SIAM Rev.
Vol. 40, No. 4, pp. 789–837, December 1998

002

FINITE ELEMENT METHODS OF LEAST-SQUARES TYPE*

PAVEL B. BOCHEV† AND MAX D. GUNZBURGER‡

Abstract. We consider the application of least-squares variational principles to the numerical solution of partial differential equations. Our main focus is on the development of least-squares finite element methods for elliptic boundary value problems arising in fields such as fluid flows, linear elasticity, and convection-diffusion. For many of these problems, least-squares principles offer numerous theoretical and computational advantages in the algorithmic design and implementation of corresponding finite element methods that are not present in standard Galerkin discretizations. Most notably, the use of least-squares principles leads to symmetric and positive definite algebraic problems and allows us to circumvent stability conditions such as the inf-sup condition arising in mixed methods for the Stokes and Navier–Stokes equations. As a result, application of least-squares principles has led to the development of robust and efficient finite element methods for a large class of problems of practical importance.

Key words. least-squares finite element methods, elliptic equations

AMS subject classification. 65N30

PII. S0036144597321156

1. Introduction.

The success of finite element methods for the numerical solution of boundary value problems for elliptic partial differential equations is, to a large extent, due to the variational principles upon which these methods are built. These principles allow us to draw upon rich mathematical foundations that influence both the analysis and the algorithmic development of finite element methods. A key ingredient in the application of variational principles is the casting of elliptic problems into a set of variational (or "weak") equations. For linear elliptic problems, weak forms usually involve bilinear forms that are continuous in some Hilbert space. In some instances, e.g., the Dirichlet problem for the Poisson equation, weak problems can be associated with the minimization of quadratic functionals. In such a case, variational principles lead to *symmetric* and *coercive* bilinear forms, i.e., forms which are equivalent to an inner product for the underlying function space. One immediate and important consequence is that the existence and uniqueness of weak solutions for such problems can be established through the application of the Riesz representation theorem in the form of the Lax–Milgram lemma. Of similar importance is that any *conforming* discretization of such weak problems, i.e., a discretization for which the finite-dimensional approximating space is a subspace of the underlying space, automatically leads to symmetric and positive definite algebraic problems. Furthermore, the equivalence of bilinear forms to inner products also implies that the discrete solutions are projections of exact solutions onto the approximating space with respect to the norms generated by these bilinear forms, i.e., approximations are optimally accurate.

When elliptic boundary value problems involve systems of partial differential equations in several variables, variational problems derived in a standard manner often correspond to saddle-point optimization problems. A typical example is given

*Received by the editors May 7, 1997; accepted for publication January 20, 1998.
http://www.siam.org/journals/sirev/40-4/32115.html

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

‡Department of Mathematics, Iowa State University, Ames IA 50011-2064 (gunzburg@iastate.edu).

by the primitive variable formulation of the Stokes problem for which a pair of approximating spaces is used for the velocity and the pressure fields. The fact that we have to deal with a saddle-point optimization problem leads to several difficulties of both a theoretical and practical nature. First, it is now well known that the spaces used for the approximation of the different unknowns, e.g., velocity and pressure, displacement and stress, cannot be chosen independently, and must satisfy strict stability conditions such as the inf-sup or Ladyzhenskaya–Babuska–Brezzi (LBB) condition; see, e.g., [27], [79], or [82]. For example, a mixed method for the Stokes problem cannot use equal-interpolation-order finite element spaces defined with respect to the same triangulation, since such spaces form unstable pairs. The saddle-point nature of the mixed method is also manifested through the indefiniteness of the associated discrete algebraic problems. Although significant progress has been made in the development of methods for such algebraic systems, their numerical solution is still challenging and computationally demanding.

As a result, in the past decade, the formulation of finite element methods that circumvent stability conditions such as the LBB condition has been the subject of intensive research efforts. Existing approaches can be broadly classified into two main categories: stabilization techniques for mixed methods and the application of least-squares principles. We stress that in this paper the term "least-squares" will be used in strict reference to bona fide least-squares methods, i.e., to methods based upon minimization of quadratic least-squares functionals, as opposed to Galerkin least-squares or stabilized mixed methods where least-squares terms are added locally or globally to mixed variational problems; see, e.g., [8], [28], [29], [84], [77], and [78].

Loosely speaking, least-squares methods can be viewed as a combination of a least-squares step at which we define a quadratic functional, and a discretization step at which we choose the form of the approximate solution. Methods for which the discretization step is invoked before the least-squares step are traditionally called *point-matching*, *collocation*, or *discrete* least-squares methods; see [56], [69], [100], [105], [119], [120], and [101]. As a rule, the study of collocation least-squares methods emphasizes algebraic principles, since they often lead to overdetermined algebraic systems that are solved through the corresponding normal equations.

In this paper, we focus attention on least-squares methods for which the discretization step is invoked *after* the least-squares functional has been defined; some earlier works refer to such methods as "continuous least-squares methods;" see [69]. The chief reason to adopt this setting is that it allows us to accentuate the variational interpretation of least-squares principles as projections in a Hilbert space with respect to problem-dependent inner products. From this point of view, the principal task in the formulation of the method becomes setting up a least-squares functional that is *norm-equivalent* in some Hilbert space. This in turn allows us to work in the variational setting of, e.g., the Lax–Milgram lemma.

From a theoretical viewpoint, such bona fide (continuous) least-squares finite element methods possess a number of significant and valuable properties, such as

- the weak problems are in general coercive;
- conforming discretizations lead to stable and, ultimately, optimally accurate methods;
- the resulting algebraic problems are symmetric and positive definite; and
- essential boundary conditions may be imposed in a weak sense.

These properties can yield the following notable computational advantages and simplifications when properly accounted for in the algorithmic design of least-squares

finite element methods:

- finite element spaces of equal interpolation order, defined with respect to the same triangulation, can be used for all unknowns;
- algebraic problems can be solved using standard and robust iterative methods, such as conjugate gradient methods; and
- methods can be implemented without any matrix assemblies, even at the element level.

In some specific nonlinear applications, e.g., the numerical solution of the incompressible Navier–Stokes equations, least-squares principles can offer the following significant added advantages:

- used in conjunction with a Newton linearization, least-squares finite element methods involve only symmetric, positive definite linear systems, at least in the neighborhood of a solution; and
- used in conjunction with properly implemented continuation techniques, e.g., with respect to the Reynolds number, a solution algorithm can be devised that will only encounter symmetric and positive definite linear systems.

In recent years this impressive list of theoretical and computational advantages has sparked a steadily growing interest in the use of least-squares ideas for the numerical solution of partial differential equations, and in particular, for the numerical solution of elliptic boundary value problems. This high level of activity in least-squares finite element methods makes it impossible to present, within a limited space, an exhaustive account of all current and past research directions. Thus, we have in mind the less ambitious goal of giving the reader a selective account of past and ongoing work that is sufficiently representative and illustrative of the developments in least-squares finite element methods.

The paper will have a strong focus on the advances made in least-squares finite element methods for the Stokes and Navier–Stokes equations. The analysis and implementation of such methods have drawn most of the attention of researchers interested in modern least-squares finite element methods, and there exists an abundant mathematical and engineering literature devoted to this subject; see, e.g., [3], [9], [11], [12], [13], [14], [15], [16], [17], [18], [19], [25], [34], [36], [47], [54], [57], [50], [51], [53], [59], [68], [87], [90], [91], [92], [93], [94], [96], [99], [105], [112], and [117] among others. As a result, least-squares finite element methods in these settings are among the best understood, studied, and tested from both the theoretical and computational viewpoints. Our discussion will also include least-squares methods for convection-diffusion and other second-order elliptic problems (see [6], [24], [26], [33], [35], [38], [42], [43], [44], [45], [48], [52], [62], [71], [72], [75], [86], [95], [106], [107], and [104]), linear elasticity (see [34], [36], and [37]), inviscid, compressible flows (see [61], [64], [67], [99], and [118]), and electromagnetics (see [46], [58], [60], [97], and [116]).

The paper is organized as follows. The rest of this section introduces notation, gives a background on finite element spaces, outlines the model problems that will be used in the discussion of least-squares finite element methods, and, for the sake of completeness and contrast, gives a short description of mixed and stabilized Galerkin methods for the Stokes equations. In section 2, we discuss a general least-squares framework that includes the formulation, analysis, and implementation of least-squares finite element methods, using an abstract boundary value problem. In section 3, we focus attention on the transformation of elliptic boundary value problems into first-order systems, which is one of the fundamental ideas in modern least-squares methods. The transformation process is illustrated using five different first-

order forms of the Stokes equations along with first-order forms for the biharmonic, convection-diffusion, and other equations. The central core of the paper is section 4, where various least-squares finite element methods for linear, elliptic partial differential equations are presented and compared. In section 5, we briefly consider the extension of least-squares finite element methodology and analysis to nonlinear problems, using as prototypes the Navier–Stokes equations and the equations of compressible, potential flow. In section 6, we include a brief review of methods such as collocation, restricted least-squares, and least-squares/optimization methods that fall outside the framework given in section 2.

1.1. Notation. Let Ω denote an open bounded domain in $\mathbb{R}^n$, $n = 2$ or 3, having a sufficiently smooth boundary Γ. Throughout, vectors will be denoted by boldface letters, e.g., $\mathbf{u}$, tensors by underlined boldface capitals, e.g., $\underline{\mathbf{T}}$, and C will denote a generic positive constant whose meaning and value changes with context. For $s \geq 0$, we use the standard notation and definition for the Sobolev spaces $H^s(\Omega)$ and $H^s(\Gamma)$ with corresponding inner products denoted by $(\cdot,\cdot)_{s,\Omega}$ and $(\cdot,\cdot)_{s,\Gamma}$ and norms by $\|\cdot\|_{s,\Omega}$ and $\|\cdot\|_{s,\Gamma}$, respectively. Whenever there is no chance for ambiguity, the measures Ω and Γ will be omitted from inner product and norm designations. We will simply denote the $L^2(\Omega)$ and $L^2(\Gamma)$ inner products by $(\cdot,\cdot)$ and $(\cdot,\cdot)_\Gamma$, respectively. We recall the space $H_0^1(\Omega)$ consisting of all $H^1(\Omega)$ functions that vanish on the boundary and the space $L_0^2(\Omega)$ consisting of all square integrable functions with zero mean with respect to Ω. Also, for negative values of s, we recall the dual spaces $H^s(\Omega)$; see, e.g., [1], for details. We also recall the notion of a Banach scale X_q (see, e.g., [98] or [102]).

By $(\cdot,\cdot)_{\mathbf{X}}$ and $\|\cdot\|_{\mathbf{X}}$ we denote inner products and norms, respectively, on the product spaces $\mathbf{X} = H^{s_1}(\Omega) \times \cdots \times H^{s_n}(\Omega)$; whenever all the indices s_i are equal we shall denote the resulting space by $[H^{s_1}(\Omega)]^n$ or $\mathbf{H}^s(\Omega)$, and simply write $(\cdot,\cdot)_{s,\Omega}$ and $\|\cdot\|_{s,\Omega}$ for the inner product and norm, respectively. Some important spaces that arise in the decomposition of vector fields are

$$H(\Omega,\mathrm{div}) = \{\mathbf{u} \in [L^2(\Omega)]^n \;\mid\; \mathrm{div}\,\mathbf{u} \in L^2(\Omega)\} \tag{1.1}$$

and

$$H(\Omega,\mathrm{curl}) = \{\mathbf{u} \in [L^2(\Omega)]^n \;\mid\; \mathrm{curl}\,\mathbf{u} \in [L^2(\Omega)]^k\}\,, \tag{1.2}$$

where $k = 1$ in two dimensions and $k = 3$ in three dimensions, along with the subspaces

$$H_0(\Omega,\mathrm{div}) = \{\mathbf{u} \in H(\Omega,\mathrm{div}) \;\mid\; \mathbf{u}\cdot\mathbf{n} = 0 \quad \text{on } \Gamma\} \tag{1.3}$$

and

$$H_0(\Omega,\mathrm{curl}) = \{\mathbf{u} \in H(\Omega,\mathrm{curl}) \;\mid\; \mathbf{u}\times\mathbf{n} = 0 \quad \text{on } \Gamma\}\,. \tag{1.4}$$

Norms corresponding to (1.1) and (1.3) and to (1.2) and (1.4) are given by

$$\|\mathbf{v}\|^2_{H(\Omega,\mathrm{div})} = \|\mathbf{v}\|_0^2 + \|\mathrm{div}\,\mathbf{v}\|_0^2 \quad \text{and} \quad \|\mathbf{v}\|^2_{H(\Omega,\mathrm{curl})} = \|\mathbf{v}\|_0^2 + \|\mathrm{curl}\,\mathbf{v}\|_0^2\,,$$

respectively. See [79] for details.

1.1.1. Finite element spaces. We let $\mathcal{T}_h$ denote a regular triangulation (see, e.g., [65]) of the domain Ω into finite elements. For example, in two dimensions, $\mathcal{T}_h$ could consist of triangles or rectangles. The parameter h is normally associated with

the size of the elements in the triangulation. Let $\mathcal{P}_k$ denote the set of all polynomials of degree less than or equal to k. For $k \geq 1$, a corresponding finite element space defined with respect to a subdivision $\mathcal{T}_h$ of Ω into triangles, or more generally, into simplices, is given by

$$P_k = \{u^h \in C^0(\overline{\Omega}) \;\mid\; u^h|_{\triangle} \in \mathcal{P}_k \quad \forall \triangle \in \mathcal{T}_h\}\,;$$

e.g., in $\mathbb{R}^2$, P_k is the space of all continuous, over $\overline{\Omega}$, piecewise polynomial functions u^h such that, in each triangle, $u^h \in \mathcal{P}_k$. Alternately, let $\mathcal{Q}_k$ denote the space of polynomial functions such that the degree of $q \in \mathcal{Q}_k$ in each coordinate direction does not exceed k. A corresponding finite element space defined with respect to a subdivision $\mathcal{T}_h$ of Ω into rectangles is given by

$$Q_k = \{u^h \in C^0(\overline{\Omega}) \;\mid\; u^h|_{\square} \in \mathcal{Q}_k \quad \forall \square \in \mathcal{T}_h\}\,.$$

Some commonly used finite element spaces are P_1 and P_2 (continuous, piecewise-linear, and quadratic elements on triangles) and Q_1 and Q_2 (continuous, piecewise-bilinear, and biquadratic elements on rectangles). Also, there are the piecewise-constant finite element spaces

$$P_0 = \{u^h|_{\triangle} \in \mathcal{P}_0 \quad \forall \triangle \in \mathcal{T}_h\} \qquad \text{and} \qquad Q_0 = \{u^h|_{\square} \in \mathcal{Q}_0 \quad \forall \square \in \mathcal{T}_h\}\,.$$

An important characteristic of every finite element space is its approximation order, i.e., the asymptotic rate of convergence of the best approximation out of the space. For the spaces P_k (or Q_k) defined above, we have the following property: for $k \geq 1$, given a function $u \in H^{k+1}(\Omega)$, there exists an element w^h in P_k (or Q_k) such that

$$\|u - w^h\|_r \leq Ch^{k+1-r}\|u\|_{k+1}, \quad r = 0, 1\,,$$

where the constant C is independent of h. For example, if $u \in H^3(\Omega)$, we can show that there exists an element $w^h \in P_2$ (or Q_2) such that

$$\|u - w^h\|_r \leq Ch^{3-r}\|u\|_3, \quad r = 0, 1\,.$$

See [65] for details.

1.2. Model problems. We now list the model problems that we will use as the context for our discussion of least-squares finite element algorithms.

1.2.1. The Poisson, Helmholtz, and biharmonic equations. Two fundamental prototype problems for second-order elliptic partial differential equations are given by the Poisson equation

$$-\triangle\phi = f \quad \text{in } \Omega \tag{1.5}$$

along with the boundary condition

$$\phi = 0 \quad \text{on } \Gamma \tag{1.6}$$

and the Helmholtz equation

$$\triangle\phi + k^2\phi = f \quad \text{in } \Omega \tag{1.7}$$

along with the boundary condition (1.6). Instead of (1.6), we can impose the inhomogeneous boundary condition

$$\phi = g \quad \text{on } \Gamma \tag{1.8}$$

or the Neumann boundary condition

$$\frac{\partial \phi}{\partial n} = \theta \quad \text{on } \Gamma \tag{1.9}$$

or combinations of (1.8) and (1.9) on disjoint parts of the boundary.

The leading prototype for higher-order problems is the fourth-order biharmonic equation

$$\triangle^2 \phi = f \quad \text{in } \Omega \tag{1.10}$$

along with the homogeneous Dirichlet boundary conditions

$$\phi = 0 \qquad \text{and} \qquad \frac{\partial \phi}{\partial n} = 0 \quad \text{on } \Gamma\,. \tag{1.11}$$

1.2.2. Convection-diffusion and potential flow. A more general problem involving second-order linear elliptic partial differential equations is given by the (generalized) convection-diffusion equation

$$-\text{div } (A(x)\,\text{grad}\,\phi) + \Lambda\phi = f \quad \text{in } \Omega\,, \tag{1.12}$$

where $A(x)$ is a symmetric, positive definite matrix and Λ is a linear differential operator of order less than or equal to one, along with the boundary condition (1.6). For (1.12), we could replace (1.6) by the inhomogeneous boundary condition (1.8) or by the flux condition

$$\mathbf{n} \cdot A(x)\,\text{grad}\,\phi = \theta \quad \text{on } \Gamma \tag{1.13}$$

or by a combination of (1.8) and (1.13) on disjoint parts of the boundary.

