4}-\frac{d}{4q}}, \quad i = 1,\ldots,m, \; n = 1,\ldots,k, \tag{4.13}$$

and then prove that

$$\|e_{h,m}^n\|_{W^{1,\infty}(\Omega)} \leqslant h^{\frac{1}{4}-\frac{d}{4q}}, \quad n = 1,\ldots,k. \tag{4.14}$$

Note that (4.11) implies (4.13) with $i = 1$ for sufficiently small h. This is a consequence of

$$\begin{aligned}
\|e_{h,0}^n\|_{W^{1,\infty}(\Omega)} &= \|e_{h,\ell}^{n-1}\|_{W^{1,\infty}(\Omega)} \\
&\leqslant \|\theta_{h,\ell}^{n-1}\|_{W^{1,\infty}(\Omega)} + \|\rho_{h,\ell}^{n-1}\|_{W^{1,\infty}(\Omega)} \\
&\leqslant ch^{-\frac{d}{q}}(\|e_{h,\ell}^{n-1}\|_{W^{1,q}(\Omega)} + \|\rho_{h,\ell}^{n-1}\|_{W^{1,q}(\Omega)}) + ch^{1-\frac{d}{q}}\|u_\ell^{n-1}\|_{C^{1,1-\frac{d}{q}}(\overline{\Omega})} \quad \text{(inverse inequality is used)} \\
&\leqslant ch^{\frac{1}{2}-\frac{d}{2q}} + ch^{1-\frac{d}{q}}\|u_\ell^{n-1}\|_{W^{2,q}(\Omega)},
\end{aligned}$$

where we have used (4.11), (4.3) and $W^{2,q}(\Omega) \hookrightarrow C^{1,1-\frac{d}{q}}(\overline{\Omega})$ in the last inequality. The boundedness of $\|u_\ell^{n-1}\|_{W^{2,q}(\Omega)}$ 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 $\|u_m^n\|_{W^{1,\infty}(\Omega)}$ and $\|u_m^n\|_{W^{2,q}(\Omega)}$ for all $0 \leqslant m \leqslant \ell$, which imply the uniform ellipticity (3.15). Under assumption (4.13), we have the following estimates (due to the local Lipschitz continuity of $\boldsymbol{f}$):

$$\big|\mathrm{D}\boldsymbol{f}(\nabla u_{h,m-1}^n) - \mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\big| \leqslant c|\nabla e_{h,m-1}^n|, \qquad n = 1,\ldots,k, \tag{4.15}$$
$$\big|\mathrm{D}\boldsymbol{f}((1-s)\nabla u_{h,m-1}^n + s\nabla u_{m-1}^n) - \mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\big| \leqslant c|\nabla e_{h,m-1}^n|, \qquad n = 1,\ldots,k. \tag{4.16}$$

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

$$\left\|\frac{1}{\tau}\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right\|_{W^{-1,s}(\Omega)} \leqslant c\|w_h\|_{W^{-1,s}(\Omega)} \qquad \forall w_h \in S_h, \; \forall 1 < s < \infty, \tag{4.17}$$
$$\left\|\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right\|_{W^{1,s}(\Omega)} \leqslant c\|w_h\|_{W^{-1,s}(\Omega)} \qquad \forall w_h \in S_h, \; \forall 1 < s < \infty. \tag{4.18}$$

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

$$\tau^{\frac{1}{p}}\|\phi_h^1/\tau\|_{W^{-1,s}(\Omega)} \leqslant \tau^{\frac{1}{p}}\|w_h\|_{W^{-1,s}(\Omega)} \quad \text{and} \quad \tau^{\frac{1}{p}}\|\phi_h^1\|_{W^{1,s}(\Omega)} \leqslant \tau^{\frac{1}{p}}\|w_h\|_{W^{-1,s}(\Omega)}$$

for $\phi_h^1 = \left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h$, which imply (4.17) and (4.18), respectively.

Substituting $v_h = \left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h$ into (4.9) yields, for $1 \leqslant n \leqslant k$,

$$\left(\theta_{h,m}^n, w_h\right) = \sum_{j=3}^{7} I_j^n\left(\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right) \quad \forall w_h \in S_h, \tag{4.19}$$

where

$$\begin{aligned}
&\left|I_3^n\left(\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right)\right| \\
&= \left|\left((\mathrm{D}\boldsymbol{f}(\nabla u_{h,m-1}^n) - \mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n))\nabla(e_{h,m}^n - e_{h,m-1}^n), \nabla\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right)\right| \\
&\leqslant \|\mathrm{D}\boldsymbol{f}(\nabla u_{h,m-1}^n) - \mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\|_{L^\infty(\Omega)}\|\nabla(e_{h,m}^n - e_{h,m-1}^n)\|_{L^q(\Omega)}\left\|\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right\|_{W^{1,q'}(\Omega)} \\
&\leqslant c\|\nabla e_{h,m-1}^n\|_{L^\infty(\Omega)}(\|\nabla e_{h,m}^n\|_{L^q(\Omega)} + \|\nabla e_{h,m-1}^n\|_{L^q(\Omega)})\|w_h\|_{W^{-1,q'}(\Omega)} \quad \text{[(4.18) is used]} \\
&\leqslant ch^{\frac{1}{4}-\frac{d}{4q}}(\|\nabla e_{h,m}^n\|_{L^q(\Omega)} + \|\nabla e_{h,m-1}^n\|_{L^q(\Omega)})\|w_h\|_{W^{-1,q'}(\Omega)} \quad \text{[(4.13) is used]}
\end{aligned} \tag{4.20}$$

and, similarly,

$$\begin{aligned}
&\left|I_4^n\left(\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right)\right| \\
&= \left|\left((\mathrm{D}\boldsymbol{f}(\nabla u_{h,m-1}^n) - \mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n))\nabla(u_m^n - u_{m-1}^n), \nabla\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right)\right| \\
&\leqslant c\|\nabla e_{h,m-1}^n\|_{L^q(\Omega)}\|\nabla(u_m^n - u_{m-1}^n)\|_{L^\infty(\Omega)}\|w_h\|_{W^{-1,q'}(\Omega)} \\
&\leqslant c\tau\|\nabla e_{h,m-1}^n\|_{L^q(\Omega)}\|w_h\|_{W^{-1,q'}(\Omega)}, \quad \text{[(3.5) is used to estimate } \|\nabla(u_m^n - u_{m-1}^n)\|_{L^\infty(\Omega)}\text{]}
\end{aligned} \tag{4.21}$$

