SIAM J. NUMER. ANAL.
Vol. 35, No. 3, pp. 990–1009, June 1998
© 1998 Society for Industrial and Applied Mathematics
006

ANALYSIS OF VELOCITY-FLUX FIRST-ORDER SYSTEM LEAST-SQUARES PRINCIPLES FOR THE NAVIER–STOKES EQUATIONS: PART I*

P. BOCHEV†, Z. CAI‡, T. A. MANTEUFFEL§, AND S. F. MCCORMICK§

Abstract. This paper develops a least-squares approach to the solution of the incompressible Navier–Stokes equations in primitive variables. As with our earlier work on Stokes equations, we recast the Navier–Stokes equations as a first-order system by introducing a *velocity-flux* variable and associated curl and *trace* equations. We show that a least-squares principle based on L^2 norms applied to this system yields optimal discretization error estimates in the H^1 norm in each variable, including the velocity flux.

An analogous principle based on the use of an H^{-1} norm for the reduced system (with no curl or trace constraints) is shown to yield similar estimates, but now in the L^2 norm for velocity-flux and pressure. Although the H^{-1} least-squares principle does not allow practical implementation, these results are critical to the analysis of a practical least-squares method for the reduced system based on a discrete equivalent of the negative norm. A practical method of this type is the subject of a companion paper. Finally, we establish optimal multigrid convergence estimates for the algebraic system resulting from the L^2 norm approach.

Key words. Navier–Stokes equations, least-squares functional, finite element methods, multigrid methods

AMS subject classifications. 76D05, 76D07, 65N12, 65N30, 65F10

PII. S0036142996313592

1. Introduction.

In [6], Cai, Manteuffel, and McCormick developed first-order system least-squares (FOSLS) functionals for formulation of the Stokes equations (generalized by a pressure-perturbed form of the continuity equation to allow for linear elasticity). Full ellipticity of the L^2-based least-squares formulation in n dimensions ($n = 2, 3$) was established by showing that the homogeneous form of the functional is equivalent to the $(H^1)^{n^2+n+1}$ norm applied to the first-order system variables (the new n^2-component velocity-flux variable, the n-component velocity variable, and the scalar pressure variable). This ellipticity immediately yields optimal discretization error estimates for standard finite elements in this H^1 product norm, as well as optimal convergence estimates for multigrid methods applied to the resulting discrete systems.

The aim of the present paper is to extend this methodology to the primitive variable form of the incompressible Navier–Stokes equations in two and three dimensions. We make this extension in the same way that the Stokes equations were reformulated based on the velocity-flux variable, but now we include the nonlinear convection term in the first-order system. We first consider an L^2-based least-squares formulation for the Navier–Stokes equations. The Euler–Lagrange equations

*Received by the editors December 12, 1996; accepted for publication March 3, 1997.
http://www.siam.org/journals/sinum/35-3/31359.html

†Department of Mathematics, Box 19408, University of Texas at Arlington, Arlington, TX 76019-0408 (bochev@utamat.uta.edu).

‡Department of Mathematics, Purdue University, 1395 Mathematical Science Building, West Lafayette, IN 47907-1395 (zcai@math.purdue.edu). This work was sponsored by the National Science Foundation under grant DMS-9619792.

§Department of Applied Mathematics, Campus Box 526, University of Colorado at Boulder, Boulder, CO 80309-0526 (tmanteuf@colorado.edu, stevem@colorado.edu). This work was sponsored by the National Science Foundation under grant DMS-9312752, and the Department of Energy under grant DE-FG03-94ER25217.

for the corresponding least-squares principle are then recast in the canonical form $F(\lambda,\mathcal{U}) \equiv \mathcal{U} + T \cdot G(\lambda,\mathcal{U}) = 0$, where T is the least-squares solution operator for the Stokes equations. This allows us to apply conventional abstract theory and our Stokes results to obtain optimal discretization and multigrid solution estimates for each variable (including velocity flux) in the H^1 norm.

These are the first H^1 product ellipticity results for the Navier–Stokes equations that admit practical velocity boundary conditions. Earlier work on the Stokes equations by Chang [8] used an *acceleration* variable analogous to our velocity flux; however, velocity was eliminated from the first-order system, which seems to prevent its extension to the Navier–Stokes equations, and, in any case, the formulation is limited to two dimensions. In [2], Bochev and Gunzburger developed a least-squares approach for the velocity-vorticity-pressure form of the Stokes equations, but showed that it does not allow H^1 product ellipticity in the velocity boundary condition case. (A mesh weighting was introduced in the functional to obtain optimal estimates.) Finally, Bochev [1] extended the velocity-vorticity-pressure methodology to the Navier–Stokes equations, but established H^1 product ellipticity only for nonstandard boundary conditions.

Next, we turn our attention to a least-squares functional in which the residual of the momentum equation is measured in the norm of the negative-order Sobolev space H^{-1}. For earlier work on H^{-1} norm functionals, we refer the reader to papers by Bramble, Lazarov, and Pasciak [4], [5] and Cai, Manteuffel, and McCormick [6]. In the present paper, our analysis again combines previous results on analogous operators for the Stokes equations with the abstract framework outlined above. Although in both cases we deal with similar abstract formulations of the least-squares principle, the use of negative-order norms yields a functional analytic setting in which ellipticity is established on a product of L^2 and H^1 spaces. As a result, the optimal discretization error estimates for the velocity flux are now derived in the L^2 norm. These are the first results concerning error analysis of H^{-1} functionals in the context of the nonlinear Navier–Stokes equations. One of the practical advantages of such functionals is their applicability to problems that do not exhibit H^2 regularity. One of the main purposes of the analysis here of H^{-1} least-squares principles is to provide the background for the study of a discrete, negative norm least-squares functional. The need to consider a discrete equivalent arises because negative norm per se involves exact solution of the Poisson equation, making it impractical computationally. Formulation and analysis of practical negative norm least-squares methods is the subject of a companion paper, referred to herein as Part II [3].

Along with discretization error estimates, we present an analysis of well-posedness of the least-squares variational problems, the importance of which stems from the fact that application of least-squares principles results in weak problems whose nonlinear terms are coupled with the Stokes operator. For both the L^2 and H^{-1} approaches, we show that a nonsingular branch of solutions of the original Navier–Stokes equations corresponds to a nonsingular branch of solutions of the least-squares variational problem.

This paper is organized as follows: in the next section, we introduce the Navier–Stokes equations and their first-order form; in section 3, we develop the associated L^2 least-squares principle; in section 4, we recast this L^2 principle in canonical form and apply a corresponding abstract theory to derive error estimates; in section 5, we develop a simple but optimal multigrid solver for the resulting discrete L^2 system; in section 6, we develop the H^{-1} least-squares approach and derive corresponding error

estimates; finally, in section 7, we prove well-posedness of the least-squares canonical forms for both the L^2 and H^{-1} principles based on well-posedness of the original Navier–Stokes equations.

Throughout the paper, we use boldface lower case font to denote vectors and underlined boldface upper case font to denote matrices.

2. Velocity-flux Navier–Stokes equations. In what follows, Ω will denote a bounded domain in $\mathbb{R}^n$, $n = 2, 3$, with Lipschitz continuous boundary Γ. The dimensionless equations governing the steady incompressible flow of a viscous fluid in domain Ω may be written in the form

$$-\nu\triangle\mathbf{u} + (\nabla\mathbf{u}^t)^t\mathbf{u} + \nabla p = \mathbf{f} \quad \text{in } \Omega, \tag{1}$$

$$\nabla^t\mathbf{u} = 0 \quad \text{in } \Omega, \tag{2}$$

where $\mathbf{u}$, p, and $\mathbf{f}$ denote velocity, pressure, and given body force, respectively, and ν is the inverse of the Reynolds number, λ. The velocity variable $\mathbf{u}$ is a column vector with scalar components u_i, so that $\nabla\mathbf{u}^t$ is a matrix with columns ∇u_i. Together with equations (1)–(2), we consider the velocity boundary condition

$$\mathbf{u} = \mathbf{0} \quad \text{on } \Gamma\,, \tag{3}$$

where Γ is the boundary of Ω. For uniqueness, we also impose the baseline pressure condition

$$\int_\Omega p d\Omega = 0. \tag{4}$$

To formulate the least-squares method, equations (1)–(2) will be transformed into an equivalent first-order system. The first step in this process is to introduce the *velocity-flux* variable

$$\underline{\mathbf{U}} = \nabla\mathbf{u}^t\,, \tag{5}$$

which is a matrix with entries $U_{ij} = \partial u_j/\partial x_i$, $1 \le i, j \le n$. Then

$$(\nabla^t\underline{\mathbf{U}})^t = \triangle\mathbf{u}$$

and it is easy to see that the new variable satisfies the identities

$$\mathrm{tr}\underline{\mathbf{U}} = 0\,, \quad \nabla\times\underline{\mathbf{U}} = \underline{\mathbf{0}} \quad \text{in } \Omega$$

and

$$\mathbf{n}\times\underline{\mathbf{U}} = \underline{\mathbf{0}} \quad \text{on } \Gamma\,, \tag{6}$$

where $\mathrm{tr}\underline{\mathbf{U}} = \sum_{i=1}^n U_{ii}$ and $\mathbf{n}$ is the outward unit normal on Γ. Furthermore, the nonlinear term in (1) takes the particularly simple form

$$(\nabla\mathbf{u}^t)^t\mathbf{u} = \underline{\mathbf{U}}^t\,\mathbf{u}\,.$$

As a result, the original Navier–Stokes system (1)–(2) can be replaced by the first-order system

$$-\nu(\nabla^t\underline{\mathbf{U}})^t + \underline{\mathbf{U}}^t\mathbf{u} + \nabla p = \mathbf{f} \quad \text{in } \Omega, \tag{7}$$

$$\nabla^t\mathbf{u} = 0 \quad \text{in } \Omega, \tag{8}$$

$$\underline{\mathbf{U}} - \nabla\mathbf{u}^t = \underline{\mathbf{0}} \quad \text{in } \Omega, \tag{9}$$

$$\nabla(\mathrm{tr}\underline{\mathbf{U}}) = \mathbf{0} \quad \text{in } \Omega, \tag{10}$$

$$\nabla\times\underline{\mathbf{U}} = \underline{\mathbf{0}} \quad \text{in } \Omega, \tag{11}$$

along with conditions (3), (4), and (6).

The second step in the formulation of a suitable first-order system is to scale the momentum equation by the Reynolds number and to replace the data $\mathbf{f}$ by functions with known boundary values. The resulting form of the equations will provide insight into the overall approach and facilitate error analysis of the corresponding least-squares method. For this purpose, we assume that $\mathbf{f} \in L^2(\Omega)^n$ and consider the unique solution $(\mathbf{u}_0, p_0)$ of the scaled Stokes problem

$$\begin{aligned} -\triangle \mathbf{u} + \nabla p &= \frac{1}{\nu}\mathbf{f} \quad \text{in } \Omega, \\ \nabla^t \mathbf{u} &= 0 \quad \text{in } \Omega, \\ \mathbf{u} &= \mathbf{0} \quad \text{on } \Gamma, \\ \int_\Omega p d\Omega &= 0. \end{aligned} \tag{12}$$

Letting $\underline{\mathbf{U}}_0 = \nabla \mathbf{u}_0^t$, then equation (7) is replaced by

$$-(\nabla^t \underline{\mathbf{U}})^t + \frac{1}{\nu}(\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t(\mathbf{u} + \mathbf{u}_0) + \nabla p = \mathbf{0} \ \text{ in } \Omega\,, \tag{13}$$

which is the principal equation that relates the perturbation $(\underline{\mathbf{U}}, \mathbf{u}, \nu p)$ to the Stokes solution $(\underline{\mathbf{U}}_0, \mathbf{u}_0^t, \nu p_0)$. To summarize, our reformulation yields the system

$$-(\nabla^t \underline{\mathbf{U}})^t + \frac{1}{\nu}(\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t(\mathbf{u} + \mathbf{u}_0) + \nabla p = \mathbf{0} \quad \text{in } \Omega, \tag{14}$$

$$\nabla^t \mathbf{u} = 0 \quad \text{in } \Omega, \tag{15}$$

$$\underline{\mathbf{U}} - \nabla \mathbf{u}^t = \underline{\mathbf{0}} \quad \text{in } \Omega, \tag{16}$$

$$\nabla(\text{tr } \underline{\mathbf{U}}) = \mathbf{0} \quad \text{in } \Omega, \tag{17}$$

$$\nabla \times \underline{\mathbf{U}} = \underline{\mathbf{0}} \quad \text{in } \Omega, \tag{18}$$

along with conditions (3), (4), and (6).