A related nonlinear, second-order problem is the velocity potential equation for steady, inviscid, irrotational, compressible flow

$$\text{div}\,(\rho \text{grad}\,\phi) = 0 \quad \text{in } \Omega\,, \tag{1.14}$$

where ϕ is the velocity potential and ρ is the fluid density; ρ is given in terms of $|\text{grad}\,\phi|$ by

$$\rho = \rho_0 \left(1 - \frac{|\text{grad}\,\phi|^2}{H_0}\right)\,, \tag{1.15}$$

where ρ_0 and H_0 denote the stagnation density and enthalpy, respectively; see, e.g., [64] or [83]. A boundary condition for (1.14) is given by

$$\rho\,\text{grad}\,\phi \cdot \mathbf{n} = 0 \quad \text{on } \Gamma\,, \tag{1.16}$$

where in this case Γ is usually an obstacle in the flow. Additional boundary conditions on a far-field boundary are also imposed for problems posed in exterior domains; see [83]. If the flow is everywhere subsonic, then (1.14) is a nonlinear elliptic equation; in regions where the flow is supersonic, (1.14) is of hyperbolic type.

1.2.3. Linear elasticity. The equations of linear elasticity provide a model for linear, second-order, elliptic systems of partial differential equations. These equations are given by

$$-\mu\triangle\mathbf{u} - (\lambda+\mu)\text{grad}\,\text{div}\,\mathbf{u} = \mathbf{f} \quad \text{in } \Omega\,, \tag{1.17}$$

where $\mathbf{u}$ denotes the displacement vector, $\mathbf{f}$ a given body force, and λ and μ are the Lamé constants. Displacement and traction boundary conditions are given by

$$\mathbf{u} = \mathbf{g} \quad \text{on } \Gamma \tag{1.18}$$

and

$$\sigma_{ij}(\mathbf{u})\mathbf{n}_j = \boldsymbol{\theta} \quad \text{on } \Gamma\,, \tag{1.19}$$

respectively, where $\sigma(\mathbf{u}) = 2\mu\varepsilon(\mathbf{u}) + \lambda\text{tr}(\varepsilon(\mathbf{u}))$ denotes the stress tensor and $\varepsilon(\mathbf{u}) = (1/2)(\text{grad}\,\mathbf{u} + \text{grad}\,\mathbf{u}^T)$ the deformation tensor. Combinations of (1.18) and (1.19) on disjoint parts of the boundary are also of interest.

If $A = \lambda\mathbf{b}\mathbf{b}^T + 2\mu B$, where $\mathbf{b} = (1,0,0,1)^T$ and $B_{11} = B_{44} = 1$, $B_{22} = B_{33} = B_{23} = B_{32} = 1/2$, with all other $B_{ij} = 0$, the system (1.17)–(1.19) takes a form very similar to (1.12) and (1.13), i.e.,

$$-\text{div}\,(A\,\text{grad}\,\mathbf{u}) = \mathbf{f} \quad \text{in } \Omega \tag{1.20}$$

and (1.18) or

$$\mathbf{n}\cdot(A\,\text{grad}\,\mathbf{u}) = \boldsymbol{\theta} \quad \text{on } \Gamma\,. \tag{1.21}$$

We also easily see that (1.17) can be rewritten in the form

$$-\frac{\mu}{\lambda+\mu}\triangle\mathbf{u} + \text{grad}\,p = \frac{1}{\lambda+\mu}\mathbf{f} \quad \text{in } \Omega \tag{1.22}$$

and

$$\text{div}\,\mathbf{u} + p = 0 \quad \text{in } \Omega\,. \tag{1.23}$$

This form of the equations of linear elasticity is merely a perturbed form of the Stokes equations introduced below in subsection 1.2.4.

1.2.4. The Stokes and Navier–Stokes equations. Most of our discussion will be in the context of the stationary Stokes problem with velocity boundary conditions as given by

$$-\nu\triangle\mathbf{u} + \text{grad}\,p = \mathbf{f} \quad \text{in } \Omega\,, \tag{1.24}$$

$$\text{div}\,\mathbf{u} = 0 \quad \text{in } \Omega\,, \tag{1.25}$$

and

$$\mathbf{u} = \mathbf{0} \quad \text{on } \Gamma\,, \tag{1.26}$$

where $\mathbf{u}$ denotes the velocity field, p the pressure, ν a given constant, and $\mathbf{f}$ a given function. We shall also consider the incompressible Navier–Stokes equations for which (1.24) is replaced by

$$-\nu\triangle\mathbf{u} + \mathbf{u}\cdot\text{grad}\,\mathbf{u} + \text{grad}\,p = \mathbf{f} \quad \text{in } \Omega\,. \tag{1.27}$$

Often, we append to (1.24)–(1.26) or (1.25)–(1.27) a zero mean constraint on the pressure field of the form

$$\int_\Omega p\, d\Omega = 0 \tag{1.28}$$

in order to assure the uniqueness of pressure. Also, in many cases, (1.26) is replaced by the inhomogeneous velocity boundary condition $\mathbf{u} = \mathbf{U}$ on Γ.

Clearly, the constant ν can be scaled out of the problem (1.24)–(1.26); however, keeping the reference to ν in (1.24) will ease the transition from the discussion of least-squares finite element methods for the linear Stokes problem to like methods for the nonlinear Navier–Stokes problem for which ν denotes the inverse of the Reynolds number.

1.2.5. Div-curl systems. We conclude the list of model problems with div-curl systems of the form

$$\operatorname{curl} \mathbf{u} = \mathbf{f} \quad \text{in } \Omega \tag{1.29}$$

and

$$\operatorname{div} \mathbf{u} = g \quad \text{in } \Omega, \tag{1.30}$$

augmented by one of the boundary conditions

$$\mathbf{u} \cdot \mathbf{n} = 0 \qquad \text{or} \qquad \mathbf{n} \times \mathbf{u} = 0 \quad \text{on } \Gamma\,. \tag{1.31}$$

The functions $\mathbf{f}$ and g are subject to compatibility conditions. For example, (1.29) requires that $\operatorname{div} \mathbf{f} = 0$, and the first boundary condition in (1.31) requires that g have zero mean with respect to Ω. A related system in three dimensions is given by (1.30) and

$$\operatorname{curl} \mathbf{u} + \nabla \phi = \mathbf{f} \quad \text{in } \Omega, \tag{1.32}$$

with perhaps an additional boundary condition on ϕ, e.g., $\phi = 0$ on Γ.

The systems (1.29)–(1.31) and (1.30)–(1.32) are representative of systems of partial differential equations that arise in electromagnetics, fluid dynamics, and other applications. Note that (1.29)–(1.30) is a first-order system of partial differential equations. Also note that, in two dimensions, the operator on the left-hand side of (1.29)–(1.30) is merely the Cauchy–Riemann operator, so that (1.29)–(1.30) can be viewed as a generalization of the Cauchy–Riemann equations.

1.3. Mixed and stabilized methods for the Stokes problem. For the sake of completeness and for contrast with bona fide least-squares finite element methods, we describe, in a very brief manner, well-known mixed-Galerkin finite element methods for the Stokes problem and some of the approaches that have been used to stabilize these methods.

1.3.1. Mixed finite element methods for the Stokes problem. A mixed variational formulation for (1.24)-(1.26) is given by (see [79], [82], or [27])

$$\text{seek } \mathbf{u} \in \mathbf{H}_0^1(\Omega) \text{ and } p \in L_0^2(\Omega) \text{ such that}$$

$$a(\mathbf{u}, \mathbf{v}) + b(p, \mathbf{v}) = (\mathbf{f}, \mathbf{v})_0 \quad \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \tag{1.33}$$

and

$$b(q,\mathbf{u}) = 0 \qquad \forall\, q \in L_0^2(\Omega)\,. \tag{1.34}$$

In (1.33)–(1.34),

$$a(\mathbf{u},\mathbf{v}) = \int_\Omega \operatorname{grad}\mathbf{u} : \operatorname{grad}\mathbf{v}\, d\Omega \quad \text{and} \quad b(q,\mathbf{u}) = -\int_\Omega q \operatorname{div}\mathbf{u}\, d\Omega\,,$$

where the colon denotes the scalar product operator between two tensors.

If (V^h, S^h), $V^h \subset \mathbf{H}_0^1(\Omega)$ and $S^h \subset L_0^2(\Omega)$, denotes a pair of discrete spaces for the approximation of the velocity and pressure fields, respectively, then it is not difficult to show that a discretization of (1.33)–(1.34) results in an algebraic problem of the form

$$\begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} U^h \\ P^h \end{pmatrix} = \begin{pmatrix} F \\ 0 \end{pmatrix}, \tag{1.35}$$

where U^h and P^h are vectors from which the discrete velocity and pressure fields may be determined. The coefficient matrix in (1.35) is symmetric but indefinite.

A consequence of the fact that the variational equations (1.33)–(1.34) correspond to a saddle-point optimization problem is that the existence and uniqueness of solutions to the discrete system (1.35) are subject to the well-known inf-sup or LBB condition (see, e.g., [79], [82], or [27]) that there exists $\beta > 0$, independent of h, such that

$$\inf_{q\in S^h} \sup_{\mathbf{v}\in V^h} \frac{b(q,\mathbf{v})}{\|q\|_0 \|\mathbf{v}\|_1} \geq \beta\,. \tag{1.36}$$

In fact, a conforming discretization of (1.33)–(1.34) is well posed if and only if the conforming finite-dimensional spaces V^h and S^h satisfy the inf-sup condition (1.36) and the form $a(.,.)$ is coercive on $Z^h \times Z^h$, where $Z^h \subset V^h$ denotes the subspace of *discretely divergence-free functions*

$$Z^h = \{v^h \in V^h \mid b(q^h, v^h) = 0 \quad \forall\, q^h \in S^h\}\,.$$

Consequently, the inclusions $V^h \subset \mathbf{H}_0^1(\Omega)$ and $S^h \subset L_0^2(\Omega)$ are not by themselves sufficient to guarantee the stability of the discrete problem.

The verification of the inf-sup condition (1.36) can be rather intricate (see, e.g., [20], [22], [79], or [82]) and some of the most obvious choices of pairs of finite element spaces do not satisfy it. Examples of such unstable spaces are the linear-constant pair (P_1-P_0) and the bilinear-constant pair (Q_1-Q_0), neither of which satisfies the LBB condition, although the inclusions $P_1 \subset H^1(\Omega)$, $Q_1 \subset H^1(\Omega)$, $P_0 \subset L^2(\Omega)$, and $Q_0 \subset L^2(\Omega)$ hold; see [21], [22], [79], or [82].

1.3.2. Stabilized Galerkin and penalty methods for the Stokes problem. Complications caused by the inf-sup condition have prompted the introduction of various stabilization techniques intended to circumvent this condition. Typically, stabilization relies on some form of modification of the discrete continuity equation. For example, in [84] that equation is modified to

$$\int_\Omega q^h \operatorname{div}\mathbf{u}^h d\Omega + \alpha \sum_{\triangle\in\mathcal{T}_h} h_\triangle^2 \int_\triangle (-\triangle\mathbf{u}^h + \operatorname{grad} p^h - \mathbf{f}) \cdot \operatorname{grad} q^h\, d\Omega = 0\,.$$

The purpose of the new term (with an appropriate choice for α) is to make the bilinear form

$$a(\mathbf{u}^h,\mathbf{v}^h)+b(p^h,\mathbf{v}^h)-b(q^h,\mathbf{u}^h)+\alpha\sum_{\triangle\in\mathcal{T}_h}h_\triangle^2\int_\triangle(-\triangle\mathbf{u}^h+\operatorname{grad}p^h)\cdot\operatorname{grad}q^h\,d\Omega$$

coercive with respect to the mesh-dependent norm $(\|\mathbf{u}^h\|_1^2+\sum_{\triangle\in\mathcal{T}_h}h_\triangle^2\|\nabla p\|_{0,\triangle}^2)^{1/2}$. This approach can be interpreted as adding a type of least-squares term; however, it can't be viewed as being, according to our definition, a bona fide least-squares method, i.e., one based on the minimization of a quadratic functional. For more examples of such Galerkin least-squares methods we refer to [8], [77], and [78].

A somewhat different approach, which targets the indefiniteness of the system (1.35), is represented by *penalty* methods; see, e.g. [10], [28], [29], [70], [85], and [114]. The main idea is to make (1.35) positive definite by "penalizing" the continuity equation, i.e., by replacing (1.35) with a system of the form

$$\begin{pmatrix}A & B\\ B^T & \varepsilon M\end{pmatrix}\begin{pmatrix}U^h\\ P^h\end{pmatrix}=\begin{pmatrix}F\\ 0\end{pmatrix}.$$

The matrix M denotes the Gram matrix of the finite element basis for the pressure-approximating space. Penalty methods are essentially a form of Tikhonov regularization since they can be derived starting from the regularized Stokes problem

$$-\nu\triangle\mathbf{u}_\varepsilon+\operatorname{grad}p_\varepsilon=\mathbf{f}\quad\text{in }\Omega,$$

$$\operatorname{div}\mathbf{u}_\varepsilon=-\varepsilon p_\varepsilon\quad\text{in }\Omega,\tag{1.37}$$

and

$$\mathbf{u}_\varepsilon=\mathbf{0}\quad\text{on }\Gamma.$$

A related penalty stabilization method is that of [28], [29]. Here, instead of $\varepsilon p_\varepsilon$ in (1.37) the continuity equation is penalized by the mesh-dependent term $h^2\triangle p^h$. Accordingly, the modified weak continuity equation takes the form

$$\int_\Omega q^h\operatorname{div}\mathbf{u}^h d\Omega+\sum_{\triangle\in\mathcal{T}_h}h^2\int_\triangle\operatorname{grad}p^h\cdot\operatorname{grad}q^h\,d\Omega=0.$$

Unlike Galerkin least-squares methods, penalty formulations introduce a *penalty* error of order proportional to the added penalty term.

Each of these classes of methods, e.g., mixed, stabilized Galerkin, and penalty, have their adherents and are used in practice; none, however, has gained universal popularity.

2. An abstract framework for least-squares problems. A specific least-squares finite element method for the numerical solution of a given elliptic boundary value problem is determined by several factors, such as the form of the system of partial differential equations and boundary conditions, the functional setting for the relevant differential and boundary operators, the choice of discretization spaces, and the choice of solution method for the discrete equations. Thus, the application of least-squares principles can result in several substantially different algorithms, even for the same problem. Despite these differences, many least-squares methods share as

a common principle the minimization of a *norm-equivalent* functional. This allows us to consolidate the formulation and analysis of a large class of methods into a single abstract variational framework. The main goal of this section is to outline such a framework for an abstract boundary value problem of the form

$$\mathcal{L}(U) = F \quad \text{in } \Omega \tag{2.1}$$

and

$$\mathcal{R}(U) = G \quad \text{on } \Gamma, \tag{2.2}$$

and then to discuss how this framework can be adapted in order to address issues such as the optimality and practicality of least-squares finite element methods.

2.1. The abstract framework. We assume that $\mathcal{L}$ is a differential operator which corresponds to a scalar equation or to a system of equations and that $\mathcal{L}$ is elliptic in the sense of Agmon, Douglis, and Nirenberg (ADN); see [2]. We recall that an important subclass of ADN elliptic operators arises in planar systems of *Petrovsky* type; see [115]. These are first-order differential operators of the form $\mathcal{L}(U) = AU_x + BU_y + CU$, where A, B, and C denote $2n$ by $2n$ matrices such that $\det(A\boldsymbol{\xi} + B\boldsymbol{\eta}) = 0$ if and only if $\boldsymbol{\xi} = \boldsymbol{\eta} = \mathbf{0}$. We further assume that there are two Hilbert scales $X_q(\Omega)$ and $Y_q(\Omega) \times Y_q(\Gamma)$ such that $(\mathcal{L}, \mathcal{R})$ has a complete set of homeomorphisms, i.e., the mapping $U \mapsto (\mathcal{L}(U), \mathcal{R}(U))$ is a homeomorphism $X_q(\Omega) \mapsto Y_q(\Omega) \times Y_q(\Gamma)$ for all q; see [109]. As a result, (2.1)–(2.2) is well posed in the indicated Hilbert scales, and the a priori estimate

$$\|U\|_{X_q} \le C \left(\|\mathcal{L}(U)\|_{Y_q(\Omega)} + \|\mathcal{R}(U)\|_{Y_q(\Gamma)} \right) \tag{2.3}$$

is valid for all q. In order to set up a norm-equivalent functional, residuals of (2.1)–(2.2) must be measured in the norms indicated by the a priori estimate (2.3), i.e., with the problem (2.1)–(2.2), we associate a quadratic least-squares functional of the form

$$\mathcal{J}(U) = \frac{1}{2} \left(\|\mathcal{L}(U) - F\|^2_{Y_q(\Omega)} + \|\mathcal{R}(U) - G\|^2_{Y_q(\Gamma)} \right). \tag{2.4}$$

It is not difficult to see that a minimizer of (2.4) solves (2.1)–(2.2) and conversely, i.e., an abstract least-squares principle for (2.1)–(2.2) is given by

$$\text{seek } U \in X_q \text{ such that } \mathcal{J}(U) \le \mathcal{J}(V) \ \forall\ V \in X_q. \tag{2.5}$$

Then, a variational problem for (2.1)–(2.2) is readily available by means of the Euler–Lagrange equation for the principle (2.5), as follows:

$$\delta \mathcal{J}(U) = \lim_{\varepsilon \to 0} \frac{d}{d\varepsilon} \mathcal{J}(U + \varepsilon V) = 0 \quad \forall\ V \in X_q\,.$$

Equivalently, we can write the necessary condition as

$$\text{seek } U \in X_q \text{ such that } \mathcal{B}(U, V) = \mathcal{F}(V) \quad \forall\ V \in X_q, \tag{2.6}$$

where $\mathcal{B}(U, V) = (\mathcal{L}(U), \mathcal{L}(V))_{Y_q(\Omega)} + (\mathcal{R}(U), \mathcal{R}(V))_{Y_q(\Gamma)}$ is a symmetric, continuous bilinear form and $\mathcal{F}(V) = (F, \mathcal{L}(V))_{Y_q(\Omega)} + (G, \mathcal{R}(V))_{Y_q(\Gamma)}$ is a continuous linear

functional. A consequence of the a priori estimate (2.3) is that the form $\mathcal{B}(\cdot,\cdot)$ is coercive on $X_q \times X_q$, i.e., a bound of the form

$$C\|U\|^2_{X_q} \le \mathcal{B}(U,U) \tag{2.7}$$

is valid for all $U \in X_q$. As a result, the existence and uniqueness of minimizers for the problem (2.6) follow by virtue of the Lax–Milgram lemma.

