SIAM J. SCI. COMPUT.
Vol. 37, No. 5, pp. S503–S525

FOSLL* FOR NONLINEAR PARTIAL DIFFERENTIAL EQUATIONS*

EUNJUNG LEE†, THOMAS A. MANTEUFFEL‡, AND CHAD R. WESTPHAL§

Abstract. In previous work, the first-order system LL^* (FOSLL*) method was developed for linear partial differential equations. This approach seeks to minimize the residual of the equations in a dual norm induced by the differential operator, yielding approximations accurate in $L^2(\Omega)$ rather than $H^1(\Omega)$ or $H(\text{Div})$. In this paper, the general framework of FOSLL* is extended to a wide range of nonlinear problems. Four approaches to propagating an inexact Newton iteration based on a FOSLL* approximation are presented, and theory for robust convergence in $L^2(\Omega)$ is established. Numerical results are presented for two formulations of the steady incompressible Navier–Stokes equations and for a diffusion equation with reduced regularity due to a discontinuous diffusion coefficient.

Key words. nonlinear partial differential equations, Newton's method, least-squares finite element methods, first-order system LL* method, Navier–Stokes

AMS subject classifications. 65N30, 65N12, 65N15

DOI. 10.1137/140974353

1. Introduction.

We consider approximating the solution to nonlinear partial differential equations (PDE) by combining least-squares minimization principles with an efficient nonlinear outer iteration. In particular, we focus on problems for which standard L^2 least-squares approaches do not produce a robust approximation in the L^2 norm. By weakening the variational problem with the first-order system LL^* (FOSLL*) approach, it is possible to enhance overall L^2 convergence by first finding a smooth preimage of the solution, then using a postprocessing step to reconstruct the approximate solution. In a nonlinear setting, this approach requires an additional level of investigation that has not been considered in the literature before.

To illustrate the discretization framework, first consider the first-order system least-squares (FOSLS) approach to linear PDE. Let $L\mathbf{u} = \mathbf{f}$ denote a first-order PDE system, possibly occurring directly from an application (e.g., the first-order stress/velocity/pressure form of Stokes equations) or possibly reformulated from a higher-order problem by introducing additional unknowns. The solution is found by minimizing the FOSLS functional, $\mathcal{F}(\mathbf{u}; \mathbf{f}) := \|L\mathbf{u} - \mathbf{f}\|^2$, in an appropriate Hilbert space, $\mathcal{V}$. The choice of $\mathcal{V}$ depends on the particular problem, but is generally chosen based on the preimage of $L^2(\Omega)$ under the operator L, that is, the set of functions, $\mathbf{u}$, such that $L\mathbf{u} \in L^2$ and that satisfy appropriate boundary conditions and assumptions required by the FOSLS framework. Typical examples are products of $H^1(\Omega)$, $H(\text{Div})$, $H(\text{Curl})$, and $H(\text{Div}) \cap H(\text{Curl})$. Discrete approximation is accomplished by

*Received by the editors July 8, 2014; accepted for publication (in revised form) June 8, 2015; published electronically October 29, 2015.
http://www.siam.org/journals/sisc/37-5/97435.html
†Department of Computational Science and Engineering, Yonsei University, Seoul 120-749, South Korea (eunjunglee@yonsei.ac.kr). This author's work was supported by Basic Science Research Program through the NRF of Korea 2013R1A1A1004836.
‡Department of Applied Mathematics, University of Colorado at Boulder, Boulder, CO 80309 (tmanteuf@colorado.edu). This author's work was supported in part by the U.S. Department of Energy under grants (Sc) DE-FC02-03ER25574 and (NNSA) DE-NA0002376, Lawrence Livermore National Laboratory under contract B600360, and by AFOSR under contract FA95501210487.
§Department of Mathematics and Computer Science, Wabash College, Crawfordsville, IN 47933 (westphac@wabash.edu). This author's work was supported by the National Science Foundation under grant DMS-1216297.

restricting the minimization to an appropriate finite-dimensional subspace, $\mathcal{V}^h \subset \mathcal{V}$.

An important property for the FOSLS framework is that the homogeneous functional norm, $\|L\mathbf{u}\|$, is coercive and continuous in the Hilbert space, $\mathcal{V}$; that is, there exist constants, $0 < c_0 \leq c_1$, such that

$$c_0\|\mathbf{u}\|_{\mathcal{V}} \leq \|L\mathbf{u}\| \leq c_1\|\mathbf{u}\|_{\mathcal{V}} \qquad \forall\ \mathbf{u} \in \mathcal{V}. \tag{1.1}$$

When L satisfies (1.1), the operator L is called $\mathcal{V}$-elliptic. This guarantees a unique solution of the associated weak problem via the Riesz representation theorem. Moreover, the discrete minimization inherits these properties and, thus, involves a symmetric positive definite linear system. This also implies that the discrete system is stable: the Ladyzhenskaya–Babuška–Brezzi (LBB) stability condition is automatically satisfied, and the finite element spaces for each variable can be chosen independently. Continuity of the functional implies that discrete error bounds may be established by standard interpolation bounds in $\mathcal{V}$. In addition, with sufficient regularity of L^*L, the L^2 norm of the error converges at an enhanced rate [35]. Background on least-squares finite element methods can be found in [9, 10, 16, 18, 19].

However, the FOSLS methodology also has some limitations. When the solution of the original problem is not sufficiently smooth, for example, due to discontinuous coefficients, nonsmooth domain boundary, or certain boundary conditions, the $\mathcal{V}$ norm equivalence in (1.1) and its associated properties may be compromised. For example, the ratio c_1/c_0 may become large or the space $\mathcal{V}$ may change. Certain components of the solution may be in $H(\mathrm{Div}) \cap H(\mathrm{Curl})$ but not in $H^1(\Omega)$ [6, 30, 31]. In the latter case, the use of standard H^1-conforming finite element spaces with a standard L^2-FOSLS functional will not result in convergence of the discrete approximation. These difficulties may be overcome by a number of remedies. The finite element space can be enhanced so that the limit is dense in $H(\mathrm{Div}) \cap H(\mathrm{Curl})$ [6, 5]. In some cases, the problem can be formulated in $H(\mathrm{Div})$ or $H(\mathrm{Curl})$ rather than $H^1(\Omega)$, allowing the use of Raviart–Thomas or Nédélec finite element spaces (see, e.g., [39, 40]). Additionally, problems with boundary singularities can be treated with a weighted least-squares approach (see, e.g., [32, 33]). Each of these approaches admits convergence in a weaker norm than $H^1(\Omega)$, but requires extra consideration in implementation and in pairing the discretization with a robust solver for the resulting algebraic systems.

As an alternative to the standard L^2 least-squares approach, many H^{-1} norm approaches have been introduced, where the error is essentially controlled in $L^2(\Omega)$ rather than in Sobolev subspaces of $L^2(\Omega)$ (e.g., see [8, 13, 12]). In practice, the negative norms must be approximated, and in many cases the resulting algebraic systems have poor conditioning relative to the analogous FOSLS approach. The FOSLL* approach introduces a dual norm induced by the problem operator and results in a computable variational problem that naturally minimizes the L^2 error. In [20, 32, 36, 21], this approach is presented and studied in the context of scalar elliptic problems and applications in electrostatics. In [21], the FOSLL* method is extended to $H(\mathrm{Div})$ formulations of second-order, elliptic problems, and a hybrid FOSLS-FOSLL* approach is given in [34], where approximations are nearly optimal with respect to a graph norm that balances the error in the L^2 norm and H^1 seminorm.

In this paper, we expand the general FOSLL* methodology to nonlinear problems, where the nonlinear PDE is linearized via Newton's method. Combining a Newton linearization and a least-squares discretization has been used successfully in many contexts (e.g., see [3, 22, 25, 37, 39, 40]). An issue of both theoretical and practical importance for such nonlinear problems is that, in general, FOSLL* yields a discontinuous approximation that is no longer in the domain of the operator. This potential lack of smoothness of the iterates is a major consideration in this paper, since a part of the premise of a FOSLL* approach is to sacrifice some smoothness of the approximation in favor of more robust L^2 convergence. We present four approaches for the continuation of the iterates, either allowing the freedom of discontinuous solutions or introducing measures to ensure continuity of iterates. Our numerical results suggest that preserving smoothness works well when the exact solution has sufficiently local smoothness, while allowing a discontinuous approximation appears superior when the solution is discontinuous.

The idea of weakening the variational problem by allowing discontinuous approximations is a popular approach in discontinuous Galerkin methods [24, 29], and the goal of seeking better direct control of the L^2 error is realized in the recent advent of discontinuous Petrov–Galerkin (DPG) [23, 26] methods. The general FOSLL* methodology shares some obvious similarities with these approaches.

The organization of this paper is as follows. Section 2 establishes basic notation and recalls some well-known theory needed for our theoretical results. In section 3, the main algorithmic details are presented, and we discuss four alternatives for continuing the nonlinear iteration. Theoretical results for nonlinear convergence and finite element convergence when the Newton iterates have reduced regularity are presented in sections 4 and 5, respectively. Numerical results are given in section 6, and brief concluding remarks are found in section 7.

2. Notation and definitions.

Let Ω be an open, bounded, Lipschitz-continuous subset of $\mathbb{R}^d$, $d = 2, 3$. For each integer, $m \geq 0$, and real p with $1 \leq p \leq \infty$, $W^{m,p}(\Omega)$ denotes the standard Sobolev space with corresponding norm $\|\cdot\|_{W^{m,p}(\Omega)}$. For fractional-order Sobolev spaces, we denote by $W^{s,p}(\Omega)$, with $s = m + \sigma$ and $0 < \sigma < 1$, a Banach space with the norm

$$\|u\|_{W^{s,p}(\Omega)} = \left(\|u\|^p_{W^{m,p}(\Omega)} + \sum_{|\alpha|=m} \int_\Omega \int_\Omega \frac{|\partial^\alpha u(\mathbf{x}) - \partial^\alpha u(\mathbf{y})|^p}{\|\mathbf{x} - \mathbf{y}\|^{d+\sigma p}} d\mathbf{x} d\mathbf{y} \right)^{\frac{1}{p}}.$$

Denoting $\mathscr{C}_0^\infty(\Omega)$ as the space of $\mathscr{C}^\infty(\Omega)$ functions with compact support, we define $W_0^{s,p}(\Omega)$ with $s > 0$ as the closure of $\mathscr{C}_0^\infty(\Omega)$ in the $W^{s,p}(\Omega)$ norm and $W^{-s,p'}(\Omega)$ as the dual space of $W_0^{s,p}(\Omega)$ with $1/p + 1/p' = 1$. The space $W^{s,2}(\Omega)$ is usually denoted as $H^s(\Omega)$, and $W^{0,p}(\Omega)$ is denoted as $L^p(\Omega)$. For $L^2(\Omega)$, we omit the subscript in the norm, $\|\cdot\|$, and the L^2 inner product, $\langle\cdot,\cdot\rangle$, which also denotes the duality pairing between spaces and their dual. We use $\mathcal{D}(\mathcal{K})$ to indicate the domain of operator $\mathcal{K}$.

We will make use of the following well-known result regarding imbedding of spaces (see, e.g., Chapter 4 of [1]).

LEMMA 1. *Let $\Omega \subset \mathbb{R}^d$ be a domain having the cone property and let $0 < r < \infty$ and $1 \leq p < \infty$.*

(i) *If $rp < d$, then $W^{r,p}(\Omega) \hookrightarrow L^q(\Omega)$ for $p \leq q \leq dp/(d - rp)$.*

(ii) *If $rp = d$, then $W^{r,p}(\Omega) \hookrightarrow L^q(\Omega)$ for $p \leq q < \infty$.*

Where not denoted explicitly, we use the c with or without subscript to denote a generic positive constant, possibly different at different occurrences.

3. Main ideas.

We consider a system of nonlinear first-order PDE,

$$\mathcal{L}(\mathbf{U}) = \mathbf{F}, \tag{3.1}$$

where $\mathcal{D}(\mathcal{L})$ incorporates boundary conditions on $\mathbf{U}$. In this section, we first briefly introduce Newton's method as an outer iteration to define a sequence of linear problems related to (3.1), and then consider the FOSLS and FOSLL* methods for numerically approximating the solutions to each linearized problem. In particular, we focus on the resulting Newton-FOSLL* approach and highlight the differences between it and the more well-studied Newton-FOSLS approach.

The Gâteaux derivative of the operator, $\mathcal{L}(\mathbf{U})$, is given by

$$\mathcal{L}'(\mathbf{U})\mathbf{W} = \lim_{\alpha \to 0} \frac{\mathcal{L}(\mathbf{U} + \alpha\mathbf{W}) - \mathcal{L}(\mathbf{U})}{\alpha}. \tag{3.2}$$

If the Gâteaux derivative is linear, this is also known as the Fréchet derivative, which is a linear operator acting on the function $\mathbf{W}$. Higher Fréchet derivatives are defined analogously. The basic idea of Newton's method is that, given an approximation, $\mathbf{U}_n$, we seek the next approximation by

$$\mathbf{U}_{n+1} = \mathbf{U}_n + \mathcal{L}'(\mathbf{U}_n)^{-1}(\mathbf{F} - \mathcal{L}(\mathbf{U}_n)), \tag{3.3}$$

which may be viewed as a fixed point iteration. By defining $\boldsymbol{\delta}\mathbf{U} = \mathbf{U}_{n+1} - \mathbf{U}_n$, we may view the process of one Newton iteration as solving

$$\mathcal{L}'(\mathbf{U}_n)\boldsymbol{\delta}\mathbf{U} = \mathbf{F} - \mathcal{L}(\mathbf{U}_n) \tag{3.4}$$

for $\boldsymbol{\delta}\mathbf{U} \in D(\mathcal{L})$, then updating with $\mathbf{U}_{n+1} = \mathbf{U}_n + \boldsymbol{\delta}\mathbf{U}$.