3. L^2 least-squares. The L^2 least-squares functional for first-order system (14)–(18), (3), (4), and (6) is defined as follows:

$$\begin{aligned} J(\underline{\mathbf{U}}, \mathbf{u}, p) = &\left\| -(\nabla^t \underline{\mathbf{U}})^t + \frac{1}{\nu}(\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t(\mathbf{u} + \mathbf{u}_0) + \nabla p \right\|_0^2 \\ &+ \|\nabla^t \mathbf{u}\|_0^2 + \|\underline{\mathbf{U}} - \nabla \mathbf{u}^t\|_0^2 + \|\nabla(\text{tr } \underline{\mathbf{U}})\|_0^2 + \|\nabla \times \underline{\mathbf{U}}\|_0^2\,. \end{aligned} \tag{19}$$

Note that our scaling of (7) by the Reynolds number is equivalent to the use of an L^2 norm weighted by λ^2 for the residual of this equation; see also [6].

To define the least-squares method, we need a suitable minimization problem. Let

$$\begin{aligned} &\mathbf{X} = \\ &\{(\underline{\mathbf{U}}, \mathbf{u}, p) \in H^1(\Omega)^{n^2} \times H^1(\Omega)^n \times H^1(\Omega) \cap L_0^2(\Omega) \,|\, \mathbf{u} = \mathbf{0}\,, \mathbf{n} \times \underline{\mathbf{U}} = \underline{\mathbf{0}} \ \text{ on } \Gamma\}\,, \end{aligned} \tag{20}$$

where $L_0^2(\Omega) = \{p \in L^2(\Omega) | \int_\Omega p d\Omega = 0\}$. Then the least-squares principle for functional (19) is

find $(\underline{\mathbf{U}}, \mathbf{u}, p) \in \mathbf{X}$ such that

$$J(\underline{\mathbf{U}}, \mathbf{u}, p) \le J(\underline{\mathbf{V}}, \mathbf{v}, q) \quad \text{for all } (\underline{\mathbf{V}}, \mathbf{v}, q) \in \mathbf{X}. \tag{21}$$

It is easy to see that the Euler–Lagrange equation for this minimization problem is given by the variational problem

find $(\underline{\mathbf{U}}, \mathbf{u}, p) \in \mathbf{X}$ such that

$$\begin{aligned}
&\mathcal{B}((\underline{\mathbf{U}}, \mathbf{u}, p), (\underline{\mathbf{V}}, \mathbf{v}, p)) \\
&\quad \equiv \Big(-(\nabla^t \underline{\mathbf{U}})^t + \frac{1}{\nu}(\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t(\mathbf{u} + \mathbf{u}_0) + \nabla p, \\
&\qquad\quad -(\nabla^t \underline{\mathbf{V}})^t + \frac{1}{\nu}\left((\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t \mathbf{v} + \underline{\mathbf{V}}^t(\mathbf{u} + \mathbf{u}_0)\right) + \nabla q \Big)_0 \\
&\quad + \left(\nabla^t \mathbf{u}, \nabla^t \mathbf{v}\right)_0 + \left(\nabla(\mathrm{tr}\underline{\mathbf{U}}), \nabla(\mathrm{tr}\underline{\mathbf{V}})\right)_0 \\
&\quad + \left(\underline{\mathbf{U}} - \nabla \mathbf{u}^t, \underline{\mathbf{V}} - \nabla \mathbf{v}^t\right)_0 + (\nabla \times \underline{\mathbf{U}}, \nabla \times \underline{\mathbf{V}})_0 = 0 \qquad (22)
\end{aligned}$$

for all $(\underline{\mathbf{V}}, \mathbf{v}, q) \in \mathbf{X}$.

Let $\mathbf{X}_h$ denote a finite-dimensional subspace of $\mathbf{X}$. Then the least-squares discretization method for the Navier–Stokes equations is defined by the following discrete variational problem:

find $(\underline{\mathbf{U}}^h, \mathbf{u}^h, p^h) \in \mathbf{X}_h$ such that

$$\mathcal{B}((\underline{\mathbf{U}}^h, \mathbf{u}^h, p^h), (\underline{\mathbf{V}}^h, \mathbf{v}^h, p^h)) = 0 \quad \text{for all } (\underline{\mathbf{V}}^h, \mathbf{v}^h, q^h) \in \mathbf{X}_h. \qquad (23)$$

It is easy to see that the discrete variational problem (23) corresponds to the necessary condition for the following discrete least-squares principle for (19):

find $(\underline{\mathbf{U}}^h, \mathbf{u}^h, p^h) \in \mathbf{X}_h$ such that

$$J(\underline{\mathbf{U}}^h, \mathbf{u}^h, p^h) \le J(\underline{\mathbf{V}}^h, \mathbf{v}^h, q^h) \quad \text{for all } (\underline{\mathbf{V}}^h, \mathbf{v}^h, q^h) \in \mathbf{X}_h. \qquad (24)$$

For space $\mathbf{X}_h$, we assume the following approximation property: there exists an integer $d \ge 1$ such that, for all $\underline{\mathbf{U}} \in H^{d+1}(\Omega)^{n^2}$, $\mathbf{u} \in H^{d+1}(\Omega)^n$, and $p \in H^{d+1}(\Omega)$, one can find $(\underline{\mathbf{U}}^h, \mathbf{u}^h, p^h) \in \mathbf{X}_h$ such that

$$\begin{aligned}
&\|\underline{\mathbf{U}} - \underline{\mathbf{U}}^h\|_\mu + \|\mathbf{u} - \mathbf{u}^h\|_\mu + \|p - p^h\|_\mu \\
&\le C h^{d+1-\mu} \left(\|\underline{\mathbf{U}}\|_{d+1} + \|\mathbf{u}\|_{d+1} + \|p\|_{d+1}\right), \qquad (25)
\end{aligned}$$

$\mu = 0, 1$. (Here and henceforth C is a generic constant that may change values with each occurrence. We take C to be independent of λ in the sense that it depends only on the maximum value in the compact set $\Lambda \subset \mathbb{R}^+$. C is also independent of h.) Note, for example, that (25) can be satisfied with $d = 1$ by choosing continuous piecewise linears for all variables.

4. Discretization error estimates. The main goal of this section is to derive error estimates for least-squares method (23). For this purpose, we show how to cast nonlinear problems (22) and (23) in the respective canonical forms

$$F(\lambda, \mathcal{U}) \equiv \mathcal{U} + T \cdot G(\lambda, \mathcal{U}) = \mathbf{0} \qquad (26)$$

and

$$F^h(\lambda, \mathcal{U}^h) \equiv \mathcal{U}^h + T_h \cdot G(\lambda, \mathcal{U}^h) = \mathbf{0}\,. \qquad (27)$$

This will allow us to apply the abstract approximation theory of [9]. The following function spaces will be needed below (with m representing some nonnegative integer):

$$\mathbf{X}^m = \left[H^{m+1}(\Omega)^{n^2} \times H^{m+1}(\Omega)^n \times H^{m+1}(\Omega)\right] \cap \mathbf{X}\,, \tag{28}$$

$$\mathbf{Y} = \mathbf{X}^*\,, \tag{29}$$

$$\mathbf{Z} = L^{3/2}(\Omega)^{n^2} \times L^{3/2}(\Omega)^n \times L^{3/2}(\Omega), \tag{30}$$

where $\mathbf{X}^*$ denotes the dual of $\mathbf{X}$ with respect to the L^2 inner product. The approximation in (27) is introduced by way of operator T_h. Therefore, the error estimates will depend largely on the nature of operator T and its approximation T_h. The basic idea is to define T to be the least-squares Stokes solution operator and T_h its finite element approximation. The approximation properties of these choices have been studied in [6]. Now, once T is known, operator G is then defined by the remaining terms in (22). The key is that the corresponding nonlinear part for T_h is also G, as we assert in our first lemma.

With this in mind, we make the identifications $\mathcal{U} = (\underline{\mathbf{U}}, \mathbf{u}, p)$, $\mathcal{U}^h = (\underline{\mathbf{U}}^h, \mathbf{u}^h, p^h)$, $\mathcal{V} = (\underline{\mathbf{V}}, \mathbf{v}, q)$, $\mathcal{V}^h = (\underline{\mathbf{V}}^h, \mathbf{v}^h, q^h)$, and $\lambda = 1/\nu$, and we assume that $\lambda \in \Lambda$, where Λ is a compact subset of $\mathbb{R}^+$. We then introduce the following:

$T : \mathbf{Y} \mapsto \mathbf{X}$ defined by $\mathcal{U} = T\mathbf{g}$ for $\mathbf{g} \in \mathbf{Y}$ if and only if

$$\begin{aligned}\mathcal{B}_S(\mathcal{U}, \mathcal{V}) \equiv{}& \left(-(\nabla^t \underline{\mathbf{U}})^t + \nabla p, -(\nabla^t \underline{\mathbf{V}})^t + \nabla q\right)_0 \\ &+ \left(\nabla^t \mathbf{u}, \nabla^t \mathbf{v}\right)_0 + (\nabla(\operatorname{tr} \underline{\mathbf{U}}), \nabla(\operatorname{tr} \underline{\mathbf{V}}))_0 \\ &+ \left(\underline{\mathbf{U}} - \nabla \mathbf{u}^t, \underline{\mathbf{V}} - \nabla \mathbf{v}^t\right)_0 + (\nabla \times \underline{\mathbf{U}}, \nabla \times \underline{\mathbf{V}})_0 \\ ={}& (\mathbf{g}_1, \underline{\mathbf{V}}) + (\mathbf{g}_2, \mathbf{v}) + (\mathbf{g}_3, q)\end{aligned} \tag{31}$$

for all $(\underline{\mathbf{V}}, \mathbf{v}, q) \in \mathbf{X}$;
$T_h : \mathbf{Y} \mapsto \mathbf{X}_h$ defined by $\mathcal{U}^h = T\mathbf{g}$ for $\mathbf{g} \in \mathbf{Y}$ if and only if

$$\mathcal{B}_S(\mathcal{U}^h, \mathcal{V}^h) = (\mathbf{g}_1, \underline{\mathbf{V}}^h) + (\mathbf{g}_2, \mathbf{v}^h) + (\mathbf{g}_3, q^h) \quad \text{for all } (\underline{\mathbf{V}}^h, \mathbf{v}^h, q^h) \in \mathbf{X}_h\,; \tag{32}$$

and

$G : \Lambda \times \mathbf{X} \to \mathbf{Y}$ with $\mathbf{g} = G(\lambda, \mathcal{U})$ for $\mathcal{U} \in \mathbf{X}$ if and only if

$$\begin{aligned}&(\mathbf{g}_1, \underline{\mathbf{V}}) + (\mathbf{g}_2, \mathbf{v}) + (\mathbf{g}_3, q) \\ &\quad= \left(-(\nabla^t \underline{\mathbf{U}})^t + \nabla p, \frac{1}{\nu}\left((\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t \mathbf{v} + \underline{\mathbf{V}}^t(\mathbf{u} + \mathbf{u}_0)\right)\right)_0 \\ &\quad+ \left(\frac{1}{\nu}(\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t(\mathbf{u} + \mathbf{u}_0)\,,\right. \\ &\qquad\left. -(\nabla^t \underline{\mathbf{V}})^t + \nabla q + \frac{1}{\nu}\left((\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t \mathbf{v} + \underline{\mathbf{V}}^t(\mathbf{u} + \mathbf{u}_0)\right)\right)_0\end{aligned} \tag{33}$$

for all $(\underline{\mathbf{V}}, \mathbf{v}, q) \in \mathbf{X}$.