The inclusion of the boundary residual in (2.4) allows the use of minimization spaces $X_q(\Omega)$ that are not constrained to satisfy the boundary condition (2.2), i.e., such conditions are enforced weakly through the variational principle. This is advantageous whenever the condition (2.2) is difficult to satisfy computationally and represents an additional beneficial feature of least-squares–based methods. If, on the other hand, (2.2) can be easily imposed, we can consider (2.4) with the boundary term omitted. Then the functions belonging to the space $X_q(\Omega)$ should be required to satisfy the boundary condition, i.e., (2.2) is enforced strongly or directly on candidate minimizers.

We now have a framework for developing a least-squares finite element method for (2.1)–(2.2); indeed, we can proceed as follows. Given the boundary value problem (2.1)–(2.2),

- choose a functional setting, i.e., two Hilbert scales $X_q(\Omega)$ and $Y_q(\Omega) \times Y_q(\Gamma)$ such that (2.3) is valid;
- choose between variational or strong enforcement of the boundary conditions and set up the relevant least-squares functional, i.e., (2.4) with or without the boundary residual term, respectively;
- fix the scale parameter q, e.g., choose $q = 0$;
- choose a finite-dimensional subspace $X_q^h \subset X_q$ parametrized by $h \to 0$; and
- restrict (2.6) to X_q^h.

This process leads to a discrete variational problem given by

$$\text{seek } U^h \in X_q^h \text{ such that } \mathcal{B}(U^h, V^h) = \mathcal{F}(V^h) \quad \forall\, V^h \in X_q^h. \tag{2.8}$$

Since $X_q^h \subset X_q$, the method defined by (2.8) is a *conforming* discretization. Such discretizations have numerous theoretical advantages. In particular, the analysis of such methods falls within the standard framework of elliptic finite element theory. However, there are instances when we prefer to use *nonconforming* discretizations for which the spaces X_q^h are not contained in X_q. Such a case is discussed in more detail in subsection 4.2.1.

2.2. Optimality and practicality. Most research activity in least-squares finite element methods has used the framework of subsection 2.1 to formulate methods that are "easy" to implement and that, at the same time, are optimally accurate, where the latter is defined to mean that there exists a constant C, independent of the parameter h, such that

$$\|U - U^h\|_{X_q} \le C \inf_{V^h \in X_q^h} \|U - V^h\|_{X_q}\,. \tag{2.9}$$

For conforming methods, i.e., if $X_q^h \subset X_q(\Omega)$, it is not difficult to see that a condition that guarantees the existence and uniqueness of minimizers and optimal discretization errors is given by the *optimality principle*, where $\mathcal{L}$, $\mathcal{R}$, $X_q(\Omega)$, and $Y_q(\Omega) \times Y_q(\Gamma)$ are such that an a priori estimate of the form (2.3) is valid. Indeed, if $\{\phi_i\}$ denotes a basis for X_q^h, it is not difficult to see that problem (2.8) corresponds to a linear algebraic

system with a symmetric coefficient matrix A having entries given by $a_{ij} = \mathcal{B}(\phi_i, \phi_j)$ and right-hand side vector B with components given by $b_i = \mathcal{F}(\phi_i)$. The inclusion $X_q^h \subset X_q$ implies that the bound (2.7), which is a consequence of (2.3), is also valid for all discrete functions U^h, i.e., the form $\mathcal{B}(\cdot, \cdot)$ is coercive on $X_q^h \times X_q^h$. As a result, the discrete variational problem will have a unique solution out of X_q^h. Similarly, the validity of (2.7) for all functions in X_q^h implies that the matrix $\mathcal{B}(\phi_i, \phi_j)$ is positive definite. Finally, standard arguments from elliptic finite element theory yield the optimal estimate (2.9).

Thus, given a particular elliptic boundary value problem, the framework of subsection 2.1 allows us to define, in relatively easy terms, a least-squares method that theoretically is optimally accurate. However, a straightforward application of this framework will not necessarily result in a practical method. To deem a least-squares finite element method as being practical, we should at least be able to obtain the discrete system without difficulty, certainly with no more difficulty than that encountered for a Galerkin method for the same problem. To meet this requirement, first $a_{ij} = \mathcal{B}(\phi_i, \phi_j)$ and $b_i = \mathcal{F}(\phi_i)$ should be computable, i.e., inner products for fractional-order Sobolev spaces should not occur in their definition. Second, discretization should be accomplished using standard, easy-to-use finite element spaces. We must, of course, also solve the discrete problem (2.8), and thus a third requirement for practicality, especially when iterative solution methods are used, is that the discrete problem should have a "manageable" condition number. In summary, we then have the *practicality principle*, where the inner products in $\mathcal{B}(\cdot, \cdot)$ are computable, discretization is accomplished employing standard $C^0(\overline{\Omega})$ finite element spaces, and the discrete problem has a manageable condition number.

For an example of a problem and a least-squares finite element method for which the optimality principle holds but is not practical, consider the problem (1.5)–(1.6). We require that the boundary condition is imposed strongly and we thus choose $X_q(\Omega) = H^2(\Omega) \cap H_0^1(\Omega)$. A least-squares functional can then be defined in a straightforward manner by taking the L^2-norm of the equation residual

$$\mathcal{J}(\phi) = \frac{1}{2} \left(\| \triangle \phi + f \|_0^2 \right) . \tag{2.10}$$

The bilinear form corresponding to the functional (2.10) is given by

$$\mathcal{B}(\phi, \psi) = \int_\Omega \triangle \phi \triangle \psi \, d\Omega \, .$$

As a result, any conforming method based on the minimization of (2.10) necessarily uses finite-dimensional subspaces of $H^2(\Omega)$, i.e., we are forced to work with impractical continuously differentiable finite element spaces. The resulting least-squares finite element method also fails to be a practical one because the condition numbers of the corresponding discrete problems are $O(h^{-4})$ compared with the $O(h^{-2})$ condition numbers that result from standard Galerkin methods for the same problem. See [6], [24], and [26] for details.

2.3. A recipe for practicality. A critical idea responsible for overcoming these flaws and rendering least-squares finite element methods into viable alternatives to Galerkin methods is the introduction of a decomposition step prior to defining a least-squares functional. The decomposition step consists of transforming the problem into a first-order system. For many problems, decomposition can be accomplished through

the introduction of physically meaningful variables such as vorticity, stresses, or fluxes, and has often been exploited in both least-squares and Galerkin methods; see, e.g., [11], [12], [13], [14], [15], [16], [17], [18], [19], [33], [34], [35], [36], [37], [38], [42], [43], [44], [45], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [59], [61], [62], [64], [66], [67], [72], [73], [86], [87], [90], [91], [92], [93], [94], [95], [96], [99], [103], [104], [107], [108], [111], [112], [113], and [117].

Among the first methods for which a decomposition (or transformation) step has been combined with least-squares principles are the methods of [72], [86], and [103]. For example, in [86], the problem (1.5)–(1.6) is transformed into a first-order div-grad system (see also [49] and [95])

$$-\operatorname{div}\mathbf{v} = f \quad \text{and} \quad \mathbf{v} = \operatorname{grad}\phi \quad \text{in } \Omega \qquad \text{and} \qquad \phi = 0 \quad \text{on } \Gamma\,. \tag{2.11}$$

We can use L^2-norms of the equation residuals to define the least-squares functional

$$\mathcal{J}(\phi,\mathbf{v}) = \frac{1}{2}\left(\|\operatorname{div}\mathbf{v} + f\|_0^2 + \|\mathbf{v} - \operatorname{grad}\phi\|_0^2\right)\,, \tag{2.12}$$

where we require that the boundary condition is imposed strongly on candidate minimizers. Then, in contrast to (2.10), the resulting least-squares method can be implemented using practical, merely continuous finite element spaces such as P_1, P_2, Q_1, or Q_2. If, however, the space $X_q(\Omega)$ is not constrained by the boundary condition in (2.11), then instead of (2.12) we have to consider the least-squares functional

$$\mathcal{J}(\phi,\mathbf{v}) = \frac{1}{2}\big(\|\operatorname{div}\mathbf{v} + f\|_0^2 + \|\mathbf{v} - \operatorname{grad}\phi\|_0^2 + \|\phi\|_{1/2,\Gamma}^2\big)\,.$$

As a result, the form $\mathcal{B}(\cdot,\cdot)$ now involves the inner product of the trace space $Y(\Gamma) = H^{1/2}(\Gamma)$, i.e., the method is still impractical despite the reduction to a first-order system. We postpone discussion of this topic until subsection 4.2.2 and turn our attention to the issue of the optimality of methods based on functionals that only involve L^2-norms of residuals of equations in first-order systems.

Evidently, the practicality principle can always be met by transforming the given partial differential equations into a first-order system and forming least-squares functionals that use only L^2-norms. We refer to least-squares methods based on this approach as *basic L^2 methods.* Then, a seemingly natural choice for $Y_q(\Omega)$ and $X_q(\Omega)$ is given by products of L^2- and H^1-spaces, respectively. The crucial question is then: does this choice of $Y_q(\Omega)$ and $X_q(\Omega)$ satisfy the optimality principle as well, i.e., is the basic L^2 least-squares functional norm equivalent to a product of H^1-norms? In what follows, we shall refer to all first-order systems having this property, i.e., systems that are well posed in products of L^2- and H^1-spaces for the data and solution, respectively, as *fully H^1-coercive* formulations. In such a case, basic L^2 least-squares functionals are quite attractive. First, conforming methods are practical and yield optimal error estimates with respect to the $H^1(\Omega)$-norm for *all variables.* Second, the resulting algebraic systems have condition numbers of $O(h^{-2})$. A third valuable trait of H^1-coercivity is that interactions between the dependent variables are subdominant, i.e., the variables become essentially decoupled. This allows us to precondition the resulting algebraic systems by products of the inverses of discrete Laplacian operators and ensures the fast convergence of multiplicative and additive multigrid methods; see [34], [35], and [36].

One example of a fully H^1-coercive problem is furnished by planar first-order systems with differential operators $\mathcal{L}$ of Petrovsky type (see subsection 2.1 and [115]).

Unfortunately, and somewhat surprisingly, many (important) first-order systems are not fully H^1-coercive. For example, the system (2.11) is well posed in the space $H_0^1(\Omega) \times H(\Omega, \text{div})$. As a result, a finite element method based on the minimization of (2.12) yields optimal error estimates in the norm of this space and not necessarily with respect to the $\mathbf{H}^1(\Omega)$-norm of the vector field $\mathbf{v}$, i.e., we cannot control all first derivatives of the error. As we shall see below, for (2.12), optimal convergence in the $\mathbf{H}^1(\Omega)$-norm requires an additional "grid decomposition property" on the finite element spaces (see [61] and [72]) or an additional "curl" constraint added to the first-order system (2.11) (see [43], [44], [48], [49], [95], and [106]). Other examples of first-order systems that are not H^1-coercive are given by particular decompositions of the biharmonic equation and the Stokes problem. These will be discussed in subsections 3.1 and 3.2, respectively.

There are many other theoretical problems that accompany the decomposition process. For example, transformation to a first-order system may lead to ambiguously defined principal parts and may affect the set of admissible boundary operators as well as the ellipticity, which can also depend on the space dimension. As a result, if an optimally accurate and practical method is desired, all of these issues must be addressed. The possibilities which exist can be distinguished as techniques affecting the functional setting of least-squares principles or techniques invoked at the discretization level, as shown below.

Functional level

Spaces

- establish the relevant a priori estimates in intermediate spaces such as $H(\Omega, \text{div})$; see [33], [35], [38], [76], [86], and [107];
- use weaker spaces from the Hilbert scale, e.g., choose negative-order Sobolev spaces; see [14], [23], [25], and [34];

Operators

- find, if possible, fully H^1-coercive first-order systems; see [36], [37], [47], and [53];

Discretization level

- replace stronger (and impractical) norms by weighted L^2-norms; see [3], [4], [11], [12], [17], [19], [64], [75], [86], and [115];
- use nonconforming discretizations; see [3].

Most modern least-squares methods rely on one or another form of these techniques in order to achieve robustness and efficiency. Most methods also rely on the transformation of elliptic boundary value problems into equivalent first-order systems, if the former are not given in the latter form in the first place. Due to the central place of this subject in modern least-squares methods, the next section is devoted to a detailed account of various transformations to first-order systems.

3. Decomposition into first-order systems. In some settings, e.g., the Maxwell equations, mathematical models of natural phenomena are most naturally set as first-order systems of partial differential equations. In many other settings, however, such models involve differential equations of higher order. Fortunately, virtually every scalar partial differential equation or system of partial differential equations of second or higher order can be reduced to a first-order system by introducing new dependent variables defined as derivatives (or linear combinations of derivatives) of the original variables. For example, in the scalar, second-order setting, if ϕ denotes the scalar dependent variable, then a standard way to introduce new dependent variables is to set $\mathbf{v} = \text{grad}\,\phi$, which leads to a "div-grad" decomposition, i.e., to a div-grad first-order system. Such decompositions will be considered in subsection 3.1.

If the dependent variable $\mathbf{u}$ is a vector-valued function, there are at least three categories of transformations to first-order systems. In addition to $\mathbf{v} = \operatorname{grad}\mathbf{u}$, two other natural choices are $\mathbf{v} = (\operatorname{grad}\mathbf{u} + \operatorname{grad}\mathbf{u}^T)/2$ or $\mathbf{v} = (\operatorname{grad}\mathbf{u} - \operatorname{grad}\mathbf{u}^T)/2$, i.e., the new variables are identified with the symmetric and the skew-symmetric parts of the velocity gradient tensor, respectively. While the first choice can be viewed as a generalization of the scalar div-grad decomposition, the other two choices do not have direct equivalents in the scalar case. Their interpretation also depends on the given problem. For example, in the context of the Stokes problem, the choice of $(\operatorname{grad}\mathbf{u} - \operatorname{grad}\mathbf{u}^T)/2$ gives rise to a vorticity-based first-order system, whereas the choice $(\operatorname{grad}\mathbf{u} + \operatorname{grad}\mathbf{u}^T)/2$ corresponds to a stress-based first-order system.

As we shall see, the well-posedness of resulting first-order boundary value problems, i.e., their ellipticity and the existence of a priori estimates, depends on factors such as the choice for the new variables, the space dimension, and the particular form of the boundary conditions. As a result, two first-order systems describing the same problem (in terms of different variables) may have substantially different properties. To illustrate these issues, the main focus (see subsections 3.2–3.3) will be on five different possibilities for expressing the Stokes problem as a first-order system. The first two systems correspond to decompositions that use the skew-symmetric (vorticity) and symmetric (stress) parts of the velocity gradient, respectively. The last three systems can be viewed as variants of the div-grad decomposition. For each system, we indicate the properties relevant to least-squares methods such as the validity of a priori estimates. In addition, we also indicate how resulting first-order systems may be extended to the incompressible Navier–Stokes equations.

Most details in the discussion below are given in the context of homogeneous boundary conditions. However, in subsection 3.4, we discuss the form of the a priori estimates in case the boundary data is inhomogeneous. Such a priori estimates are relevant to least-squares functionals for which the essential boundary conditions are enforced in a weak sense.

3.1. Scalar div-grad systems. We present several examples of the reduction of scalar, higher-order problems to first-order systems involving the divergence and gradient operators. The new variables introduced to effect the reduction are the components of the gradient of the scalar dependent variable of the original higher-order equation.