Note that this outer linearization scheme is formulated on the continuous level. We now consider approaches for approximating $\boldsymbol{\delta}\mathbf{U}$ on the discrete level.

3.1. FOSLS. A FOSLS approach to approximating $\boldsymbol{\delta}\mathbf{U}$ is to define the functional,

$$\mathcal{F}(\boldsymbol{\delta}\mathbf{U}) := \|\mathcal{L}'(\mathbf{U}_n)\boldsymbol{\delta}\mathbf{U} - (\mathbf{F} - \mathcal{L}(\mathbf{U}_n))\|^2, \tag{3.5}$$

and seek a numerical approximation, $\boldsymbol{\delta}\mathbf{U}^h \in \mathcal{V}^h$, where $\mathcal{V}^h$ is a finite element subspace of $\mathcal{D}(\mathcal{L})$, such that $\boldsymbol{\delta}\mathbf{U}^h$ minimizes $\mathcal{F}$ over $\mathcal{V}^h$.

Suppose that the norm induced by $\mathcal{L}'(\mathbf{U}_n)$ is equivalent to the H^1 norm, that is, that there are positive constants, c_0 and c_1, such that

$$c_0\|\mathbf{V}\|_{H^1(\Omega)} \le \|\mathcal{L}'(\mathbf{U}_n)\mathbf{V}\| \le c_1\|\mathbf{V}\|_{H^1(\Omega)} \quad \forall\ \mathbf{V} \in H^1(\Omega) \cap D(\mathcal{L}). \tag{3.6}$$

Further, assume that $\mathcal{V}^h$ admits an interpolation operator, $\mathcal{I}^h$, with the approximation property

$$\|\mathbf{V} - \mathcal{I}^h\mathbf{V}\|_{H^k(\Omega)} \le ch^{1+s-k}\|\mathbf{V}\|_{H^{1+s}(\Omega)} \tag{3.7}$$

for $0 \le k \le s \le q$ and all $\mathbf{V} \in H^{1+s}(\Omega)$, for some $q > 0$, where $h > 0$ is a discretization parameter [11, 14]. The discretized Newton step is, thus, well defined. Given $\mathbf{U}_n$, denote $\boldsymbol{\delta}\mathbf{U}$ as the exact Newton increment and $\mathbf{U}_{n+1}$ as the exact next iterate. Error bounds on their finite element approximations, $\boldsymbol{\delta}\mathbf{U}^h$ and $\mathbf{U}^h_{n+1}$, follow directly from (3.6) and the minimization of $\mathcal{F}$:

$$\begin{aligned}\|\mathbf{U}^h_{n+1} - \mathbf{U}_{n+1}\|_{H^1(\Omega)} = \|\boldsymbol{\delta}\mathbf{U}^h - \boldsymbol{\delta}\mathbf{U}\|_{H^1(\Omega)} &\le c_0^{-1}\|\mathcal{L}'(\mathbf{U}_n)(\boldsymbol{\delta}\mathbf{U}^h - \boldsymbol{\delta}\mathbf{U})\| \\ &\le c_1/c_0\|\mathcal{I}^h\boldsymbol{\delta}\mathbf{U} - \boldsymbol{\delta}\mathbf{U}\|_{H^1(\Omega)} \le c\,h^s\|\boldsymbol{\delta}\mathbf{U}\|_{H^{1+s}(\Omega)}\end{aligned}$$

for $0 < s \le q$. Similar results can be established when $\mathcal{V}$ ellipticity is in a product norm that includes $H(\mathrm{Div})$ or $H(\mathrm{Curl})$ [16, 39, 40].

Now, if $\mathbf{U}_n \in \mathcal{V}^h \subset \mathcal{D}(\mathcal{L})$, then $\mathbf{U}_{n+1} \in \mathcal{V}^h \subset \mathcal{D}(\mathcal{L})$ and the next iteration is well defined. This Newton-FOSLS approach has been applied in a variety of settings [4, 3, 22, 25, 37, 39, 40] and tends to be successful when the original system is sufficiently regular and the equivalence between $\|\cdot\|_{H^1(\Omega)}$ and $\|\mathcal{L}'(\mathbf{U}_n)\cdot\|$ is strong (i.e., c_0 and c_1 are near 1). The absence of these favorable conditions motivates us to consider a FOSLL* discretization of (3.4). In particular, we seek an approach that yields more direct control of the error in the L^2 norm, rather than the H^1 norm through (3.6), or the $\mathcal{V}$ norm through (1.1) when $\mathcal{V}$ is a weaker norm.

3.2. FOSLL*. To define the FOSLL* approach to approximating $\boldsymbol{\delta}\mathbf{U}$, we first define $\mathcal{L}'(\mathbf{U}_n)^*$ as the adjoint of $\mathcal{L}'(\mathbf{U}_n)$, where

$$\langle \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}, \mathbf{Z}\rangle = \langle \mathbf{W}, \mathcal{L}'(\mathbf{U}_n)\mathbf{Z}\rangle$$

for all $\mathbf{Z} \in \mathcal{D}(\mathcal{L}'(\mathbf{U}_n))$ and for all $\mathbf{W} \in \mathcal{D}(\mathcal{L}'(\mathbf{U}_n)^*)$. Rather than to directly seek an approximation to $\boldsymbol{\delta}\mathbf{U}$, we first seek $\mathbf{W}$ such that $\mathcal{L}'(\mathbf{U}_n)^*\mathbf{W} = \boldsymbol{\delta}\mathbf{U}$, which will have a solution under the very weak assumption that $\mathcal{L}'(\mathbf{U}_n)$ is coercive in L^2 [34]. To this end, define the functional,

$$\begin{aligned} \mathcal{G}(\mathbf{W}) &:= \|\mathcal{L}'(\mathbf{U}_n)^*\mathbf{W} - \boldsymbol{\delta}\mathbf{U}\|^2 \\ &= \|\mathcal{L}'(\mathbf{U}_n)^*\mathbf{W} - \mathcal{L}'(\mathbf{U}_n)^{-1}(\mathbf{F} - \mathcal{L}(\mathbf{U}_n))\|^2. \end{aligned} \tag{3.8}$$

The minimization of $\mathcal{G}$ over $\mathcal{D}(\mathcal{L}'(\mathbf{U}_n)^*)$ is equivalent to finding $\mathbf{W} \in \mathcal{D}(\mathcal{L}'(\mathbf{U}_n)^*)$ such that

$$\begin{aligned} \langle \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}, \mathcal{L}'(\mathbf{U}_n)^*\mathbf{Z}\rangle &= \langle \boldsymbol{\delta}\mathbf{U}, \mathcal{L}'(\mathbf{U}_n)^*\mathbf{Z}\rangle = \langle \mathcal{L}'(\mathbf{U}_n)\boldsymbol{\delta}\mathbf{U}, \mathbf{Z}\rangle \\ &= \langle \mathbf{F} - \mathcal{L}(\mathbf{U}_n), \mathbf{Z}\rangle \quad \forall\ \mathbf{Z} \in \mathcal{D}(\mathcal{L}'(\mathbf{U}_n)^*). \end{aligned} \tag{3.9}$$

An important feature to recognize here is that, in the functional and variational problem above, $\boldsymbol{\delta}\mathbf{U}$ represents the exact Newton step, but $\mathcal{G}$ is minimized without explicitly using $\boldsymbol{\delta}\mathbf{U}$. The Newton update is given by

$$\mathbf{U}_{n+1} = \mathbf{U}_n + \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}.$$

Under proper smoothness assumptions, the FOSLL* functional, $\mathcal{G}(\mathbf{W})$, will be equivalent to the L^2 norm of the error, and approximations will have strong control over the error in the L^2 norm.

The discrete problem is analogous. Define a finite element space, $\mathcal{W}^h \subset \mathcal{D}(\mathcal{L}'(\mathbf{U}_n)^*)$, where $\mathcal{W}^h$ satisfies an approximation property as in (3.7). Minimizing $\mathcal{G}$ over $\mathcal{W}^h$ is done by finding $\mathbf{W}^h \in \mathcal{W}^h$ such that

$$\langle \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h, \mathcal{L}'(\mathbf{U}_n)^*\mathbf{Z}^h\rangle = \langle \mathbf{F} - \mathcal{L}(\mathbf{U}_n), \mathbf{Z}^h\rangle \quad \forall\ \mathbf{Z}^h \in \mathcal{W}^h. \tag{3.10}$$

The discrete Newton step is, thus, given by

$$\mathbf{U}_{n+1} = \mathbf{U}_n + \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h.$$

Despite its merit in finding an L^2 approximation to problems with low regularity, continuation of the Newton iteration by using an approximate solution to the discrete FOSLL* system has a limitation. Suppose that $\mathcal{W}^h \subset \mathscr{C}^0$, but $\mathcal{W}^h \not\subset \mathscr{C}^1$, which is the case with standard finite element spaces. Then, in general, the Newton-FOSLL* approach described above will have $\mathbf{U}_{n+1} \notin \mathscr{C}^0$ and, in general, not in $\mathcal{D}(\mathcal{L})$. Continuing to the next step requires the evaluation of $\mathcal{L}(\mathbf{U}_{n+1})$, which may not be well defined globally. To address this issue, we consider several modifications that result in a well-defined iteration.

3.2.1. $\mathscr{C}^1$ elements. For many nonlinear PDE, taking $\mathbf{U}_n \in \mathscr{C}^0$ and using $\mathcal{W}^h \subset \mathscr{C}^1$ ensures that $\mathbf{U}_{n+1} = \mathbf{U}_n + \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h \in \mathscr{C}^0$, and the iteration can proceed normally since $\mathcal{L}(\mathbf{U}_n)$ will always be well defined. This can be accomplished in practice by using specialized elements such as the Argyris triangular element, the Bogner–Fox–Schmidt rectangular element, or others described in [14, 11, 38, 41].

Let $\delta\mathbf{U} = \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}$ be the exact Newton step such that $\mathbf{U} = \mathbf{U}_n + \delta\mathbf{U}$, and let $\mathbf{W}^h \in \mathcal{W}^h$ be the solution of (3.10) so that the approximation of the correction satisfies $\delta\mathbf{U}^h = \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h \in \mathscr{C}^0$. Assuming $\mathcal{L}'(\mathbf{U}_n)^*$ satisfies a continuity bound like (1.1), where $H^1(\Omega) \subset \mathcal{V}$, we have

$$\begin{aligned} \|\delta\mathbf{U} - \delta\mathbf{U}^h\| &= \|\mathcal{L}'(\mathbf{U}_n)^*(\mathbf{W} - \mathbf{W}^h)\| \le \|\mathcal{L}'(\mathbf{U}_n)^*(\mathbf{W} - \mathcal{I}^h\mathbf{W})\| \\ &\le c\,\|\mathbf{W} - \mathcal{I}^h\mathbf{W}\|_{H^1(\Omega)} \le c\,h^s\|\mathbf{W}\|_{H^{1+s}(\Omega)} \le c\,h^s\|\delta\mathbf{U}\|_{H^s(\Omega)} \end{aligned} \tag{3.11}$$

for $0 < s \le q$, where $\mathcal{I}^h$ is an interpolation operator into $\mathcal{W}^h$.

This approach has the obvious advantage of being able to control the smoothness of the iteration directly through the approximation properties of the finite element spaces. However, such spaces are not common in practice, and furthermore, they generally lead to more dense linear systems that are more difficult for iterative solvers to handle efficiently than those from more standard spaces. We, therefore, focus on the more practical alternative approaches below.

3.2.2. L^2 projections. Once $\boldsymbol{\delta}\mathbf{U}^h = \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h \notin \mathscr{C}^0$ is computed, it may be projected into $\mathcal{D}(\mathcal{L})$ via an L^2-orthogonal projection. By choosing a finite element space $\mathcal{X}^h \subset \mathcal{D}(\mathcal{L})$ in which to represent the Newton step, we define $\Pi^h\boldsymbol{\delta}\mathbf{U}^h$ to satisfy

$$\|\Pi^h\boldsymbol{\delta}\mathbf{U}^h - \boldsymbol{\delta}\mathbf{U}^h\| \le \|\mathbf{V}^h - \boldsymbol{\delta}\mathbf{U}^h\| \quad \forall\ \mathbf{V}^h \in \mathcal{X}^h,$$

and the iteration proceeds with

$$\mathbf{U}^h_{n+1} = \mathbf{U}_n + \Pi^h\boldsymbol{\delta}\mathbf{U}^h = \mathbf{U}_n + \Pi^h\mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h.$$

This approach ensures continuity of iterates, but requires additional work to perform the projection.