We then have the following equivalence.

LEMMA 1. *Assume that T, T_h, and G are defined by* (31), (32), *and* (33), *respectively. Then nonlinear problem* (22) *is equivalent to* (26) *and discrete nonlinear problem* (23) *is equivalent to* (27).

Proof. Assume that $\mathcal{U} = (\underline{\mathbf{U}}, \mathbf{u}, p)$ solves problem (26) with T and G given by (31) and (33), respectively. Then $\mathcal{U} = -T\mathbf{g}$ if and only if

$$\mathcal{B}_S(\mathcal{U}, \mathcal{V}) = (\mathbf{g}, \mathcal{V}) \quad \text{for all} \quad \mathcal{V} \in \mathbf{X}\,,$$

and $\mathbf{g} = G(\lambda,\mathcal{U})$ if and only if (33) holds. It follows that $\mathcal{U}$ also solves variational problem (22). Conversely, if $\mathcal{U}$ solves (22), let $\mathbf{g}$ be defined by (33). Then $\mathcal{B}_S(\mathcal{U}, \mathcal{V}) = (\mathbf{g}, \mathcal{V})$ for all $\mathcal{V} \in \mathbf{X}$, i.e., $\mathcal{U} = -T\mathbf{g}$. Thus, (22) and (26) are equivalent. Proof of the equivalence of (23) and (27) is identical. ∎

Error estimates for least-squares method (23) will now be derived from the abstract approximation theory of [9]. Below we state the main result of this theory for general T and T_h, but otherwise specialized to our needs. Here we let $D_{\mathcal{U}}G(\lambda,\mathcal{U})$ and $D_{\mathcal{U}}F(\lambda,\mathcal{U})$ denote the Fréchet derivative of G and F with respect to $\mathcal{U}$. We refer to $\{(\lambda,\mathcal{U}(\lambda)) \,|\, \lambda \in \Lambda\}$ as a *regular branch of solutions* of (26) if $\mathcal{U} = \mathcal{U}(\lambda)$ is a weak solution of (26) for each $\lambda \in \Lambda$, $\lambda \mapsto \mathcal{U}(\lambda)$ is a continuous map $\Lambda \mapsto \mathbf{X}$, and $D_{\mathcal{U}}F(\lambda,\mathcal{U})$ is an isomorphism of $\mathbf{X}$.

THEOREM 1. *Let $F(\lambda,\mathcal{U}) = \mathbf{0}$ denote abstract form* (26) *and assume that $\{(\lambda,\mathcal{U}(\lambda)) \,|\, \lambda \in \Lambda\}$ is a branch of regular solutions of* (26). *Furthermore, assume that $T \in L(\mathbf{Y}, \mathbf{X})$, that G is a C^2 map $\Lambda \times \mathbf{X} \mapsto \mathbf{Y}$ such that all second derivatives of G are bounded on bounded subsets of $\Lambda \times \mathbf{X}$, and that there exists a space $\mathbf{Z} \subset \mathbf{Y}$, with continuous imbedding, such that $D_{\mathcal{U}}G(\lambda,\mathcal{U}) \in L(\mathbf{X}, \mathbf{Z})$ for all $\lambda \in \Lambda$ and $\mathcal{U} \in \mathbf{X}$. If approximate problem* (27) *is such that*

$$\lim_{h\to 0} \|(T - T_h)\mathbf{g}\|_{\mathbf{X}} = 0 \tag{34}$$

for all $\mathbf{g} \in \mathbf{Y}$ and

$$\lim_{h\to 0} \|T - T_h\|_{L(\mathbf{Z},\mathbf{X})} = 0. \tag{35}$$

Then:

1. *there exists a neighborhood $\mathcal{O}$ of the origin in $\mathbf{X}$ and, for h sufficiently small, a unique C^2 function $\lambda \mapsto \mathcal{U}^h(\lambda) \in \mathbf{X}_h$ such that $\{(\lambda,\mathcal{U}^h(\lambda)) \,|\, \lambda \in \Lambda\}$ is a branch of regular solutions of discrete problem* (27) *and $\mathcal{U}(\lambda) - \mathcal{U}^h(\lambda) \in \mathcal{O}$ for all $\lambda \in \Lambda$;*

2. *for all $\lambda \in \Lambda$ we have*

$$\|\mathcal{U}^h(\lambda) - \mathcal{U}(\lambda)\|_{\mathbf{X}} \le C\|(T - T^h)G(\lambda,\mathcal{U}(\lambda))\|_{\mathbf{X}}\,; \tag{36}$$

3. *if the regular branch is such that $\mathcal{U}(\lambda) \in \mathbf{X}^m$ for some integer $m \ge 1$ and $\tilde{d} \equiv \min\{d, m\}$, where d is the largest integer satisfying* (25), *then*

$$\begin{aligned} &\|\underline{\mathbf{U}}(\lambda) - \underline{\mathbf{U}}^h(\lambda)\|_1 + \|\mathbf{u}(\lambda) - \mathbf{u}^h(\lambda)\|_1 + \|p(\lambda) - p^h(\lambda)\|_1 \\ &\qquad \le Ch^{\tilde{d}} \left(\|\underline{\mathbf{U}}(\lambda)\|_{\tilde{d}+1} + \|\mathbf{u}(\lambda)\|_{\tilde{d}+1} + \|p(\lambda)\|_{\tilde{d}+1}\right)\,. \end{aligned} \tag{37}$$

In the next few lemmas, we verify the hypotheses of Theorem 1 for our least-squares formulation. We begin by establishing essential properties of operators T and T_h, which we assume, for this and the next section, are defined by (31) and (32), respectively.

LEMMA 2. *$T \in L(\mathbf{Y}, \mathbf{X})$ and $T_h \in L(\mathbf{Y}, \mathbf{X}_h)$.*

Proof. Form $\mathcal{B}_S(\cdot,\cdot)$ is continuous and coercive on $\mathbf{X} \times \mathbf{X}$ (see [6]) and, by virtue of the inclusion $\mathbf{X}_h \subset \mathbf{X}$, it is also continuous and coercive on $\mathbf{X}_h \times \mathbf{X}_h$. Furthermore, for each $\mathbf{g} \in \mathbf{Y}$, $(\mathbf{g}, \mathcal{V})$ defines a continuous functional on $\mathbf{X}$. Thus, the Lax–Milgram

theorem implies that, for all $\mathbf{g} \in \mathbf{Y}$, variational problems (31) and (32) have unique respective solutions $\mathcal{U} \in \mathbf{X}$ and $\mathcal{U}^h \in \mathbf{X}_h$, i.e., $T : \mathbf{Y} \mapsto \mathbf{X}$ and $T_h : \mathbf{Y} \mapsto \mathbf{X}_h$ are well-defined linear operators. From

$$C\|\mathcal{U}\|_{\mathbf{X}}^2 \leq \mathcal{B}_S(\mathcal{U},\mathcal{U}) = (\mathbf{g},\mathcal{U}) \leq \|\mathbf{g}\|_{\mathbf{Y}}\|\mathcal{U}\|_{\mathbf{X}},$$

it follows that

$$\|T\mathbf{g}\|_{\mathbf{X}} = \|\mathcal{U}\|_{\mathbf{X}} \leq C\|\mathbf{g}\|_{\mathbf{Y}};$$

i.e., T is in $L(\mathbf{Y},\mathbf{X})$. The proof that $T_h \in L(\mathbf{Y},\mathbf{X}_h)$ is similar. □

Before continuing with the approximation properties of T_h, consider the choice of $\mathbf{Y}$ and $\mathbf{Z}$ in (29) and (30). When $\mathbf{Z} \subset \mathbf{Y}$ with compact imbedding, the proof of (35) in Theorem 1 can be simplified. First, note that $\mathbf{Y}$ is not identical to a product of $H^{-1}(\Omega)$ spaces. For instance, with $\underline{\mathbf{U}}_i$ denoting the ith column of $\underline{\mathbf{U}}$, then $\underline{\mathbf{U}}_i \in \mathbf{H}_t^1(\Omega) = \{\mathbf{v} \in H^1(\Omega)^n \ | \mathbf{n} \times \mathbf{v} = 0 \quad \text{on } \Gamma\}$, whose dual is not $H^{-1}(\Omega)^n$. As a result, $\mathbf{Z}$ will be compactly imbedded in $\mathbf{Y}$ if $L^{3/2}(\Omega)$ is compactly imbedded in the duals of $H_0^1(\Omega)$, $\mathbf{H}_t^1(\Omega)$, and $H^1(\Omega)$. The first imbedding follows from Sobolev's imbedding theorem; see, e.g., [9]. Compactness of the other two imbeddings can be shown along the following lines. Since components of $\mathbf{H}_t^1(\Omega)$ and the space $H^1(\Omega)$ are compactly imbedded in $L^3(\Omega)$ and the adjoint of a compact operator is compact, it follows that $L^{3/2}(\Omega)^n$ and $L^{3/2}(\Omega)$ are compactly imbedded in the dual spaces of $\mathbf{H}_t^1(\Omega)$ and $H^1(\Omega)$.

LEMMA 3. *Convergence properties* (34) *and* (35) *hold. If, in addition,* $\mathbf{g} \in \mathbf{Y}$ *is such that* $T\mathbf{g} \in \mathbf{X}^m$ *for some* $m \geq 1$ *and* $\tilde{d} = \min(d, m)$*, where* d *is the largest integer satisfying* (25)*, then*

$$\|(T - T_h)\mathbf{g}\|_{\mathbf{X}} \leq Ch^{\tilde{d}} \|T\mathbf{g}\|_{\mathbf{X}^{\tilde{d}+1}}. \tag{38}$$

Proof. First note that (35) follows from (34) when the imbedding $\mathbf{Z} \subset \mathbf{Y}$ is compact. It thus suffices to establish (34), i.e., that

$$\|(T - T_h)\mathbf{g}\|_{\mathbf{X}} = \|\underline{\mathbf{U}} - \underline{\mathbf{U}}^h\|_1 + \|\mathbf{u} - \mathbf{u}^h\|_1 + \|p - p^h\|_1 \to 0$$

when $h \to 0$. Recall that $T : \mathbf{Y} \mapsto \mathbf{X}$. Therefore, from $\mathbf{g} \in \mathbf{Y}$ it follows that $\mathcal{U} \in \mathbf{X}$, i.e., that $\underline{\mathbf{U}} \in H^1(\Omega)^{n^2}$, $\mathbf{u} \in H^1(\Omega)^n$, and $p \in H^1(\Omega)$. Then the above limit follows from the definition of $\mathbf{X}_h$, (25), Cea's lemma, and the standard approximation result for $v \in H^1(\Omega)$:

$$\lim_{h \to 0} \inf_{v^h} \|v - v^h\|_1 = 0.$$

(See [7] for an analogous result for scalar elliptic equations.)