and

$$\begin{aligned}
&\left|I_7^n\left(\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right)\right| \\
&\leqslant \left\|\int_0^1(\mathrm{D}\boldsymbol{f}((1-s)\nabla u_{h,m-1}^n + s\nabla u_{m-1}^n) - \mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n))\mathrm{d}s\right\|_{L^\infty(\Omega)}\|\nabla e_{h,m-1}^n\|_{L^q(\Omega)}\left\|\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right\|_{W^{1,q'}(\Omega)} \\
&\leqslant c\|\nabla e_{h,m-1}^n\|_{L^\infty(\Omega)}\|\nabla e_{h,m-1}^n\|_{L^q(\Omega)}\|w_h\|_{W^{-1,q'}(\Omega)} \quad \text{[(4.18) is used]} \\
&\leqslant ch^{\frac{1}{4}-\frac{d}{4q}}\|\nabla e_{h,m-1}^n\|_{L^q(\Omega)}\|w_h\|_{W^{-1,q'}(\Omega)}. \quad \text{[(4.13) is used]}
\end{aligned} \tag{4.22}$$

Moreover, we have

$$\begin{aligned}
\left|I_5^n\left(\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right)\right| &= \left|\left(e_{h,\ell}^{n-1}, \frac{1}{\tau}\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right)\right| \\
&\leqslant \|e_{h,\ell}^{n-1}\|_{W^{1,q}(\Omega)}\left\|\frac{1}{\tau}\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right\|_{W^{-1,q'}(\Omega)} \\
&\leqslant c\|e_{h,\ell}^{n-1}\|_{W^{1,q}(\Omega)}\|w_h\|_{W^{-1,q'}(\Omega)}, \quad \text{[(4.17) is used]}
\end{aligned} \tag{4.23}$$

and

$$\begin{aligned}
\left|I_6^n\left(\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right)\right| &= \left|\left(\mathrm{D}\boldsymbol{f}(\nabla u_{m-1}^n)\nabla\rho_{h,m}^n, \nabla\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right)\right| \\
&\leqslant c\|\rho_{h,m}^n\|_{W^{1,q}(\Omega)}\left\|\left(\frac{1}{\tau} - A_{h,m-1}^n\right)^{-1} w_h\right\|_{W^{1,q'}(\Omega)}
\end{aligned}$$

$$\leqslant c\|\rho^n_{h,m}\|_{W^{1,q}(\Omega)}\|w_h\|_{W^{-1,q'}(\Omega)} \qquad [(4.18) \text{ is used}]$$
$$\leqslant ch\|u^n_m\|_{W^{2,q}(\Omega)}\|w_h\|_{W^{-1,q'}(\Omega)}. \qquad [(4.3) \text{ is used}] \tag{4.24}$$

Substituting the estimates of I^n_j, $j = 3,4,5,6,7$, into (4.19) yields

$$|(\theta^n_m, w_h)| \leqslant \left[c(\tau + h^{\frac{1}{4}-\frac{d}{4q}})(\|e^n_{h,m}\|_{W^{1,q}(\Omega)} + \|e^n_{h,m-1}\|_{W^{1,q}(\Omega)}) + c\|e^{n-1}_{h,\ell}\|_{W^{1,q}(\Omega)} + ch\right]\|w_h\|_{W^{-1,q'}(\Omega)},$$

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

$$|(\theta^n_m, w)| = |(\theta^n_m, P_h w)| \leqslant \left[c(\tau + h^{\frac{1}{4}-\frac{d}{4q}})(\|e^n_{h,m}\|_{W^{1,q}(\Omega)} + \|e^n_{h,m-1}\|_{W^{1,q}(\Omega)}) + c\|e^{n-1}_{h,\ell}\|_{W^{1,q}(\Omega)} + ch\right]\|w\|_{W^{-1,q'}(\Omega)},$$

which implies (via the duality argument)

$$\|\theta^n_m\|_{W^{1,q}(\Omega)} \leqslant c(\tau + h^{\frac{1}{4}-\frac{d}{4q}})(\|e^n_{h,m}\|_{W^{1,q}(\Omega)} + \|e^n_{h,m-1}\|_{W^{1,q}(\Omega)}) + c\|e^{n-1}_{h,\ell}\|_{W^{1,q}(\Omega)} + ch,$$

whence

$$\begin{aligned}\|e^n_{h,m}\|_{W^{1,q}(\Omega)} &\leqslant \|\theta^n_{h,m}\|_{W^{1,q}(\Omega)} + \|\rho^n_{h,m}\|_{W^{1,q}(\Omega)} \\ &\leqslant c(\tau + h^{\frac{1}{4}-\frac{d}{4q}})(\|e^n_{h,m}\|_{W^{1,q}(\Omega)} + \|e^n_{h,m-1}\|_{W^{1,q}(\Omega)}) + c\|e^{n-1}_{h,\ell}\|_{W^{1,q}(\Omega)} + ch.\end{aligned} \tag{4.25}$$

For sufficiently small τ and h, the first term on the right-hand side of (4.25) can be absorbed by the left-hand side. Thus, estimate (4.25) reduces to

$$\|e^n_{h,m}\|_{W^{1,q}(\Omega)} \leqslant c(\tau + h^{\frac{1}{4}-\frac{d}{4q}})\|e^n_{h,m-1}\|_{W^{1,q}(\Omega)} + c\|e^{n-1}_{h,\ell}\|_{W^{1,q}(\Omega)} + ch. \tag{4.26}$$

Iterating estimate (4.26) yields

$$\begin{aligned}&\|e^n_{h,m}\|_{W^{1,q}(\Omega)} \\ &\leqslant c(\tau + h^{\frac{1}{4}-\frac{d}{4q}})\|e^n_{h,m-1}\|_{W^{1,q}(\Omega)} + c\|e^{n-1}_{h,\ell}\|_{W^{1,q}(\Omega)} + ch \\ &\leqslant c^2(\tau + h^{\frac{1}{4}-\frac{d}{4q}})^2\|e^n_{h,m-2}\|_{W^{1,q}(\Omega)} + c[1 + c(\tau + h^{\frac{1}{4}-\frac{d}{4q}})](\|e^{n-1}_{h,\ell}\|_{W^{1,q}(\Omega)} + h) \\ &\cdots \\ &\leqslant c^m(\tau + h^{\frac{1}{4}-\frac{d}{4q}})^m\|e^n_{h,0}\|_{W^{1,q}(\Omega)} + c\sum_{j=1}^{m} c^{j-1}(\tau + h^{\frac{1}{4}-\frac{d}{4q}})^{j-1}(\|e^{n-1}_{h,\ell}\|_{W^{1,q}(\Omega)} + h) \\ &\leqslant c^m(\tau + h^{\frac{1}{4}-\frac{d}{4q}})^m\|e^n_{h,0}\|_{W^{1,q}(\Omega)} + c(\|e^{n-1}_{h,\ell}\|_{W^{1,q}(\Omega)} + h) \\ &\leqslant \|e^{n-1}_{h,\ell}\|_{W^{1,q}(\Omega)} + ch,\end{aligned} \tag{4.27}$$