Let $\boldsymbol{\delta}\mathbf{U} = \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W} \in H^r(\Omega)$ for $r > 0$ be the exact Newton step such that $\mathbf{U}_{n+1} = \mathbf{U}_n + \delta\mathbf{U}$, and let $\mathbf{W}^h \in \mathcal{W}^h$ be the solution of (3.10) so that the approximation of the correction satisfies $\delta\mathbf{U}^h = \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h$. Using approximation properties of $\mathcal{W}^h$, assuming $\mathcal{X}^h$ admits a bound like (3.7), and assuming $\mathcal{L}'(\mathbf{U}_n)^*$ is H^1-elliptic, we have

$$\begin{aligned} \|\boldsymbol{\delta}\mathbf{U} - \Pi^h\boldsymbol{\delta}\mathbf{U}^h\| &\le \|\boldsymbol{\delta}\mathbf{U} - \boldsymbol{\delta}\mathbf{U}^h\| + \|\boldsymbol{\delta}\mathbf{U}^h - \Pi^h\boldsymbol{\delta}\mathbf{U}^h\| \\ &\le \|\mathcal{L}'(\mathbf{U}_n)^*(\mathbf{W} - \mathbf{W}^h)\| + \|(I - \Pi^h)\mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h\| \\ &\le c_1\,h^r\|\mathbf{W}\|_{H^{1+r}(\Omega)} + c_2\,h^s\|\mathbf{W}^h\|_{H^{1+s}(\Omega)} \\ &\le c\,h^r\|\boldsymbol{\delta}\mathbf{U}\|_{H^r(\Omega)} + c_2\,h^s\|\mathbf{W}^h\|_{H^{1+s}(\Omega)} \end{aligned}$$

for some $0 < r, s \le q$, which indicates that the projection preserves the convergence to the solution. Even though the projected approximation, $\mathbf{U}^h_{n+1}$, is in $H^1(\Omega)$, the convergence remains in $L^2(\Omega)$. The implication of this is that, even if the exact solution is not in $H^1(\Omega)$, the approximations still converge in $L^2(\Omega)$, which is the motivating reason for using FOSLL*. Similar bounds hold when $\mathcal{V}$ ellipticity holds in a weaker norm than $H^1(\Omega)$.

3.2.3. Extended system. If, as previously, we wish to have the discrete representation of the Newton step in a $\mathscr{C}^0$ finite element space, then we may couple the FOSLL* solution for $\mathbf{W}$ with the L^2 projection for $\boldsymbol{\delta}\mathbf{U}$, approximating these simultaneously.

Define the extended functional

$$\mathcal{E}(\mathbf{W}, \boldsymbol{\delta}\mathbf{U}) = \|\mathcal{L}'(\mathbf{U}_n)^*\mathbf{W} - \mathcal{L}'(\mathbf{U}_n)^{-1}(\mathbf{F} - \mathcal{L}(\mathbf{U}_n))\|^2 + \|\mathcal{L}'(\mathbf{U}_n)^*\mathbf{W} - \boldsymbol{\delta}\mathbf{U}\|^2,$$

to be minimized for $\mathbf{W} \in \mathcal{D}(\mathcal{L}'(\mathbf{U}_n)^*)$ and $\boldsymbol{\delta}\mathbf{U} \in \mathcal{D}(\mathcal{L})$. If $\mathbf{W}$ and $\boldsymbol{\delta}\mathbf{U}$ are the minimizers, then $\boldsymbol{\delta}\mathbf{U} = \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}$. The discrete version is to minimize over $\mathbf{W}^h \in \mathcal{W}^h \subset \mathcal{D}(\mathcal{L}'(\mathbf{U}_n)^*)$ and $\boldsymbol{\delta}\mathbf{U}^h \in \mathcal{X}^h \subset \mathcal{D}(\mathcal{L})$.

Under the same assumptions used in subsection 3.2.2, an error bound can be derived as follows:

$$\begin{aligned}
&\|\boldsymbol{\delta}\mathbf{U} - \boldsymbol{\delta}\mathbf{U}^h\|^2 = \|\boldsymbol{\delta}\mathbf{U} - \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h + \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h - \boldsymbol{\delta}\mathbf{U}^h\|^2 \\
&\le 2\left(\|\mathcal{L}'(\mathbf{U}_n)^{-1}(\mathbf{F} - \mathcal{L}(\mathbf{U}_n)) - \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h\|^2 + \|\mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h - \boldsymbol{\delta}\mathbf{U}^h\|^2\right) \\
&= 2\mathcal{E}(\mathbf{W}^h, \boldsymbol{\delta}\mathbf{U}^h) \ \le\ 2\mathcal{E}(\mathcal{I}^h\mathbf{W}, \mathcal{I}^h(\boldsymbol{\delta}\mathbf{U})) \\
&= 2\left(\|\mathcal{L}'(\mathbf{U}_n)^*(\mathbf{W} - \mathcal{I}^h\mathbf{W})\|^2 + \|\mathcal{L}'(\mathbf{U}_n)^*\mathcal{I}^h\mathbf{W} - \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W} + \boldsymbol{\delta}\mathbf{U} - \mathcal{I}^h(\delta\mathbf{U})\|^2\right) \\
&\le c_1\left(h^s\|\mathbf{W}\|_{H^{1+s}(\Omega)} + h^s\|\boldsymbol{\delta}\mathbf{U}\|_{H^s(\Omega)}\right) \le c\, h^s\|\boldsymbol{\delta}\mathbf{U}\|_{H^s(\Omega)}
\end{aligned}$$

for some $0 < s \le q$, where $\mathcal{I}^h$ is an interpolation operator. Again, we note that the approximations are in $\mathcal{D}(\mathcal{L})$, but the convergence is in $L^2(\Omega)$.

3.2.4. Weak functional and *weak*-weak form. Subsections 3.2.2–3.2.3 above examine ways in which smoothness of iterates may be preserved in the Newton-FOSLL* framework. Here we consider a type of formulation that does not require the special considerations discussed above. If the system (3.1) can be written as

$$\mathcal{L}(\mathbf{U}) \;=\; L_0\mathbf{U} + L_1\mathcal{N}_1(\mathbf{U}) + \mathcal{N}_2(\mathbf{U}), \tag{3.12}$$

where L_0, L_1 are linear differential operators and $\mathcal{N}_1(\mathbf{U}), \mathcal{N}_2(\mathbf{U})$ are nonlinear operators which involve no derivatives of $\mathbf{U}$, then we can restate the minimization. First, note that

$$\mathcal{L}'(\mathbf{U})\mathbf{V} \;=\; L_0\mathbf{V} + L_1\mathcal{N}_1'(\mathbf{U})\mathbf{V} + \mathcal{N}_2'(\mathbf{U})\mathbf{V}, \tag{3.13}$$

where $\mathcal{N}_1'(\mathbf{U}), \mathcal{N}_2'(\mathbf{U})$ involve no derivatives of $\mathbf{U}$. Thus, the minimization of (3.8) is equivalent to finding $\mathbf{W} \in \mathcal{D}(\mathcal{L}'(\mathbf{U}_n)^*)$ such that

$$\langle\mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}, \mathcal{L}'(\mathbf{U}_n)^*\mathbf{Z}\rangle = \langle\mathbf{F} - \mathcal{N}_2(\mathbf{U}_n), \mathbf{Z}\rangle - \langle\mathbf{U}_n, L_0^*\mathbf{Z}\rangle - \langle\mathcal{N}_1(\mathbf{U}_n), L_1^*\mathbf{Z}\rangle \tag{3.14}$$

for all $\mathbf{Z} \in \mathcal{D}(\mathcal{L}'(\mathbf{U}_n)^*)$. Then, as above,

$$\mathbf{U}_{n+1} = \mathbf{U}_n + \delta\mathbf{U}_n = \mathbf{U}_n + \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}. \tag{3.15}$$

Discrete approximation follows in the natural way by restricting (3.14) to a finite element space, $\mathcal{W}^h$. If $\mathcal{W}^h$ is a typical $\mathscr{C}^0$ finite element space of piecewise polynomials, then $\delta\mathbf{U}_n = \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h \in H^{1/2-\nu}(\Omega)$ for $\nu > 0$. In this case, the sequence of approximate solutions is of reduced regularity, that is, $\mathbf{U}_n \in H^{\epsilon}(\Omega)$ for some $\epsilon > 0$. (We remark that, if $\mathbf{U}_n^h$ are piecewise smooth, then products of dependent variables are also in $H^{1/2-\nu}(\Omega)$.)

We now pose the broader question: given $\mathbf{U}_n \in H^{\epsilon}(\Omega)$, under what conditions can the exact Newton iteration be continued, and under what conditions will it converge? If $\mathbf{U}_n \in H^{\epsilon}(\Omega)$, $\mathbf{F} - \mathcal{N}_2(\mathbf{U}_n) \in H^{\epsilon-1}(\Omega)$, and $\mathcal{N}_1(\mathbf{U}_n) \in H^{\epsilon}(\Omega)$, then the right-hand side of (3.14),

$$\mathbf{b}(\mathbf{Z}) = \langle \mathbf{F} - \mathcal{N}_2(\mathbf{U}_n), \mathbf{Z}\rangle - \langle \mathbf{U}_n, L_0^*\mathbf{Z}\rangle - \langle \mathcal{N}_1(\mathbf{U}_n), L_1^*\mathbf{Z}\rangle, \tag{3.16}$$

is a bounded linear functional on $H^{1-\epsilon}(\Omega)$. If $\mathcal{L}'(\mathbf{U}_n)^*$ has sufficient regularity, then we can expect $\mathbf{W} \in H^{1+\epsilon}(\Omega)$, which yields $\delta\mathbf{U}_n = \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W} \in H^{\epsilon}(\Omega)$, and the iterates remain in $H^{\epsilon}(\Omega)$. In section 4, we establish conditions and prove convergence of the Newton-FOSLL* iteration under reduced regularity of the iterates. In section 5, convergence of the discrete form of Newton-FOSLL* is established.

In many cases it is possible to choose the formulation of the PDE system to match (3.12), so that the additional considerations in controlling the smoothness of iterates are not necessary. The numerical tests in section 6 describe three semilinear PDE systems, which are of the form (3.12) with $\mathcal{N}_1(\mathbf{U}) = 0$, and for which $\mathcal{N}_2(\mathbf{U}_n^h) \in H^{1/2-\nu}(\Omega)$.

4. Convergence of the Newton-FOSLL* iteration.

In this section, we examine the convergence of the Newton-FOSLL* iteration in the H^{ϵ} norm under appropriate conditions on the linearized problem and its adjoint. In two dimensions, convergence is established for $0 < \epsilon \leq 1$, while in three dimensions, convergence is established for $1/2 \leq \epsilon \leq 1$. As discussed above, $\mathscr{C}^0$ finite element spaces yield iterates in $H^{1/2-\nu}(\Omega)$ for $\nu > 0$. This gap in the theory for three-dimensional problems remains an open question.

Let $\mathbf{U}$ be the solution of (3.1), and define the open ball, $B^{\epsilon}_{R}(\mathbf{U}) = \{\mathbf{V} \in H^{\epsilon}(\Omega) : \|\mathbf{U} - \mathbf{V}\|_{H^{\epsilon}(\Omega)} < R\}$. We assume that there is an $R > 0$ such that for $\mathbf{U}_n \in B^{\epsilon}_{R}(\mathbf{U})$, where $\epsilon \in (0, 1]$, $\mathcal{L}(\mathbf{U}_n) \in H^{\epsilon-1}(\Omega)$, $\mathcal{L}'(\mathbf{U}_n)$ is bijective from $H^{\epsilon}(\Omega)$ to $H^{\epsilon-1}(\Omega)$, and $\mathcal{L}'(\mathbf{U}_n)^*$ is bijective from $H^{\epsilon+1}(\Omega)$ to $H^{\epsilon}(\Omega)$. That is, assume for $\mathbf{V} \in H^{\epsilon}(\Omega) \cap \mathcal{D}(\mathcal{L}'(\mathbf{U}_n))$ and $\mathbf{W} \in H^{\epsilon+1} \cap \mathcal{D}(\mathcal{L}'(\mathbf{U}_n)^*)$ that there exist positive constants α_0, α_1 and β_0, β_1, dependent only on ϵ and Ω, such that

$$\alpha_0\|\mathbf{V}\|_{H^{\epsilon}(\Omega)} \leq \|\mathcal{L}'(\mathbf{U}_n)\mathbf{V}\|_{H^{\epsilon-1}(\Omega)} \leq \alpha_1\|\mathbf{V}\|_{H^{\epsilon}(\Omega)}, \tag{4.1}$$

$$\beta_0\|\mathbf{W}\|_{H^{\epsilon+1}(\Omega)} \leq \|\mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}\|_{H^{\epsilon}(\Omega)} \leq \beta_1\|\mathbf{W}\|_{H^{\epsilon+1}(\Omega)}, \tag{4.2}$$

uniformly for $\mathbf{U}_n \in B^{\epsilon}_{R}(\mathbf{U})$.

In addition, we assume that for any $\mathbf{U}_n \in B^{\epsilon}_{R}$, $\mathcal{L}(\mathbf{U})$ admits a Taylor expansion

$$\mathbf{F} = \mathcal{L}(\mathbf{U}) = \mathcal{L}(\mathbf{U}_n) + \mathcal{L}'(\mathbf{U}_n)(\mathbf{U} - \mathbf{U}_n) + \frac{1}{2}\mathcal{L}''(\widetilde{\mathbf{U}})[\mathbf{U} - \mathbf{U}_n, \mathbf{U} - \mathbf{U}_n]$$

for $\widetilde{\mathbf{U}} = \gamma\mathbf{U} + (1-\gamma)\mathbf{U}_n$ with $\gamma \in [0, 1]$ such that the second Fréchet derivative of $\mathcal{L}$ is bounded in the following way: there is a positive constant c, independent of $\mathbf{U}_n$, such that

$$||\mathcal{L}''(\widetilde{\mathbf{U}})[\mathbf{Z}, \mathbf{Z}]||_{H^{\epsilon-1}(\Omega)} \;\leq\; c\,||\mathbf{Z}^2||_{H^{\epsilon-1}(\Omega)} \tag{4.3}$$

for any $\mathbf{Z} \in H^{\epsilon}(\Omega)$.