3.1.1. Poisson, biharmonic, and higher-order equations. The first example of a div-grad decomposition has already been given by (2.11). This decomposition is also probably the first one used in the context of least-squares methods; see [72], [86], and [103]. A functional setting for the system (2.11) is provided by the space

$$X = \{(\phi, \mathbf{v}) \in H_0^1(\Omega) \times H(\Omega, \operatorname{div})\}\,, \tag{3.1}$$

i.e., the system (2.11) is not fully H^1-coercive, and the a priori estimate relevant to the least-squares method is given by

$$\|\phi\|_1 + \|\mathbf{v}\|_{H(\Omega,\operatorname{div})} \le C\left(\|\mathbf{v} - \operatorname{grad}\phi\|_0 + \|\operatorname{div}\mathbf{v}\|_0\right)\,. \tag{3.2}$$

In view of the equation $\mathbf{v} = \operatorname{grad}\phi$, we can augment (2.11) by a compatibility condition known as the *curl constraint* (see, e.g., [55], [49], [48], and [95]) as shown:

$$\operatorname{curl}\mathbf{v} = 0\,. \tag{3.3}$$

The addition of (3.3) allows us to control both $\|\operatorname{curl}\mathbf{v}\|_0$ and $\|\operatorname{div}\mathbf{v}\|_0$. As a result, we can establish full H^1-coercivity of the system (2.11) and (3.3) as follows. Since $\phi = 0$ on Γ is specified, the new variable $\mathbf{v}$ satisfies $\mathbf{v}\times\mathbf{n} = 0$ on Γ, i.e., $\mathbf{v} \in H(\Omega,\operatorname{div})\cap H_0(\Omega,\operatorname{curl})$. Then, full H^1-coercivity follows from the well-known result (see [79]) that the space $H(\Omega,\operatorname{div})\cap H_0(\Omega,\operatorname{curl})$ is topologically and algebraically equivalent to $[H^1(\Omega)]^n$. A div-grad decomposition with very similar properties can be defined in an obvious manner for the Helmholtz equation (1.7). If, instead of (1.6), $\partial\phi/\partial n = \mathbf{v}\cdot\mathbf{n} = 0$ is specified on Γ, then $\mathbf{v}\in H_0(\Omega,\operatorname{div})\cap H(\Omega,\operatorname{curl})$ and again full H^1-coercivity can be established if the system $\operatorname{div}\mathbf{v} = f$ and $\mathbf{v} = \operatorname{grad}\phi$ is augmented by (3.3).

The fourth-order biharmonic problem (1.10)–(1.11) can be recast as a first-order system in a very similar manner. We introduce the new dependent variables as follows [76]:

$$\phi_0 = \phi\,,\quad \mathbf{v}_0 = \operatorname{grad}\phi_0\,,\quad \phi_1 = \operatorname{div}\operatorname{grad}\phi_0 = \triangle\phi_0\,,\quad \mathbf{v}_1 = \operatorname{grad}\triangle\phi_0\,.$$

Then, a first-order problem corresponding to (1.10)–(1.11) is given by

$$\mathbf{v}_0 - \operatorname{grad}\phi_0 = 0 \quad \text{in } \Omega\,, \tag{3.4}$$

$$\mathbf{v}_1 - \operatorname{grad}\phi_1 = 0 \quad \text{in } \Omega\,, \tag{3.5}$$

$$\operatorname{div}\mathbf{v}_0 - \phi_1 = 0 \quad \text{in } \Omega\,, \tag{3.6}$$

$$\operatorname{div}\mathbf{v}_1 = f \quad \text{in } \Omega\,, \tag{3.7}$$

and

$$\phi_0 = 0 \quad \text{and} \quad \mathbf{v}_0\cdot\mathbf{n} = 0 \quad \text{on } \Gamma\,. \tag{3.8}$$

A function setting appropriate for this first-order system is provided by

$$X = \{(\phi_0,\phi_1,\mathbf{v}_0,\mathbf{v}_1)\in H_0^1(\Omega)\times H^1(\Omega)\times H_0(\Omega,\operatorname{div})\times H(\Omega,\operatorname{div})\}\,. \tag{3.9}$$

An a priori estimate relevant to the least-squares method for (3.4)–(3.8) is given by

$$\begin{aligned}&\|\phi_0\|_1 + \|\phi_1\|_1 + \|\mathbf{v}_0\|_{H(\Omega,\mathrm{hdiv})} + \|\mathbf{v}_1\|_{H(\Omega,\mathrm{hdiv})}\\ &\quad\le C(\|\mathbf{v}_0 - \operatorname{grad}\phi_0\|_0 + \|\mathbf{v}_1 - \operatorname{grad}\phi_1\|_0 + \|\operatorname{div}\mathbf{v}_0 - \phi_1\|_0 + \|\operatorname{div}\mathbf{v}_1\|_0)\,.\end{aligned} \tag{3.10}$$

The transformation scheme applied to the biharmonic problem (1.10)–(1.11) can be generalized to scalar elliptic problems of order $2m$; see [76]. Starting with $\phi_0 = \phi$ and $\mathbf{v}_0 = \operatorname{grad}\phi_0$, we introduce the new variables

$$\mathbf{v} = (\mathbf{v}_0,\dots,\mathbf{v}_{m-1}) \quad \text{and} \quad \phi = (\phi_0,\dots,\phi_{m-1})$$

recursively according to the formulas

$$\phi_i = \operatorname{div}\mathbf{v}_{i-1}\,,\quad i = 1,\dots,m-1\,,$$

and

$$\mathbf{v}_i = \operatorname{grad}\phi_i\,,\quad i = 1,\dots,m-1\,.$$

The resulting first-order system has the structure

$$\begin{pmatrix} I & G \\ D & C \end{pmatrix} \begin{pmatrix} V \\ \Phi \end{pmatrix} = F, \tag{3.11}$$

where G is matrix of gradients, D is matrix of divergences, and C is a bounded (with respect to L^2) multiplication operator.

The two first-order systems (3.4)–(3.8) and (3.11) are very similar to the system (2.11) arising from the decomposition of the Poisson equation, both in structure and ellipticity properties (compare (3.2) and (3.10)). Likewise, (3.4)–(3.8) and (3.11) can be augmented by curl constraints to obtain fully H^1-coercive first-order systems. For example, the curl constraints for (3.4)–(3.8) are given by

$$\operatorname{curl} \mathbf{v}_0 = 0 \qquad \text{and} \qquad \operatorname{curl} \mathbf{v}_1 = 0\,.$$

3.1.2. Compressible flows and convection-diffusion problems. The transformation of both (1.12)–(1.13) and (1.14)–(1.16) can also be handled by introducing $\operatorname{grad}\phi$ as a new dependent variable, i.e., by using a scalar div-grad decomposition. For example, a first-order form for (1.14)–(1.16) is given by (see [61], [64], [67], and [99])

$$\mathbf{u} - \operatorname{grad}\phi = 0 \quad \text{in } \Omega\,, \tag{3.12}$$

$$\operatorname{div} \rho\mathbf{u} = 0 \quad \text{in } \Omega\,, \tag{3.13}$$

and

$$\rho\mathbf{u} \cdot \mathbf{n} = 0 \quad \text{on } \Gamma\,, \tag{3.14}$$

where

$$\rho = \rho_0 \left(1 - \frac{1}{H_0}|\mathbf{u}|^2\right)\,. \tag{3.15}$$

Equation (3.12) is a consequence of the irrotationality assumption, and (3.13) is a result of mass balance. Similarly, (1.12) can be cast into the following first-order system (see, e.g., [33], [38], [43], [44], [106], and [107]):

$$\mathbf{v} - A(x)\operatorname{grad}\phi = 0 \quad \text{in } \Omega \tag{3.16}$$

and

$$\operatorname{div}\mathbf{v} + \Lambda\phi = f \quad \text{in } \Omega\,. \tag{3.17}$$

In both cases, if the flow is subsonic and since A is positive definite, respectively, the appropriate function spaces for the original variable ϕ and the new dependent variable $\mathbf{v}$ are given by $H^1(\Omega)$ and $H(\Omega, \operatorname{div})$, respectively, i.e., the relevant a priori estimates are similar to (3.2). Likewise, in view of (3.12) and (3.16), the systems (3.12)–(3.13) and (3.16)–(3.17) can be augmented by curl constraints of the form

$$\operatorname{curl}\mathbf{v} = 0 \quad \text{in } \Omega \tag{3.18}$$

and

$$\operatorname{curl}\left(A^{-1}(x)\mathbf{v}\right) = 0 \quad \text{in } \Omega\,, \tag{3.19}$$

respectively; see [43], [44], [106], and [107].

3.2. Vorticity and stress decompositions of the Stokes problem. We now turn to the Stokes equations, the setting for which least-squares finite element methods are most developed and have been most studied. In this section, we consider two first-order formulations of the Stokes equations that are not fully H^1-coercive. In subsection 3.3 we will consider three additional first-order formulations for the Stokes system, two of which are fully H^1-coercive.

3.2.1. The velocity-vorticity-pressure formulation of the Stokes problem. Despite the fact that it is not a fully H^1-coercive system, the *velocity-vorticity-pressure* formulation of the Stokes equations is, by a large margin, the most widely used (and studied) formulation in the context of least-squares finite element methods for fluid flows; see, e.g., [11], [12], [15], [16], [17], [18], [34], [50], [51], [54], [57], [59], [87], [90], [91], [92], [93], [94], [96], [99], [112], and [113], among others.

We recall the curl operator in three dimensions

$$\operatorname{curl} \mathbf{u} = \nabla \times \mathbf{u}\,,$$

and its two-dimensional counterparts

$$\operatorname{curl} \phi = \begin{pmatrix} \phi_y \\ -\phi_x \end{pmatrix} \qquad \text{and} \qquad \operatorname{curl} \mathbf{u} = u_{2x} - u_{1y}\,.$$

The context should make clear which operator is relevant.

Using the *vorticity* field $\boldsymbol{\omega} = \operatorname{curl} \mathbf{u}$ as a new dependent variable and the vector identity

$$\operatorname{curl}\operatorname{curl} \mathbf{u} = -\triangle \mathbf{u} + \operatorname{grad}\operatorname{div} \mathbf{u}\,,$$

and in view of the incompressibility constraint (1.25), the Stokes equations (1.24)–(1.25) can be cast into the first-order system

$$\nu \operatorname{curl} \boldsymbol{\omega} + \operatorname{grad} p = \mathbf{f} \qquad \text{in } \Omega\,, \tag{3.20}$$

$$\operatorname{div} \mathbf{u} = 0 \qquad \text{in } \Omega\,, \tag{3.21}$$

and

$$\operatorname{curl} \mathbf{u} - \boldsymbol{\omega} = \mathbf{0} \qquad \text{in } \Omega\,. \tag{3.22}$$

Remark. For the linear elasticity problem (1.22)–(1.23), the corresponding first-order system is given by (see [34])

$$\frac{\mu}{\lambda+\mu} \operatorname{curl} \boldsymbol{\omega} + \left(1 + \frac{\mu}{\lambda+\mu}\right) \operatorname{grad} p = \frac{1}{\lambda+\mu} \mathbf{f} \qquad \text{in } \Omega\,,$$

$$\operatorname{div} \mathbf{u} + p = 0 \qquad \text{in } \Omega\,,$$

and (3.22).

In two dimensions, the system (3.20)–(3.22) contains four equations and four unknowns and is uniformly elliptic of total order four. In three dimensions, the number of equations and unknowns increases to seven, and the resulting system is not elliptic in the sense of ADN. By adding a seemingly redundant equation

$$\operatorname{div} \boldsymbol{\omega} = 0 \qquad \text{in } \Omega \tag{3.23}$$

and the gradient of a "slack" variable ϕ to (3.22) we get

$$\operatorname{curl}\mathbf{u} - \boldsymbol{\omega} + \operatorname{grad}\phi = 0 \qquad \text{in } \Omega \tag{3.24}$$

and uniform ellipticity can be restored; see [55] and [49]. The augmented system (3.20), (3.21), (3.23), and (3.24) has total order eight, in contrast to the total order of the Stokes problem in primitive variables, which is six in three dimensions. It should be noted that we also impose homogeneous boundary conditions for the slack variable ϕ and that we can then show that $\phi \equiv 0$ so that, a posteriori, (3.24) is identical to (3.22). In fact, the addition of ϕ is needed only for the purpose of analysis; it is not needed in the development or implementation of least-squares–based algorithms for which we can safely use the system (3.20)–(3.23). However, the addition of (3.23) is crucial to the stability and accuracy of least-squares finite element methods for the Stokes problem in three dimensions.

To extend the velocity-vorticity-pressure formulation to the Navier–Stokes equations, we have to choose a particular form for the nonlinear term in (1.27). One possibility is to keep the nonlinear term in a form involving only the velocity field, i.e., to replace (3.20) by

$$\nu\operatorname{curl}\boldsymbol{\omega} + \mathbf{u}\cdot\operatorname{grad}\mathbf{u} + \operatorname{grad}p = \mathbf{f} \qquad \text{in } \Omega\,. \tag{3.25}$$

Another possibility is to use the vector identity

$$\mathbf{u}\cdot\operatorname{grad}\mathbf{u} = \frac{1}{2}\operatorname{grad}|\mathbf{u}|^2 - \mathbf{u}\times\operatorname{curl}\mathbf{u} = \frac{1}{2}\operatorname{grad}|\mathbf{u}|^2 - \mathbf{u}\times\boldsymbol{\omega}$$

to replace (3.20) by

$$\nu\operatorname{curl}\boldsymbol{\omega} + \boldsymbol{\omega}\times\mathbf{u} + \operatorname{grad}P = \mathbf{f} \qquad \text{in } \Omega\,, \tag{3.26}$$

where $P = p + 1/2|\mathbf{u}|^2$ denotes the total pressure.

For simplicity, in what follows, we shall discuss properties of the velocity-vorticity-pressure Stokes equations in two dimensions; most of the relevant results can be easily extended to the augmented system, i.e., including (3.23), in three dimensions.

Recall that the existence of a priori estimates for (3.20)–(3.22) along with the boundary condition (1.26) and the zero mean condition (1.28) is of paramount importance for the development of least-squares methods. The existence of such estimates is essentially equivalent to the well-posedness of elliptic boundary value problems in appropriate Sobolev spaces. As we shall see, the choice of boundary conditions is very important to the validity of a priori estimates. To determine functional settings in which elliptic boundary value problems are well posed, we may rely on the elliptic regularity theory of Agmon, Douglis, and Nirenberg [2]. In particular, well-posed problems are characterized as having uniformly elliptic principal parts and boundary conditions which satisfy the celebrated *complementing condition*.

Analyses based on ADN elliptic theory (see [17]), show that the velocity-vorticity-pressure system admits two different principal parts given by

$$\mathcal{L}_1^p = \begin{pmatrix} \nu\operatorname{curl}\boldsymbol{\omega} & + & \operatorname{grad}p \\ & & \operatorname{curl}\mathbf{u} \\ & & \operatorname{div}\mathbf{u} \end{pmatrix} \tag{3.27}$$

and

$$\mathcal{L}_2^p = \begin{pmatrix} \nu\operatorname{curl}\boldsymbol{\omega} & + & \operatorname{grad}p \\ -\boldsymbol{\omega} & + & \operatorname{curl}\mathbf{u} \\ & & \operatorname{div}\mathbf{u} \end{pmatrix}. \tag{3.28}$$