where the last inequality requires τ and h to be sufficiently small (independent of n, m, k and ℓ), and we have used the identity $e^n_{h,0} = e^{n-1}_{h,\ell}$. The constant c is independent of m and ℓ. Substituting the induction assumption (4.11) into (4.27) yields

$$\|e^n_{h,m}\|_{W^{1,q}(\Omega)} \leqslant ch^{\frac{d}{2q}+\frac{1}{2}}, \tag{4.28}$$

and the triangle inequality implies

$$\|\theta^n_{h,m}\|_{W^{1,q}(\Omega)} \leqslant \|e^n_{h,m}\|_{W^{1,q}(\Omega)} + \|\rho^n_{h,m}\|_{W^{1,q}(\Omega)} \leqslant ch^{\frac{d}{2q}+\frac{1}{2}}. \tag{4.29}$$

The inverse inequality of the finite element space gives

$$\begin{aligned}\|e^n_{h,m}\|_{W^{1,\infty}(\Omega)} &\leqslant \|\theta^n_{h,m}\|_{W^{1,\infty}(\Omega)} + \|\rho^n_{h,m}\|_{W^{1,\infty}(\Omega)} \\ &\leqslant ch^{-\frac{d}{q}}\|\theta^n_{h,m}\|_{W^{1,q}(\Omega)} + \|\rho^n_{h,m}\|_{W^{1,\infty}(\Omega)} \\ &\leqslant ch^{-\frac{d}{q}}\|\theta^n_{h,m}\|_{W^{1,q}(\Omega)} + ch^{1-\frac{d}{q}}\|u^n_m\|_{W^{2,q}(\Omega)} \leqslant ch^{\frac{1}{2}-\frac{d}{2q}} + ch^{1-\frac{d}{q}}.\end{aligned}$$

For sufficiently small h, this estimate implies (4.14). This completes the second loop of mathematical induction. Thus, (4.14) holds for $m=\ell$, i.e.,

$$\|e_{h,\ell}^n\|_{W^{1,\infty}(\Omega)} \leqslant h^{\frac{1}{4}-\frac{d}{4q}}, \quad n=1,\ldots,k. \tag{4.30}$$

Substituting $m=\ell$ into (4.9), we rewrite the resulting equation as

$$\left(\delta_\tau \theta_{h,\ell}^n - A_{h,\ell-1}^n \theta_{h,\ell}^n, v_h\right) = I_3^n(v_h) + I_4^n(v_h) + I_6^n(v_h) + I_7^n(v_h) \quad \forall v_h \in S_h, \tag{4.31}$$

where $I_5^n(v_h)$ from (4.9) has been absorbed by the left-hand side of (4.31). Note that I_j^n, $j=3,4,6,7$, can be identified with linear functionals on $W_0^{1,q}(\Omega)$. Then

$$\begin{aligned}
\|I_3^n\|_{W^{-1,q}(\Omega)} &\leqslant \|\mathrm{D}\boldsymbol{f}(\nabla u_{h,\ell-1}^n) - \mathrm{D}\boldsymbol{f}(\nabla u_{\ell-1}^n)\|_{L^\infty(\Omega)} \|\nabla(e_{h,\ell}^n - e_{h,\ell-1}^n)\|_{L^q(\Omega)} \\
&\leqslant c\|\nabla e_{h,\ell-1}^n\|_{L^\infty(\Omega)} (\|\nabla e_{h,\ell}^n\|_{L^q(\Omega)} + \|\nabla e_{h,\ell-1}^n\|_{L^q(\Omega)}) \\
&\leqslant ch^{\frac{1}{4}-\frac{d}{4q}} (\|\nabla e_{h,\ell}^n\|_{L^q(\Omega)} + \|\nabla e_{h,\ell-1}^n\|_{L^q(\Omega)}) \qquad \text{[(4.13) is used]}.
\end{aligned}$$

This estimate agrees with the estimate of I_3^n in (4.20). Similarly, from the estimates of I_4^n, I_6^n and I_7^n in (4.21), (4.24) and (4.22) we see that

$$\|I_4^n\|_{W^{-1,q}(\Omega)} + \|I_6^n\|_{W^{-1,q}(\Omega)} + \|I_7^n\|_{W^{-1,q}(\Omega)} \leqslant c(\tau + h^{\frac{1}{4}-\frac{d}{4q}}) \|\nabla e_{h,\ell-1}^n\|_{L^q(\Omega)} + ch.$$

By denoting $\mathscr{A}(\cdot,t_n) = \mathrm{D}\boldsymbol{f}(\nabla u_{\ell-1}^n)$ and constructing $\mathscr{A}(\cdot,t)$ as the piecewise linear interpolant of $\mathscr{A}(\cdot,t_n)$, $n=0,1,\ldots,N$, the resulting matrix $\mathscr{A}(\cdot,t)$ satisfies the conditions of Theorem 2.2 in view of (3.5). Since $\theta_{h,\ell}^0 = 0$, applying (2.15) of Theorem 2.2 to (4.31) yields, for $1<p<\infty$,

$$\begin{aligned}
&\left\|\left(\delta_\tau \theta_{h,\ell}^n\right)_{n=1}^k\right\|_{L^p(W^{-1,q}(\Omega))} + \left\|(\theta_{h,\ell}^n)_{n=1}^k\right\|_{L^p(W^{1,q}(\Omega))} \\
&\leqslant c\|(I_3^n)_{n=1}^k\|_{L^p(W^{-1,q}(\Omega))} + c\|(I_4^n)_{n=1}^k\|_{L^p(W^{-1,q}(\Omega))} \\
&\quad + c\|(I_6^n)_{n=1}^k\|_{L^p(W^{-1,q}(\Omega))} + c\|(I_7^n)_{n=1}^k\|_{L^p(W^{-1,q}(\Omega))} \\
&\leqslant c(\tau + h^{\frac{1}{4}-\frac{d}{4q}})(\|(e_{h,\ell}^n)_{n=1}^k\|_{L^p(W^{1,q}(\Omega))} + \|(e_{h,\ell-1}^n)_{n=1}^k\|_{L^p(W^{1,q}(\Omega))}) + ch \\
&\leqslant c(\tau + h^{\frac{1}{4}-\frac{d}{4q}})(\|(\theta_{h,\ell}^n)_{n=1}^k\|_{L^p(W^{1,q}(\Omega))} + \|(\theta_{h,\ell-1}^n)_{n=1}^k\|_{L^p(W^{1,q}(\Omega))}) + ch,
\end{aligned} \tag{4.32}$$