To prove the second part of the lemma, suppose $\mathcal{U} = T\mathbf{g} \in \mathbf{X}^m$. Then an immediate consequence of (25) and the continuity and coercivity of $\mathcal{B}_S(\cdot,\cdot)$ is the Stokes error estimate

$$\|(T-T_h)\mathbf{g}\|_{\mathbf{X}} = \|\underline{\mathbf{U}}-\underline{\mathbf{U}}^h\|_1+\|\mathbf{u}-\mathbf{u}^h\|_1+\|p-p^h\|_1 \leq Ch^{\tilde{d}} \left(\|\underline{\mathbf{U}}\|_{\tilde{d}+1} + \|\mathbf{u}\|_{\tilde{d}+1} + \|p\|_{\tilde{d}+1}\right).$$

□

The only hypotheses of Theorem 1 that remain to be verified are the assumptions concerning the nonlinear operator G. For this purpose, we need the weak and strong forms of the first Fréchet derivative $D_{\mathcal{U}}G(\lambda,\mathcal{U})$ and the weak form of the second

Fréchet derivative $D^2_{\mathcal{U}}G(\lambda,\mathcal{U})$. To determine the weak form of $D_{\mathcal{U}}G(\lambda,\mathcal{U})$, let $\hat{\mathcal{U}} \in \mathbf{X}$, substitute $\mathcal{U}+\hat{\mathcal{U}}$ into (33), and expand about $\mathcal{U}$. This yields the following weak representation of $D_{\mathcal{U}}G(\lambda,\mathcal{U})$:

$D_{\mathcal{U}}G(\lambda,\mathcal{U}) : \Lambda \times \mathbf{X} \to \mathbf{Y}$ defined by $\mathbf{g} = D_{\mathcal{U}}G(\lambda,\mathcal{U})\hat{\mathcal{U}}$ for $\mathcal{U} \in \mathbf{X}$ if and only if

$$
\begin{aligned}
&(\mathbf{g}_1, \underline{\mathbf{V}}) + (\mathbf{g}_2, \mathbf{v}) + (\mathbf{g}_3, q) \\
&\quad = \left(-(\nabla^t \underline{\mathbf{U}})^t + \nabla p, \frac{1}{\nu}\left(\hat{\underline{\mathbf{U}}}^t \mathbf{v} + \underline{\mathbf{V}}^t \hat{\mathbf{u}} \right) \right)_0 \\
&\quad + \left(-(\nabla^t \hat{\underline{\mathbf{U}}})^t + \nabla \hat{p}, \frac{1}{\nu}\left((\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t \mathbf{v} + \underline{\mathbf{V}}^t (\mathbf{u} + \mathbf{u}_0) \right) \right)_0 \\
&\quad + \left(\frac{1}{\nu}(\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t (\mathbf{u} + \mathbf{u}_0), \frac{1}{\nu}\left(\hat{\underline{\mathbf{U}}}^t \mathbf{v} + \underline{\mathbf{V}}^t \hat{\mathbf{u}} \right) \right)_0 \\
&\quad + \left(\frac{1}{\nu}\left((\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t \hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t (\mathbf{u} + \mathbf{u}_0) \right), \right. \\
&\qquad \left. -(\nabla^t \underline{\mathbf{V}})^t + \nabla q + \frac{1}{\nu}\left((\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t \mathbf{v} + \underline{\mathbf{V}}^t (\mathbf{u} + \mathbf{u}_0) \right) \right)_0
\end{aligned} \tag{39}
$$

for all $(\underline{\mathbf{V}}, \mathbf{v}, q) \in \mathbf{X}$.

The strong form of $D_{\mathcal{U}}G(\lambda,\mathcal{U})\hat{\mathcal{U}}$ can be found from (39) using standard integration by parts:

$$
\begin{aligned}
\mathbf{g}_1 &= \frac{1}{\nu}\hat{\mathbf{u}}\left(-(\nabla^t \underline{\mathbf{U}})^t + \nabla p + \frac{1}{\nu}(\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t (\mathbf{u} + \mathbf{u}_0) \right)^t \\
&\quad + \frac{1}{\nu}(\mathbf{u} + \mathbf{u}_0)\left(-(\nabla^t \hat{\underline{\mathbf{U}}})^t + \nabla \hat{p} + \frac{1}{\nu}\left((\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t \hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t (\mathbf{u} + \mathbf{u}_0) \right) \right)^t \\
&\quad + \frac{1}{\nu}\nabla\left((\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t \hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t (\mathbf{u} + \mathbf{u}_0) \right)^t,
\end{aligned} \tag{40}
$$

$$
\begin{aligned}
\mathbf{g}_2 &= \frac{1}{\nu}\hat{\underline{\mathbf{U}}}\left(-(\nabla^t \underline{\mathbf{U}})^t + \nabla p + \frac{1}{\nu}(\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t (\mathbf{u} + \mathbf{u}_0) \right) \\
&\quad + \frac{1}{\nu}(\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)\left(-(\nabla^t \hat{\underline{\mathbf{U}}})^t + \nabla \hat{p} + \frac{1}{\nu}\left((\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t \hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t (\mathbf{u} + \mathbf{u}_0) \right) \right),
\end{aligned} \tag{41}
$$

$$
\mathbf{g}_3 = -\frac{1}{\nu}\nabla^t\left((\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t \hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t (\mathbf{u} + \mathbf{u}_0) \right) \tag{42}
$$

for all $(\underline{\mathbf{V}}, \mathbf{v}, q) \in \mathbf{X}$.

Finally, the weak form of the second Fréchet derivative is

$D^2_{\mathcal{U}}G(\lambda,\mathcal{U}) : \Lambda \times [\mathbf{X} \times \mathbf{X}] \to \mathbf{Y}$ defined by $\mathbf{g} = D^2_{\mathcal{U}}G(\lambda,\mathcal{U})[\hat{\mathcal{U}}, \hat{\hat{\mathcal{U}}}]$ for $\mathcal{U} \in \mathbf{X}$ if and only if

$$
\begin{aligned}
&(\mathbf{g}_1, \underline{\mathbf{V}}) + (\mathbf{g}_2, \mathbf{v}) + (\mathbf{g}_3, q) \\
&\quad = \left(-(\nabla^t \hat{\underline{\mathbf{U}}})^t + \nabla \hat{\hat{p}} + \frac{1}{\nu}\left(\hat{\hat{\underline{\mathbf{U}}}}^t (\mathbf{u} + \mathbf{u}_0) + (\underline{\mathbf{U}} + \underline{\mathbf{U}}_0)^t \hat{\hat{\mathbf{u}}} \right) \right.,
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{\nu}\,(\hat{\underline{\mathbf{U}}}^t\mathbf{v}+\underline{\mathbf{V}}^t\hat{\mathbf{u}})\Big)_0\\
&+\frac{1}{\nu}\left(-(\nabla^t\hat{\underline{\mathbf{U}}})^t+\nabla\hat{p}+\hat{\underline{\mathbf{U}}}^t(\mathbf{u}+\mathbf{u}_0)+(\underline{\mathbf{U}}+\underline{\mathbf{U}}_0)^t\hat{\mathbf{u}},\,\frac{1}{\nu}(\hat{\hat{\underline{\mathbf{U}}}}^t\mathbf{v}+\underline{\mathbf{V}}^t\hat{\mathbf{u}})\right)_0\\
&+\left(\frac{1}{\nu}\left(\hat{\underline{\mathbf{U}}}^t\hat{\mathbf{u}}+\hat{\hat{\underline{\mathbf{U}}}}^t\hat{\mathbf{u}}\right),\right.\\
&\qquad\left.-(\nabla^t\underline{\mathbf{V}})^t+\nabla q+\frac{1}{\nu}\left((\underline{\mathbf{U}}+\underline{\mathbf{U}}_0)^t\mathbf{u}+\underline{\mathbf{V}}^t(\mathbf{u}+\mathbf{u}_0)\right)\right)_0 \qquad (43)
\end{aligned}
$$

for all $(\underline{\mathbf{V}},\mathbf{v},q)\in\mathbf{X}$.

The next lemma summarizes technical results that we use below.

LEMMA 4. *Let D_i denote the derivative with respect to the ith coordinate variable in $\mathbb{R}^n$, $1\le i\le n$, and assume that u, v, w, and z are in $H^1(\Omega)$. Then*

$$\left|\int_\Omega D_i u\,v\,w\,d\Omega\right|\le C\|u\|_1\,\|v\|_1\,\|w\|_1\,, \qquad (44)$$

$1\le i\le n$, and

$$\left|\int_\Omega u\,v\,w\;z\,d\Omega\right|\le C\|u\|_1\,\|v\|_1\,\|w\|_1\,\|z\|_1\,. \qquad (45)$$

Moreover, $(u,v)\mapsto uv$ is a continuous bilinear mapping from $L^2(\Omega)\times H^1(\Omega)$ into $L^{3/2}(\Omega)$ and $(u,v,w)\mapsto uvw$ is a continuous trilinear mapping from $H^1(\Omega)\times H^1(\Omega)\times H^1(\Omega)$ into $L^{3/2}(\Omega)$; i.e.,

$$\|uv\|_{0,3/2}\le C\|u\|_{0,2}\|v\|_{1,2}\quad \textit{for all } u\in L^2(\Omega) \textit{ and } v\in H^1(\Omega)\,, \qquad (46)$$

$$\|uvw\|_{0,3/2}\le C\|u\|_{1,2}\|v\|_{1,2}\|w\|_{1,2}\quad \textit{for all } u,v,w\in H^1(\Omega)\,. \qquad (47)$$

Proof. The first part of the lemma follows easily from the imbedding $H^1(\Omega)\subset L^4(\Omega)$ in two and three dimensions and the Hölder inequality. The second part follows directly from a result in [9] (see Corollary 1.1, p. 5). □

For a more general version of (44) and (45), see [11].

In the next lemma, we establish properties of G that are required for the validity of the approximation result in Theorem 1.

LEMMA 5. *Assume that mapping G is defined by* (33). *For $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ given by* (20), (29), *and* (30), *respectively, the following are true.*

1. *For all $\mathcal{U}\in\mathbf{X}$, $D_{\mathcal{U}}G(\lambda,\mathcal{U})\in L(\mathbf{X},\mathbf{Z})$.*
2. *The second Fréchet derivative $D^2_{\mathcal{U}}G(\lambda,\mathcal{U})$ is bounded on bounded subsets of $\Lambda\times\mathbf{X}$.*

Proof. To prove 1, consider strong form (40)–(42) of $D_{\mathcal{U}}G(\lambda,\mathcal{U})$. By assumption, $\mathcal{U}\in\mathbf{X}$; i.e., $\underline{\mathbf{U}}\in H^1(\Omega)^{n^2}$, $\mathbf{u}\in H^1(\Omega)^n$, and $p\in H^1(\Omega)$. Now each equation (40), (41), and (42) consists of terms of the form $D_i u\,v$ and uvw, where u, v, and w belong to $H^1(\Omega)$, so the second part of Lemma 4 implies that $(\mathbf{g}_1,\mathbf{g}_2,\mathbf{g}_3)\in\mathbf{Z}$. Using (46) and (47), it also follows that

$$\|D_{\mathcal{U}}G(\lambda,\mathcal{U})\hat{\mathcal{U}}\|_{\mathbf{Z}}\le C\|\hat{\mathcal{U}}\|_{\mathbf{X}}\,,$$

i.e., that $D_{\mathcal{U}}G(\lambda,\mathcal{U})\in L(\mathbf{X},\mathbf{Z})$.

To prove 2, consider weak form (43) of the second Fréchet derivative. Assume that $(\lambda,\mathcal{U})$ belongs to a bounded subset of $\Lambda\times\mathbf{X}$ and let $\hat{\mathcal{U}}, \hat{\tilde{\mathcal{U}}} \in \mathbf{X}$ be arbitrary. Then it is not difficult to see that weak form (43) involves only terms of the form $D_i uvw$ and $uvwz$, where u, v, w, and z belong to $H^1(\Omega)$. Thus, each term can be estimated using (44) or (45):

$$|(\mathbf{g}_1, \underline{\mathbf{V}})| \le C_1(\lambda,\mathcal{U},\mathcal{U}_0)(\|\hat{\mathcal{U}}\|_{\mathbf{X}} + \|\hat{\tilde{\mathcal{U}}}\|_{\mathbf{X}})\|\underline{\mathbf{V}}\|_1,$$

$$|(\mathbf{g}_2, \mathbf{u})| \le C_2(\lambda,\mathcal{U},\mathcal{U}_0)(\|\hat{\mathcal{U}}\|_{\mathbf{X}} + \|\hat{\tilde{\mathcal{U}}}\|_{\mathbf{X}})\|\mathbf{u}\|_1,$$

$$|(\mathbf{g}_3, q)| \le C_3(\lambda,\mathcal{U},\mathcal{U}_0)(\|\hat{\mathcal{U}}\|_{\mathbf{X}} + \|\hat{\tilde{\mathcal{U}}}\|_{\mathbf{X}})\|q\|_1,$$

where C_i is polynomial function of λ, $\|\mathcal{U}\|_{\mathbf{X}}$, and $\|\mathcal{U}_0\|_{\mathbf{X}}$. In combination with the fact that λ and $\|\mathcal{U}\|_{\mathbf{X}}$ are in bounded subsets of $\Lambda \times \mathbf{X}$, and that $\|\mathcal{U}_0\|_{\mathbf{X}}$ is fixed, it follows that $D^2_{\mathcal{U}} G(\lambda,\mathcal{U})$ is bounded in the norm of $L(\mathbf{X}, L(\mathbf{X}, \mathbf{Y}))$. □

This completes verification of all assumptions of Theorem 1. As a result, we can conclude that error estimates (36) and (37) hold for the least-squares finite element approximation as long as problem (22) has a regular branch of solutions with sufficient regularity.