In view of the boundary condition (1.26) and the zero mean condition (1.28), the function spaces corresponding to these principal parts are given by

$$X_q = H^{q+1}(\Omega) \times H^{q+1}(\Omega) \cap L_0^2(\Omega) \times \mathbf{H}^{q+1}(\Omega) \cap \mathbf{H}_0^1(\Omega) \tag{3.29}$$

and

$$Y_q = H^q(\Omega) \times H^q(\Omega) \times \mathbf{H}^q(\Omega) \tag{3.30}$$

for (3.27) and

$$X_q = H^{q+1}(\Omega) \times H^{q+1}(\Omega) \cap L_0^2(\Omega) \times \mathbf{H}^{q+2}(\Omega) \cap \mathbf{H}_0^1(\Omega) \tag{3.31}$$

and

$$Y_q = H^q(\Omega) \times H^q(\Omega) \times \mathbf{H}^{q+1}(\Omega) \tag{3.32}$$

for (3.28), where X_q denotes the function space for the unknowns $(\boldsymbol{\omega}, p, \mathbf{u})$ and Y_q denotes the function space for the data or equation residuals. To each pair of function spaces there corresponds a specific a priori estimate given by

$$\begin{aligned} &\|\boldsymbol{\omega}\|_{q+1} + \|p\|_{q+1} + \|\mathbf{u}\|_{q+1} \\ &\qquad \le C\left(\|\nu \operatorname{curl} \boldsymbol{\omega} + \operatorname{grad} p\|_q + \|\operatorname{curl} \mathbf{u} - \boldsymbol{\omega}\|_q + \|\operatorname{div} \mathbf{u}\|_q\right) \end{aligned} \tag{3.33}$$

and

$$\begin{aligned} &\|\boldsymbol{\omega}\|_{q+1} + \|p\|_{q+1} + \|\mathbf{u}\|_{q+2} \\ &\qquad \le C\left(\|\nu \operatorname{curl} \boldsymbol{\omega} + \operatorname{grad} p\|_q + \|\operatorname{curl} \mathbf{u} - \boldsymbol{\omega}\|_{q+1} + \|\operatorname{div} \mathbf{u}\|_{q+1}\right), \end{aligned} \tag{3.34}$$

respectively. Although both the principal parts (3.27) and (3.28) are uniformly elliptic operators of total order four, not all boundary conditions for the system (3.20)–(3.22) will satisfy the complementing condition for both principal parts. For example, the boundary condition (1.26) on the velocity vector satisfies the complementing condition only with the principal part (3.28). As a result, the a priori estimate for the system (3.20)–(3.22), (1.26), and (1.28) relevant to the least-squares methods is given by (3.34). In fact, we can show that the estimate (3.33) cannot hold with the velocity boundary condition. For an example (see [17]), consider Ω given by the unit square and let $\nu = 1$, $q = 0$, $\boldsymbol{\omega}_n = -\cos(nx)\exp(ny)$, $p_n = \sin(nx)\exp(ny)$, and $\mathbf{u}_n \equiv \mathbf{0}$. Then, (3.33) would imply that

$$\begin{aligned} O(\exp(n)) \sim \|\operatorname{curl} \boldsymbol{\omega}_n + \operatorname{grad} p_n\|_0 + \|\operatorname{curl} \mathbf{u}_n - \boldsymbol{\omega}_n\|_0 + \|\operatorname{div} \mathbf{u}_n\|_0 \\ \ge C(\|\mathbf{u}_n\|_1 + \|\boldsymbol{\omega}_n\|_1 + \|p_n\|_1) \sim O(n \exp(n)), \end{aligned}$$

which is a contradiction. This counterexample can also be extended to three dimensions; see [12]. An example of a boundary condition for which (3.33) is valid is provided by the *pressure–normal-velocity* boundary condition

$$p = 0 \quad \text{and} \quad \mathbf{u} \cdot \mathbf{n} = 0 \quad \text{on } \Gamma\,. \tag{3.35}$$

The fact that (3.34) is not valid for velocity boundary conditions indicates that the corresponding boundary value problem is not well posed in the spaces (3.29)–(3.30). This can also be seen by considering the principal part (3.27) along with the velocity

boundary condition. The corresponding boundary value problem then uncouples into two ill-posed problems given by

$$\left\{ \nu \operatorname{curl} \boldsymbol{\omega} + \operatorname{grad} p = \mathbf{f} \right\} \quad \text{and} \quad \left\{ \begin{array}{r} \operatorname{curl} \mathbf{u} = 0 \\ \operatorname{div} \mathbf{u} = 0 \\ \mathbf{u}_{|\Gamma} = \mathbf{0} \end{array} \right\};$$

the first is underdetermined and the second is overdetermined. In contrast, the same principal part with (3.35) uncouples into the following two well-posed problems:

$$\left\{ \begin{array}{r} \nu \operatorname{curl} \boldsymbol{\omega} + \operatorname{grad} p = \mathbf{f} \\ p_{|\Gamma} = 0 \end{array} \right\} \quad \text{and} \quad \left\{ \begin{array}{r} \operatorname{curl} \mathbf{u} = 0 \\ \operatorname{div} \mathbf{u} = 0 \\ \mathbf{u} \cdot \mathbf{n}_{|\Gamma} = 0 \end{array} \right\}.$$

To summarize, the full H^1-coercivity of the velocity-vorticity-pressure Stokes system depends on the particular set of boundary conditions. Even more interestingly, we can find examples of boundary conditions for which full H^1-coercivity also depends on the space dimension; see [12]. One such example is given by the tangential-velocity–pressure boundary condition

$$\mathbf{n} \times \mathbf{u} \times \mathbf{n} = \mathbf{0} \quad \text{and} \quad p = 0 \quad \text{on } \Gamma.$$

In two dimensions, this boundary operator satisfies the complementing condition with either of the principal parts (3.27) or (3.28), whereas in three dimensions it satisfies the same condition only with the principal part (3.28). As a result, the estimate (3.33) is valid only in two dimensions.

Despite the fact that the velocity-vorticity-pressure formulation (3.20)–(3.22) along with (1.26) is not fully H^1-coercive, it is the most widely used formulation for least-squares finite element methods for the Stokes equations. The corresponding formulations (3.21), (3.22), and (3.25) or (3.26) along with (1.26) are likewise the most used for the Navier–Stokes system. The reasons for the popularity of these formulations in engineering circles are that they require relatively few unknowns and that the lack of full H^1-coercivity does not cause a catastrophic loss in accuracy, at least for the approximations of the velocity field; see [54] and Table 1 in subsection 4.1.1.

3.2.2. The velocity-pressure-stress formulation of the Stokes problem. A first-order system with substantially different properties is obtained when the stress tensor scaled by $\sqrt{\nu/2}$:

$$\underline{\mathbf{T}} = \sqrt{2\nu}\epsilon(\mathbf{u}), \quad \text{where} \quad \epsilon(\mathbf{u}) \equiv \frac{1}{2}\left(\operatorname{grad} \mathbf{u} + (\operatorname{grad} \mathbf{u})^T\right),$$

is used in the transformation of (1.24)–(1.25) into a first-order system. Here, the relevant vector identity is given by

$$\operatorname{div} \underline{\mathbf{T}} = \sqrt{2\nu}\left(\triangle \mathbf{u} + \operatorname{grad} \operatorname{div} \mathbf{u}\right),$$

where $\operatorname{div} \underline{\mathbf{T}}$ denotes the vector whose components are the divergences of the corresponding rows of $\underline{\mathbf{T}}$. Then, in view of incompressibility constraint (1.25), the system

(1.24)–(1.26) can be replaced by the *velocity-pressure-stress* system

$$
\begin{aligned}
\sqrt{2\nu}\,\operatorname{div}\underline{\mathbf{T}} - \operatorname{grad} p &= \mathbf{f} && \text{in } \Omega,\\
\operatorname{div}\mathbf{u} &= 0 && \text{in } \Omega,\\
\underline{\mathbf{T}} - \sqrt{2\nu}\,\epsilon(\mathbf{u}) &= \underline{\mathbf{0}} && \text{in } \Omega,\\
\mathbf{u} &= \mathbf{0} && \text{on } \Gamma\,.
\end{aligned}
\tag{3.36}
$$

The inclusion of the nonlinear term $\mathbf{u}\cdot\operatorname{grad}\mathbf{u}$ in the first equation of (3.36) provides an extension of the velocity-pressure-stress system to the Navier–Stokes equations. As before, uniqueness of solutions to (3.36) can be guaranteed by imposing the zero mean constraint on the pressure (1.28).

In two dimensions, the velocity-pressure-stress system has six equations and unknowns. In three dimensions, the number of unknowns and equations increases to ten. Analysis of this system based on the ADN theory (see [19]) indicates that the principal part of the differential operator in (3.36) is given, in two dimensions, by

$$
\mathcal{L}^p U = \begin{pmatrix}
\sqrt{2\nu}\left(\dfrac{\partial T_1}{\partial x_1} + \dfrac{\partial T_2}{\partial x_2}\right) - \dfrac{\partial p}{\partial x_1}\\[2ex]
\sqrt{2\nu}\left(\dfrac{\partial T_2}{\partial x_1} + \dfrac{\partial T_3}{\partial x_2}\right) - \dfrac{\partial p}{\partial x_2}\\[2ex]
\dfrac{\partial u_1}{\partial x_1} + \dfrac{\partial u_2}{\partial x_2}\\[2ex]
T_1 - \sqrt{2\nu}\dfrac{\partial u_1}{\partial x_1}\\[2ex]
2T_2 - \sqrt{2\nu}\left(\dfrac{\partial u_1}{\partial x_2} + \dfrac{\partial u_2}{\partial x_1}\right)\\[2ex]
T_3 - \sqrt{2\nu}\dfrac{\partial u_2}{\partial x_2}
\end{pmatrix}, \quad \text{where} \quad \underline{\mathbf{T}} = \begin{pmatrix} T_1 & T_2\\ T_2 & T_3 \end{pmatrix}.
\tag{3.37}
$$

In contrast to the velocity-vorticity-pressure equations, the principal part (3.37) is unambiguously defined, and the total order of (3.36) coincides with the total order of the Stokes problem in primitive variables in both two and three dimensions; see [19]. The corresponding functional setting in two dimensions for the problem (3.36) is given by

$$
X_q = [H^{q+1}(\Omega)]^3 \times H^{q+1}(\Omega) \cap L^2_0(\Omega) \times [H^{q+2}(\Omega) \cap H^1_0(\Omega)]^2
$$

for the unknowns $(\underline{\mathbf{T}}, p, \mathbf{u})$ and

$$
Y_q = [H^q(\Omega)]^2 \times H^{q+1}(\Omega) \times [H^{q+1}(\Omega)]^3
$$

for the data or equation residuals. As a result, the a priori estimate relevant to least-squares methods is given by

$$
\begin{aligned}
&\|\underline{\mathbf{T}}\|_{q+1} + \|p\|_{q+1} + \|\mathbf{u}\|_{q+2}\\
&\qquad \le C\big(\|\sqrt{2\nu}\,\operatorname{div}\underline{\mathbf{T}} - \operatorname{grad} p\|_q + \|\operatorname{div}\mathbf{u}\|_{q+1} + \|\underline{\mathbf{T}} - \sqrt{2\nu}\,\epsilon(\mathbf{u})\|_{q+1}\big)\,.
\end{aligned}
\tag{3.38}
$$

Note that the estimate (3.38) implies that regardless of the choice of boundary operators, the system (3.36) cannot be H^1-coercive, i.e., in two dimensions (3.36) is not of Petrovsky type.

3.3. Div-grad decompositions of the Stokes problem. A fully H^1-coercive first-order system is quite appealing from a least-squares perspective: a basic L^2 method is optimal, discrete problems have condition numbers of order $O(h^{-2})$, and these problems can be solved by efficient multilevel techniques; see [35] and [36]. However, as we have seen in subsection 3.2, when the skew-symmetric or symmetric parts of the velocity gradient are used as new dependent variables, the resulting first-order systems are not always fully H^1-coercive. Two of the first-order Stokes systems presented in this section are guaranteed to be fully H^1-coercive, regardless of the choice of boundary conditions. Interestingly, both formulations introduce all derivatives of the velocity field as new dependent variables and augment the resulting first-order systems with additional constraints, i.e., they can be viewed as vector counterparts of the scalar div-grad or div-grad-curl decompositions in subsection 3.1. Currently it appears that, for the Stokes problem, full H^1-coercivity that is independent of space dimension and boundary conditions can be achieved solely by means of augmented div-grad decompositions.

3.3.1. Two velocity–gradient-velocity-pressure formulations of the Stokes problem. To define the first *velocity–gradient-velocity-pressure formulation,* we introduce *all* first derivatives of the velocity components as new dependent variables, i.e., we set $\underline{\mathbf{V}} = (\operatorname{grad}\mathbf{u})^t$ so that $V_{ij} = (\partial u_i/\partial x_j)$. In terms of $\underline{\mathbf{V}}$, the Stokes problem (1.24)–(1.26) is given by

$$-\nu \operatorname{div}\underline{\mathbf{V}} + \operatorname{grad} p = \mathbf{f} \quad \text{in } \Omega\,, \tag{3.39}$$

$$\operatorname{div}\mathbf{u} = 0 \quad \text{in } \Omega\,, \tag{3.40}$$

$$\underline{\mathbf{V}} - (\operatorname{grad}\mathbf{u})^t = \underline{\mathbf{0}} \quad \text{in } \Omega\,, \tag{3.41}$$

and

$$\mathbf{u} = \mathbf{0} \quad \text{on } \Gamma\,, \tag{3.42}$$

where $\operatorname{div}\underline{\mathbf{V}}$ denotes the vector whose components are the divergences of the corresponding rows of $\underline{\mathbf{V}}$. The system (3.39)–(3.42) is not fully H^1-coercive. The relevant a priori estimate associated with this system is given by

$$\begin{aligned} &\|\underline{\mathbf{V}}\|_{q+1} + \|\mathbf{u}\|_{q+2} + \|p\|_{q+1} \\ &\qquad \le C\left(\| - \nu \operatorname{div}\underline{\mathbf{V}} + \operatorname{grad} p\|_q + \|\underline{\mathbf{V}} - (\operatorname{grad}\mathbf{u})^t\|_{q+1} + \|\operatorname{div}\mathbf{u}\|_{q+1}\right)\,. \end{aligned} \tag{3.43}$$

In [36], the new variables $\underline{\mathbf{V}}$ are called "velocity fluxes"; since that terminology is usually reserved for a different physical quantity, we have adopted the terminology "velocity gradient."

The main idea of [36] is that full H^1-coercivity can be obtained by augmenting (3.39)–(3.42) with additional constraints. In particular, in view of the identity $\operatorname{tr}\underline{\mathbf{V}} = \operatorname{div}\mathbf{u}$, the definition of $\underline{\mathbf{V}}$, and the boundary condition (1.26), we can add to (3.39)–(3.42) the equations

$$\operatorname{grad}(\operatorname{tr}\underline{\mathbf{V}}) = \mathbf{0} \quad \text{in } \Omega \tag{3.44}$$

and

$$\operatorname{curl}\underline{\mathbf{V}} = \underline{\mathbf{0}} \quad \text{in } \Omega \tag{3.45}$$

and the boundary condition

$$\underline{\mathbf{V}} \times \mathbf{n} = \underline{\mathbf{0}} \quad \text{on } \Gamma\,, \tag{3.46}$$

where curl $\underline{\mathbf{V}}$ denotes the vector whose components are the curls of the corresponding rows of $\underline{\mathbf{V}}$ and $\underline{\mathbf{V}} \times \mathbf{n}$ denotes the vector whose components are the vector product of the rows of $\underline{\mathbf{V}}$ with the unit outer normal vector $\mathbf{n}$. Note that (3.45) generalizes the curl constraints (3.3) and (3.18) to the tensor case.

The resulting system (3.39)–(3.42) and (3.44)–(3.46) is overdetermined, but consistent. In 2 dimensions, the number of unknowns equals 7, and the number of equations equals 11. In 3 dimensions, we have 13 unknowns and 25 equations. A functional setting relevant to the least-squares method for the augmented system (3.39)–(3.42) and (3.44)–(3.46) is given by

$$X_q = \widetilde{\mathbf{H}}^{q+1}(\Omega) \times \mathbf{H}^{q+1}(\Omega) \cap \mathbf{H}_0^1(\Omega) \times H^{q+1}(\Omega) \cap L_0^2(\Omega)$$

for the unknowns $(\underline{\mathbf{V}}, \mathbf{u}, p)$, where $\widetilde{\mathbf{H}}^{q+1}(\Omega) = [H^{q+1}(\Omega)]^{n^2}$ constrained by (3.46), and

$$Y_q = \mathbf{H}^q(\Omega) \times H^q(\Omega) \times [H^q(\Omega)]^{n^2} \times \mathbf{H}^q(\Omega) \times \mathbf{H}^q(\Omega)$$

for the equation residuals; the corresponding a priori estimate is given by

$$\begin{aligned} &\|\underline{\mathbf{V}}\|_{q+1} + \|\mathbf{u}\|_{q+1} + \|p\|_{q+1} \\ &\le C\big(\| - \nu \operatorname{div} \underline{\mathbf{V}} + \operatorname{grad} p\|_q + \|\operatorname{div} \mathbf{u}\|_q \\ &+ \|\underline{\mathbf{V}} - (\operatorname{grad} \mathbf{u})^t\|_q + \|\operatorname{grad}(tr\underline{\mathbf{V}})\|_q + \|\operatorname{curl} \underline{\mathbf{V}}\|_q\big)\,. \end{aligned} \tag{3.47}$$

The two velocity–gradient-velocity-pressure formulations can be easily extended to the Navier–Stokes equations. In terms of the new variable $\underline{\mathbf{V}}$, the nonlinear term in (1.27) can be expressed as $\underline{\mathbf{V}} \cdot \mathbf{u}$ so that, for the Navier–Stokes problem, (3.39) is replaced by

$$-\nu \operatorname{div} \underline{\mathbf{V}} + \underline{\mathbf{V}} \cdot \mathbf{u} + \operatorname{grad} p = \mathbf{f} \quad \text{in } \Omega\,.$$

The (generalized) div-grad decomposition used above can also be used for the linear elasticity equations (1.17)–(1.19); see [37]. The new variables are again $\underline{\mathbf{V}} = (\operatorname{grad} \mathbf{u})^t$ and are called "displacement fluxes" in [37], but perhaps should be more properly referred to as "displacement gradients." The first-order displacement-gradient system corresponding to the form (1.20)–(1.21) of the equations of linear elasticity is given by

$$-\operatorname{div} A\underline{\mathbf{V}} = \mathbf{f} \quad \text{in } \Omega\,, \tag{3.48}$$

$$\underline{\mathbf{V}} - (\operatorname{grad} \mathbf{u})^t = 0 \quad \text{in } \Omega\,, \tag{3.49}$$

and

$$\mathbf{u} = \mathbf{0} \quad \text{and} \quad \mathbf{n} \cdot A\underline{\mathbf{V}} = \mathbf{0} \quad \text{on } \Gamma\,. \tag{3.50}$$

Similarly to (3.39)–(3.41), to achieve full H^1-coercivity, the system (3.48)–(3.50) should be augmented by a curl constraint of the form

$$\operatorname{curl} \underline{\mathbf{V}} = 0 \quad \text{in } \Omega\,.$$

3.3.2. The constrained-velocity–gradient-pressure formulation of the Stokes problem. This approach was suggested in [47] and here we present it in the case of two space dimensions. The new variables introduced to effect the transformation to a first-order system are the entries of the velocity gradient constrained by the incompressibility constraint (1.25), i.e., they are given by

$$\underline{\mathbf{G}} = \begin{pmatrix} v_1 & v_2 \\ v_3 & -v_1 \end{pmatrix}, \tag{3.51}$$

where

$$v_1 = \frac{\partial u_1}{\partial x_1} = -\frac{\partial u_2}{\partial x_2}, \qquad v_2 = \frac{\partial u_1}{\partial x_2}, \qquad \text{and} \qquad v_3 = \frac{\partial u_2}{\partial x_1},$$

and where u_1 and u_2 denote the components of the velocity $\mathbf{u}$. Using the new variables and the equality of second mixed derivatives, the Stokes problem (1.24)–(1.26) in two dimensions can be written in the form [47]

$$\begin{aligned} -\nu \operatorname{div} \underline{\mathbf{G}} + \operatorname{grad} p &= \mathbf{f} \quad \text{in } \Omega, \\ \operatorname{curl} \underline{\mathbf{G}} &= \mathbf{0} \quad \text{in } \Omega, \\ \underline{\mathbf{G}} \times \mathbf{n} &= \mathbf{0} \quad \text{on } \Gamma. \end{aligned} \tag{3.52}$$

In [47], the new variables (3.51) are called "accelerations" and the system (3.52) the "acceleration-velocity" formulation of the Stokes equations. However, the new variables are not components of the acceleration vector so that, instead, we call the system (3.52) the *constrained-velocity–gradient-pressure* formulation of the Stokes problem.