The following result establishes quadratic convergence of the Newton-FOSLL* iteration.

THEOREM 1. *Let Ω be an open, bounded, Lipschitz-continuous subset of $\mathbb{R}^d, d = 2, 3$. Let $\mathbf{U}$ be the solution of (3.1), and let $\mathbf{U}_n$ be the current approximation of* *the Newton-FOSLL* iteration in* $B^{\epsilon}_{R}(\mathbf{U}) = \{\mathbf{V} \in H^{\epsilon}(\Omega) : \|\mathbf{U}-\mathbf{V}\|_{H^{\epsilon}(\Omega)} < R\}$ *for* $0 < \epsilon \leq 1$ *if* $d = 2$ *and* $1/2 \leq \epsilon \leq 1$ *if* $d = 3$. *Assume that* (4.1), (4.2), *and* (4.3) *hold uniformly for every* $\mathbf{V} \in B^{\epsilon}_{R}(\mathbf{U})$. *Let*

$$\mathbf{U}_{n+1} = \mathbf{U}_n + \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W},$$

where $\mathbf{W} \in H^{1+\epsilon}(\Omega) \cap \mathcal{D}(\mathcal{L}'(\mathbf{U}_n)^*)$ *satisfies*

$$\begin{aligned}\langle \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}, \mathcal{L}'(\mathbf{U}_n)^*\mathbf{Z}\rangle &= \langle \mathbf{F} - \mathcal{L}(\mathbf{U}_n), \mathbf{Z}\rangle \\ &= \langle \mathbf{F} - \mathcal{N}_2(\mathbf{U}_n), \mathbf{Z}\rangle - \langle \mathbf{U}_n, L_0^*\mathbf{Z}\rangle - \langle \mathcal{N}_1(\mathbf{U}_n), L_1^*\mathbf{Z}\rangle\end{aligned} \tag{4.4}$$

for all $\mathbf{Z} \in H^{1+\epsilon}(\Omega) \cap \mathcal{D}(\mathcal{L}'(\mathbf{U}_n)^*)$. *Then, for any initial approximation,* $\mathbf{U}_0$, *sufficiently close to* $\mathbf{U}$ *in the* $H^{\epsilon}(\Omega)$ *norm, the sequence* $\{\mathbf{U}_n\}_{n=0}^{\infty}$ *is quadratically convergent to* $\mathbf{U}$ *in the* $H^{\epsilon}(\Omega)$ *norm.*

Proof. The Taylor expansion of $\mathcal{L}(\mathbf{U})$ around the current approximation $\mathbf{U}_n$ yields

$$\mathbf{F} = \mathcal{L}(\mathbf{U}) = \mathcal{L}(\mathbf{U}_n) + \mathcal{L}'(\mathbf{U}_n)(\mathbf{U}-\mathbf{U}_n) + \frac{1}{2}\mathcal{L}''(\widetilde{\mathbf{U}})[\mathbf{U}-\mathbf{U}_n, \mathbf{U}-\mathbf{U}_n],$$

where $\widetilde{\mathbf{U}} = \gamma\mathbf{U} + (1-\gamma)\mathbf{U}_n$ for some $\gamma \in [0,1]$. Since $\mathbf{W}$ is the solution of (4.4), the inequalities (4.1) and (4.3) imply

$$\begin{aligned}\|\mathbf{U}-\mathbf{U}_{n+1}\|_{H^{\epsilon}(\Omega)} &= \|\mathbf{U}-\mathbf{U}_n - \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}\|_{H^{\epsilon}(\Omega)} \\ &= \|\mathbf{U}-\mathbf{U}_n - \mathcal{L}'(\mathbf{U}_n)^{-1}(\mathbf{F}-\mathcal{L}(\mathbf{U}_n))\|_{H^{\epsilon}(\Omega)} \\ &= \left\|\frac{1}{2}\mathcal{L}'(\mathbf{U}_n)^{-1}\mathcal{L}''(\widetilde{\mathbf{U}})[\mathbf{U}-\mathbf{U}_n, \mathbf{U}-\mathbf{U}_n]\right\|_{H^{\epsilon}(\Omega)} \\ &\leq c_1 \left\|\mathcal{L}''(\widetilde{\mathbf{U}})[\mathbf{U}-\mathbf{U}_n, \mathbf{U}-\mathbf{U}_n]\right\|_{H^{\epsilon-1}(\Omega)} \\ &\leq c_2 \|(\mathbf{U}-\mathbf{U}_n)^2\|_{H^{\epsilon-1}(\Omega)} \\ &= c_2 \sup_{\substack{\mathbf{V} \in H^{1-\epsilon}(\Omega) \\ \mathbf{V} \neq 0}} \frac{\left|\left\langle (\mathbf{U}-\mathbf{U}_n)^2, \mathbf{V}\right\rangle\right|}{\|\mathbf{V}\|_{H^{1-\epsilon}(\Omega)}}.\end{aligned} \tag{4.5}$$

Hölder's inequality implies

$$\left|\left\langle (\mathbf{U}-\mathbf{U}_n)^2, \mathbf{V}\right\rangle\right| \leq \left(\int_{\Omega} |\mathbf{U}-\mathbf{U}_n|^{2s} dx\right)^{\frac{1}{s}} \left(\int_{\Omega} |\mathbf{V}|^{t} dx\right)^{\frac{1}{t}} = \|\mathbf{U}-\mathbf{U}_n\|^2_{L^{2s}(\Omega)} \|\mathbf{V}\|_{L^t(\Omega)}$$

with $1/s + 1/t = 1$. In $\mathbb{R}^2$, we take $s = 2/(2-\epsilon)$ and $t = 2/\epsilon$, and in $\mathbb{R}^3$, we choose $s = 6/(5-2\epsilon)$ and $t = 6/(1+2\epsilon)$. If $0 < \epsilon \leq 1$ in $\mathbb{R}^2$ or $1/2 \leq \epsilon \leq 1$ in $\mathbb{R}^3$, Lemma 1 can be applied and yields

$$\|\mathbf{V}\|_{L^t(\Omega)} \leq c_3 \|\mathbf{V}\|_{H^{1-\epsilon}(\Omega)} \quad \text{and} \quad \|\mathbf{U}-\mathbf{U}_n\|_{L^{2s}(\Omega)} \leq c_4 \|\mathbf{U}-\mathbf{U}_n\|_{H^{\epsilon}(\Omega)}. \tag{4.6}$$

By applying (4.6) to (4.5) we get

$$\|\mathbf{U}-\mathbf{U}_{n+1}\|_{H^{\epsilon}(\Omega)} \leq c\,\|\mathbf{U}-\mathbf{U}_n\|^2_{H^{\epsilon}(\Omega)} \leq c\,R^2.$$

Thus, for any $\mathbf{U}_n \in B^{\epsilon}_{R}(\mathbf{U})$ with $R \leq \theta/c$ for some fixed $\theta \in (0,1)$, we may be assured that $\mathbf{U}_{n+1} \in B^{\epsilon}_{R}(\mathbf{U})$ and that, as $n \to \infty$, the iteration converges quadratically. ∎

The above theorem shows that we have convergence of the Newton iteration with less-regular approximations than the conventional Newton iteration requires. The convergence of the Newton-FOSLL* iteration in a finite-dimensional subspace is derived in the following section.

5. Finite element convergence. In this section we show that, for sufficiently small h, the discrete iteration stays in the attraction basin of Newton's method. As above, assume the hypotheses of Theorem 1 hold, let $\mathbf{U}_n \in H^\epsilon(\Omega)$ be the current approximation to $\mathbf{U}$, and let $\mathbf{W} \in H^{1+\epsilon}(\Omega) \cap \mathcal{D}(\mathcal{L}'(\mathbf{U}_n)^*)$ satisfy

$$\langle \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}, \mathcal{L}'(\mathbf{U}_n)^*\mathbf{Z}\rangle = \langle \mathbf{F} - \mathcal{N}_2(\mathbf{U}_n), \mathbf{Z}\rangle - \langle \mathbf{U}_n, L_0^*\mathbf{Z}\rangle - \langle \mathcal{N}_1(\mathbf{U}_n), L_1^*\mathbf{Z}\rangle \tag{5.1}$$

for all $\mathbf{Z} \in H^{1+\epsilon}(\Omega) \cap \mathcal{D}(\mathcal{L}'(\mathbf{U}_n)^*)$. Then, let the exact Newton iterate be given by

$$\mathbf{U}_{n+1} = \mathbf{U}_n + \delta\mathbf{U}_n = \mathbf{U}_n + \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}. \tag{5.2}$$

Now, let $\mathbf{W}^h \in \mathcal{W}^h \subset H^{1+\epsilon}(\Omega)$ satisfy

$$\left\langle \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h, \mathcal{L}'(\mathbf{U}_n)^*\mathbf{Z}^h\right\rangle = \langle \mathbf{F} - \mathcal{N}_2(\mathbf{U}_n), \mathbf{Z}^h\rangle - \langle \mathbf{U}_n, L_0^*\mathbf{Z}^h\rangle - \langle \mathcal{N}_1(\mathbf{U}_n), L_1^*\mathbf{Z}^h\rangle \tag{5.3}$$

for all $\mathbf{Z}^h \in \mathcal{W}^h$, and define

$$\mathbf{U}_{n+1}^h = \mathbf{U}_n + \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h \tag{5.4}$$

as the approximate Newton iterate. Using (3.7), (4.2),

$$\begin{aligned}\|\mathbf{W} - \mathbf{W}^h\|_{H^1(\Omega)} &\le \beta_0^{-1}\|\mathcal{L}'(\mathbf{U}_n)^*(\mathbf{W} - \mathbf{W}^h)\| \le \beta_0^{-1}\|\mathcal{L}'(\mathbf{U}_n)^*(\mathbf{W} - \mathcal{I}^h\mathbf{W})\|\\ &\le \frac{\beta_1}{\beta_0}\|\mathbf{W} - \mathcal{I}^h\mathbf{W}\|_{H^1(\Omega)} \le c\, h^s\|\mathbf{W}\|_{H^{1+s}(\Omega)}\end{aligned}$$

for $0 < s \le q$. Since $\mathbf{U}_{n+1} - \mathbf{U}_{n+1}^h = \mathcal{L}'(\mathbf{U}_n)^*(\mathbf{W} - \mathbf{W}^h)$, it immediately follows that

$$\|\mathbf{U}_{n+1} - \mathbf{U}_{n+1}^h\| \le c\, h^s\|\mathbf{W}\|_{H^{1+s}(\Omega)}. \tag{5.5}$$

In the following theorem, we show that, if h is sufficiently small, the discrete iterate remains in the basin of attraction of Newton's method in the H^ϵ norm.

THEOREM 2. *Assume that Ω, $\mathbf{U}$, $B_R^\epsilon(\mathbf{U})$, ϵ, $\mathbf{U}_n$, and $\mathbf{U}_{n+1}$ are the same as in Theorem 1, and let $\mathbf{U}_{n+1}^h$ be the discrete iterate given in (5.3) and (5.4). For sufficiently small h,*

$$\|\mathbf{U} - \mathbf{U}_{n+1}^h\|_{H^\epsilon(\Omega)} < R, \tag{5.6}$$

where $R > 0$ is established in the proof of Theorem 1.

Proof. By the proof of Theorem 1 and the triangle inequality, we have

$$\|\mathbf{U} - \mathbf{U}_{n+1}^h\|_{H^\epsilon(\Omega)} \le \tilde{c}\, R^2 + \|\mathbf{U}_{n+1} - \mathbf{U}_{n+1}^h\|_{H^\epsilon(\Omega)}.$$

The triangle inequality, (4.2), an inverse inequality (see, e.g., Chapter 4 of [14]), (5.5), and (3.7) imply

$$\begin{aligned}\|\mathbf{U}_{n+1} - \mathbf{U}_{n+1}^h\|_{H^\epsilon(\Omega)} &= \|\mathcal{L}'(\mathbf{U}_n)^*(\mathbf{W} - \mathbf{W}^h)\|_{H^\epsilon(\Omega)}\\ &\le \|\mathcal{L}'(\mathbf{U}_n)^*(\mathbf{W} - \mathcal{I}^h\mathbf{W})\|_{H^\epsilon(\Omega)} + \|\mathcal{L}'(\mathbf{U}_n)^*(\mathcal{I}^h\mathbf{W} - \mathbf{W}^h)\|_{H^\epsilon(\Omega)}\\ &\le \beta_1\|\mathbf{W} - \mathcal{I}^h\mathbf{W}\|_{H^{1+\epsilon}(\Omega)} + c_1 h^{-\epsilon}\|\mathcal{L}'(\mathbf{U}_n)^*(\mathcal{I}^h\mathbf{W} - \mathbf{W}^h)\|\\ &\le c_2 h^{s-\epsilon}\|\mathbf{W}\|_{H^{1+s}(\Omega)} + c_1 h^{-\epsilon}\|\mathcal{L}'(\mathbf{U}_n)^*(\mathcal{I}^h\mathbf{W} - \mathbf{W} + \mathbf{W} - \mathbf{W}^h)\|\\ &\le c_2 h^{s-\epsilon}\|\mathbf{W}\|_{H^{1+s}(\Omega)} + c_3 h^{-\epsilon}\|\mathcal{L}'(\mathbf{U}_n)^*(\mathbf{W} - \mathcal{I}^h\mathbf{W})\|\\ &\le c h^{s-\epsilon}\|\mathbf{W}\|_{H^{1+s}(\Omega)}.\end{aligned}$$

Suppose h is small enough so that $h \leq R^{2/(s-\epsilon)}$. Then, we have

$$\|\mathbf{U} - \mathbf{U}^h_{n+1}\|_{H^\epsilon(\Omega)} \leq \tilde{c}\, R^2 + c\, R^2 \|\mathbf{W}\|_{H^{1+s}(\Omega)},$$

and choosing R such that $(\tilde{c} + c\|\mathbf{W}\|_{H^{1+s}(\Omega)})\, R \leq 1$ yields the result (5.6). ∎

For the projection technique (subsection 3.2.2) and the extended system (subsection 3.2.3) the iterates will be finite element functions, and, in general, $\mathbf{U}^h \in H^{3/2-\nu}(\Omega)$ for $\nu > 0$. If $\mathbf{U} \in H^{1+\epsilon}(\Omega)$, then the error, $\mathbf{U} - \mathbf{U}^h$, will be in the larger of those two spaces. In either case, Theorem 2 guarantees that the iteration can proceed. In this context, the functional is minimized over a finite element space. Convergence to that minimum will be linear with rate $O(h^s)$. Further background in the convergence of inexact Newton methods in the nonlinear least-squares context can be found in Chapter 10 of [27].