5. Multigrid solver. Here we consider a simple iterative method applied to (27) and show that it converges linearly with bound uniform in h and λ. Our approach rests on using a multigrid preconditioner for T_h and observing that the operator in (27) is well conditioned uniformly in h and λ. The development is greatly simplified by basing the analysis on inner product $\mathcal{B}_S(\cdot,\cdot)$ defined in (31) and by choosing elements of the multigrid-based algorithm that are very easy to analyze. (Most assumptions are made only for convenience; more general conditions can be handled with more cumbersome but straightforward arguments. However, allowing for the more effective *direct* treatment of the nonlinearity within the multigrid process would require much more sophisticated analysis tools than we use here.)

Let M_h be defined so that $\mathcal{U}^h = M_h\mathbf{g}$ represents one or more cycles of (additive or multiplicative) multigrid applied to problem (32), starting from the initial guess $\mathcal{U}^h = 0$. For simplicity, assume that M_h is symmetric in the $\mathcal{B}_S(\cdot,\cdot)$ inner product (e.g., M_h may consist of one relaxation of point Gauss–Seidel with a given ordering before coarsening and one relaxation with the reverse ordering afterwards). Again, for simplicity, assume that M_h is so effective that

$$\delta\mathcal{B}_S(T_h\mathcal{V}^h,\mathcal{V}^h) \le \mathcal{B}_S(M_h\mathcal{V}^h,\mathcal{V}^h) \le \mathcal{B}_S(T_h\mathcal{V}^h,\mathcal{V}^h) \tag{48}$$

for all $\mathcal{V}^h \in \mathbf{X}_h$ and for some positive constant δ independent of h and λ. The upper bound can be assured simply by dividing the result of the usual multigrid cycle by 2, and the lower bound follows from the product H^1 equivalence of $\mathcal{B}_S(\cdot,\cdot)$ established in [6]. Assume that

$$\{(\lambda,\mathcal{U}(\lambda)) \,|\, \lambda \in \Lambda\}$$

is a branch of regular solutions of (26), and let $F^h(\lambda,\mathcal{U}^h) = 0$ denote canonical form (27). Then it is easy to see that there exists a neighborhood $\mathcal{O}$ of the origin in $\mathbf{X}$ and positive constants γ and ρ, independent of h and λ, such that

$$\gamma\mathcal{B}_S(\mathcal{V}^h,\mathcal{V}^h) \le \mathcal{B}_S(D_{\mathcal{U}}F^h(\lambda,\mathcal{U})\mathcal{V}^h,\mathcal{V}^h) \le \rho\mathcal{B}_S(\mathcal{V}^h,\mathcal{V}^h) \tag{49}$$

for all $\mathcal{V}^h \in \mathbf{X}_h$, where $(\lambda,\mathcal{U})$ is any element of $\Lambda \times \mathbf{X}_h$ for which $\mathcal{U}(\lambda) - \mathcal{U} \in \mathcal{O}$. The lower bound follows from our regular branch assumption, and the upper bound follows from Lemma 2 and property 1 of Lemma 5.

The iterative method that we consider for solving (27) is given by the expression

$$\mathcal{U}^h \leftarrow \mathcal{U}^h - sM_h \nabla J(\mathcal{U}^h), \tag{50}$$

where $J(\mathcal{U}^h)$ is the functional in (19) and $s = 1/\rho$. Suppose for the moment that $M_h = T_h$. Then the proof of local linear convergence of (50) in the $\mathcal{B}_S(\cdot,\cdot)$ norm with linear factor bounded by $\sqrt{1-\gamma/\rho}$ would follow from linearizing $\nabla J(\mathcal{U}^h)$ about the solution of (27), from the relation $T_h \nabla J(\mathcal{U}^h) = F^h(\lambda,\mathcal{U}^h)$, and from the symmetry of $D_{\mathcal{U}} F^h(\lambda,\mathcal{U})$ in the $\mathcal{B}_S(\cdot,\cdot)$ inner product. For (50) with general M_h, we can then use (48) to prove local linear convergence in the $\mathcal{B}_S(\cdot,\cdot)$ norm with factor bounded by $\sqrt{1-\delta\gamma/\rho}$.

This establishes optimality of our simple iterative method based on a multigrid Stokes preconditioner. It is straightforward to extend this result to a full-multigrid-like scheme, where an approximation to the solution of the Navier–Stokes equations is achieved with accuracy up to discretization error at the cost of a few fine grid operator evaluations.

6. H^{-1} least-squares. Here we take as a starting point for the development of a least-squares method for the Navier–Stokes equations another functional that uses norms of negative-order Sobolev spaces. As with the previous method, this functional has its origin in the Stokes problem. Thus, among other things, we demonstrate that a least-squares method for the linear Stokes equations involving negative-order norms can be successfully extended to a method for the Navier–Stokes equations along the same lines as in sections 3 and 4. In contrast to L^2 least-squares, the negative norm approach does not immediately admit a practical implementation. The culprit here is the negative norm whose evaluation requires the exact solution of a Poisson problem (54)–(55). Nevertheless, practical methods can be developed by replacing these norms with discrete negative norms, which involves approximate Poisson solvers. In Part II [3] of this paper, we develop practical negative norm methods of this type. There we establish error estimates based on the fundamental results obtained here in Part I. Thus, the main focus of this section is on applying the abstract nonlinear theory to the new H^{-1} functional.

Let $H^{-1}(\Omega)$ denote the dual of $H_0^1(\Omega)$. Using the equivalence of the seminorm and norm on $H_0^1(\Omega)$, we equip $H^{-1}(\Omega)$ with the norm

$$|f|_{-1} = \sup_{\phi \in H_0^1(\Omega)} \frac{(f,\phi)}{|\phi|_1} \quad \forall f \in H^{-1}(\Omega)\,, \tag{51}$$

for which the following representation result holds (cf. [4]).

LEMMA 6. *For all $f \in H^{-1}(\Omega)$, we have*

$$|f|_{-1} = (Sf, f) \tag{52}$$

and

$$\|Sf\|_1 \le C|f|_{-1}\,, \tag{53}$$

where $S : H^{-1}(\Omega) \mapsto H_0^1(\Omega)$ is the solution operator of the Dirichlet problem

$$-\triangle u = f \quad \text{in } \Omega, \tag{54}$$

$$u = 0 \quad \text{on } \Gamma; \tag{55}$$

that is, $u = Sf$ is the solution of (54)–(55).

The inner product associated with norm (51) *is given by*

$$(f,g)_{-1} = (Sf,g) = (f,Sg) \quad \forall f,g \in H^{-1}(\Omega)\,. \tag{56}$$

Regularity properties of inverse Laplace operator S are summarized in the next theorem, the proof of which can be found in [10].

THEOREM 2. *Let $\Omega \in \mathbb{R}^n$ be a bounded open set with C^{k+1} boundary Γ. Assume that $f \in W^{k,p}(\Omega)$, $1 < p < \infty$. Then the solution u of* (54)–(55) *satisfies $u \in W^{k+2,p}(\Omega)$ and*

$$\|u\|_{k+2,p} \le C\|f\|_{k,p}\,.$$

When $\Omega \in \mathbb{R}^2$ is a bounded polygon with no reentrant corners, there exists a real $p_\Omega > 2$ depending on the greatest inner angle of Γ such that $u \in W^{k,p}(\Omega)$, $1 < p < p_\Omega$, whenever $f \in L^p(\Omega)$. If Ω is a bounded convex polyhedron in $\mathbb{R}^3$, then this result is valid for the homogeneous Dirichlet problem.

Using a negative-order Sobolev norm in the least-squares approach enables us to restrict the functional to the simplest first-order system for the Navier-Stokes equations, namely, (7)–(9). (It can be advantageous to include terms involving inverse-order norms applied to trace (10) and curl (11) (cf. Remark 3.2 in [6]); however, we restrict ourselves here to the simpler system to avoid further complications in the discussion.) We are also able to make the more general assumption that body force $\mathbf{f} \in H^{-1}(\Omega)^n$. Accordingly, we define

$$\begin{aligned} J(\underline{\mathbf{U}},\mathbf{u},p) = &\left\| -(\nabla^t \underline{\mathbf{U}})^t + \nabla p + \frac{1}{\nu}\left(\underline{\mathbf{U}}^t\mathbf{u} - \mathbf{f}\right)\right\|_{-1}^2 \\ &+ \|\nabla^t \mathbf{u}\|_0^2 + \|\underline{\mathbf{U}} - \nabla \mathbf{u}^t\|_0^2\,. \end{aligned} \tag{57}$$

A necessary condition for minimization of J over the space

$$\mathbf{X} = L^2(\Omega)^{n^2} \times H_0^1(\Omega)^n \times L_0^2(\Omega) \tag{58}$$

is

find $(\underline{\mathbf{U}},\mathbf{u},p) \in \mathbf{X}$ such that

$$\begin{aligned} &\mathcal{B}((\underline{\mathbf{U}},\mathbf{u},p),(\underline{\mathbf{V}},\mathbf{v},p)) \\ &\equiv \left(-(\nabla^t \underline{\mathbf{U}})^t + \nabla p + \frac{1}{\nu}\left(\underline{\mathbf{U}}^t\mathbf{u} - \mathbf{f}\right), -(\nabla^t \underline{\mathbf{V}})^t + \frac{1}{\nu}\left(\underline{\mathbf{U}}^t\mathbf{v} + \underline{\mathbf{V}}^t\mathbf{u}\right) + \nabla q\right)_{-1} \\ &+ \left(\nabla^t\mathbf{u}, \nabla^t\mathbf{v}\right)_0 + \left(\underline{\mathbf{U}} - \nabla\mathbf{u}^t, \underline{\mathbf{V}} - \nabla\mathbf{v}^t\right)_0 = 0 \end{aligned} \tag{59}$$

for all $(\underline{\mathbf{V}},\mathbf{v},q) \in \mathbf{X}$.

Assume that $\mathbf{X}_h$ is a finite-dimensional subspace of $\mathbf{X}$. Then the least-squares method based on (57) is defined by restricting variational problem (59) to $\mathbf{X}_h$. However, due to the choice of $\mathbf{X}$, space $\mathbf{X}_h$ must possess an approximation property that is different than what we needed for the L^2 least-squares method: there exists an integer $d \ge 1$ such that, for all $\underline{\mathbf{U}} \in H^d(\Omega)^{n^2}$, $\mathbf{u} \in H^{d+1}(\Omega)^n$, and $p \in H^d(\Omega)$, one can find $(\underline{\mathbf{U}}^h,\mathbf{u}^h,p^h) \in \mathbf{X}_h$ such that

$$\|\underline{\mathbf{U}} - \underline{\mathbf{U}}^h\|_0 + \|\mathbf{u} - \mathbf{u}^h\|_1 + \|p - p^h\|_0 \le Ch^d\left(\|\underline{\mathbf{U}}\|_d + \|\mathbf{u}\|_{d+1} + \|p\|_d\right)\,. \tag{60}$$

Note, for example, that (60) can be satisfied with $d = 1$ by choosing either continuous piecewise linears for all unknowns or piecewise constants for $\underline{\mathbf{U}}^h$ and p^h and linears for $\mathbf{u}^h$.