The planar system (3.52) has four equations and four unknowns, and we can show that it is elliptic in the sense of Petrovsky so that, with $\widetilde{\mathbf{H}}^{q+1}(\Omega) = [H^{q+1}(\Omega)]^3$ constrained by the boundary condition in (3.52), the function setting is given by

$$X_q = \widetilde{\mathbf{H}}^{q+1}(\Omega) \times H^{q+1}(\Omega) \cap L_0^2(\Omega) \qquad \text{and} \qquad Y_q = [H^q(\Omega)]^2 \times [H^q(\Omega)]^2$$

for the unknowns $(\underline{\mathbf{G}}, p)$ and the equation residuals, respectively. The a priori estimate relevant to least-squares methods is given by

$$\|\underline{\mathbf{G}}\|_{q+1} + \|p\|_{q+1} \le C \left(\| - \nu \operatorname{div} \underline{\mathbf{G}} + \operatorname{grad} p\|_q + \|\operatorname{curl} \underline{\mathbf{G}}\|_q \right). \tag{3.53}$$

The velocity has been eliminated from (3.52); it is recovered by solving the additional div-curl system

$$\begin{aligned} \operatorname{curl} \mathbf{u} &= v_3 - v_2 \quad \text{in } \Omega, \\ \operatorname{div} \mathbf{u} &= 0 \quad \text{in } \Omega, \\ \mathbf{u} \cdot \mathbf{n} &= 0 \quad \text{on } \Gamma. \end{aligned} \tag{3.54}$$

Although it is not obvious that the solution of (3.54) satisfies the boundary condition (1.26), it can be shown that this is indeed the case.

Although the system (3.52) is fully H^1-coercive, owing to the elimination of the velocity field, this system cannot be extended to the Navier–Stokes equations. Elimination of the velocity field in (3.52) can be considered as an artifact since we can simply consider (3.52) together with (3.54). Such a first-order system is studied in [53], where the new variables are called "stresses" and the corresponding first-order system is called the "stress-velocity-pressure" Stokes system despite the fact that the new variables are not the components of the stress tensor. This is not to be confused with the formulation of subsection 3.2.2 for which the true stresses are used.

3.3.3. First-order Stokes formulations: Concluding remarks. We have presented five different first-order systems that can be derived from the Stokes equations by introducing new dependent variables. In all five cases the new variables involve derivatives of the velocity field. When new variables represent linear combinations of these derivatives, such as the vorticity or stresses, resulting systems are not always fully H^1-coercive. This is due to the fact that interdependencies between the new variables and the velocity field remain coupled, i.e., formulations "remember" that some of the variables are actually velocity derivatives. To uncouple the variables, the velocity-gradient and the constrained velocity-gradient approaches use the components of the velocity gradient as new dependent variables, and add new constraints until the dependencies between the variables become subdominant. This may lead to an overdetermined, but consistent, problem.

3.4. Inhomogeneous boundary conditions. The a priori estimates presented above are valid for homogeneous boundary conditions imposed on the function spaces. Below we give a necessarily brief account of the relevant a priori estimates for inhomogeneous essential boundary conditions.

To determine the appropriate norms for a given boundary operator we may rely again on the elliptic regularity theory of [2], or on various trace theorems relating boundary and interior norms of functions. For example, a result of [102] states that for every $g \in H^{1/2}(\Gamma)$ there is a unique $u \in H^1(\Omega)$ such that $\triangle u = 0$ in Ω, $u = g$ on Γ, and $\|u\|_1 \leq C\|g\|_{1/2,\Gamma}$. As a result, for the div-grad system with inhomogeneous Dirichlet boundary conditions given by

$$-\operatorname{div}\mathbf{v} = f \quad \text{and} \quad \mathbf{v} = \operatorname{grad}\phi \quad \text{in } \Omega \qquad \text{and} \qquad \phi = g \quad \text{on } \Gamma\,,$$

the relevant a priori estimate is given by

$$\|\phi\|_1 + \|\mathbf{v}\|_{H(\Omega,\operatorname{div})} \leq C\left(\|\mathbf{v} - \operatorname{grad}\phi\|_0 + \|\operatorname{div}\mathbf{v}\|_0 + \|\phi\|_{1/2,\Gamma}\right)\,. \tag{3.55}$$

Similar estimates can be derived for the other div-grad–type first-order systems.

Recall that the velocity-vorticity-pressure Stokes problem has an ambiguously defined principal part and that, as a result, there are two possible functional settings for this problem, given by (3.29)–(3.30) and (3.31)–(3.32). When this problem is augmented with inhomogeneous boundary conditions, the data spaces are given by $Y_q \times Z_q$, where Z_q is a trace space defined on Γ. The specific form of Z_q can be determined with the help of the elliptic regularity theory of [2]. More precisely, given a particular boundary operator, the form of Z_q will depend on the principal part, which assures the validity of the complementing condition for this boundary operator. For example, for the velocity-vorticity-pressure formulation of the Stokes equations with the pressure–normal-velocity boundary condition (3.35) this space is given by

$$Z_q = H^{q+1/2}(\Gamma) \times H^{q+1/2}(\Gamma)$$

for $(\mathbf{u}\cdot\mathbf{n}, p)$, whereas for the velocity boundary condition (1.26) we have that

$$Z_q = [H^{q+3/2}(\Gamma)]^n, \quad n = 2 \text{ or } 3\,,$$

for $\mathbf{u}$. As a result, the relevant a priori estimates corresponding to the two principal parts (3.27) and (3.28) are now given by

$$\begin{aligned}&\|\omega\|_{q+1} + \|p\|_{q+1} + \|\mathbf{u}\|_{q+1}\\ &\leq C\left(\|\nu\operatorname{curl}\omega + \operatorname{grad}p\|_q + \|\operatorname{curl}\mathbf{u} - \omega\|_q + \|\operatorname{div}\mathbf{u}\|_q + \|\mathbf{u}\cdot\mathbf{n}\|_{q+1/2,\Gamma} + \|p\|_{q+1/2,\Gamma}\right)\end{aligned} \tag{3.56}$$

and

(3.57)
$$\|\omega\|_{q+1} + \|p\|_{q+1} + \|\mathbf{u}\|_{q+2} \leq C\left(\|\nu\mathrm{curl}\,\omega + \mathrm{grad}\,p\|_q + \|\mathrm{curl}\,\mathbf{u} - \omega\|_{q+1} + \|\mathrm{div}\,\mathbf{u}\|_{q+1} + \|\mathbf{u}\|_{q+3/2,\Gamma}\right),$$

respectively. A priori estimates for other first-order Stokes problems with inhomogeneous boundary conditions can be derived in a similar manner. For example, when the first-order Stokes problem is H^1-coercive, e.g., the velocity–gradient-velocity-pressure formulation, the space Z_q for the inhomogeneous velocity boundary condition is given by $[H^{q+1/2}(\Gamma)]^n$, $n = 2$ or 3. If the system is not H^1-coercive, e.g., the velocity-stress-pressure formulation, then Z_q is given by $[H^{q+3/2}(\Gamma)]^n$, $n = 2$ or 3.

The estimates (3.55), (3.56), and (3.57) can be used to define norm-equivalent least-squares functionals when a variational enforcement of the essential boundary conditions is desired. In particular, these estimates indicate the appropriate norms that should be used to measure the residuals of the boundary data.

In conclusion, we note that the ADN theory also allows us to determine the form of the boundary data space Z_q when the boundary condition involves differential operators. Such boundary conditions for the Stokes problem are, however, outside the scope of this paper.

4. Least-squares methods. We now turn to specific least-squares finite element methods for elliptic boundary value problems. We divide these methods into three classes according to the analytical and computational approaches employed to satisfy the optimality and practicality principles of section 2. Thus, we use the following nomenclature to classify least-squares finite element methods:

- *Basic L^2 methods*: methods that use only L^2-norms in the least-squares functional;
- *Weighted L^2 methods*: methods that use mesh-dependent or other weighted L^2-norms in the least-squares functional; and
- *H^{-1} methods*: methods that use negative Sobolev space norms in the functional setting and in the least-squares functional.

Most of the methods discussed in this section are illustrated using the Stokes equations as a model problem, although other model problems are also used.

4.1. Basic L^2 least-squares methods. The principal appeal of basic L^2 methods is their straightforward formulation and ease of implementation. Given a first-order system, a quadratic least-squares functional for the system is readily available by summing up the L^2-norms of the residual equations. As a result of the use of first-order differential equations in the functionals, the corresponding variational problems can be discretized by means of standard finite element spaces such as P_k or Q_k, i.e., spaces consisting of functions that are merely continuous across element edges. However, for such methods, optimal discretization error estimates in L^2- and H^1-norms can only be established provided the first-order system is fully H^1-coercive. Moreover, if the system lacks the latter property, we can find smooth solutions for which convergence rates are suboptimal. Thus, the main theoretical and practical problems experienced by basic L^2 formulations can be attributed to the lack of such coercivity in the first-order system. These problems will be discussed in subsection 4.1.1 in the context of the Stokes equations. In subsection 4.1.2, fully H^1-coercive basic L^2 methods for div-curl systems are considered.

The need for full H^1-coercivity reflects the desire to establish Sobolev space error estimates that provide control over all first derivatives. In subsection 4.1.3 we

consider basic L^2 methods that adhere to a different philosophy with respect to error estimates. These methods take advantage of the fact that a first-order system, although not fully H^1-coercive, may nonetheless be coercive in a setting which involves "intermediate" function spaces such as $H(\Omega, \mathrm{div})$. Then, optimal error estimates for a basic L^2 method can still be established, but for some of the variables these estimates will be in the weaker norm of the intermediate space and we do not establish full control over all first derivatives. Such methods allow us to exploit, in the analysis, the ellipticity properties that are inherent in the first-order system without adding additional constraints. Finally, in subsection 4.1.3, we also discuss methods for div-grad systems that result in optimal H^1 error estimates by means of adding additional constraints or by employing special grids.

4.1.1. Basic L^2 least-squares methods for the Stokes equations. For the Stokes problem in two dimensions, we can consider the following five least-squares functionals associated with the first-order systems discussed in subsections 3.2–3.3.
Velocity-vorticity-pressure functional:

$$\mathcal{J}(\boldsymbol{\omega}, \mathbf{u}, p) = \frac{1}{2}\big(\|\nu \mathrm{curl}\, \boldsymbol{\omega} + \mathrm{grad}\, p - \mathbf{f}\|_0^2 + \|\mathrm{curl}\, \mathbf{u} - \boldsymbol{\omega}\|_0^2 + \|\mathrm{div}\, \mathbf{u}\|_0^2\big); \tag{4.1}$$

Velocity-pressure-stress functional:

$$\mathcal{J}(\underline{\mathbf{T}}, \mathbf{u}, p) = \frac{1}{2}\big(\|\underline{\mathbf{T}} - \sqrt{2\nu}\, \epsilon(\mathbf{u})\|_0^2 + \|\mathrm{div}\, \mathbf{u}\|_0^2 + \|\sqrt{2\nu}\, \mathrm{div}\, \underline{\mathbf{T}} - \mathrm{grad}\, p - \mathbf{f}\|_0^2\big); \tag{4.2}$$

Constrained-velocity–gradient-pressure functional:

$$\mathcal{J}(\underline{\mathbf{G}}, p) = \frac{1}{2}\big(\| - \nu \mathrm{div}\, \underline{\mathbf{G}} + \mathrm{grad}\, p - \mathbf{f}\|_0^2 + \|\mathrm{curl}\, \underline{\mathbf{G}}\|_0^2\big); \tag{4.3}$$

Velocity–gradient-velocity-pressure functional I*:*

$$\mathcal{J}(\underline{\mathbf{V}}, \mathbf{u}, p) = \frac{1}{2}\big(\| - \nu \mathrm{div}\, \underline{\mathbf{V}} + \mathrm{grad}\, p - \mathbf{f}\|_0^2 + \|\mathrm{div}\, \mathbf{u}\|_0^2 + \|\underline{\mathbf{V}} - (\mathrm{grad}\, \mathbf{u})^t\|_0^2\big); \tag{4.4}$$

Velocity–gradient-velocity-pressure functional II*:*

$$\begin{aligned}\mathcal{J}(\underline{\mathbf{V}}, \mathbf{u}, p) = &\frac{1}{2}\big(\| - (\mathrm{div}\, \underline{\mathbf{V}})^t + \mathrm{grad}\, p - \mathbf{f}\|_0^2 + \|\mathrm{div}\, \mathbf{u}\|_0^2 \\ &+ \|\underline{\mathbf{V}} - \mathrm{grad}\, \mathbf{u}^t\|_0^2 + \|\mathrm{grad}\, (tr\underline{\mathbf{V}})\|_0^2 + \|\mathrm{curl}\, \underline{\mathbf{V}}\|_0^2\big)\,.\end{aligned} \tag{4.5}$$

With each one of the above functionals we associate a least-squares finite element method in the usual manner, as described in section 2. To obtain optimally accurate discretization error estimates for the resulting methods we have to establish the coercivity of respective bilinear forms, i.e., show that the least-squares functionals (4.1)–(4.5) with zero data are norm-equivalent to the H^1-norm. This, however, is not true for all five functionals.

Consider, for example, the functional (4.1). When the first-order system (3.20)–(3.22) is augmented by the normal-velocity–pressure boundary condition (3.35), the corresponding boundary value problem is elliptic in the sense of Petrovsky. As a result, the system (3.20)–(3.22) is fully H^1-coercive, and the relevant a priori estimate is given by (3.33). Analyses based on standard elliptic finite element theory can be used to establish optimal discretization errors for this method (see [17]) of the form

$$\begin{aligned}&\|\mathbf{u} - \mathbf{u}^h\|_r + \|\boldsymbol{\omega} - \boldsymbol{\omega}^h\|_r + \|p - p^h\|_r \\ &\qquad \le C h^{k+1-r}\big(\|\mathbf{u}\|_{k+1} + \|\boldsymbol{\omega}\|_{k+1} + \|p\|_{k+1}\big), \quad r = 0, 1\,.\end{aligned} \tag{4.6}$$

TABLE 1
Rates of convergence of the L^2- and H^1-errors in the least-squares finite element solution with velocity (V) *and normal-velocity–pressure* (NVP) *boundary conditions compared to the best approximation rates* (BA).

	L^2-error rates			H^1-error rates		
Variable	BA	V	NVP	BA	V	NVP
u	3.00	2.71	3.11	2.00	2.03	2.04
v	3.00	2.37	3.10	2.00	2.06	2.02
w	3.00	2.20	3.00	2.00	1.64	1.93
p	3.00	2.34	2.98	2.00	1.64	1.97

This estimate is valid, e.g., if the finite element spaces P_k or Q_k (see section 1) are used for all variables. Let us now suppose that (3.20)–(3.22) is instead augmented by the velocity boundary condition (1.26). The corresponding boundary value problem is elliptic in the more general sense of ADN, but fails to be elliptic in the sense of Petrovsky. As a result, the system (3.20)–(3.22) is not fully H^1-coercive, and the relevant a priori estimate is now given by (3.34). This fact by itself does not immediately imply that the method is not optimal; it only indicates that standard finite element analysis cannot be used to show that the optimally accurate error estimates given by (4.6) are valid with the velocity boundary condition. A more careful analysis of this method does however reveal that it is indeed suboptimal; suboptimal convergence rates can be observed computationally as well. Using the exact solution from subsection 3.2.1 with $n = 1$, the functional (4.1), and discretization by quadratic elements on triangles, we have computationally obtained the (approximate) convergence rates as given in Table 1; we can conclude that the rates for the velocity boundary condition case are suboptimal.