With the *weak*-weak FOSLL* method (subsection 3.2.4), on the other hand, the correction step yields $\boldsymbol{\delta}\mathbf{U}^h = \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h$. With standard finite element spaces, $\boldsymbol{\delta}\mathbf{U} \in H^{1/2-\nu}(\Omega)$. Thus, the error will also satisfy $\mathbf{U} - \mathbf{U}^h \in H^{1/2-\nu}(\Omega)$. As in the projected and extended methods, convergence to the minimum over finite element spaces will be linear with rate $O(h^s)$.

In two dimensions, Theorem 2 holds for $\mathbf{U} - \mathbf{U}^h \in H^\epsilon(\Omega)$ for any $\epsilon > 0$, which guarantees that if the discrete problem is solved to sufficient accuracy, the iteration can proceed. However, in three dimensions, the theorem holds only for $\epsilon \geq 1/2$. This gap may only be theoretical; the theorem may not be sharp. In any case, we leave that issue for future research.

6. Computational examples. In this section, we present three examples that illustrate the general methodology of the proposed Newton-FOSLL* iteration. These examples illuminate the process of setting up the adjoint system and demonstrate the strengths and weaknesses of some of the options of the approaches.

The first computational example uses a vorticity formulation of the steady incompressible Navier–Stokes equations and has full regularity and a known exact solution. This allows direct computation of the approximation error and confirms the theory developed above, namely, quadratic convergence of Newton's method and finite element discretization convergence rates. We also directly compare the accuracy and efficiency of the *weak*-weak, projection, and extended functional approaches to the nonlinear iteration.

The second computational example develops a mixed $H^1 \times H(\mathrm{Div})$ finite element approach to a velocity gradient formulation for Navier–Stokes, applying the projection approach to continuation. A test problem of flow over a backwards facing step shows excellent mass conservation, a difficult challenge for standard least-squares methods. The size of the salient corner recirculation eddy is computed for a range of Reynolds numbers and is found to be in good agreement with results in the literature. Since all three methods perform essentially the same, only results for the projection method are presented.

The third computational example involves a problem with low regularity and illustrates how the FOSLL* approach can approximate discontinuous solutions. The operator has a discontinuous diffusion coefficient and nonlinearity in a convection-like term. This example also illustrates the use of a nonlinear nested iteration approach for increased efficiency. In this problem, the exact solution is discontinuous. Projection onto $\mathcal{D}(\mathcal{L})$ would require enforcing a jump condition across the discontinuities. The *weak*-weak form does not require knowledge of the discontinuities; thus, results for only this approach are presented.

6.1. Two-dimensional Navier–Stokes: Vorticity formulation. For the first example, we consider a vorticity formulation of the steady, incompressible Navier–Stokes equations. This form admits a reformulation with the nonlinearity in zeroth-order terms, thus allowing for direct comparison of the *weak*-weak, the *projection*, and the *extended functional* approaches outlined in subsections 3.2.4, 3.2.2, and 3.2.3, respectively. Further, to clearly demonstrate the convergence predicted in our theory, this problem is posed on a straightforward convex domain with a known smooth exact solution, making it possible to directly compute convergence rates. Consider

$$\begin{cases} \Delta \mathbf{u} - \lambda \mathbf{u} \cdot \nabla \mathbf{u} - \nabla p = \mathbf{f} & \text{in } \Omega, \\ \nabla \cdot \mathbf{u} = 0 & \text{in } \Omega, \end{cases} \tag{6.1}$$

equipped with boundary conditions $\mathbf{n} \cdot \mathbf{u} = 0$ and $\nabla \times \mathbf{u} = 0$ on $\partial\Omega$. Here, $\mathbf{u}$ is the velocity, p is the pressure, $\mathbf{f}$ represents body forces, λ is the dimensionless Reynolds number, and $\mathbf{n}$ is the outward unit normal to the domain boundary $\partial\Omega$. Using the vorticity field, $\omega = -\nabla \times \mathbf{u}$, and the total pressure, $P = (\lambda/2)|\mathbf{u}|^2 + p$, we can rewrite (6.1) in the form $\mathcal{L}(\mathbf{U}) = L\mathbf{U} + \mathcal{N}(\mathbf{U}) = \mathcal{F}$ with

$$\begin{bmatrix} I & \nabla\times & 0 \\ \nabla^\perp & \mathbf{0} & -\nabla \\ 0 & \nabla\cdot & 0 \end{bmatrix} \begin{bmatrix} \omega \\ \mathbf{u} \\ P \end{bmatrix} + \lambda \begin{bmatrix} 0 \\ \begin{bmatrix} -\omega u_2 \\ \omega u_1 \end{bmatrix} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{f} \\ 0 \end{bmatrix}, \tag{6.2}$$

where $\mathbf{u} = (u_1, u_2)^t$ and $\nabla^\perp$ denotes the adjoint of $\nabla\times$ in $\mathbb{R}^2$. The boundary conditions are thus $\mathbf{n} \cdot \mathbf{u} = 0$ and $\omega = 0$ on $\partial\Omega$. Taking the Fréchet derivative of the operator yields

$$\mathcal{L}'(\mathbf{U})\mathbf{V} = L\mathbf{V} + \mathcal{N}'(\mathbf{U})\mathbf{V},$$

where

$$\mathcal{N}'(\mathbf{U})\mathbf{V} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -u_2 & 0 & -\omega & 0 \\ u_1 & \omega & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix},$$

which can be shown to satisfy (4.1) if we also require $p \in L_0^2(\Omega)$. Now,

$$\mathcal{N}''(\mathbf{U})[\mathbf{V}, \mathbf{Z}] = [0, -(v_1 z_3 + v_3 z_1), (v_1 z_2 + v_2 z_3), 0]^t\,,$$

which clearly satisfies (4.3). The corresponding adjoint operator is

$$\mathcal{L}'(\mathbf{U})^*\mathbf{W} = \begin{bmatrix} I & \nabla\times & 0 \\ \nabla^\perp & \mathbf{0} & -\nabla \\ 0 & \nabla\cdot & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \begin{bmatrix} w_2 \\ w_3 \end{bmatrix} \\ w_4 \end{bmatrix} + \lambda \begin{bmatrix} 0 & -u_2 & u_1 & 0 \\ 0 & 0 & \omega & 0 \\ 0 & -\omega & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix},$$

with boundary conditions $w_1 = 0$ and $\mathbf{n} \cdot (w_2, w_3)^t = 0$ on $\partial\Omega$, which satisfy (4.2) if we also require $w_4 \in L_0^2(\Omega)$. Following (3.14), the *weak*-weak form becomes the following: given iterate $\mathbf{U}_n$, find $\mathbf{W}^h \in \mathcal{W}^h$ such that

$$\left\langle \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h, \mathcal{L}'(\mathbf{U}_n)^*\mathbf{Z}^h \right\rangle = \left\langle \mathbf{F} - \mathcal{N}(\mathbf{U}_n), \mathbf{Z}^h \right\rangle - \left\langle \mathbf{U}_n, L^*\mathbf{Z}^h \right\rangle$$

for every $\mathbf{Z}^h \in \mathcal{W}^h$. Then, the next iterate is given by

$$\mathbf{U}_{n+1} = \mathbf{U}_n + \mathcal{L}'(\mathbf{U}_n)^* \mathbf{W}^h.$$

For this test problem, define $\Omega = (0,2) \times (0,1)$ and let

$$\mathbf{u} = \begin{bmatrix} \sin(\pi x)\cos(\pi y) \\ -\cos(\pi x)\sin(\pi y) \end{bmatrix} \quad \text{and} \quad \mathbf{f} = \begin{bmatrix} -2\pi^2 u_1 - \lambda\pi \sin(\pi x)\cos(\pi x) \\ -2\pi^2 u_2 - \lambda\pi \sin(\pi y)\cos(\pi y) \end{bmatrix}$$

be the exact solution to system (6.1) and corresponding function, $\mathbf{f}$. For this test we use $\lambda = 10$ and discretize Ω with a uniform triangulation of mesh size, h. Options for finite element spaces for $\mathbf{W}^h$ and $\mathbf{U}^h$ are described below.

Weak-weak form. Following the method outlined in subsection 3.2.4, we approximate $\mathbf{W}^h = (w_1, w_2, w_3, w_4)^t$ using $P2$ elements (conforming piecewise quadratics on triangles) and represent each $\mathbf{U} = (\omega, u_1, u_2, P)^t$ using $P1_{\mathrm{dc}}$ elements (discontinuous piecewise linears). With these choices, the error bound predicted in (5.5) is $\|\mathbf{U} - \mathbf{U}^h\| \sim \mathcal{O}(h^2)$.

Projection. For the projection method outlined in subsection 3.2.2, we use $P2$ elements for $\mathbf{W}^h$ at each iteration and project $\mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h$ onto the space of $P2$ elements to continue to the next iteration, as described in subsection 3.2.2. Since this projection has $\mathcal{O}(h^3)$ interpolation error, we should still expect an overall error of $\|\mathbf{U} - \mathbf{U}^h\| \sim \mathcal{O}(h^2)$.

Extended. For the extended approach, as presented in subsection 3.2.3, we simultaneously solve for $\mathbf{W}^h$ and $\boldsymbol{\delta}\mathbf{U}^h$ using $P2$ elements. Again, we expect to see an overall error of $\|\mathbf{U} - \mathbf{U}^h\| \sim \mathcal{O}(h^2)$.

To compare these three approaches we consider a sequence of Newton iterates on different mesh resolutions. For each iterate we compute the L^2 norm of the error ($\texttt{error} = \|\mathbf{U}_n^h - \mathbf{U}\|$), based on the exact solution. Using a measure of computational cost relative to one Newton iteration of the *weak*-weak approach on the coarsest mesh ($h^{-1} = 16$), we take the relative cost of the projection and extended approaches as twice that of the *weak*-weak approach since each has twice the number of unknowns in the resulting linear system solve. With this we define a rough estimate of efficiency as $\texttt{eff} = (\texttt{error} * \texttt{cost})^{-1}$ (larger is better).

Table 1 shows the results for six nonlinear steps per level, each computed independently from an initial guess of zero. We find that each approach reflects well the theory in sections 4 and 5. On each mesh, the nonlinear iteration converges rapidly to within discretization error, and the discretization accuracy is $\mathcal{O}(h^2)$ in each case, though the errors in the *weak*-weak approximations are larger by nearly an order of magnitude. For this problem, where the exact solution is very smooth, the extra smoothness imposed by the projection and extended approaches seems to be a good investment in computational resources, with the projection approach showing the highest peak efficiency, even though, compared with the *weak*-weak approach, each iteration is more expensive and at least one more nonlinear iteration is required to converge to the level of discretization error.