where the last inequality is due to

$$\begin{aligned}
\|(e_{h,\ell}^n)_{n=1}^k\|_{L^p(W^{1,q}(\Omega))} &\leqslant \|(\theta_{h,\ell}^n)_{n=1}^k\|_{L^p(W^{1,q}(\Omega))} + \|(\rho_{h,\ell}^n)_{n=1}^k\|_{L^p(W^{1,q}(\Omega))} \\
&\leqslant \|(\theta_{h,\ell}^n)_{n=1}^k\|_{L^p(W^{1,q}(\Omega))} + ch.
\end{aligned}$$

For sufficiently small τ and h, estimate (4.32) reduces to

$$\begin{aligned}
&\left\|\left(\delta_\tau \theta_{h,\ell}^n\right)_{n=1}^k\right\|_{L^p(W^{-1,q}(\Omega))} + \left\|(\theta_{h,\ell}^n)_{n=1}^k\right\|_{L^p(W^{1,q}(\Omega))} \\
&\leqslant c(\tau + h^{\frac{1}{4}-\frac{d}{4q}}) \|(\theta_{h,\ell-1}^n)_{n=1}^k\|_{L^p(W^{1,q}(\Omega))} + ch.
\end{aligned} \tag{4.33}$$

In view of the relation $\theta_{h,m}^n = e_{h,m}^n - \rho_{h,m}^n$, setting $m=\ell-1$ in (4.27) we have

$$\|\theta_{h,\ell-1}^n\|_{W^{1,q}(\Omega)} \leqslant c\|\theta_{h,\ell}^{n-1}\|_{W^{1,q}(\Omega)} + ch.$$

Substituting this estimate into (4.33), we obtain

$$\begin{aligned}
&\left\|\left(\delta_\tau \theta_{h,\ell}^n\right)_{n=1}^k\right\|_{L^p(W^{-1,q}(\Omega))} + \left\|(\theta_{h,\ell}^n)_{n=1}^k\right\|_{L^p(W^{1,q}(\Omega))} \\
&\leqslant c(\tau + h^{\frac{1}{4}-\frac{d}{4q}}) \|(\theta_{h,\ell}^{n-1})_{n=1}^k\|_{L^p(W^{1,q}(\Omega))} + ch.
\end{aligned} \tag{4.34}$$

For sufficiently small τ and h, the first term on the right-hand side of (4.34) can be absorbed by the left-hand side (since $\theta_{h,\ell}^0 = 0$). This yields

$$\left\|\left(\delta_\tau \theta_{h,\ell}^n\right)_{n=1}^k\right\|_{L^p(W^{-1,q}(\Omega))} + \left\|(\theta_{h,\ell}^n)_{n=1}^k\right\|_{L^p(W^{1,q}(\Omega))} \leqslant ch \quad \forall\, 1<p<\infty. \tag{4.35}$$

The inverse inequality of the finite element space gives

$$\left\|\left(\delta_\tau \theta_{h,\ell}^n\right)_{n=1}^k\right\|_{L^p(W^{1,q}(\Omega))} \leqslant ch^{-2} \left\|\left(\delta_\tau \theta_{h,\ell}^n\right)_{n=1}^k\right\|_{L^p(W^{-1,q}(\Omega))} \leqslant ch^{-1}. \tag{4.36}$$

Then, the Sobolev interpolation inequality implies

$$\begin{aligned}
&\big\|(\theta^n_{h,\ell})^k_{n=1}\big\|_{L^\infty(W^{1,q}(\Omega))} \\
&\leqslant c\big\|(\theta^n_{h,\ell})^k_{n=1}\big\|^{1-\frac{1}{p}}_{L^p(W^{1,q}(\Omega))}\big\|\big(\delta_\tau\theta^n_{h,\ell}\big)^k_{n=1}\big\|^{\frac{1}{p}}_{L^p(W^{1,q}(\Omega))} \quad \text{(see Appendix C)} \\
&\leqslant (ch)^{1-\frac{1}{p}}(ch^{-1})^{\frac{1}{p}} \leqslant ch^{1-\frac{2}{p}} \quad \forall\, 1<p<\infty. \qquad (4.37)
\end{aligned}$$

Thus (by the triangle inequality)

$$\begin{aligned}
\big\|(e^n_{h,\ell})^k_{n=1}\big\|_{L^\infty(W^{1,q}(\Omega))} &\leqslant \big\|(\theta^n_{h,\ell})^k_{n=1}\big\|_{L^\infty(W^{1,q}(\Omega))} + \big\|(\rho^n_{h,\ell})^k_{n=1}\big\|_{L^\infty(W^{1,q}(\Omega))} \\
&\leqslant ch^{1-\frac{2}{p}} \quad \forall\, 1<p<\infty. \qquad (4.38)
\end{aligned}$$

For sufficiently small τ and h, estimate (4.38) implies

$$\big\|(e^n_{h,\ell})^k_{n=1}\big\|_{L^\infty(W^{1,q}(\Omega))} \leqslant h^{\frac{d}{2q}+\frac{1}{2}}. \qquad (4.39)$$

This proves (4.12) and thus completes the first loop of mathematical induction.

To conclude, (4.38) implies (4.1).