In the present context, we make the following identification:

$$\mathbf{Y} = L^2(\Omega)^{n^2} \times H^{-1}(\Omega)^n \times L^2(\Omega)\,. \tag{61}$$

We then define operators T, T_h, and G analogous to those in (31), (32), and (33):

$T : \mathbf{Y} \mapsto \mathbf{X}$ defined by $\mathcal{U} = T\mathbf{g}$ for $\mathbf{g} \in \mathbf{Y}$ if and only if

$$\begin{aligned} \mathcal{B}_S(\mathcal{U},\mathcal{V}) &\equiv \left(-(\nabla^t \underline{\mathbf{U}})^t + \nabla p, -(\nabla^t \underline{\mathbf{V}})^t + \nabla q\right)_{-1} \\ &\quad + \left(\nabla^t \mathbf{u}, \nabla^t \mathbf{v}\right)_0 + \left(\underline{\mathbf{U}} - \nabla \mathbf{u}^t, \underline{\mathbf{V}} - \nabla \mathbf{v}^t\right)_0 \\ &= (\mathbf{g}_1, \underline{\mathbf{V}}) + (\mathbf{g}_2, \mathbf{v}) + (\mathbf{g}_3, q) \end{aligned} \tag{62}$$

for all $(\underline{\mathbf{V}}, \mathbf{v}, q) \in \mathbf{X}$;

$T_h : \mathbf{Y} \mapsto \mathbf{X}_h$ defined by $\mathcal{U}^h = T_h\mathbf{g}$ for $\mathbf{g} \in \mathbf{Y}$ if and only if

$$\mathcal{B}_S(\mathcal{U}^h,\mathcal{V}^h) = (\mathbf{g}_1, \underline{\mathbf{V}}^h) + (\mathbf{g}_2, \mathbf{v}^h) + (\mathbf{g}_3, q^h) \quad \text{for all } (\underline{\mathbf{V}}^h, \mathbf{v}^h, q^h) \in \mathbf{X}_h\,; \tag{63}$$

and

$G : \Lambda \times \mathbf{X} \to \mathbf{Y}$ with $\mathbf{g} = G(\lambda,\mathcal{U})$ for $\mathcal{U} \in \mathbf{X}$ if and only if

$$\begin{aligned} &(\mathbf{g}_1, \underline{\mathbf{V}}) + (\mathbf{g}_2, \mathbf{v}) + (\mathbf{g}_3, q) \\ &\quad = \left(-(\nabla^t \underline{\mathbf{U}})^t + \nabla p, \frac{1}{\nu}\left(\underline{\mathbf{U}}^t \mathbf{v} + \underline{\mathbf{V}}^t \mathbf{u}\right)\right)_{-1} \\ &\quad + \left(\frac{1}{\nu}\left(\underline{\mathbf{U}}^t \mathbf{u} - \mathbf{f}\right), -(\nabla^t \underline{\mathbf{V}})^t + \nabla q + \frac{1}{\nu}\left(\underline{\mathbf{U}}^t \mathbf{v} + \underline{\mathbf{V}}^t \mathbf{u}\right)\right)_{-1} \end{aligned} \tag{64}$$

for all $(\underline{\mathbf{V}}, \mathbf{v}, q) \in \mathbf{X}$.

Along the same lines as in Lemma 1, we can show that problem (59) and its discrete counterpart admit canonical representations of the respective forms (26) and (27). Thus, our goal now is to establish a result similar to Theorem 1, which in the present setting specializes as follows.

THEOREM 3. *Let $F(\lambda,\mathcal{U})$ denote abstract form* (26) *and assume that T, T_h, and G are defined by* (62), (63), *and* (64), *respectively. Suppose also that $\{(\lambda,\mathcal{U}(\lambda)) \,|\, \lambda \in \Lambda\}$ is a branch of regular solutions of* (26). *Furthermore, assume that $T \in L(\mathbf{Y},\mathbf{X})$, that G is a C^2 map $\Lambda \times \mathbf{X} \mapsto \mathbf{Y}$ such that all second derivatives of G are bounded on bounded subsets of $\Lambda \times \mathbf{X}$, and that there exists a space $\mathbf{Z} \subset \mathbf{Y}$, with continuous imbedding, such that $D_{\mathcal{U}}G(\lambda,\mathcal{U}) \in L(\mathbf{X},\mathbf{Z})$ for all $\lambda \in \Lambda$ and $\mathcal{U} \in \mathbf{X}$. If approximate problem* (27) *is such that* (34) *and* (35) *hold, then:*

1. *there exists a neighborhood $\mathcal{O}$ of the origin in $\mathbf{X}$ and, for h sufficiently small, a unique C^2 function $\lambda \mapsto \mathcal{U}^h(\lambda) \in \mathbf{X}_h$ such that $\{(\lambda,\mathcal{U}^h(\lambda)) \,|\, \lambda \in \Lambda\}$ is a branch of regular solutions of discrete problem* (27) *and $\mathcal{U}(\lambda) - \mathcal{U}^h(\lambda) \in \mathcal{O}$ for all $\lambda \in \Lambda$;*

2. *for all $\lambda \in \Lambda$ we have*

$$\|\mathcal{U}^h(\lambda) - \mathcal{U}(\lambda)\|_{\mathbf{X}} \le C\|(T - T^h)G(\lambda,\mathcal{U}(\lambda))\|_{\mathbf{X}}\,; \tag{65}$$

3. *if the regular branch is such that $\mathcal{U}(\lambda) \in H^m(\Omega)^{n^2} \times H^{m+1}(\Omega)^n \times H^m(\Omega)$ for some integer $m \ge 1$ and $\tilde{d} \equiv \min\{d, m\}$, where d is the largest integer satisfying* (60),

we have

$$\begin{aligned} &\|\underline{\mathbf{U}}(\lambda) - \underline{\mathbf{U}}^h(\lambda)\|_0 + \|\mathbf{u}(\lambda) - \mathbf{u}^h(\lambda)\|_1 + \|p(\lambda) - p^h(\lambda)\|_0 \\ &\le Ch^{\tilde{d}} \left(\|\underline{\mathbf{U}}(\lambda)\|_{\tilde{d}} + \|\mathbf{u}(\lambda)\|_{\tilde{d}+1} + \|p(\lambda)\|_{\tilde{d}} \right) . \end{aligned} \tag{66}$$

To establish the conditions of abstract Theorem 3 for our specific application, we need the additional technical results summarized in the next lemma.

LEMMA 7.

1. *For $2 \le q \le 6$, we have $H^1(\Omega) \subset L^q(\Omega)$ with compact imbedding.*
2. *For $p > 6/5$, we have $L^p(\Omega) \subset H^{-1}(\Omega)$ with compact imbedding.*
3. *For any $u \in L^2(\Omega)$ and $v \in H^1(\Omega)$, we have*

$$|\nabla u|_{-1} \le \|u\|_0, \tag{67}$$

$$|uv|_{-1} \le C\|u\|_0\|v\|_1 . \tag{68}$$

4. *$\{u, v\} \mapsto uv$ is a continuous bilinear mapping from $H^1(\Omega) \times H^1(\Omega)$ into $W^{1,2-\varepsilon}(\Omega)$ for any $0 < \varepsilon \le 1$; i.e.,*

$$\|uv\|_{1,2-\varepsilon} \le C\|u\|_1\|v\|_1 \qquad \text{for } u, v \in H^1(\Omega) . \tag{69}$$

Proof. The first two statements follow directly from the Sobolev imbedding theorem. To prove 3, we use definition (51), bound (46), and the imbedding of H^{-1} in L^3:

$$\begin{aligned} |\nabla u|_{-1} &= \sup_{\boldsymbol{\phi} \in H_0^1(\Omega)^n} \frac{(\nabla u, \boldsymbol{\phi})}{|\boldsymbol{\phi}|_1} = \sup_{\boldsymbol{\phi} \in H_0^1(\Omega)^n} \frac{(u, \nabla^t \boldsymbol{\phi})}{|\boldsymbol{\phi}|_1} \\ &\le \sup_{\boldsymbol{\phi} \in H_0^1(\Omega)^n} \frac{\|u\|_0 \|\nabla^t \boldsymbol{\phi}\|_0}{|\boldsymbol{\phi}|_1} \le \|u\|_0 \end{aligned}$$

and

$$\begin{aligned} |uv|_{-1} &= \sup_{\phi \in H_0^1(\Omega)} \frac{(uv, \phi)}{|\phi|_1} \le \sup_{\phi \in H_0^1(\Omega)} \frac{\|uv\|_{0,3/2} \|\phi\|_{0,3}}{|\phi|_1} \\ &\le C \sup_{\phi \in H_0^1(\Omega)} \frac{\|u\|_0 \|v\|_1 |\phi|_1}{|\phi|_1} \le \|u\|_0 \|v\|_1 . \end{aligned}$$

Finally, 4 follows from a general result on multiplication in Sobolev spaces (see [9, Corollary 1.1]). ∎

The aim now is to verify the assumptions of Theorem 3 for our negative-order least-squares approach applied to the Navier–Stokes equations. In the present setting, we make the following identification:

$$\mathbf{Z} = W^{1,3/2}(\Omega)^{n^2} \times L^{3/2}(\Omega)^n \times W^{1,3/2}(\Omega) . \tag{70}$$

By virtue of compactness of the imbeddings $W^{1,3/2}(\Omega) \subset L^2(\Omega)$ and $L^{3/2}(\Omega) \subset H^{-1}(\Omega)$, space $\mathbf{Z}$ is compactly imbedded in space $\mathbf{Y}$.

The assumptions of Theorem 3 that concern (62) are established in [6]. Since $\mathbf{Z}$ is compactly imbedded in $\mathbf{Y}$, then (35) follows again from (34). Hence, it remains to

verify that $D_{\mathcal{U}}G(\lambda,\mathcal{U}) \in L(\mathbf{X},\mathbf{Z})$ when $\mathbf{Z}$ is chosen as in (70), and that $D^2_{\mathcal{U}}G(\lambda,\mathcal{U})$ is bounded on all bounded subsets of $\Lambda \times \mathbf{X}$.

LEMMA 8. *Assume that G, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ are defined by* (64), (58), (61), *and* (70), *respectively. Then $D_{\mathcal{U}}G(\lambda,\mathcal{U}) \in L(\mathbf{X},\mathbf{Z})$ for all $\mathcal{U} \in \mathbf{X}$. Moreover, the second Fréchet derivative $D^2_{\mathcal{U}}G(\lambda,\mathcal{U})$ is bounded on bounded subsets of $\Lambda \times \mathbf{X}$.*