6.2. Two-dimensional Navier–Stokes: Velocity gradient formulation. We now consider a different formulation of the Navier–Stokes equations, one that is more practical than the previous example for many applications. Consider again system (6.1). Introducing the velocity gradient, $\boldsymbol{\sigma} = \nabla \mathbf{u}$, (6.1) can be reformulated

Table 1

Convergence of Newton-FOSLL for system (6.1) with $\lambda = 10$ (h is the meshsize parameter, n is the nonlinear step,* `error` $= \|\mathbf{U}_n^h - \mathbf{U}\|$ *is the L^2 error,* `cost` *is the cumulative relative computational cost, and* `eff` *is an efficiency measure). Convergence rate is computed between approximations on the two finest levels.*

		Weak-weak			Projection			Extended		
h^{-1}	n	`error`	`cost`	`eff`	`error`	`cost`	`eff`	`error`	`cost`	`eff`
16	1	3.95	1	0.25	3.95	2	0.13	3.95	2	0.13
16	2	0.695	2	0.72	0.518	4	0.48	0.565	4	0.44
16	3	0.244	3	**1.4**	0.0746	6	2.2	0.104	6	1.6
16	4	0.235	4	1.1	0.0393	8	**3.2**	0.0496	8	**2.5**
16	5	0.234	5	0.86	0.0372	10	2.7	0.0396	10	**2.5**
16	6	0.234	6	0.71	0.0368	12	2.3	0.0377	12	2.2
32	1	3.95	4	0.063	3.95	8	0.032	3.95	8	0.032
32	2	0.186	8	0.67	0.158	16	0.40	0.165	16	0.38
32	3	0.0610	12	**1.4**	0.0140	24	3.0	0.0232	24	1.8
32	4	0.0608	16	1.0	0.00943	32	**3.3**	0.0124	32	2.5
32	5	0.0608	20	0.82	0.00904	40	2.8	0.00978	40	**2.6**
32	6	0.0608	24	0.69	0.00894	48	2.3	0.00927	48	2.3
64	1	3.95	16	0.016	3.95	32	0.0079	3.95	32	0.0079
64	2	0.0473	32	0.66	0.0418	64	0.37	0.0430	64	0.36
64	3	0.0153	48	**1.4**	0.00350	96	3.0	0.00595	96	1.8
64	4	0.0153	64	1.0	0.00236	128	**3.3**	0.00318	128	**2.5**
64	5	0.0153	80	0.82	0.00225	160	2.8	0.00246	160	**2.5**
64	6	0.0153	96	0.68	0.00223	192	2.3	0.00232	192	2.2
Rate		1.99			2.00			2.00		

as

$$\begin{cases} \nabla \cdot \boldsymbol{\sigma} - \nabla p - \lambda \boldsymbol{\sigma} \mathbf{u} = \mathbf{f} & \text{in } \Omega, \\ -\nabla \mathbf{u} + \boldsymbol{\sigma} = \mathbf{0} & \text{in } \Omega, \\ \nabla \cdot \mathbf{u} = 0 & \text{in } \Omega. \end{cases} \tag{6.3}$$

Defining $\mathbf{U} = (\mathbf{u}, \boldsymbol{\sigma}, p)$, this can be written as the 7×7, formally self-adjoint system

$$\mathcal{L}(\mathbf{U}) = L\mathbf{U} + \mathcal{N}(\mathbf{U}) = \begin{bmatrix} \mathbf{0} & \nabla\cdot & -\nabla \\ -\nabla & I & \mathbf{0} \\ \nabla\cdot & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\sigma} \\ p \end{bmatrix} + \begin{bmatrix} -\lambda \boldsymbol{\sigma} \mathbf{u} \\ \mathbf{0} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \\ 0 \end{bmatrix} = \mathbf{F}.$$

On the inflow boundary, Γ_I, a parabolic velocity profile, $\mathbf{u} = \mathbf{g}$, is imposed; on the walls, Γ_W, a no slip condition, $\mathbf{u} = \mathbf{0}$, is imposed; and on the outflow, Γ_O, a developed gradient, $\sigma_{11} = \sigma_{21} = p = 0$, is imposed.

By the structure of the system, it is natural to take $\mathbf{U} = (\mathbf{u}, \boldsymbol{\sigma}, p) \in \mathcal{V} = (H^1)^2 \times H(\mathrm{Div})^2 \times H^1$, with appropriate boundary conditions. This is similar to systems developed in [21, 22, 17], but does not share the same $\mathcal{V}$ ellipticity. The tests below indicate that it is sufficient for FOSLL*.

We denote the structure of dual variables similarly,

$$\mathbf{Z} = \left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \begin{bmatrix} z_3 & z_4 \\ z_5 & z_6 \end{bmatrix}, z_7 \right) \quad \text{and} \quad \mathbf{W} = \left(\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \begin{bmatrix} w_3 & w_4 \\ w_5 & w_6 \end{bmatrix}, w_7 \right),$$

and the adjoint $L^* = L$, thus, satisfies

$$\langle L\mathbf{U}, \mathbf{Z} \rangle = \langle \mathbf{U}, L^*\mathbf{Z} \rangle + \oint (\mathbf{n} \cdot \boldsymbol{\sigma}) \cdot \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} - \oint \left(\mathbf{n} \cdot \begin{bmatrix} z_3 & z_4 \\ z_5 & z_6 \end{bmatrix} \right) \cdot \mathbf{u} - \oint \left(\mathbf{n} \cdot \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \right) p + \oint (\mathbf{n} \cdot \mathbf{u})\, z_7.$$

Velocity boundary conditions require $w_1 = w_2 = 0$ on the inflow and walls, while developed gradient requires $w_3 = w_5 = w_7 = 0$ on the outflow, which yields a self-adjoint system. Recalling (3.9), the FOSLL* variational problem also requires the addition of two boundary terms when $\mathbf{u}$ is nonzero on the boundary. The variational problem we solve at each step is given by

$$\left\langle \mathcal{L}'(\mathbf{U}_n)^*\mathbf{W}^h, \mathcal{L}'(\mathbf{U}_n)^*\mathbf{Z}^h \right\rangle = \left\langle \mathbf{F} - \mathcal{N}(\mathbf{U}_n), \mathbf{Z}^h \right\rangle - \left\langle \mathbf{U}_n, L^*\mathbf{Z}^h \right\rangle + \int_{\Gamma_I} \left(\mathbf{n} \cdot \begin{bmatrix} z_3^h & z_4^h \\ z_5^h & z_6^h \end{bmatrix} \right) \cdot \mathbf{g} - \int_{\Gamma_I} (\mathbf{n} \cdot \mathbf{g}) z_7^h.$$

While all three continuation approaches detailed in the previous example are appropriate, in this example their numerical behavior was very similar. The L^2 projection approach was slightly more efficient. For the sake of brevity, only those results are presented. For the numerical approximation, we choose $P2$ elements for w_1, w_2, and w_7 and RT1 elements for (w_3, w_4) and (w_5, w_6) (i.e., the next-to-lowest-order Raviart–Thomas elements). We impose $w_1 = w_2 = 0$ on $\Gamma_I \cup \Gamma_W$, and $w_3 = w_5 = w_7 = 0$ on Γ_O. For the FOSLS comparison, we use $P2$ elements for $\mathbf{u}, p$, and RT1 for σ, imposing $\mathbf{u} = \mathbf{g}$ on Γ_I, $\mathbf{u} = \mathbf{0}$ on Γ_W, and $\mathbf{n} \cdot \boldsymbol{\sigma} = \mathbf{0}$, $p = 0$ on Γ_O.

As a test problem, we consider fluid flow over a backwards facing step as illustrated by Figure 1, where the channel lengths used are $L_1 = L_2 = 4$. The diameter of the outflow channel is chosen as $H = 1.94$, which matches the domain geometry used in [7]. As stated above, inflow boundary conditions are a fully developed parabolic profile with a mean velocity of 1, while outflow boundary conditions are developed gradient flow.

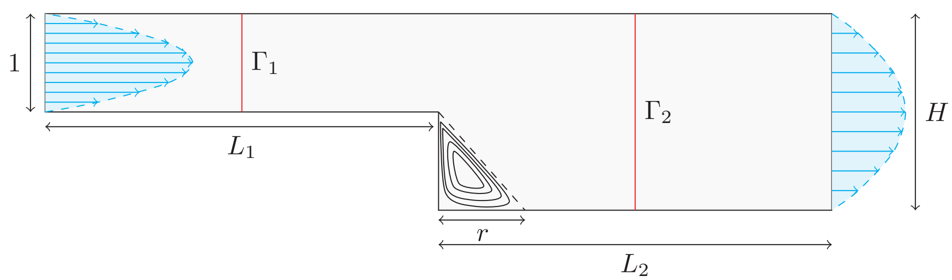


FIG. 1. *Domain and boundary conditions for example 2, flow across a backwards facing step. In the tests below, $L_1 = L_2 = 4$, $H = 1.94$, the step is located at $x = 0$, Γ_1, Γ_2 are located at -2 and 2, respectively, and r is the reattachment point.*

The domain is discretized with a total of 12860 triangles of mesh size $h_{max} = 1/16$ and $h_{min} = 1/64$, where the finer mesh is placed near the step. Using a zero initial approximation for $\mathbf{U}$, the nonlinear iteration coincides with a first approximation consistent with Stokes equations, and we find that for $\lambda \leq 50$ it is sufficient to use five linearization steps for the approximation to converge.

Standard H^1 or $H(\mathrm{Div})$ least-squares formulations for incompressible flow problems tend to do a poor job of conserving mass in approximations, especially on coarse mesh resolutions or when relatively low-order elements are used. FOSLL* formulations do much better on coarse meshes since they are designed to more directly minimize the L^2 norm of the error rather than the H^1 or $H(\mathrm{Div})$ seminorm of the error. To illustrate the strong mass conservation in our approach, we compute the mass loss across the center of the upstream and downstream channels (i.e., Γ_1 and

Table 2

Reattachment point r of salient corner recirculation eddy and mass loss at Γ_1 and Γ_2 versus Reynolds number for flow across a backwards facing step for nonlinear FOSLL. Mass loss for a comparable FOSLS solution is shown for comparison.*

$Re = 2\lambda$	0	1	10	50	100
r	0.29	0.33	0.53	1.59	2.06
FOSLL*					
mass loss Γ_1	0.019%	0.021%	0.057%	−0.204%	0.756%
mass loss Γ_2	0.007%	−0.015%	−0.029%	0.756%	0.822%
FOSLS					
mass loss Γ_1	2.22%	2.12%	1.49%	0.61%	0.61%
mass loss Γ_2	4.64%	4.44%	3.16%	1.68%	1.57%

Γ_2 in Figure 1) relative to the mass flux at the inflow. Negative mass loss indicates a higher flux than at the inflow. Table 2 shows the results for increasing λ. For comparison we include mass loss for a comparable FOSLS solution for each λ. Using the same mesh and approximation spaces for the unknowns, we solve according to the Newton-FOSLS approach described in subsection 3.1 and observe a much higher mass loss than the FOSLL* method for small and moderate λ.

At larger values of λ, the effect of strong L^2 control (at the expense of higher norm control) begins to show. The quality of the FOSLL* approximations approaches that of the FOSLS formulation. For both FOSLS and FOSLL*, further mesh refinement will yield a solution that is accurate in L^2 and, thus, demonstrates accurate mass conservation. For larger λ, a hybrid FOSLS-FOSLL* approach, as in [34], could be expected to perform better since it would impose a balance between the L^2 and the H^1 seminorm measure of the error.

Since an analytic solution is not available for this problem we also compare a measurement of the reattachment point, r, of the salient corner recirculation eddy to results published in [7]. In that work, the Reynolds number used is computed relative to a characteristic length of twice the inflow channel height, allowing a comparison here with $Re := 2\lambda$. Table 2 summarizes r for a range of Reynolds numbers, which are in good agreement with the streamline plots in Figure 3 of [7].

6.3. Semilinear Div-Curl system with reduced regularity. In this test we consider an elliptic-like problem that has two distinct challenges: (1) a principal part that induces a nonsmooth solution, and (2) a nonlinearity in first-order terms. We apply the Newton-FOSLL* approach to a Div-Curl formulation of this problem, examine the resulting convergence, and present some specific algorithmic implementation issues.

Consider the second-order system

$$\begin{cases} -\nabla \cdot a\nabla p + \lambda |a\nabla p|^2 = f & \text{in } \Omega, \\ p = g & \text{on } \partial\Omega, \end{cases} \tag{6.4}$$

where $\Omega = (-1,1) \times (-1,1)$, $f \in L^2(\Omega)$, and $g \in H^{1/2}(\partial\Omega)$. The diffusion coefficient, a, is chosen to be discontinuous across the horizontal and vertical axes in the following way: for $\varepsilon > 0$,

$$a = \begin{cases} \varepsilon, & xy \leq 0 \quad \text{(NW and SE)}, \\ 1, & xy > 0 \quad \text{(SW and NE)}, \end{cases} \tag{6.5}$$

where NE, NW, SW, and SE correspond to the northeast, northwest, southwest, and southeast quadrants of Ω. For a general $f \in L^2(\Omega)$, the solution p behaves like r^α

near the origin, where

$$\varepsilon = \tan^2(\alpha\pi/4), \tag{6.6}$$

and $r = \sqrt{x^2+y^2}$, $\phi = \arctan(y/x)$ are polar coordinates. Thus, $p \in H^s(\Omega)$ for $s < 1+\alpha$. Note that $\alpha > 0$ can be made as small as desired by choosing the appropriate value of ϵ. Background on singular solutions of elliptic equations can be found in [6, 31].

The first equation in (6.4) can be written in first-order form as

$$\begin{cases} \mathbf{u} - \nabla p = 0, \\ -\nabla \cdot a\mathbf{u} + \lambda a^2 |\mathbf{u}|^2 = f, \\ \nabla \times \mathbf{u} = 0, \end{cases}$$

with the additional boundary condition

$$\tau \cdot \mathbf{u} = g' \quad \text{on } \partial\Omega. \tag{6.7}$$

It is clear that, in general, $\mathbf{u} \in H^s(\Omega)$ for $s < \alpha$ near the origin. In addition, since $a\mathbf{u} \in H(\mathrm{Div})$, u_1 is discontinuous across the vertical axis, while u_2 is discontinuous across the horizontal axis. This indicates that $\mathbf{u} \in H^s(\Omega)$ for $s < 1/2$.

While the entire system can be solved by minimizing a least-squares functional based on the above first-order system, it can also be addressed in two stages. The first stage is to approximately solve the Div-Curl system

$$\begin{cases} -\nabla \cdot a\mathbf{u} + \lambda a^2 |\mathbf{u}|^2 = f, \\ \nabla \times \mathbf{u} = 0, \end{cases} \tag{6.8}$$

with boundary condition (6.7). The general solution to linear Div-Curl systems has been studied in the context of FOSLS in [6, 5].