4.3 *Error estimate in $L^q(\Omega)$*

Note that (4.31) can be rewritten as

$$\big(\delta_\tau e^n_{h,\ell} - A^n_{h,\ell-1}e^n_{h,\ell}, v_h\big) = I^n_3(v_h) + I^n_4(v_h) + I^n_7(v_h) \quad \forall\, v_h \in S_h, \qquad (4.40)$$

where $I^n_6(v_h)$ in (4.31) has been absorbed into the left-hand side. Let $\eta^n_h \in S_h$ be the solution of the finite element equation

$$\big(\delta_\tau \eta^n_h - A^n_{h,\ell-1}\eta^n_h, v_h\big) = I^n_3(v_h) + I^n_4(v_h) + I^n_7(v_h) \quad \forall\, v_h \in S_h, \qquad (4.41)$$

with zero initial value $\eta^0_h = 0$. Then η^n_h satisfies the following estimate (cf. (2.15) of Theorem 2.2)

$$\begin{aligned}
&\big\|\big(\delta_\tau\eta^n_h\big)^N_{n=1}\big\|_{L^p(W^{-1,q/2}(\Omega))} + \|(\eta^n_h)^N_{n=1}\|_{L^p(W^{1,q/2}(\Omega))} \\
&\leqslant c(\|(I^n_3+I^n_4+I^n_7)^N_{n=1}\|_{L^p(W^{-1,q/2}(\Omega))} \\
&\leqslant c\|\nabla e^n_{h,\ell-1}\|_{L^p(L^q(\Omega))}\|\nabla(e^n_{h,\ell}-e^n_{h,\ell-1})\|_{L^\infty(L^q(\Omega))} \\
&\quad + c\|\nabla e^n_{h,\ell-1}\|_{L^p(L^q(\Omega))}\|\nabla(u^n_\ell - u^n_{\ell-1})\|_{L^\infty(L^q(\Omega))} \\
&\quad + c\|\nabla e^n_{h,\ell-1}\|_{L^p(L^q(\Omega))}\|\nabla e^n_{\ell-1}\|_{L^\infty(L^q(\Omega))} \qquad \text{[similar to (4.20)–(4.22)]} \\
&\leqslant c\|\nabla e^n_{h,\ell}\|^2_{L^\infty(L^q(\Omega))} + c\|\nabla e^n_{h,\ell-1}\|^2_{L^\infty(L^q(\Omega))} + c\|\nabla e^n_{h,\ell-1}\|_{L^\infty(L^q(\Omega))}\tau,
\end{aligned}$$

where we have used the estimate $\|\nabla(u^n_\ell - u^n_{\ell-1})\|_{L^\infty(L^q(\Omega))} \leqslant c\tau$, which is a consequence of (3.5). Setting $m = \ell-1$ in (4.27) yields

$$\|e^n_{h,\ell-1}\|_{W^{1,q}(\Omega)} \leqslant c\|e^{n-1}_{h,\ell}\|_{W^{1,q}(\Omega)} + ch.$$

The last two estimates together imply

$$\begin{aligned}
&\big\|\big(\delta_\tau\eta^n_h\big)^N_{n=1}\big\|_{L^p(W^{-1,q/2}(\Omega))} + \|(\eta^n_h)^N_{n=1}\|_{L^p(W^{1,q/2}(\Omega))} \\
&\leqslant c\|(\nabla e^n_{h,\ell})^N_{n=1}\|^2_{L^\infty(L^q(\Omega))} + c\|(\nabla e^{n-1}_{h,\ell})^N_{n=1}\|^2_{L^\infty(L^q(\Omega))} + c\|(\nabla e^{n-1}_{h,\ell})^N_{n=1}\|_{L^\infty(L^q(\Omega))}\tau \\
&\leqslant ch^{1-\frac{2}{p}}(\tau + h^{1-\frac{2}{p}}) \qquad \forall\, 1<p<\infty,
\end{aligned}$$

where the last inequality is due to (4.38). The inverse inequality of the finite element space yields

$$\big\|\big(\delta_\tau\eta^n_h\big)^N_{n=1}\big\|_{L^p(W^{1,q/2})} \leqslant ch^{-2}\big\|\big(\delta_\tau\eta^n_h\big)^N_{n=1}\big\|_{L^p(W^{-1,q/2})} \leqslant ch^{-1-\frac{2}{p}}(\tau + h^{1-\frac{2}{p}}).$$

Since $\eta^0_h = 0$, the Sobolev interpolation inequality implies

$$\big\|(\eta^n_h)^N_{n=1}\big\|_{L^\infty(W^{1,q/2})} \leqslant c\big\|(\eta^n_h)^N_{n=1}\big\|^{1-\frac{1}{p}}_{L^p(W^{1,q/2})}\big\|\big(\delta_\tau\eta^n_h\big)^N_{n=1}\big\|^{\frac{1}{p}}_{L^p(W^{1,q/2})} \quad \text{(see Appendix C)}$$

$$\leqslant ch^{1-\frac{4}{p}}(\tau+h^{1-\frac{2}{p}}) \qquad \forall 1<p<\infty. \tag{4.42}$$

Since $W^{1,q/2}(\Omega) \hookrightarrow L^q(\Omega)$ for $q>d$, it follows that

$$\big\|(\eta_h^n)_{n=1}^N\big\|_{L^\infty(L^q(\Omega))} \leqslant ch^{1-\frac{4}{p}}(\tau+h^{1-\frac{2}{p}}) \leqslant c(\tau^2+h^{2-\frac{8}{p}}) \qquad \forall 1<p<\infty. \tag{4.43}$$

Subtracting (4.41) from (4.40), we have

$$\big(\delta_\tau(u_{h,\ell}^n-\eta_h^n-u_\ell^n),\upsilon_h\big)+\big(\mathrm{D}\boldsymbol{f}(\nabla u_{\ell-1}^n)\nabla(u_{h,\ell}^n-\eta_h^n-u_\ell^n),\nabla\upsilon_h\big)=0 \quad \forall \upsilon_h\in S_h, \tag{4.44}$$

whence $u_{h,\ell}^n-\eta_h^n$ and u_ℓ^n play the same roles as ϕ_h^n and ϕ in Theorem 2.2, respectively. Thus, (2.18) implies

$$\|(u_{h,\ell}^n-\eta_h^n-u_\ell^n)_{n=1}^N\|_{L^p(L^q(\Omega))} \leqslant ch^2.$$

and therefore

$$\|(u_{h,\ell}^n-\eta_h^n-P_hu_\ell^n)_{n=1}^N\|_{L^p(L^q(\Omega))} \leqslant \|(u_{h,\ell}^n-\eta_h^n-u_\ell^n)_{n=1}^N\|_{L^p(L^q(\Omega))}+\|(u_\ell^n-P_hu_\ell^n)_{n=1}^N\|_{L^p(L^q(\Omega))} \leqslant ch^2. \tag{4.45}$$