Consider next the functional (4.2). From subsection 3.2.2, we know that the associated boundary value problem is not fully H^1-coercive, regardless of the choice of boundary conditions. Similarly, the first-order system (3.39)–(3.42) is not fully H^1-coercive and estimate (3.43) implies that the functional (4.4) is not norm-equivalent. Thus, in both cases, the optimality of the resulting methods cannot be established using standard elliptic arguments. In fact, in both cases, we can devise counterexamples, similar to the one in subsection 3.2.1, that will result in suboptimal convergence rates.

Lastly, consider the constrained velocity gradient functional (4.3) and the augmented velocity gradient functional (4.5). In both cases, the functionals are associated with fully H^1-coercive systems given by (3.52) and (3.39)–(3.42), (3.44)–(3.46), respectively. The a priori estimates (3.53) and (3.47) that are valid for these systems imply that both functionals are norm-equivalent to a product of H^1 spaces. As a result, optimally accurate discretization error estimates in the $H^1(\Omega)$-norm can be established for all dependent variables using standard elliptic finite element theory; see [36] and [47].

4.1.2. Basic L^2 least-squares methods for div–curl type equations. In two dimensions, the div-curl first-order system (1.29)–(1.30) has two equations and two unknowns and is elliptic in the sense of ADN. Moreover, along with one of the boundary conditions (1.31), this system is fully H^1-coercive, provided Γ is of class $C^{1,1}$ or is piecewise smooth with no re-entrant corners. This follows from the algebraic and topological equivalence of each the spaces $H_0(\Omega, \mathrm{div}) \cap H(\Omega, \mathrm{curl})$ and $H(\Omega, \mathrm{div}) \cap H_0(\Omega, \mathrm{curl})$ with the space $\mathbf{H}^1(\Omega)$ for such domains; see [79]. Thus, in two dimensions,

the quadratic functional

$$\mathcal{J}(\mathbf{u}) = \frac{1}{2}\left(\|\operatorname{curl}\mathbf{u} - \mathbf{f}\|_0^2 + \|\operatorname{div}\mathbf{u} - g\|_0^2\right) \tag{4.7}$$

leads to an optimal least-squares method.

In three dimensions, the system (1.29)–(1.30) consists of four equations and three unknowns, i.e., it is seemingly overdetermined. Thus, along with the velocity-gradient–velocity-pressure system, this system is another example of the utility of least-squares methods for solution of systems having different numbers of unknowns and equations. The functional (4.7) can be used in three dimensions as well; however, analyses of resulting methods can vary significantly depending on the mathematical framework chosen. For example, in [74], existence and uniqueness of minimizers to (4.7) and optimal discretization error estimates are established using the integral identity

$$\int_\Omega \left(|\operatorname{curl}\mathbf{u}|^2 + |\operatorname{div}\mathbf{u}|^2\right) d\Omega = \int_\Omega |\operatorname{grad}\mathbf{u}|^2 .$$

Unfortunately, this identity holds for special cases such as rectangular domains, but is not true in general; see [55].

A more general mathematical framework for the analysis of least-squares methods based on (4.7) has been suggested in [55]. This framework uses the ADN elliptic theory. Since in three dimensions the system (1.29)–(1.30) is not elliptic in the sense of ADN, it must be modified before the ADN theory can be applied. For this purpose, equation (1.29) is modified according to (1.32), where ϕ is a "slack" variable subject to the boundary condition $\phi = 0$ on Γ (compare with subsection 3.2.1). The slack variable ϕ is used only for the analysis; we can show that this variable is identically zero and therefore it can be completely ignored in computations. Thus, using, e.g., P_1 elements, we have the optimal error estimates (see [55])

$$\|\mathbf{u} - \mathbf{u}^h\|_r \le Ch^{2-r}\|\mathbf{u}\|_2, \quad r = 0, 1 .$$

Methods described in this section can be extended to problems arising in electromagnetic applications. For further details on such methods, see, e.g., [46], [58], [60], [97], and [116].

4.1.3. Basic L^2 least-squares methods for div-grad–type equations. We now turn to a class of basic L^2 methods that is principally associated with the scalar div-grad decomposition of subsection 3.1, e.g., methods that arise in biharmonic-, convection-diffusion–, and reaction-diffusion–type problems. We first consider least-squares methods that are customarily set in $H(\Omega, \operatorname{div})$ spaces.

The simplest setting for such methods is given by Poisson's equation (1.5) with the homogeneous Dirichlet boundary condition (1.6). Here, the scalar div-grad decomposition yields the first-order system (2.11) and the corresponding least-squares functional is given by (2.12). From the a priori estimate (3.2), it follows that (2.12) is norm-equivalent on the space X, where X is given by (3.1). As a result, the bilinear form associated with this functional is coercive on $X \times X$. Then, standard elliptic finite element theory can be used to establish error estimates of the form

$$\|\phi - \phi^h\|_1 + \|\mathbf{v} - \mathbf{v}^h\|_{H(\Omega,\operatorname{div})} \le Ch^k\left(\|\phi\|_{k+1} + \|\mathbf{v}\|_{k+1}\right) \tag{4.8}$$

if, e.g., the finite element spaces P_k or Q_k are used for all variables.

Next, consider the biharmonic problem (1.10)–(1.11) and the associated first-order problem (3.4)–(3.8). We recall that a function setting for this first-order problem is given by (3.9) and that the a priori estimate relevant to least-squares methods is given by (3.10). As a result, a least-squares functional for (3.4)–(3.8) can be defined by (see [76])

$$\begin{aligned}&\mathcal{J}(\phi_0,\phi_1,\mathbf{v}_0,\mathbf{v}_1)\\ &\quad=\frac{1}{2}\big(\|\mathbf{v}_0-\phi_0\|_0^2+\|\mathbf{v}_1-\phi_1\|_0^2+\|\operatorname{div}\mathbf{v}_0-\phi_1\|_0^2+\|\operatorname{div}\mathbf{v}_1-f\|_0^2\big)\,.\end{aligned}\tag{4.9}$$

According to (3.10), this functional is norm-equivalent on a space X where X is now given by (3.9). Correspondingly, the bilinear form associated with the functional (4.9) is not coercive on a product of $H^1(\Omega)$ spaces; instead, this form is coercive on $X\times X$. As a result, the optimal error estimates for least-squares finite element approximations of (3.4)–(3.8) are very similar to (4.8), i.e., we can find optimal upper bounds for

$$\|\mathbf{v}_0-\mathbf{v}_0^h\|_{H(\Omega,\operatorname{div})}+\|\mathbf{v}_1-\mathbf{v}_1^h\|_{H(\Omega,\operatorname{div})}+\|\phi_0-\phi_0^h\|_1+\|\phi_1-\phi_1^h\|_1$$

but not for

$$\|\mathbf{v}_0-\mathbf{v}_0^h\|_1+\|\mathbf{v}_1-\mathbf{v}_1^h\|_1+\|\phi_0-\phi_0^h\|_1+\|\phi_1-\phi_1^h\|_1\,.$$

Lastly, consider the convection-diffusion problem (1.12)–(1.13). A least-squares functional for this problem can be defined using the first-order system (3.16)–(3.17) and is given by

$$\mathcal{J}(\mathbf{v},\phi)=\frac{1}{2}\big(\|\mathbf{v}-A(x)\operatorname{grad}\phi\|_0^2+\|\operatorname{div}\mathbf{v}+\Lambda\phi-f\|_0^2\big)\,.\tag{4.10}$$

It can be shown (see [33], [35], and [107]) that the functional (4.10) is equivalent to a norm on the space $X=H_0^1(\Omega)\times H(\Omega,\operatorname{div})$ for the unknown $(\phi,\mathbf{v})$, i.e., the associated bilinear form is coercive on $X\times X$. As a result, it can be shown (see [33]) that least-squares finite element approximations satisfy the error estimate (4.8).

The need to carry on the analyses of the above least-squares methods in a setting involving the space $H(\Omega,\operatorname{div})$ stems from the lack of full H^1-coercivity in div-grad–type systems. However, in most cases such coercivity can be achieved by adding curl constraints. For example, if the system (2.11) is augmented by (3.3), a least-squares functional given by

$$\mathcal{J}(\phi,\mathbf{v})=\frac{1}{2}\big(\|\operatorname{div}\mathbf{v}+f\|_0^2+\|\mathbf{v}-\operatorname{grad}\phi\|_0^2+\|\operatorname{curl}\mathbf{v}\|_0\big)\tag{4.11}$$

allows us to establish optimal error estimates in the norm of $H^1(\Omega)$ for all dependent variables; see [49]. The curl constraint (3.3) can also be used to augment the div-grad system corresponding to the Helmholtz equation; see, e.g., [48] and [95].

The same approach can be applied to the convection-diffusion problem (1.12)–(1.13). In this case, the appropriate form of the curl constraint is given by (3.19), i.e., instead of (4.10) we can consider the minimization of the following functional similar to (4.11) (see [43] and [44]):

$$\mathcal{J}(\phi,\mathbf{v})=\frac{1}{2}\big(\|\mathbf{v}-A(x)\operatorname{grad}\phi\|_0^2+\|\operatorname{div}\mathbf{v}+\Lambda\phi-f\|_0^2+\|\operatorname{curl}A^{-1}(x)\mathbf{v}\|_0^2\big)\,.\tag{4.12}$$

Functional (4.12) can also be applied to the self-adjoint case, i.e., when the term $\Lambda\phi$ is omitted from (1.12)–(1.13); see [106].

In conclusion, it should be mentioned that optimal rates of convergence in $H^1(\Omega)$-norms can still be obtained without adding a curl constraint to the first-order system. This, however, requires a restrictive condition on the triangulation known as the "grid-decomposition property"; see [72]. An example of a grid that satisfies this property is furnished by the "criss-cross" grid; see [73]. Furthermore, it can be shown (see [61]) that the grid-decomposition property is a necessary and sufficient condition for stability and optimal discretization errors.

4.2. Weighted least-squares methods. The use of weights in least-squares functionals has long been among the preferred ways to handle issues such as lack of full H^1-coercivity, inhomogeneous boundary conditions, singular solutions, and computations in regions with corners. Loosely speaking, the idea of weighted least-squares functionals can be described as follows. We are given a least-squares functional which is equivalent to a norm on some Sobolev space, but this norm is not convenient from a computational point of view or is not appropriate for a particular class of solutions. When this functional is restricted to a finite element space, we can appeal to the fact that all norms on a finite-dimensional space are equivalent. Thus, essentially all norms can be replaced by L^2-norms weighted by the respective equivalence constants. Here, we consider three typical examples of weighted least-squares methods. In the first, weights are used to replace H^1-norms by L^2-norms; in the second, weights are used to replace inconvenient boundary norms by more convenient ones; and, in the third, weights are used to handle singularities in the solution.

4.2.1. Weighted methods: Part 1. We consider the first-order system (3.20)–(3.22) along with the boundary condition (1.26). In this case, the a priori estimate relevant to the least-squares methods is given by (3.34). Setting $q = 0$ in (3.34) implies that a norm-equivalent functional should be defined as

$$\mathcal{J}(\boldsymbol{\omega}, \mathbf{u}, p) = \frac{1}{2}\big(\|\nu \operatorname{curl} \boldsymbol{\omega} + \operatorname{grad} p - \mathbf{f}\|_0^2 + \|\operatorname{curl} \mathbf{u} - \boldsymbol{\omega}\|_1^2 + \|\operatorname{div} \mathbf{u}\|_1^2\big)\,. \tag{4.13}$$

Because (4.13) is norm-equivalent on the space (3.31) (with $q = 0$), conforming discretization of this functional yields a formally optimal method. However, the use of H^1-norms in (4.13) calls for discretization of *second-order* terms such as $(\operatorname{grad}\operatorname{div}\mathbf{u})$ and $(\operatorname{grad}\operatorname{curl}\mathbf{u})$. Conforming discretizations of such terms can be handled using subspaces of $H^2(\Omega)$. In the finite element setting, this essentially requires the use of finite element spaces that are continuously differentiable across the element faces. Unfortunately, in two and three dimensions, such elements are impractical, which offsets the potential advantages of a least-squares formulation based on (4.13).

The fact that a first-order system may lead to a least-squares method that still requires discretization by subspaces of H^2 was first pointed out in [17]. In that paper some ideas of [3] were extended to the system (3.20)–(3.22). To avoid this problem, the relevant observation is that for finite element functions the equivalence constant between the L^2- and H^1-norms is given by h^{-1}. This can be seen either by a scale argument or by using inverse inequalities; see [65]. Then, instead of (4.13), we can consider the weighted functional given by

$$\mathcal{J}_h(\boldsymbol{\omega}, \mathbf{u}, p) = \frac{1}{2}\big(\|\nu \operatorname{curl} \boldsymbol{\omega} + \operatorname{grad} p - \mathbf{f}\|_0^2 + h^{-2}\|\operatorname{curl} \mathbf{u} - \boldsymbol{\omega}\|_0^2 + h^{-2}\|\operatorname{div} \mathbf{u}\|_0^2\big)\,. \tag{4.14}$$

The functional (4.14) involves only first-order derivatives of the unknown functions. As a result, the associated variational problem can be discretized using finite element

spaces like P_k or Q_k. The use of weighted L^2-norms does, however, introduce some interesting features into the resulting least-squares method. First, we note that the functionals (4.13) and (4.14) are not equivalent unless the functions $U = (\boldsymbol{\omega}, \mathbf{u}, p)$ are restricted to a finite-dimensional space. Furthermore, since discretization is performed using finite element spaces such as P_k and Q_k, it is not conforming with respect to the spaces for which (4.13) is norm-equivalent (these spaces include $\mathbf{H}^2(\Omega)$ for the velocity field). As a result, the bilinear form associated with (4.14) is not coercive in the usual sense; instead, it can be shown (see [17]) that it satisfies a "stability" estimate of the form

$$\|\boldsymbol{\omega} - \boldsymbol{\omega}^h\|_{q+1} + \|p - p^h\|_{q+1} + \|\mathbf{u} - \mathbf{u}^h\|_{q+2} \le C\, h^{-q}\, \mathcal{B}^h(U - U^h, U - U^h)^{\frac{1}{2}}\,, \tag{4.15}$$

where $q \le -1$. Together with a standard "continuity" estimate, (4.15) yields an error estimate (compare with (4.6))

$$\|\omega - \omega^h\|_0 + \|p - p^h\|_0 + \|\mathbf{u} - \mathbf{u}^h\|_1 \le C\, h^k \big(\|\boldsymbol{\omega}\|_k + \|p\|_k + \|\mathbf{u}\|_{k+1}\big) \tag{4.16}$$

that is valid for $k \ge 2$ if we use, e.g., the finite element spaces P_k or Q_k for the velocity and P_{k-1} and Q_{k-1} for the pressure and vorticity. Note that, in (4.16), the error in the approximation is measured in norms corresponding to (3.34) with $q = -1$. As a result, for the approximation of the pressure and the vorticity we can use finite element spaces with interpolation order of one degree less than that used for the velocity approximation. This also means that (4.16) is not optimal if equal-order interpolation is used for all dependent variables.

An obvious candidate for a similar treatment is the velocity-pressure-stress system (3.36). Recall that this system is not fully H^1-coercive, i.e., the basic L^2-functional (4.2) is not norm-equivalent. As a result, we can find smooth solutions such that the basic L^2-functional (4.2) for (3.36) yields suboptimal convergence rates.

At the same time, using (3.38) with $q = 0$ to define a norm-equivalent least-squares functional will lead to impractical methods. An argument similar to the one used to define the functional (4.14) now leads to the following weighted functional for (3.36):

$$\begin{aligned} \mathcal{J}_h(\underline{\mathbf{T}}, \mathbf{u}, p) = &\frac{1}{2}\big(h^{-2}\|\underline{\mathbf{T}} - \sqrt{2\nu}\,\epsilon(\mathbf{u})\|_0^2 \\ &+ h^{-2}\|\operatorname{div}\mathbf{u}\|_0^2 + \|\sqrt{2\nu}\operatorname{div}\underline{\mathbf{T}} - \operatorname{grad} p - \mathbf{f}\|_0^2\big)\,; \end{aligned} \tag{4.17}$$

see [19]. The resulting finite element method shares many common properties with the one for the velocity-vorticity-pressure system, including optimal error estimates in which the error in the approximations of $\underline{\mathbf{T}}$, $\mathbf{u}$, and p is measured in norms corresponding to (3.38) with $q = -1$, i.e.,

$$\|\underline{\mathbf{T}} - \underline{\mathbf{T}}^h\|_0 + \|p - p^h\|_0 + \|\mathbf{u} - \mathbf{u}^h\|_1 \le C\, h^k \big(\|\underline{\mathbf{T}}\|_k + \|p\|_k + \|\mathbf{u}\|_{k+1}\big), \tag{4.18}$$

which is valid for $k \ge 2$ if we use, e.g., the finite element spaces P_k or Q_k for the velocity and P_{k-1} and Q_{k-1} for the pressure and stress. As with (4.16), the estimate (4.18) is not optimal if equal-order interpolation is used for all dependent variables.