Given an approximate solution to (6.8), say $\mathbf{u}^h$, one can then find p^h in an appropriate H^1-conforming finite element space, $\mathcal{P}^h$, such that

$$p^h = \operatorname*{arg\,min}_{q^h \in \mathcal{P}^h} \|\nabla q^h - \mathbf{u}^h\|.$$

The approximation, p^h, inherits the accuracy of $\mathbf{u}^h$ subject to the properties of $\mathcal{P}^h$.

The purpose of this test is to demonstrate the behavior of the Newton-FOSLL* method on problems of reduced regularity. Towards that end, write (6.8) as

$$\mathcal{L}(\mathbf{u}) = L\mathbf{u} + \mathcal{N}(\mathbf{u}) = \begin{bmatrix} -\partial_x a & -\partial_y a \\ -\partial_y & \partial_x \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \lambda a^2 \begin{bmatrix} u_1^2 + u_2^2 \\ 0 \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix} = \mathbf{F}.$$

Linearization about $\mathbf{u}$ yields

$$\mathcal{L}'(\mathbf{u})\mathbf{v} = L\mathbf{v} + \mathcal{N}'(\mathbf{u})\mathbf{v} = \begin{bmatrix} -\partial_x a & -\partial_y a \\ -\partial_y & \partial_x \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + 2\lambda a^2 \begin{bmatrix} u_1 & u_2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

The adjoint operator applied to $\mathbf{z}$ is

$$\mathcal{L}'(\mathbf{u})^*\mathbf{z} = L^*\mathbf{z} + \mathcal{N}'(\mathbf{u})^*\mathbf{z} = \begin{bmatrix} a\partial_x & \partial_y \\ a\partial_y & -\partial_x \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + 2\lambda a^2 \begin{bmatrix} u_1 & 0 \\ u_2 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}. \tag{6.9}$$

Notice that the discontinuous coefficient appears outside of the derivative in $\mathcal{L}'(\mathbf{u})^*$. Also, note that the leading term has formally orthogonal columns if $\mathbf{z}$ has support that does not intersect a discontinuity of a.

Adjoint boundary conditions are determined by the differential part as follows:

$$\langle \nabla \cdot a\mathbf{u}, z_1 \rangle = -\langle \mathbf{u}, a\nabla z_1 \rangle + \oint a(\mathbf{n} \times \mathbf{u}) z_1, \tag{6.10}$$

$$\begin{aligned} \langle \nabla \times \mathbf{u}, z_2 \rangle &= \left\langle \mathbf{u}, \nabla^{\perp} z_1 \right\rangle + \oint (\tau \cdot \mathbf{u}) z_2 \\ &= \left\langle \mathbf{u}, \nabla^{\perp} z_1 \right\rangle + \oint \left(\frac{\partial g}{\partial \tau} \right) z_2. \end{aligned} \tag{6.11}$$

Since $\tau \cdot \mathbf{u}$ is known, while $\mathbf{n} \times \mathbf{u}$ is not known, the adjoint boundary conditions are

$$z_1 = 0 \quad \text{on } \partial\Omega. \tag{6.12}$$

With the additional requirement that $z_2 \in L^2_0(\Omega)$, it follows that $\mathcal{L}'(\mathbf{U})$ and $\mathcal{L}'(\mathbf{U})^*$ satisfy (4.1), (4.2), and (4.3).

To test FOSLL* on the above system, consider the singular function associated with this problem,

$$\psi(r, \phi) = r^{\alpha} \begin{cases} \cos(\alpha(\phi - \pi/4)) & \text{NE}, \\ -\beta \sin(\alpha(\phi - 3\pi/4)) & \text{NW}, \\ -\cos(\alpha(\phi - 5\pi/4)) & \text{SW}, \\ \beta \sin(\alpha(\phi - 7\pi/4)) & \text{SE}, \end{cases} \tag{6.13}$$

where $\beta = \cot(\alpha\pi/4)$. The singular function, $\psi(r, \phi)$, satisfies

$$\nabla \cdot a\nabla\psi = 0. \tag{6.14}$$

For this test, we let $p = \psi$, which yields

$$\mathbf{u} = \nabla\psi = \begin{bmatrix} \cos(\phi) & -\frac{1}{r}\sin(\phi) \\ \sin(\phi) & \frac{1}{r}\cos(\phi) \end{bmatrix} \begin{bmatrix} \partial_r \\ \partial_\phi \end{bmatrix} \psi(r, \phi) = \alpha r^{\alpha - 1} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix},$$

where

$$v_1 = \begin{cases} \cos(\phi)\cos(\alpha(\phi - \pi/4)) + \sin(\phi)\sin(\alpha(\phi - \pi/4)) & \text{NE}, \\ -\beta(\cos(\phi)\sin(\alpha(\phi - 3\pi/4)) - \sin(\phi)\cos(\alpha(\phi - 3\pi/4))) & \text{NW}, \\ -\cos(\phi)\cos(\alpha(\phi - 5\pi/4)) - \sin(\phi)\sin(\alpha(\phi - 5\pi/4)) & \text{SW}, \\ \beta(\cos(\phi)\sin(\alpha(\phi - 7\pi/4)) + \sin(\phi)\sin(\alpha(\phi - 7\pi/4))) & \text{SE} \end{cases}$$

and

$$v_2 = \begin{cases} \sin(\phi)\cos(\alpha(\phi - \pi/4)) - \cos(\phi)\sin(\alpha(\phi - \pi/4)) & \text{NE}, \\ -\beta(\sin(\phi)\sin(\alpha(\phi - 3\pi/4)) + \cos(\phi)\cos(\alpha(\phi - 3\pi/4))) & \text{NW}, \\ -\sin(\phi)\cos(\alpha(\phi - 5\pi/4)) + \cos(\phi)\sin(\alpha(\phi - 5\pi/4)) & \text{SW}, \\ \beta(\sin(\phi)\sin(\alpha(\phi - 7\pi/4)) - \cos(\phi)\sin(\alpha(\phi - 7\pi/4))) & \text{SE}. \end{cases}$$

Using (6.14), the right-hand side can be computed as

$$f = -\nabla \cdot a\nabla\psi + \lambda a^2 |\nabla\psi|^2 = \lambda a^2 |\nabla\psi|^2, \tag{6.15}$$

which behaves like $r^{2(\alpha-1)}$ near the origin. Note that $f \in L^2(\Omega)$ only if $\alpha > 1/2$, which we enforce in the tests below.

Since the boundary conditions on $\mathbf{u}$ are not homogeneous, they must be incorporated either through superposition or the addition of a boundary term in the weak form. Here, we appeal to (6.11) and add the term

$$\oint (\tau \cdot \mathbf{u}) z_2 = \oint (\tau \cdot \nabla \psi) z_2, \tag{6.16}$$

to the weak form. For more details on FOSLL* with nonhomogeneous boundary conditions, see [20, 36].

While all three methods described in section 3 are applicable, the *weak*-weak approach from subsection 3.2.4 outperformed the projected and extended approaches. Thus, only results for the *weak*-weak form are presented.

Given current iterate $\mathbf{u}_n^h$, we solve the discrete *weak*-weak problem based on (3.14), (3.15), (6.9), and (6.15), which is to find $\mathbf{w}^h \in \mathcal{W}^h$ such that

$$\langle \mathcal{L}'(\mathbf{u}_n^h)^* \mathbf{w}^h, \mathcal{L}'(\mathbf{u}_n^h)^* \mathbf{z}^h \rangle = \langle \mathbf{F} - \mathcal{N}(\mathbf{u}_n^h), \mathbf{z}^h \rangle - \langle \mathbf{u}_n^h, L^* \mathbf{z}^h \rangle + \oint (\tau \cdot \nabla \psi) z_2^h \tag{6.17}$$

for all $\mathbf{z}^h \in \mathcal{W}^h$. The next iterate is then computed directly according to

$$\mathbf{u}_{n+1}^h = \mathbf{u}_n + \mathcal{L}'(\mathbf{u}_n)^* \mathbf{w}^h.$$

A uniform triangulation of Ω is used, in which coefficient discontinuities are aligned with the mesh. As in the previous numerical example, for $\mathcal{W}^h$ we use standard $P2$ elements with boundary conditions as described above. The approximation for the Newton iterates, $\mathbf{u}_n^h$, uses $P1_{dc}$ elements, and the initial approximation is a piecewise smooth function that satisfies the discontinuous boundary conditions and is zero at the origin. The *weak*-weak continuation approach is clearly the most appropriate for this problem. Using either the projection or extended system approaches, as in the previous example, would require special finite element spaces that enforce a jump discontinuity across the axes. For this reason, we present results only for the *weak*-weak method.

For this example we also demonstrate the use of a nested iteration approach, where four nonlinear steps are taken on the initial coarse mesh ($N = 32$ elements per side), the mesh is then uniformly refined, and the approximation is interpolated to the new mesh. In this way, the initial nonlinear steps move the approximation into the basin of attraction for Newton's method, and each subsequent step can be performed on a refined mesh; see Figure 4. More intricate nested iteration schemes are certainly possible, but this illustrates the idea that it is generally most efficient to perform as much computational work as possible on coarse mesh levels. (Note: if a robust error estimator were available, it would also be possible to refine the mesh adaptively (cf. [2, 15, 28, 34]), though we employ uniform refinement here for simplicity. The projected and extended methods have a term that could potentially provide such an estimate. We leave this to future development.)

Figure 2 shows the domain and initial triangulation Ω^h. We also define subdomains $\Omega_1 = \Omega \backslash \{(x, y) : |x| < 0.125, |y| < 0.125\}$ and $\Omega_2 = \{(x, y) \in \Omega : |x| > 0.125, |y| > 0.125\}$, as pictured in Figure 2. Measuring the error on these domains, we are able to see the overall convergence rate, the rate excluding the singularity at the origin, and the rate excluding the discontinuities along the coordinate axes.

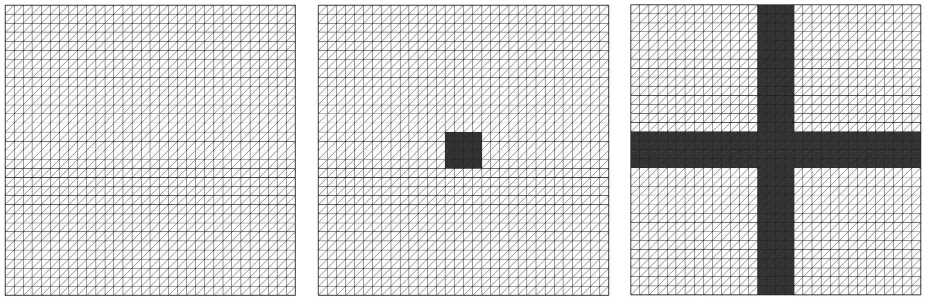


FIG. 2. *Discretization of* Ω *with* $N = 32$ *(left) and subdomains* Ω_1 *(center) and* Ω_2 *(right) in which the dark shaded parts are the regions excluded from* Ω.

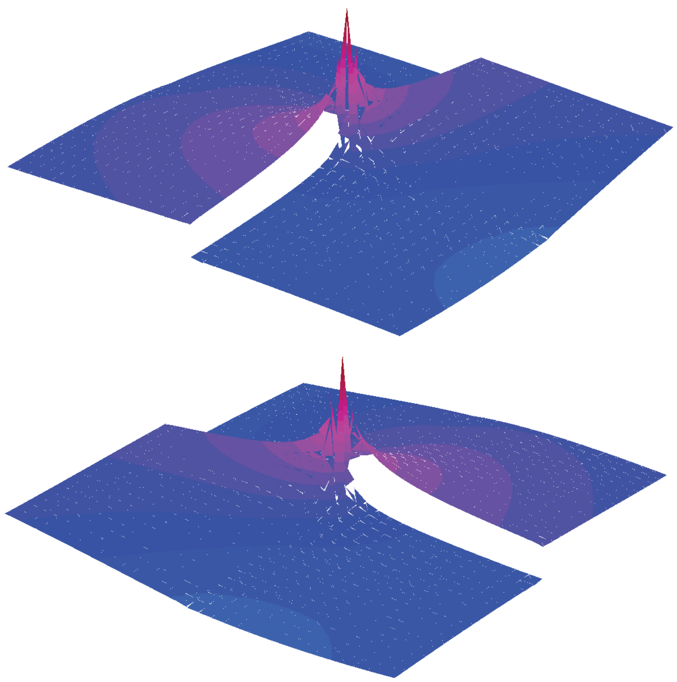
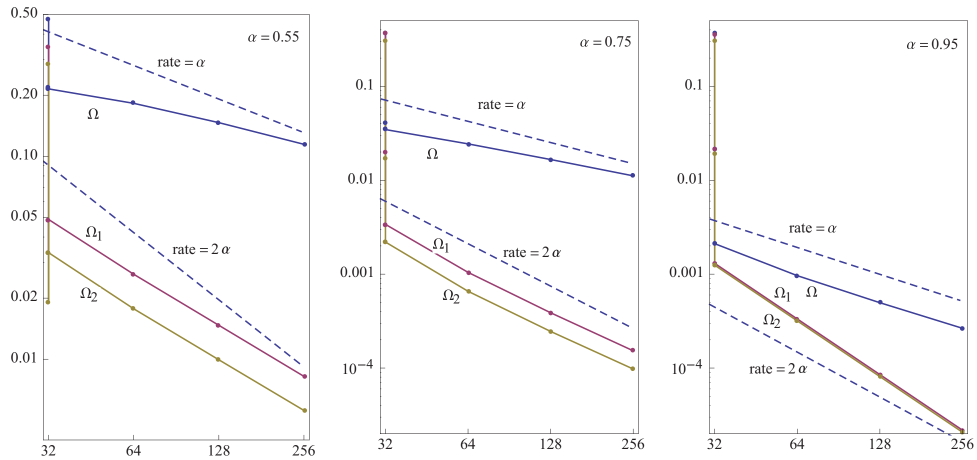


FIG. 3. *Numerical approximation for* u_1^h *(top) and* u_2^h *(bottom) at the third nonlinear step for* $\alpha = 0.75$ *and* $N = 32$ *elements per side (*2048 *total elements).*

Figure 3 shows a surface plot of a typical numerical approximation for $\mathbf{u}^h$, showing the singular behavior at the origin and the discontinuities along the axes. The advantage of FOSLL* can be clearly seen here. The PDE solve in each step is the solution of a symmetric and positive definite weak problem, done with simple conforming piecewise polynomial elements (i.e., $\mathbf{w}^h$), while the subsequent approximation of $\mathbf{u}^h$ allows optimally accurate representation of the discontinuities. In contrast to discontinuous Galerkin approaches, which also allow a discontinuous solution, the FOSLL* approach admits discontinuities in the postprocessing step rather than building the discontinuity into the variational problem.