Then (4.44) implies

$$\begin{aligned}
|(\delta_\tau(u_{h,\ell}^n-\eta_h^n-P_hu_\ell^n),\upsilon_h)| &\leqslant \big|\big(\mathrm{D}\boldsymbol{f}(\nabla u_{\ell-1}^n)\nabla(u_{h,\ell}^n-\eta_h^n-u_\ell^n),\nabla\upsilon_h\big)\big| \\
&\leqslant \big|\big(\mathrm{D}\boldsymbol{f}(\nabla u_{\ell-1}^n)\nabla(u_{h,\ell}^n-\eta_h^n-P_hu_\ell^n),\nabla\upsilon_h\big)\big|+\big|\big(\mathrm{D}\boldsymbol{f}(\nabla u_{\ell-1}^n)\nabla(u_\ell^n-P_hu_\ell^n),\nabla\upsilon_h\big)\big| \\
&\leqslant ch^{-2}\|u_{h,\ell}^n-\eta_h^n-P_hu_\ell^n\|_{L^q(\Omega)}\|\upsilon_h\|_{L^{q'}(\Omega)}+\|u_\ell^n\|_{W^{2,q}(\Omega)}\|\upsilon_h\|_{L^{q'}(\Omega)} \\
&\leqslant ch^{-2}\|u_{h,\ell}^n-\eta_h^n-P_hu_\ell^n\|_{L^q(\Omega)}\|\upsilon_h\|_{L^{q'}(\Omega)}+c\|\upsilon_h\|_{L^{q'}(\Omega)},
\end{aligned}$$

where we have used (3.5) with $m=\ell$. By using the duality argument and (4.45), we obtain

$$\|\delta_\tau(u_{h,\ell}^n-\eta_h^n-P_hu_\ell^n)\|_{L^q(\Omega)} \leqslant ch^{-2}\|u_{h,\ell}^n-\eta_h^n-P_hu_\ell^n\|_{L^q(\Omega)}+c \leqslant c. \tag{4.46}$$

Since $u_{h,\ell}^0-\eta_h^0-P_hu_\ell^0=0$, the Sobolev interpolation inequality gives

$$\begin{aligned}
&\big\|(u_{h,\ell}^n-\eta_h^n-P_hu_\ell^n)_{n=1}^N\big\|_{L^\infty(L^q(\Omega))} \qquad\qquad (4.47)\\
&\leqslant c\big\|(u_{h,\ell}^n-\eta_h^n-P_hu_\ell^n)_{n=1}^N\big\|_{L^p(L^q(\Omega))}^{1-\frac{1}{p}}\big\|\delta_\tau(u_{h,\ell}^n-\eta_h^n-P_hu_\ell^n)_{n=1}^N\big\|_{L^p(L^q(\Omega))}^{\frac{1}{p}} \quad \text{(see Appendix C)} \\
&\leqslant (ch^2)^{1-\frac{1}{p}}c^{\frac{1}{p}} \leqslant ch^{2-\frac{2}{p}} \qquad \forall 1<p<\infty.
\end{aligned}$$

Estimates (4.43) and (4.47) yield

$$\big\|(u_{h,\ell}^n-u_\ell^n)_{n=1}^N\big\|_{L^\infty(L^q(\Omega))} \leqslant c_p(\tau^2+h^{2-\frac{8}{p}}) \quad \forall 1<p<\infty, \tag{4.48}$$

which implies (4.2). This completes the proof of Proposition 4.1. $\square$

5. Numerical test

To support our theoretical analysis, we present a numerical example by solving the initial and boundary value problem

$$\begin{cases}
\partial_t u=\nabla\cdot\left(\dfrac{\nabla u}{\sqrt{1+|\nabla u|^2}}\right)+g & \text{in } \Omega\times(0,T], \\
u=0 & \text{on } \partial\Omega\times(0,T], \\
u(\cdot,0)=u_0 & \text{in } \Omega,
\end{cases} \tag{5.1}$$

with the proposed method (2.9) in the domain $\Omega=[0,1]\times[0,1]$ up to time $T=1$, where the function g and initial data u_0 are chosen corresponding to the exact solution

$$u(x,y,t)=e^t\sin(\pi x)\sin(\pi y). \tag{5.2}$$

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 $\tau = 10^{-3}$.

h	$\|u_h^N - u(t_N)\|_{L^2(\Omega)}$	$|u_h^N - u(t_N)|_{H^1(\Omega)}$
1/8	6.7246E-02	1.1894E-01
1/16	1.7987E-02	5.9414E-01
1/32	4.2704E-03	2.9663E-01
convergence rate	$O(h^{2.07})$	$O(h^{1.01})$

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

τ	$\|u_h^N - u(t_N)\|_{L^2(\Omega)}$	$|u_h^N - u(t_N)|_{H^1(\Omega)}$
1/8	5.0080E-02	2.2621E-01
1/16	2.4763E-02	1.1302E-01
1/32	1.2308E-02	5.8460E-02
convergence rate	$O(\tau^{1.00})$	$O(\tau^{0.95})$

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 Ω 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 Ω 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 $\nabla_p \boldsymbol{f}(p)$ 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 $\nabla_p \boldsymbol{f}(p) = \nabla_p^2 F(p)$ 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 $u_0 \in W^{2,q}(\Omega) \cap W_0^{1,q}(\Omega)$ 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

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

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 $A(t_k)\phi^n := \nabla\cdot(\mathscr{A}(\cdot,t_k)\nabla\phi^n)$. Then, operator $A(t_k)$ has maximal L^p-regularity (cf. [22, Lemma 2.1]), i.e., the solution of the initial value problem

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

with the coefficient frozen at $t = t_k$, satisfies

$$\|\partial_t\psi\|_{L^p(\mathbb{R}_+;W^{-1,s}(\Omega))} + \|\psi\|_{L^p(\mathbb{R}_+;W^{1,s}(\Omega))} \leqslant c\|\chi\|_{L^p(\mathbb{R}_+;W^{-1,s}(\Omega))} \qquad \forall\, p,s\in(1,\infty), \tag{B.2}$$

$$\|\partial_t\psi\|_{L^p(\mathbb{R}_+;L^s(\Omega))} + \|A(t_k)\psi\|_{L^p(\mathbb{R}_+;L^s(\Omega))} \leqslant c\|\chi\|_{L^p(\mathbb{R}_+;L^s(\Omega))} \qquad \forall\, p,s\in(1,\infty). \tag{B.3}$$

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

$$\delta_\tau\phi^n - A(t_k)\phi^n = g^n \tag{B.4}$$

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

$$\delta_\tau\phi^n - A(t_k)\phi^n = g^n - (A(t_k)-A(t_n))\phi^n, \tag{B.5}$$

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

$$B_0 = 0 \quad\text{and}\quad B_n = \|(\phi^n)_{m=1}^n\|^p_{L^p(W^{1,s}(\Omega))} = \sum_{m=1}^n \|\phi^m\|^p_{W^{1,s}(\Omega)} \quad \text{for } 1\leqslant n\leqslant N$$