Proof. We omit derivation of the Fréchet derivative $D_{\mathcal{U}}G(\lambda,\mathcal{U})$ for (64) because it is analogous to that for (33). Let $D_{\mathcal{U}}G(\lambda,\mathcal{U})\hat{\mathcal{U}} = \mathbf{g} = (g_1, g_2, g_3)$. Then, using (56), the equation corresponding to test function $\underline{\mathbf{V}}$ can be written as

$$\begin{aligned}
(g_1,\underline{\mathbf{V}}) &= \left(\frac{1}{\nu}S(\underline{\mathbf{U}}^t\hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t\mathbf{u}), -(\nabla^t\underline{\mathbf{V}})^t + \frac{1}{\nu}\underline{\mathbf{V}}^t\mathbf{u}\right)_0 \\
&\quad + \left(S\left(-(\nabla^t\underline{\mathbf{U}})^t + \nabla p + \frac{1}{\nu}\left(\underline{\mathbf{U}}^t\mathbf{u} - \mathbf{f}\right)\right), \frac{1}{\nu}\underline{\mathbf{V}}^t\hat{\mathbf{u}}\right)_0 \\
&\quad + \left(S(-(\nabla^t\hat{\underline{\mathbf{U}}})^t + \nabla\hat{p}), \frac{1}{\nu}\underline{\mathbf{V}}^t\mathbf{u}\right)_0 \\
&= \left(\frac{1}{\nu}S(\underline{\mathbf{U}}^t\hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t\mathbf{u}), -(\nabla^t\underline{\mathbf{V}})^t\right)_0 + \left(\frac{1}{\nu}S(\underline{\mathbf{U}}^t\hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t\mathbf{u})\cdot\mathbf{u}^t, \frac{1}{\nu}\underline{\mathbf{V}}\right)_0 \\
&\quad + \left(S\left(-(\nabla^t\underline{\mathbf{U}})^t + \nabla p + \frac{1}{\nu}\left(\underline{\mathbf{U}}^t\mathbf{u} - \mathbf{f}\right)\right)\cdot\hat{\mathbf{u}}^t, \frac{1}{\nu}\underline{\mathbf{V}}\right)_0 \\
&\quad + \left(S(-(\nabla^t\hat{\underline{\mathbf{U}}})^t + \nabla\hat{p})\cdot\mathbf{u}^t, \frac{1}{\nu}\underline{\mathbf{V}}\right)_0 \\
&= \left(\frac{1}{V}\nabla(S(\underline{\mathbf{U}}^t\hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t\mathbf{u})), \underline{\mathbf{V}}\right)_0 \\
&\quad + \left(\frac{1}{\nu}S\left(-(\nabla^t\hat{\underline{\mathbf{U}}})^t + \nabla\hat{p} + \frac{1}{\nu}\left(\underline{\mathbf{U}}^t\hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t\mathbf{u}\right)\right)\cdot\mathbf{u}^t, \underline{\mathbf{V}}\right)_0 \\
&\quad + \left(\frac{1}{\nu}S\left(-(\nabla^t\underline{\mathbf{U}})^t + \nabla p + \frac{1}{\nu}\left(\underline{\mathbf{U}}^t\mathbf{u} - \mathbf{f}\right)\right)\cdot\hat{\mathbf{u}}^t, \underline{\mathbf{V}}\right)_0.
\end{aligned}$$

Here we have used the fact that Sf satisfies the homogeneous Dirichlet boundary condition. Hence, g_1 can be identified as

$$\begin{aligned}
g_1 &= \nabla\frac{1}{V}S(\underline{\mathbf{U}}^t\hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t\mathbf{u}) \\
&\quad + \frac{1}{\nu}S\left(-(\nabla^t\hat{\underline{\mathbf{U}}})^t + \nabla\hat{p} + \frac{1}{\nu}\left(\underline{\mathbf{U}}^t\hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t\mathbf{u}\right)\right)\cdot\mathbf{u}^t \\
&\quad + \frac{1}{\nu}S\left(-(\nabla^t\underline{\mathbf{U}})^t + \nabla p + \frac{1}{\nu}\left(\underline{\mathbf{U}}^t\mathbf{u} - \mathbf{f}\right)\right)\cdot\hat{\mathbf{u}}^t. \qquad (71)
\end{aligned}$$

From Lemma 4, it follows that $\underline{\mathbf{U}}^t\hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t\mathbf{u} \in L^{3/2}(\Omega)$, so regularity of operator S (see Theorem 2) implies that $\nabla S(\underline{\mathbf{U}}^t\hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t\mathbf{u}) \in W^{1,3/2}(\Omega)^{n^2}$. For the remaining terms in (71), we have that

$$\begin{aligned}
S\left(-(\nabla^t\hat{\underline{\mathbf{U}}})^t + \nabla\hat{p} + \frac{1}{\nu}\left(\underline{\mathbf{U}}^t\hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t\mathbf{u}\right)\right) &\in H_0^1(\Omega)^n, \\
S\left(-(\nabla^t\underline{\mathbf{U}})^t + \nabla p + \frac{1}{\nu}\left(\underline{\mathbf{U}}^t\mathbf{u} - \mathbf{f}\right)\right) &\in H_0^1(\Omega)^n.
\end{aligned}$$

Using 4 of Lemma 7 with $\varepsilon = 1/2$, we conclude that the last two terms in (71) are also in $W^{1,3/2}(\Omega)^n$.

Similarly, g_3 can be identified as

$$g_3 = -\frac{1}{\nu}\nabla^t(S(\underline{\mathbf{U}}^t\hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t\mathbf{u})),$$

so $g_3 \in W^{1,3/2}(\Omega)$.

Finally, for g_2 we find

$$\begin{aligned} g_2 = &\frac{1}{\nu}S\left(-(\nabla^t\hat{\underline{\mathbf{U}}})^t + \nabla\hat{p} + \frac{1}{\nu}\left(\underline{\mathbf{U}}^t\hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t\mathbf{u}\right)\right)\cdot\underline{\mathbf{U}} \\ &+\frac{1}{\nu}S\left(-(\nabla^t\underline{\mathbf{U}})^t + \nabla p + \frac{1}{\nu}\left(\underline{\mathbf{U}}^t\mathbf{u} - \mathbf{f}\right)\right)\cdot\hat{\underline{\mathbf{U}}}. \end{aligned}$$

Since $\underline{\mathbf{U}}$ and $\hat{\underline{\mathbf{U}}}$ are in $L^2(\Omega)^{n^2}$ and the terms with S are in $H^1(\Omega)$, we can conclude that $g_2 \in L^{3/2}(\Omega)^n$. Thus, $\mathbf{g} \in \mathbf{Z}$ and the first assertion is proved.

Let $D^2_{\mathcal{U}}G(\lambda,\mathcal{U}) : \Lambda \times \mathbf{X} \mapsto L(\mathbf{X}, L(\mathbf{X},\mathbf{Y}))$ denote the second Fréchet derivative of G, that is, $D^2_{\mathcal{U}}G(\lambda,\mathcal{U})[\hat{\mathcal{U}},\hat{\hat{\mathcal{U}}}] = \mathbf{g} = (g_1, g_2, g_3) \in \mathbf{Y}$ for $\hat{\mathcal{U}},\hat{\hat{\mathcal{U}}} \in \mathbf{X}$. We want to establish that $D^2_{\mathcal{U}}G(\lambda,\mathcal{U})$ is bounded in the norm of $L(\mathbf{X}, L(\mathbf{X},\mathbf{Y}))$ on all bounded subsets of $\Lambda \times \mathbf{X}$. For this purpose, it suffices to show that $\|g_1\|_0$, $\|g_2\|_{-1}$, and $\|g_3\|_0$ are bounded from above by $\|\hat{\mathcal{U}}\|_{\mathbf{X}}$, $\|\hat{\hat{\mathcal{U}}}\|_{\mathbf{X}}$, and λ, respectively. For brevity, we only establish the bound for $\|g_3\|_0$; the other bounds follow in a similar fashion. As in (43), we find that

$$(g_3, q) = \frac{1}{\nu}\left(\hat{\underline{\mathbf{U}}}^t\hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t\hat{\hat{\mathbf{u}}}, \nabla q\right)_{-1}. \tag{72}$$

Choosing $q = g_3$ in (72) and using equivalent representation (56), the imbedding of H^1 in L^3, a priori estimate (53), and bound (67), we obtain

$$\begin{aligned} \|g_3\|_0^2 &= \frac{1}{\nu}\left(\hat{\underline{\mathbf{U}}}^t\hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t\hat{\hat{\mathbf{u}}}, S(\nabla g_3)\right)_0 \\ &\le \frac{1}{\nu}\left(\|\hat{\underline{\mathbf{U}}}^t\hat{\mathbf{u}}\|_{0,3/2} + \|\hat{\underline{\mathbf{U}}}^t\hat{\hat{\mathbf{u}}}\|_{0,3/2}\right)\|S(\nabla g_3)\|_{0,3} \\ &\le \frac{1}{\nu}\left(\|\hat{\underline{\mathbf{U}}}\|_0\|\hat{\mathbf{u}}\|_1 + \|\hat{\underline{\mathbf{U}}}\|_0\|\hat{\hat{\mathbf{u}}}\|_1\right)\|g_3\|_0. \qquad \square \end{aligned}$$

Lemma 8 completes verification of all assumptions of Theorem 3. We have thus demonstrated that a least-squares approach based on the Sobolev space H^{-1} can be successfully extended to the nonlinear Navier–Stokes equations. The next step towards development of a numerical method based on this approach involves formulation of a computationally feasible discrete analogue of the H^{-1} inner product. Such a method and its analysis are considered in Part II of this work [3].

7. Well-posedness. In this section, we address the question of the well-posedness of least-squares formulations (22) and (59). More precisely, our aim is to show that if $\{(\lambda, (\mathbf{u}(\lambda), p(\lambda))) \,|\, \lambda \in \Lambda\}$ is a branch of regular solutions of original velocity-pressure Navier-Stokes problem (1)–(4), then

$$\{(\lambda, (\underline{\mathbf{U}}(\lambda), \mathbf{u}(\lambda), p(\lambda))) \,|\, \lambda \in \Lambda\}$$

is a regular branch for variational problems (22) and (59). This is an important question not only because application of Theorems 1 and 3 requires a regular branch,

but also because it would assert that the least-squares formulation does not introduce bifurcation phenomena that are not already present in the original equations. The question is also nontrivial since nonlinear variational problems (22) and (59) involve coupling between the velocity-flux Stokes operator and the convective term $\underline{\mathbf{U}}^t\mathbf{u}$. As a result, the equivalent strong form of (22) now involves derivatives of nonlinear equations (1)–(2).

Assume that $(\mathbf{u}(\lambda), p(\lambda)) \in H_0^1(\Omega)^n \times L_0^2(\Omega)$ yields a regular branch of solutions of (1)–(4); i.e., for every $\lambda \in \Lambda$, the pair $(\mathbf{u}(\lambda), p(\lambda))$ is a nonsingular (weak) solution of the Navier–Stokes equations. We recall the result of [9] that $(\mathbf{u}, p)$ is a nonsingular solution if and only if the linearized problem

$$-\nu\triangle\hat{\mathbf{u}} + (\nabla\hat{\mathbf{u}}^t)^t\mathbf{u} + (\nabla\mathbf{u}^t)^t\hat{\mathbf{u}} + \nabla\hat{p} = \mathbf{f}^* \quad \text{in } \Omega, \tag{73}$$

$$\nabla^t\hat{\mathbf{u}} = 0 \quad \text{in } \Omega, \tag{74}$$

$$\hat{\mathbf{u}} = \mathbf{0} \quad \text{on } \Gamma, \tag{75}$$

$$\int_\Omega \hat{p}d\Omega = 0 \tag{76}$$

has a unique (weak) solution $(\hat{\mathbf{u}}, \hat{p}) \in H_0^1(\Omega)^n \times L_0^2(\Omega)$ for each $\mathbf{f}^* \in H^{-1}(\Omega)^n$.

Under this assumption, well-posedness of (22) and (59) would follow if we could establish that $\mathcal{U}(\lambda) = (\underline{\mathbf{U}}(\lambda), \mathbf{u}(\lambda), p(\lambda))$ with $\underline{\mathbf{U}}(\lambda) = \nabla\mathbf{u}(\lambda)^t$ is a nonsingular solution of (22) and (59) for all $\lambda \in \Lambda$. In terms of canonical representation (26), this amounts to showing that the linearized mapping $D_{\mathcal{U}}F(\lambda,\mathcal{U})$ is an isomorphism of $\mathbf{X}$, i.e., that the linearized equation

$$D_{\mathcal{U}}F(\lambda,\mathcal{U})\hat{\mathcal{U}} = (I + T \cdot D_{\mathcal{U}}G(\lambda,\mathcal{U}))\hat{\mathcal{U}} = \mathcal{V} \tag{77}$$

has a unique solution $\hat{\mathcal{U}} \in \mathbf{X}$ for all $\mathcal{V} \in \mathbf{X}$. To show this, we first establish that operator $(I + T \cdot D_{\mathcal{U}}G(\lambda,\mathcal{U}))$ is a compact perturbation of unity; that is, the Fredholm alternative applies to (77). Compactness of $T : \mathbf{Z} \mapsto \mathbf{X}$ follows from (35), which asserts that it is a uniform limit of compact operators T_h. Assuming for the moment that $D_{\mathcal{U}}G(\lambda,\mathcal{U}) \in L(\mathbf{X}, \mathbf{Z})$, it follows that operator $T \cdot D_{\mathcal{U}}G(\lambda,\mathcal{U}) : \mathbf{X} \mapsto \mathbf{X}$ is compact. Thus, the Fredholm alternative applies to (77), and we can assert that $D_{\mathcal{U}}F(\lambda,\mathcal{U})$ is indeed an isomorphism of $\mathbf{X}$ if and only if the homogeneous equation

$$D_{\mathcal{U}}F(\lambda,\mathcal{U})\hat{\mathcal{U}} = (I + T \cdot D_{\mathcal{U}}G(\lambda,\mathcal{U}))\hat{\mathcal{U}} = \mathbf{0} \tag{78}$$

has only trivial solution $\hat{\mathcal{U}} = \mathbf{0}$ in $\mathbf{X}$. Note that this argument is carried out entirely in terms of abstract problem (26). Since (35) and the property $D_{\mathcal{U}}G(\lambda,\mathcal{U}) \in L(\mathbf{X}, \mathbf{Z})$ (see Lemmas 5 and 8) are valid for both (22) and (59), then the above conclusions remain valid for these problems; that is, in both cases associated operator $T \cdot D_{\mathcal{U}}G(\lambda,\mathcal{U})$ is compact. Hence, well-posedness of (22) and (59) would follow if we could show that corresponding problem (78) has only the trivial solution. This will be established in the next two respective lemmas using our nonsingularity assumption on $(\mathbf{u}(\lambda), p(\lambda))$.