Figure 4 shows the full numerical results for $\lambda = 1$ and $\alpha = 0.55$, 0.75, and 0.95, corresponding to a range of nonsmooth solutions. Convergence rates are ap-



	$\alpha = 0.55$			$\alpha = 0.75$			$\alpha = 0.95$		
N	Ω	Ω_1	Ω_2	Ω	Ω_1	Ω_2	Ω	Ω_1	Ω_2
32	0.471	0.348	0.286	0.371	0.364	0.307	0.364	0.363	0.311
32	0.219	0.0338	0.0190	0.0409	0.0201	0.0174	0.0215	0.0214	0.0189
32	0.215	0.0488	0.0334	0.0347	0.00337	0.00217	0.00211	0.00130	0.00125
32	0.215	0.0488	0.0335	0.0347	0.00339	0.00220	0.00211	0.00130	0.00124
64	0.182	0.0262	0.0177	0.0244	0.00105	6.57e-4	9.64e-4	3.33e-4	3.18e-4
128	0.146	0.0147	0.00994	0.0166	3.88e-4	2.45e-4	4.96e-4	8.52e-5	8.11e-5
256	0.115	0.00823	0.00559	0.0111	1.53e-4	9.76e-5	2.65e-4	2.18e-5	2.07e-5
Rate	0.35	0.84	0.83	0.58	1.35	1.33	0.90	1.97	1.97

FIG. 4. *L^2 convergence in Ω, Ω_1, and Ω_2, versus mesh resolution, N, for $\alpha = 0.55$, 0.75, and 0.95. A nested iteration approach was employed, where four nonlinear steps were computed on the initial coarse mesh, $N = 32$, while only one Newton step was required on each of the finer grids.*

proximately $\mathcal{O}(h^\alpha)$ for the L^2 error in Ω, and $\mathcal{O}(h^{2\alpha})$ in both Ω_1 and Ω_2. Because of the discontinuities along the axes, $\mathbf{u} \in H^s(\Omega)$ for $s < 1/2$, which in turn implies $\mathbf{w} \in H^s(\Omega)$ for $s < 3/2$. The bound (5.5) would suggest $\mathcal{O}(h^{1/2})$ convergence. Placing the finite element boundaries along the discontinuities of a allows for faster convergence. Convergence appears to be determined by the power of the singularity at the origin. However, the presence of the discontinuities does impact the overall convergence to some extent. The overall convergence is slightly slower than would be anticipated if only the singularity at the origin were present. Note also that, as α gets closer to 1.0, this diminished rate disappears, and convergence is $\mathcal{O}(h^\alpha)$ in Ω and slightly faster than $\mathcal{O}(h^{2\alpha})$ in Ω_1 and Ω_2.

The approach here allows computation using relatively simple finite element spaces and introduces the discontinuity through the postprocessing step. We note that previous work in least-squares methods using mixed $H(\mathrm{Div}) \times H^1(\Omega)$ based functionals can also effectively treat problems with similar discontinuous solutions (see, e.g., [39, 40]). Such approaches have a computable error indicator based on the least-squares functional, which can be used in an adaptive refinement algorithm.

7. Conclusion. This paper presents a natural framework for combining a Newton linearization and a FOSLL* discretization approach for nonlinear PDE. This approach provides a potentially discontinuous approximation that more directly controls the L^2 norm of the error than a typical least-squares approach. While a Newton linearization provides fast convergence for the nonlinear iteration, it ostensibly requires more smoothness than may seem available in such an approach. Several ways to overcome this difficulty were presented in this paper. Numerical results shed insight into the relative merits of three ways of continuing Newton's method. The projected method (subsection 3.2.2) and extended method (subsection 3.2.3) were more effective in the first example, for which the solution was relatively smooth, while the *weak*-weak form (subsection 3.2.4) was the natural choice in the third example, for which the solution was discontinuous and of reduced regularity. The second example illustrates how strong L^2 error control yields strong mass conservation in incompressible flow problems. Performance of a nonlinear nested iteration, combining mesh refinement and Newton's method, was also shown to be very effective in the third example.

As in [34], a hybrid FOSLS-FOSLL* approach is also possible for nonlinear problems. Such a coupling between the two approaches has the potential of providing an optimal balance between L^2 and H^1 seminorm accuracy and is the subject of a forthcoming study.

REFERENCES

[1] R. A. Adams and J. J. F. Fournier, *Sobolev Spaces*, 2nd ed., Elsevier/Academic Press, Amsterdam, 2003.

[2] J. H. Adler, T. A. Manteuffel, S. F. McCormick, J. W. Nolting, J. W. Ruge, and L. Tang, *Efficiency based adaptive local refinement for first-order system least-squares formulations*, SIAM J. Sci. Comput., 33 (2011), pp. 1–24.

[3] J. H. Adler, T. A. Manteuffel, S. F. McCormick, and J. W. Ruge, *First-order system least squares for incompressible resistive magnetohydrodynamics*, SIAM J. Sci Comput., 32 (2010), pp. 229–248.

[4] J. H. Adler, J. Brannick, C. Liu, T. Manteuffel, and L. Zikatanov, *First-order system least squares and the energetic variational approach for two-phase flow*, J. Comput. Phys., 230 (2011), pp. 6647–6663.

[5] M. Berndt, T. A. Manteuffel, and S. F. McCormick, *Analysis of first-order system least squares (FOSLS) for elliptic problems with discontinuous coefficients: Part* II, SIAM J. Numer. Anal., 43 (2005), pp. 409–436.

[6] M. Berndt, T. A. Manteuffel, S. F. McCormick, and G. Starke, *Analysis of first-order system least squares (FOSLS) for elliptic problems with discontinuous coefficients: Part* I, SIAM J. Numer. Anal., 43 (2005), pp. 386–408.

[7] G. Biswas, M. Breuer, and F. Durst, *Backward-facing step flows for various expansion ratios at low and moderate Reynolds numbers*, J. Fluids Eng., 126 (2004), pp. 362–374.

[8] P. Bochev, *Negative norm least-squares methods for the velocity-vorticity-pressure Navier-Stokes equations*, Numer. Methods Partial Differential Equations, 15 (1999), pp. 237–256.

[9] P. B. Bochev and M. D. Gunzburger, *Finite element methods of least-squares type*, SIAM Rev., 40 (1998), pp. 789–837.

[10] P. B. Bochev and M. D. Gunzburger, *Least-Squares Finite Element Methods*, Springer, New York, 2009.

[11] D. Braess, *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*, Cambridge University Press, Cambridge, UK, 2001.

[12] J. Bramble, R. Lazarov, and J. Pasciak, *A least-squares approach based on a discrete minus one inner product for first order systems*, Math. Comp., 66 (1997), pp. 935–955.

[13] J. Bramble and J. Pasciak, *Least-squares methods for Stokes equations based on a discrete minus one inner product*, J. Comput. Appl. Math., 74 (1996), pp. 155–173.

[14] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, Springer-Verlag, New York, 1994.

[15] M. Brezina, J. Garcia, T. Manteuffel, S. McCormick, J. Ruge, and L. Tang, *Parallel adaptive mesh refinement for first-order system least squares*, Numer. Linear Algebra Appl., 19 (2012), pp. 343–366.

[16] Z. Cai, R. Lazarov, T. A. Manteuffel, and S. F. McCormick, *First-order system least squares for second-order partial differential equations: Part* I, SIAM J. Numer. Anal., 31 (1994), pp. 1785–1799.

[17] Z. Cai, B. Lee, and P. Wang, *Least-squares methods for incompressible Newtonian fluid flow: Linear stationary problems*, SIAM J. Numer. Anal., 42 (2004), pp. 843–859.

[18] Z. Cai, T. A. Manteuffel, and S. F. McCormick, *First-order system least squares for second-order partial differential equations: Part* II, SIAM J. Numer. Anal., 34 (1997), pp. 425–454.

[19] Z. Cai, T. A. Manteuffel, and S. F. McCormick, *First-order system least squares for the Stokes equations, with application to linear elasticity*, SIAM J. Numer. Anal., 34 (1997), pp. 1727–1741.

[20] Z. Cai, T. A. Manteuffel, S. F. McCormick, and J. Ruge, *First-order system $\mathcal{LL}^*$ (FOSLL*): Scalar elliptic partial differential equations*, SIAM J. Numer. Anal., 39 (2001), pp. 1418–1445.

[21] Z. Cai, R. Falgout, and S. Zhang, *Div first-order system LL* (FOSLL*) for second-order elliptic partial differential equations*, SIAM J. Numer. Anal., 53 (2015), pp. 405–420.

[22] Z. Cai and C. R. Westphal, *An adaptive mixed least-squares finite element method for viscoelastic fluids of Oldroyd type*, J. Non-Newtonian Fluid. Mech., 159 (2009), pp. 72–80.

[23] P. Causin, R. Sacco, and C. L. Bottasso, *Flux-upwind stabilization of the discontinuous Petrov-Galerkin formulation with Lagrange multipliers for advection-diffusion problems*, M2AN Math. Model. Numer. Anal., 39 (2005), pp. 1087–1114.

[24] B. Cockburn, B. Dong, and J. Guzmán, *Optimal convergence of the original DG method for the transport-reaction equation on special meshes*, SIAM J. Numer. Anal., 46 (2008), pp. 1250–1265.

[25] A. L. Codd, T. A. Manteuffel, and S. F. McCormick, *Multilevel first-order system least squares for nonlinear elliptic partial differential equations*, SIAM J. Numer. Anal., 41 (2003), pp. 2197–2209.

[26] L. Demkowicz and J. Gopalakrishnan, *A class of discontinuous Petrov-Galerkin methods. Part* II: *Optimal test functions*, Numer. Methods Partial Differential Equations, 27 (2011), pp. 70–105.

[27] J. E. Dennis, Jr., and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM, Philadelphia, 1996.

[28] H. DeSterck, T. Manteuffel, S. McCormick, J. Nolting, J. Ruge, and L. Tang, *Efficiency-based h- and hp-refinement strategies for finite element methods*, Numer. Linear Algebra Appl., 15 (2008), pp. 89–114.

[29] H. Egger and J. Schöberl, *A hybrid mixed discontinuous Galerkin finite-element method for convection-diffusion problems*, IMA J. Numer. Anal., 30 (2010), pp. 1206–1234.

[30] V. Girault and P.-A. Raviart, *Finite Element Methods for Navier-Stokes Equations: Theory and Algorithms*, Springer-Verlag, Berlin, 1986.

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

[32] E. Lee and T. A. Manteuffel, *FOSLL* method for the eddy current problem with three-dimensional edge singularities*, SIAM J. Numer. Anal., 45 (2007), pp. 787–809.

[33] E. Lee, T. A. Manteuffel, and C. R. Westphal, *Weighted-norm first-order system least squares (FOSLS) for problems with corner singularities*, SIAM J. Numer. Anal., 44 (2006), pp. 1974–1996.

[34] K. Liu, T. A. Manteuffel, S. F. McCormick, J. W. Ruge, and L. Tang, *Hybrid first-order system least squares finite element methods with application to Stokes equations*, SIAM J. Numer. Anal., 51 (2013), pp. 2214–2237.

[35] T. Manteuffel, S. McCormick, and C. Pflaum, *Improved discretization error estimates for first-order system least squares*, J. Numer. Math., 11 (2003), pp. 163–177.

[36] T. A. Manteuffel, S. F. McCormick, J. Ruge, and J. G. Schmidt, *First-order system $\mathcal{LL}^*$ (FOSLL*) for general scalar elliptic problems in the plane*, SIAM J. Numer. Anal., 43 (2005), pp. 2098–2120.

[37] T. A. Manteuffel, S. F. McCormick, J. G. Schmidt, and C. R. Westphal, *First-order system least squares for geometrically nonlinear elasticity*, SIAM J. Numer. Anal., 44 (2006), pp. 2057–2081.

[38] S. A. Papanicolopulos and A. Zervos, *A method for creating a class of triangular C^1 finite elements*, Internat. J. Numer. Methods Engrg., 89 (2011), pp. 1437–1450.

[39] G. Starke, *Gauss-Newton multilevel methods for least-squares finite element computations of variably saturated subsurface flow*, Computing, 64 (2000), pp. 323–338.

[40] G. Starke, *Least-squares mixed finite element solution of variably saturated subsurface flow problems*, SIAM J. Sci. Comput., 21 (2000), pp. 1869–1885.

[41] O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method for Solid and Structural Mechanics*, 6th ed., Elsevier, Oxford, UK, 2005.