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

$$\begin{aligned}
&\|(\delta_\tau\phi^n)_{n=1}^k\|^p_{L^p(W^{-1,s}(\Omega))} + \|(\phi^n)_{n=1}^k\|^p_{L^p(W^{1,s}(\Omega))} \\
&\leqslant c\|(g^n)_{n=1}^k\|^p_{L^p(W^{-1,s}(\Omega))} + c\|((\mathscr{A}(\cdot,t_k)-\mathscr{A}(\cdot,t_n))\nabla\phi^n)_{n=1}^k\|^p_{L^p(L^s(\Omega))} \\
&\leqslant c\|(g^n)_{n=1}^k\|^p_{L^p(W^{-1,s}(\Omega))} + c\|((t_k-t_n)\phi^n)_{n=1}^k\|^p_{L^p(W^{1,s}(\Omega))} \\
&= c\|(g^n)_{n=1}^k\|^p_{L^p(W^{-1,s}(\Omega))} + c\sum_{n=1}^k |t_k-t_n|^p\|\phi^n\|^p_{W^{1,s}(\Omega)} \\
&= c\|(g^n)_{n=1}^k\|^p_{L^p(W^{-1,s}(\Omega))} + c\sum_{n=1}^k |t_k-t_n|^p(B_n-B_{n-1}) \\
&= c\|(g^n)_{n=1}^k\|^p_{L^p(W^{-1,s}(\Omega))} + c\sum_{n=1}^{k-1} (|t_k-t_n|^p-|t_k-t_{n+1}|^p)B_n \qquad \text{(summation by parts is used)} \\
&= c\|(g^n)_{n=1}^k\|^p_{L^p(W^{-1,s}(\Omega))} + c\sum_{n=1}^{k-1} (|t_k-t_n|^p-|t_k-t_{n+1}|^p)\|(\phi^n)_{m=1}^n\|^p_{L^p(W^{1,s}(\Omega))}
\end{aligned}$$

$$\leqslant c\|(g^n)_{n=1}^k\|^p_{L^p(W^{-1,s}(\Omega))} + c\sum_{n=1}^{k-1}\tau\|(\phi^n)_{m=1}^n\|^p_{L^p(W^{1,s}(\Omega))}. \tag{B.6}$$

With a discrete Gronwall inequality, we obtain from (B.6)

$$\|(\delta_\tau\phi^n)_{n=1}^k\|^p_{L^p(W^{-1,s}(\Omega))} + \|(\phi^n)_{n=1}^k\|^p_{L^p(W^{1,s}(\Omega))} \leqslant c\|(g^n)_{n=1}^k\|^p_{L^p(W^{-1,s}(\Omega))}. \tag{B.7}$$

This proves (2.16) for the nonautonomous equation (2.12).

We have shown that the autonomous equation (B.4) satisfies (2.17). For the nonautonomous equation (2.12), estimate (2.17) can be proved by the same perturbation argument as (B.5)–(B.6). □

Proof of (2.15) *and* (2.18). Let k be fixed, and define the operator $A_h(t_k) : S_h \to S_h$ by

$$(A_h(t_k)\varphi_h, \upsilon_h) = -(\mathscr{A}(\cdot,t_k)\nabla\varphi_h, \nabla\upsilon_h) \quad \forall\, \upsilon_h \in S_h. \tag{B.8}$$

Then, [22, (2.13)] and [31, Lemma 4.c] imply the R-boundedness (uniformly with respect to h) of the semigroup generated by the operator $A_h(t_k)$. Equivalently, the set of operators $\{z(z-A_h(t_k))^{-1} : \mathrm{Re}(z) \geqslant 0\}$ is R-bounded (cf. [30, Theorem 4.2]). Since the L^2 projection operator P_h is bounded with respect to the $W^{-1,q}(\Omega)$-norm, the R-boundedness of $\{z(z-A_h(t_k))^{-1}P_h : \mathrm{Re}(z) \geqslant 0\}$ and [15, Theorem 6.1] imply that the solution of the autonomous equation

$$\big(\delta_\tau\phi_h^n, \upsilon_h\big) + \big(\mathscr{A}(\cdot,t_k)\nabla\phi_h^n, \nabla\upsilon_h\big) = (g^n, \upsilon_h) \quad \forall\, \upsilon_h \in S_h, \tag{B.9}$$

with $\phi_h^0 = 0$, with the coefficient frozen at $t = t_k$, satisfies the discrete maximal L^p-regularity

$$\|(\delta_\tau\phi_h^n)_{n=1}^k\|_{L^p(L^s(\Omega))} + \|(A_h(t_k)\phi_h^n)_{n=1}^k\|_{L^p(L^s(\Omega))} \leqslant c\|(g^n)_{n=1}^k\|_{L^p(L^s(\Omega))} \quad \forall\, p,s \in (1,\infty). \tag{B.10}$$

Estimates (2.16) and (B.10) imply (2.15) for the autonomous equation (B.9); see [23, Lemma 3.5]. Then, a perturbation argument like (B.5)–(B.6) yields (2.15) for the nonautonomous equation (2.13).

Estimate (2.18) is a consequence of (2.15) and (2.17) via a duality argument; see [23, pp. 539–541]. □

C. Discrete Sobolev interpolation inequality

In (4.37), (4.42) and (4.47), we have used the following version of temporal discrete Sobolev interpolation inequality in the case $\theta_{h,\ell}^0 = \theta^0 = 0$:

$$\big\|(\theta_{h,\ell}^n)_{n=1}^k\big\|_{L^\infty(X)} \leqslant c\big\|(\theta_{h,\ell}^n)_{n=1}^k\big\|^{1-\frac{1}{p}}_{L^p(X)}\big\|\big(\delta_\tau\theta_{h,\ell}^n\big)_{n=1}^k\big\|^{\frac{1}{p}}_{L^p(X)} \quad \forall\, 1<p<\infty, \tag{C.1}$$

where X is a Banach space. This can be proved by applying the continuous version of the Sobolev interpolation inequality to the piecewise linear interpolant $\theta(t), t \in [0,T]$, of $\theta_{h,\ell}^0, \ldots, \theta_{h,\ell}^k$,

$$\theta(t) = \frac{t_n - t}{\tau}\theta_{h,\ell}^{n-1} + \frac{t-t_{n-1}}{\tau}\theta_{h,\ell}^n \quad \text{for } t \in [t_{n-1}, t_n],$$

which satisfies $\theta(0) = 0$. We have

$$\max_{t\in[0,t_k]}\|\theta(t)\|_{W^{1,q}(\Omega)} \leqslant c\|\theta\|^{1-\frac{1}{p}}_{L^p(0,t_k;W^{1,q}(\Omega))}\|\partial_t\theta\|^{\frac{1}{p}}_{L^p(0,t_k;W^{1,q}(\Omega))}, \tag{C.2}$$

which furthermore implies (C.1) because

$$\|\theta\|_{L^p(0,t_k;W^{1,q}(\Omega))} \leqslant c\big\|(\theta_{h,\ell}^n)_{n=1}^k\big\|_{L^p(W^{1,q}(\Omega))}$$

and

$$\|\partial_t\theta\|_{L^p(0,t_k;W^{1,q}(\Omega))} \leqslant c\big\|\big(\delta_\tau\theta_{h,\ell}^n\big)_{n=1}^k\big\|_{L^p(W^{1,q}(\Omega))}.$$