LEMMA 9. *Assume that* $(\mathbf{u}, p)$ *is such that linearized equations* (73)–(76) *have a unique solution for each* $\mathbf{f}^* \in H^{-1}(\Omega)^n$. *Then homogeneous problem* (78) *corresponding to* (22) *has only the trivial solution.*

Proof. Specialized to our needs, the nonsingularity assumption asserts that the problem

$$-\nu\triangle\hat{\mathbf{u}} + (\nabla\hat{\mathbf{u}}^t)^t(\mathbf{u}+\mathbf{u}_0) + (\nabla(\mathbf{u}^t+\mathbf{u}_0^t))^t\hat{\mathbf{u}} + \nabla\hat{p} = \mathbf{f}^* \quad \text{in } \Omega, \tag{79}$$

$$\nabla^t\hat{\mathbf{u}} = 0 \quad \text{in } \Omega, \tag{80}$$

$$\hat{\mathbf{u}} = \mathbf{0} \quad \text{on } \Gamma, \tag{81}$$

$$\int_\Omega \hat{p}d\Omega = 0 \tag{82}$$

has a unique (weak) solution $(\hat{\mathbf{u}},\hat{p}) \in H_0^1(\Omega)^n \times L_0^2(\Omega)$ for each $\mathbf{f}^* \in H^{-1}(\Omega)^n$, where $(\mathbf{u}_0, p_0)$ solves Stokes problem (12) with the original data $\mathbf{f}$. Furthermore, using definitions (31) and (33), one can easily verify that (78) is equivalent to the variational problem

find $(\underline{\hat{\mathbf{U}}},\hat{\mathbf{u}},\hat{p}) \in \mathbf{X}$ such that

$$\begin{aligned}
&\mathcal{B}((\underline{\hat{\mathbf{U}}},\hat{\mathbf{u}},\hat{p}),(\underline{\mathbf{V}},\mathbf{v},p)) \\
&\quad= \Big(-(\nabla^t\underline{\hat{\mathbf{U}}})^t + \frac{1}{\nu}((\underline{\mathbf{U}}+\underline{\mathbf{U}}_0)^t\hat{\mathbf{u}} + \underline{\hat{\mathbf{U}}}^t(\mathbf{u}+\mathbf{u}_0)) + \nabla\hat{p}, \\
&\qquad -(\nabla^t\underline{\mathbf{V}})^t + \frac{1}{\nu}\left((\underline{\mathbf{U}}+\underline{\mathbf{U}}_0)^t\mathbf{v} + \underline{\mathbf{V}}^t(\mathbf{u}+\mathbf{u}_0)\right) + \nabla q\Big)_0 \\
&\quad+ \left(\nabla^t\hat{\mathbf{u}},\nabla^t\mathbf{v}\right)_0 + \Big(\nabla(\mathrm{tr}\ \underline{\hat{\mathbf{U}}}),\nabla(\mathrm{tr}\ \underline{\mathbf{V}})\Big)_0 \\
&\quad+ \Big(\underline{\hat{\mathbf{U}}} - \nabla\hat{\mathbf{u}}^t, \underline{\mathbf{V}} - \nabla\mathbf{v}^t\Big)_0 + \Big(\nabla\times\underline{\hat{\mathbf{U}}}, \nabla\times\underline{\mathbf{V}}\Big)_0 = 0
\end{aligned} \tag{83}$$

for all $(\underline{\mathbf{V}},\mathbf{v},q) \in \mathbf{X}$,

where the space $\mathbf{X}$ is given by (20).

Variational problem (83) is evidently the Euler–Lagrange equation for the minimization problem

find $(\underline{\hat{\mathbf{U}}},\hat{\mathbf{u}},\hat{p}) \in \mathbf{X}$ such that

$$J_l(\underline{\hat{\mathbf{U}}},\hat{\mathbf{u}},\hat{p}) \le J_l(\underline{\mathbf{V}},\mathbf{v},q) \quad \text{for all } (\underline{\mathbf{V}},\mathbf{v},q) \in \mathbf{X}, \tag{84}$$

where

$$\begin{aligned}
J_l(\underline{\hat{\mathbf{U}}},\hat{\mathbf{u}},\hat{p}) = &\left\|-(\nabla^t\underline{\hat{\mathbf{U}}})^t + \frac{1}{\nu}((\underline{\mathbf{U}}+\underline{\mathbf{U}}_0)^t\hat{\mathbf{u}} + \underline{\hat{\mathbf{U}}}^t(\mathbf{u}+\mathbf{u}_0)) + \nabla\hat{p}\right\|_0^2 \\
&+ \|\nabla^t\hat{\mathbf{u}}\|_0^2 + \|\underline{\hat{\mathbf{U}}} - \nabla\hat{\mathbf{u}}^t\|_0^2 \\
&+ \|\nabla(\mathrm{tr}\ \underline{\hat{\mathbf{U}}})\|_0^2 + \|\nabla\times\underline{\hat{\mathbf{U}}}\|_0^2.
\end{aligned} \tag{85}$$

Thus, nonsingularity of $(\underline{\mathbf{U}},\mathbf{u},p)$ would follow if we could show that (84) has no nontrivial minimizers. Assume the contrary. Then the nontrivial minimizer $(\underline{\hat{\mathbf{U}}},\hat{\mathbf{u}},\hat{p})$ satisfies

$$-(\nabla^t\underline{\hat{\mathbf{U}}})^t + \frac{1}{\nu}((\underline{\mathbf{U}}+\underline{\mathbf{U}}_0)^t\hat{\mathbf{u}} + \underline{\hat{\mathbf{U}}}^t(\mathbf{u}+\mathbf{u}_0)) + \nabla\hat{p} = \mathbf{0}, \tag{86}$$

$$\underline{\hat{\mathbf{U}}} - \nabla\hat{\mathbf{u}}^t = \underline{\mathbf{0}}, \tag{87}$$

$$\nabla^t\hat{\mathbf{u}} = 0\,. \tag{88}$$

Then from equations (86), (87), and identities $\underline{\mathbf{U}} = \nabla \mathbf{u}^t$, $\underline{\mathbf{U}}_0 = \nabla \mathbf{u}_0^t$, we conclude that the pair $(\hat{\mathbf{u}}, \hat{p})$ satisfies

$$-\nu \triangle \hat{\mathbf{u}} + (\nabla(\mathbf{u}^t + \mathbf{u}_0^t))^t \hat{\mathbf{u}} + (\nabla \hat{\mathbf{u}}^t)^t (\mathbf{u} + \mathbf{u}_0) + \nabla \hat{p} = \mathbf{0}\,.$$

Now the premise that $(\hat{\underline{\mathbf{U}}}, \hat{\mathbf{u}}, \hat{p})$ is nontrivial, together with (87), implies that $(\hat{\mathbf{u}}, \hat{p})$ is nontrivial. Since (88) is also satisfied, then $(\hat{\mathbf{u}}, \hat{p})$ is also a nontrivial solution of (79)–(82), which is a contradiction. □

LEMMA 10. *Assume the conditions of Lemma* 9. *Then homogeneous problem* (78) *corresponding to* (59) *has only the trivial solution.*

Proof. As in the proof of Lemma 9, we find that, in the present context, abstract homogeneous problem (78) corresponds to the Euler–Lagrange equation of the minimization problem

find $(\hat{\underline{\mathbf{U}}}, \hat{\mathbf{u}}, \hat{p}) \in \mathbf{X}$ such that

$$J_l(\hat{\underline{\mathbf{U}}}, \hat{\mathbf{u}}, \hat{p}) \le J_l(\underline{\mathbf{V}}, \mathbf{v}, q) \quad \text{for all } (\underline{\mathbf{V}}, \mathbf{v}, q) \in \mathbf{X}, \tag{89}$$

where the space $\mathbf{X}$ is now given by (58) and

$$\begin{aligned} J_l(\hat{\underline{\mathbf{U}}}, \hat{\mathbf{u}}, \hat{p}) = &\left\| -(\nabla^t \hat{\underline{\mathbf{U}}})^t + \frac{1}{\nu}(\underline{\mathbf{U}}^t \hat{\mathbf{u}} + \hat{\underline{\mathbf{U}}}^t \mathbf{u}) + \nabla \hat{p} \right\|_{-1}^2 \\ &+ \|\nabla^t \hat{\mathbf{u}}\|_0^2 + \|\hat{\underline{\mathbf{U}}} - \nabla \hat{\mathbf{u}}^t\|_0^2\,. \end{aligned} \tag{90}$$

Assuming again that $(\hat{\underline{\mathbf{U}}}, \hat{\mathbf{u}}, \hat{p})$ is a nontrivial minimizer of (90), we immediately obtain that the pair $(\hat{\mathbf{u}}, \hat{p})$ is a nontrivial solution of (73)–(76) with $\mathbf{f}^* \equiv \mathbf{0}$, which is a contradiction. □

REFERENCES

[1] P. BOCHEV, *Analysis of least-squares finite element methods for the Navier–Stokes equations*, SIAM J. Numer. Anal., 34 (1997), pp. 1817–1844.

[2] P. BOCHEV AND M. D. GUNZBURGER, *Analysis of least squares finite element methods for the Stokes equations*, Math. Comp., 63 (1994) pp. 479–506.

[3] P. BOCHEV, T. A. MANTEUFFEL, AND S. F. MCCORMICK, *Analysis of Velocity-Flux First-Order System Least-Squares Principles for the Navier–Stokes Equations: Part* II, SIAM J. Numer. Anal., submitted.

[4] J. BRAMBLE, R. LAZAROV, AND J. PASCIAK, *A Least Squares Approach Based on a Discrete Minus One Inner Product for First Order Systems*, Tech. report 94-32, Mathematical Science Institute, Cornell University, Ithaca, NY, 1994.

[5] J. H. BRAMBLE AND J. E. PASCIAK, *Least-squares method for Stokes equations based on a discrete minus one inner product*, submitted.

[6] 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.

[7] 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.

[8] C. L. CHANG, *A mixed finite element method for the Stokes problem: An acceleration pressure formulation*, Appl. Math. Comp., 36 (1990), pp. 135–146.

[9] V. GIRAULT AND P.-A. RAVIART, *Finite Element Methods for Navier-Stokes Equations*, Springer-Verlag, Berlin, 1986.

[10] P. GRISVARD, *Boundary Value Problems in Non-Smooth Domains*, Dept. of Math. Lecture Notes 19, University of Maryland, College Park, 1980.

[11] R. TÉMAM, *Nonlinear Functional Analysis and Navier-Stokes Equations*, SIAM, Philadelphia, 1983.