Acknowledgement

Work partially supported by the Research Grants Council of Hong Kong (Project No. 15301818).

REFERENCES

[1] G. Akrivis and B. Li, *Maximum norm analysis of implicit–explicit backward difference formulae for nonlinear parabolic equations*, IMA J. Numer. Anal. **38** (2018) 75–101.

[2] G. Akrivis, B. Li, and C. Lubich, *Combining maximal regularity and energy estimates for time discretizations of quasilinear parabolic equations*, Math. Comp. **86** (2017) 1527–1552.

[3] W. Arendt, C. J. Batty, M. Hieber, and F. Neubrander, *Vector-valued Laplace Transforms and Cauchy Problems*, second ed., Birkhäuser, Basel, 2011.

[4] A. Ashyralyev, S. Piskarev, and L. Weis, *On well-posedness of difference schemes for abstract parabolic equations in $L_p([0,T];E)$ spaces*, Numer. Funct. Anal. Optim. **23** (2002) 669–693.

[5] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, third ed., Texts in Applied Mathematics, Springer–Verlag, New York, 2008.

[6] M. Dauge, *Neumann and mixed problems on curvilinear polyhedra*, Integr. Equat. Oper. Th. **15** (1992) 227–261.

[7] L. C. Evans, *Partial Differential Equations*, second ed., Graduate Studies in Mathematics 19, AMS, Providence, RI, 2010.

[8] X. Feng, M. von Oehsen, and A. Prohl, *Rate of convergence of regularization procedures and finite element approximations for the total variation flow*, Numer. Math. **100** (2005) 441–456.

[9] X. Feng and A. Prohl, *Analysis of total variation flow and its finite element approximations*, ESAIM Math. Model. Numer. Anal. **37** (2003) 533–556.

[10] M. Geissert, *Discrete maximal L^p regularity for finite element operators*, SIAM J. Numer. Anal. **44** (2006) 677–698.

[11] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*. Pitman, Boston, MA, 1985.

[12] F. Hecht, *New development in FreeFEM++*, J. Numer. Math. **20** (2012) 251–265.

[13] T. Kemmochi, *Discrete maximal regularity for abstract Cauchy problems*, Studia Math. **234** (2016) 241–263.

[14] T. Kemmochi and N. Saito, Discrete maximal regularity and the finite element method for parabolic equations, Numer. Math. **138** (2018) 905–937.

[15] B. Kovács, B. Li, and C. Lubich, *A-stable discretizations preserve maximal parabolic regularity*, SIAM J. Numer. Anal. **54** (2016) 3600–3624.

[16] P. C. Kunstmann, B. Li, and C. Lubich, *Runge–Kutta time discretization of nonlinear parabolic equations studied via discrete maximal parabolic regularity*, Found. Comput. Math. **18** (2018) 1109–1130.

[17] D. Leykekhman and B. Vexler, *Discrete maximal parabolic regularity for Galerkin finite element methods*, Numer. Math. **135** (2017) 923–952.

[18] D. Leykekhman and B. Vexler, *Discrete maximal parabolic regularity for Galerkin finite element methods for non-autonomous parabolic problems*, SIAM J. Numer. Anal. **56** (2018) 2178–2202.

[19] B. Li, *Maximum-norm stability and maximal L^p regularity of FEMs for parabolic equations with Lipschitz continuous coefficients*, Numer. Math. **131** (2015) 489–516.

[20] B. Li, *Analyticity, maximal regularity and maximum-norm stability of semidiscrete finite element solutions of parabolic equations in nonconvex polyhedra*, Math. Comp. **88** (2019) 1–44.

[21] B. Li and W. Sun, *Linearized FE approximations to a nonlinear gradient flow*, SIAM J. Numer. Anal. **52** (2014) 2623–2646.

[22] B. Li and W. Sun, *Maximal L^p analysis of finite element solutions for parabolic equations with nonsmooth coefficients in convex polyhedra*, Math. Comp. **86** (2017) 1071–1102.

[23] B. Li and W. Sun, *Maximal regularity of fully discrete finite element solutions of parabolic equations*, SIAM J. Numer. Anal. **55** (2017) 521–542.

[24] A. Lichnewsky and R. Temam, *Pseudosolutions of the time-dependent minimal surface problem*, J. Differential Equations **30** (1978) 340–364.

[25] A. Lunardi, *Analytic Semigroups and Optimal Regularity in Parabolic Problems*, Birkhäuser Verlag, Basel, 1995.

[26] D. Meidner and B. Vexler, *Optimal error estimates for fully discrete Galerkin approximations of semilinear parabolic equations*, ESAIM: Math. Model. Numer. Anal. **52** (2018) 2307–2325.

[27] E.-M. Ouhabaz, *Gaussian estimates and holomorphy of semigroups*, Proc. Amer. Math. Soc. **123** (1995) 1465–1474.

[28] E. Spadaro and U. Stefanelli, *A variational view at the time-dependent minimal surface equation*, J. Evol. Equ. **11** (2011) 793–809.

[29] L. B. Wahlbin, *Local behavior in finite element methods*, "Handbook of Numerical Analysis II: Finite Element Methods", pp. 353–522, edited by P. G. Ciarlet and J. L. Lions. North-Holland, Amsterdam, 1991

[30] L. Weis, *Operator-valued Fourier multiplier theorems and maximal L_p-regularity*, Math. Ann. **319** (2001) 735–758.

[31] L. Weis, *A new approach to maximal L_p-regularity*, Lecture Notes in Pure and Applied Mathematics 215: Evolution Equations and Their Applications in Physical and Life Sciences, pp. 195–214, edited by G. Lumer. Marcel Dekker, New York, 2001.