We can also show that the weights in (4.17) are necessary for the optimal convergence rates in (4.18). For example, consider the following exact solution (see [19]):

$$\begin{aligned} u_1 = u_2 &= \sin(\pi x)\sin(\pi y), \\ T_1 = T_2 = T_3 &= \sin(\pi x)\exp(\pi y), \\ p &= \cos(\pi x)\exp(\pi y), \end{aligned}$$

TABLE 2
Rates of convergence of the H^1- and L^2-errors in the least-squares finite element solution with the weights (WLS) *and without the weights* (LS) *compared to best approximation rates* (BA).

	L^2-error rates			H^1-error rates		
Variable	WLS	LS	BA	WLS	LS	BA
u	3.59	1.11	3.00	2.85	1.00	2.00
v	3.13	1.28	3.00	2.77	1.17	2.00
T_1	2.42	1.25	2.00	0.99	0.94	1.00
T_2	2.48	1.14	2.00	1.01	0.99	1.00
T_3	2.34	1.26	2.00	1.05	0.76	1.00
p	2.40	0.94	2.00	1.10	0.92	1.00

and a method based on (4.17) implemented using P_1 elements for $\underline{\mathbf{T}}$ and p, and P_2 elements for the velocity. Numerical estimates of convergence rates for (4.19) with and without the weights are summarized in Table 2.

4.2.2. Weighted methods: Part 2. In this section, we consider least-squares methods where weights are used to replace trace norms on the boundary by computable, weighted-boundary L^2-norms. One example is provided by the least-squares theory for planar elliptic systems of Petrovsky type developed in [115].

We consider first-order systems in the plane. It is assumed that these systems can be cast into the standard form (see subsection 2.1)

$$AU_x + BU_y + CU = F \quad \text{in } \Omega\,, \tag{4.19}$$

where $U = (u_1, \ldots, u_{2n})$, $F = (F_1, \ldots, F_{2n})$, and A, B, and C are $2n \times 2n$ matrices. Along with the system (4.19), we consider a boundary condition of the form

$$RU = G \quad \text{on } \Gamma\,, \tag{4.20}$$

where R is a full-rank $n \times 2n$ matrix. It is further assumed that the boundary value problem (4.19)–(4.20) is elliptic in the sense of Petrovsky, that is, A and B satisfy the algebraic condition of subsection 2.1, which is equivalent to the uniform ellipticity of (4.19), and that (4.20) satisfies the Lopatinskii condition; see [115]. As a result, the following a priori estimate is valid for the problem (4.19)–(4.20):

$$\|U\|_{q+1} \le C\left(\|AU_x + BU_y + CU\|_q + \|RU\|_{q+1/2,\Gamma}\right)\,. \tag{4.21}$$

In the terminology of section 2, the system (4.19)–(4.20) is fully H^1-coercive (set $q = 0$ in (4.21)). Let us assume for the moment that (4.20) can be satisfied exactly by the approximating spaces. Then, an optimal least-squares method for (4.19)–(4.20) can be defined using only L^2-norms in the functional. In some cases, however, it might be advantageous to impose (4.20) in a weak sense by using a least-squares functional of the form

$$\mathcal{J}(U) = \frac{1}{2}\left(\|AU_x + BU_y + CU - F\|_0^2 + \|RU - G\|_{1/2,\Gamma}^2\right)\,. \tag{4.22}$$

In contrast with the functional (4.13), the difficulty now lies with the computability of the boundary norm $\|\cdot\|_{1/2,\Gamma}$. Using a scale argument, we can infer that for finite element functions, an appropriate equivalence constant for such a norm and the $L^2(\Gamma)$-norm is $h^{-1/2}$. The corresponding weighted least-squares functional is then given by

$$\mathcal{J}_h(U) = \frac{1}{2}\left(\|AU_x + BU_y + CU - F\|_0^2 + h^{-1}\|RU - G\|_{0,\Gamma}^2\right)\,. \tag{4.23}$$

The minimization of the functional (4.23) without the h^{-1}-coefficient results in suboptimal convergence rates; minimization of (4.23) yields optimal rates; see [115].

Evidently, the concepts presented above are not limited to elliptic systems of Petrovsky type, and can be extended to other types of first-order systems. One particular example is given by the method of [86] for Poisson's equation (1.5) with inhomogeneous Dirichlet boundary condition $\phi = g$ on Γ. The least-squares functional considered in [86] is given by

$$\mathcal{J}_h(\phi, \mathbf{v}) = \frac{1}{2}\left(\|\operatorname{div}\mathbf{v} + f\|_0^2 + \|\mathbf{v} - \operatorname{grad}\phi\|_0^2 + h^{-1}\|\phi - g\|_{0,\Gamma}^2\right). \tag{4.24}$$

To obtain the functional (4.24), we can first use (3.55) to define a norm-equivalent quadratic functional similar to (4.22)

$$\mathcal{J}(\phi, \mathbf{v}) = \frac{1}{2}\left(\|\operatorname{div}\mathbf{v} + f\|_0^2 + \|\mathbf{v} - \operatorname{grad}\phi\|_0^2 + \|\phi - g\|_{1/2,\Gamma}^2\right)$$

and then replace the boundary norm as in (4.23). The corresponding least-squares method is similar to the basic L^2 methods of subsection 4.1.3 in the sense that the error estimate for the new variable $\mathbf{v}$ is derived in the norm of $H(\Omega, \operatorname{div})$. Further examples of least-squares methods using weighted $L^2(\Gamma)$-norms can be found in [3] and [4].

Another possibility involves the use of mesh-dependent weights for the residuals of both the partial differential equations and the boundary conditions. The resulting methods combine features of the weighted methods in subsection 4.2.1 and this section. More details about such methods can be found in [3].

4.2.3. Weighted methods: Part 3. In this section we consider yet a third application of mesh-dependent weights in least-squares functionals. In the first two applications, the need to consider weights in functionals stemmed from computability and/or practicality considerations. Here, the purpose of the weights will be to handle singularities in the solutions of boundary value problems. These singularities may be caused by, e.g., shock waves in transonic flows or cracks and corners in the computational domain. Here, we consider the latter case; the case of shock waves is considered in subsection 5.2. It is known that least-squares-principle–based methods tend to be more sensitive to singularities in the solutions, and that mesh refinement alone is not a sufficient remedy; see [66]. Thus, the need for weights in the present context is motivated entirely by the salient features of the solution.

For an example, we consider a least-squares approximation to the Dirichlet problem (1.5)–(1.6) in a planar region Ω having a corner with an interior angle $\theta \in [\pi, 2\pi]$; see [75]. In this case, solutions to (1.5)–(1.6) exhibit a corner singularity which behaves as r^α, where r is the distance to the corner and the choice of α depends on θ. To formulate a least-squares method, we consider again the div-grad decomposition (2.11). Because of the presence of singularities in the solution, the functional (2.12) is not appropriate anymore. The main idea of [75] is to consider an alternative least-squares approximation set up in weighted function spaces. In particular, we can consider a weighted analogue $H_\alpha(\Omega, \operatorname{div})$ of the space $H(\Omega, \operatorname{div})$ defined as the closure of $[C^\infty(\Omega)]^2$ with respect to the norm

$$\|\mathbf{v}\|_{H_\alpha(\Omega,\operatorname{div})}^2 = \|r^{\alpha/2}\operatorname{div}\mathbf{v}\|_0^2 + \|\mathbf{v}\|_0^2.$$

Note that the new space is defined using weights which reflect the strength of the singularity. As a result, the appropriate least-squares functional for (1.5)–(1.6), when

the region Ω has a corner, is given by

$$\mathcal{J}_h(\phi, \mathbf{v}) = \frac{1}{2}\big(\|r^{\alpha/2}(\operatorname{div}\mathbf{v} + f)\|_0^2 + \|\mathbf{v} - \operatorname{grad}\phi\|_0^2\big)\,.$$

It has been shown in [75] that if α is bounded from below by $4 - 2\pi/\theta$, then we can derive the error estimates for $\mathbf{v}$ and ϕ in *unweighted* L^2-norms, i.e.,

$$\|\mathbf{v} - \mathbf{v}^h\|_0 \le Ch\|f\|_1$$

and

$$\|\phi - \phi^{\tilde{h}}\|_0 \le C(\tilde{h}^s + h)^2\|f\|_1\,,$$

respectively. In the above estimates h and $\tilde{h}$ denote the grid sizes used in the approximation of $\mathbf{v}$ and ϕ, respectively, and $s = \pi/\theta$. Note that for optimal accuracy, we must take $\tilde{h} = h^{1/s}$, i.e., the grid for the scalar field ϕ must be finer than that for $\mathbf{v}$.

4.3. H^{-1} least-squares methods. As we have seen in subsection 4.1, a basic L^2 least-squares method is optimally accurate provided the first-order system is fully H^1-coercive (or equivalently, elliptic in the sense of Petrovsky in two dimensions). The lack of full H^1-coercivity essentially implies that we cannot use the same norm to measure all residuals of the first-order system. Let us consider again the velocity-vorticity-pressure system (3.20)–(3.22) with the boundary condition (1.26). Recall that setting $q = 0$ in the a priori estimate (3.34) leads to the (impractical) functional (4.13), which has been used to motivate the weighted functional (4.14). If, on the other hand, we choose $q = -1$ in the a priori estimate (3.34), the corresponding norm-equivalent least-squares functional is given by

$$\mathcal{J}(\boldsymbol{\omega}, \mathbf{u}, p)_{-1} = \frac{1}{2}\left(\|\nu\operatorname{curl}\boldsymbol{\omega} + \operatorname{grad}p - \mathbf{f}\|_{-1}^2 + \|\operatorname{curl}\mathbf{u} - \boldsymbol{\omega}\|_0^2 + \|\operatorname{div}\mathbf{u}\|_0^2\right)\,. \tag{4.25}$$

Functionals such as (4.25) form the basis of the *minus-one* or *negative norm* or H^{-1}-*norm* least-squares methods. H^{-1} methods are a relatively recent development suggested in [23] and [25].

Because H^{-1}-norms are not easy to compute, the functional (4.25) is not any more practical than (4.13). Thus, we still have to consider replacement of the negative norm by a discrete, computable equivalent. One simple way to accomplish this is to use a scale argument. In particular, we can consider a weighted functional of the form

$$\mathcal{J}_h(\boldsymbol{\omega}, \mathbf{u}, p) = \frac{1}{2}\big(h^2\|\nu\operatorname{curl}\boldsymbol{\omega} + \operatorname{grad}p - \mathbf{f}\|_0^2 + \|\operatorname{curl}\mathbf{u} - \boldsymbol{\omega}\|_0^2 + \|\operatorname{div}\mathbf{u}\|_0^2\big)\,, \tag{4.26}$$

i.e., we replace $\|\cdot\|_{-1}$ by the weighted L^2-norm $h\|\cdot\|_0$. We note that (4.26) can be obtained from (4.14) simply by scaling the latter functional with the common (and unimportant for the minimization) factor h^2. Thus, computationally, both functionals should have similar properties.

A more sophisticated approach (see [23] and [25]) involves replacement of the H^{-1}-norm by a discrete negative norm, defined using a preconditioner for the Laplace operator. The main idea of this approach is as follows. Consider the following equivalent norm on $H^{-1}(\Omega)$:

$$\|f\|_{-1}^2 = \sup_{\phi \in H_0^1(\Omega)} \frac{(f, \phi)}{|\phi|_1}\,.$$

Now, let $S : H^{-1}(\Omega) \mapsto H_0^1(\Omega)$ denote the solution operator of the Dirichlet problem (1.5)–(1.6). Then, $\|f\|_{-1}^2 = (Sf, f)\ \forall\ f \in H^{-1}(\Omega)$; see [23]. Furthermore, the inner product associated with the norm $\|\cdot\|_{-1}$ can be expressed as $(f, g)_{-1} = (Sf, g) = (f, Sg)$. Lastly, let $S^h : H^{-1}(\Omega) \mapsto X^h$ denote a discrete approximation to S defined using, e.g., a standard Galerkin method. Then, a computable, discrete negative seminorm can be defined as $\|\phi\|_{-1,h}^2 = (S^h\phi, \phi)_0$. However, the cost of computing S^h may still be prohibitive. Thus, S^h is further replaced by a preconditioner B^h that is a symmetric and positive semidefinite operator on L^2 that is spectrally equivalent to S^h in the sense that

$$C_0(S^h\phi, \phi)_0 \le (B^h\phi, \phi)_0 \le C_1(S^h\phi, \phi)_0\,.$$

The main consideration in the choice of B^h is computational cost. The cost of computing $B^h\phi$ must be significantly lower than the cost of computing $S^h\phi$. Once B^h is chosen, a discrete minus-one norm can be defined as follows (see [23]):

$$\|\phi\|_{-1,h} = (\widetilde{S^h}\phi, \phi)_0\,, \qquad \text{where} \qquad \widetilde{S^h} = h^2 I + B^h$$

and where I denotes the identity matrix. The corresponding least-squares functional for the Stokes equations is given by

$$\begin{aligned}\mathcal{J}_{-1,h}(\boldsymbol{\omega}, \mathbf{u}, p) =& \frac{1}{2}\big(\big(\widetilde{\mathcal{S}^h}(\nu\operatorname{curl}\boldsymbol{\omega} + \operatorname{grad} p - \mathbf{f}), \nu\operatorname{curl}\boldsymbol{\omega} + \operatorname{grad} p - \mathbf{f}\big)_0 \\ &+ \|\operatorname{curl}\mathbf{u} - \boldsymbol{\omega}\|_0^2 + \|\operatorname{div}\mathbf{u}\|_0^2\big)\,,\end{aligned} \tag{4.27}$$

where $\widetilde{\mathcal{S}^h}$ is the block-diagonal matrix with diagonal blocks given by $\widetilde{S^h}$. Note that with the trivial choice $B^h \equiv 0$, (4.27) reduces to the weighted functional (4.26). It should be noted that the use of discrete negative norms in (4.27) leads to algebraic problems with dense matrices. As a result, a practical implementation of corresponding finite element methods is necessarily restricted to the use of iterative solvers that do not require matrix assembly.

5. Least-squares methods for nonlinear problems. In this section we briefly study two examples of the application of least-squares principles to the approximate solution of nonlinear problems. First, we consider the Navier–Stokes system of incompressible, viscous flow and then the potential equation of inviscid, irrotational, compressible flow.

5.1. Least-squares methods for the Navier–Stokes equations. A least-squares method for the Stokes equations can be easily extended, at least in principle, to the nonlinear Navier–Stokes equations. Indeed, given a least-squares functional for a first-order Stokes problem, the corresponding functional for the Navier–Stokes equations is readily available by simply including an appropriate form of the nonlinear term into the residual of the momentum equation. From a practical point of view, the resulting methods differ from their Stokes counterparts in two aspects. First, the associated discrete problem now constitutes a nonlinear system of algebraic equations that must be solved in an iterative manner using, e.g., a Newton linearization. Second, solving the discrete system may not be straightforward for high values of the Reynolds number since it is well known that the attraction ball for, e.g., Newton's method, decreases as the Reynolds number increases.

Most existing least-squares methods for the Navier–Stokes equations are based on the velocity-vorticity-pressure form of this problem; see, e.g., [11], [12], [15], [18],

[87], [91], [92], [93], [94], [96], [99], [112], and [113]. Exceptions include [13] and [14], which consider velocity gradient methods, and [92], where a stress-based method is discussed. The differences among various least-squares methods involve the choice of the discretization spaces, the treatment of the nonlinear term, and the method used for solution of the nonlinear discrete equations. For example, the methods of [87], [91], [92], and [94] use basic L^2-functionals, discretization by piecewise-linear finite elements, and the $\mathbf{u} \cdot \operatorname{grad} \mathbf{u}$ form of the nonlinear term. Other authors use instead the $\omega \times \mathbf{u}$ form of the nonlinear term. Solution of the nonlinear discrete equations is by Newton linearization and solution of the linearized equations is by the conjugate gradient method with Jacobi preconditioning. The method of [99] is very similar; however, solution of the linearized problem now involves the conjugate gradient method preconditioned by incomplete Cholesky factorization. The p-version of the finite element method has been used in [96]. The methods of [11] and [18] use weighted least-squares functionals similar to (4.14), where in addition to the mesh-dependent weights h^{-2}, the residual of the momentum equation is weighted by the Reynolds number. To handle large values of the Reynolds number, these methods use Newton linearization combined with continuation with respect to the Reynolds number.

The nonlinearity also considerably complicates the mathematical analysis of corresponding least-squares methods. At present, analyses available are limited to methods based on the velocity-vorticity-pressure (see [11], [18], and [12]) and velocity gradient (see [13] and [14]) forms of the Navier–Stokes equations. In both cases, analyses are based on the abstract approximation theory of [30] or its modifications. Since discussion of these results would require a substantial amount of theoretical and technical background about the theory of [30], it is beyond the scope of this paper. Thus, in what follows we only outline the main idea of the error analysis.

It can be shown that the Euler–Lagrange equation associated with a least-squares functional for the Navier–Stokes equations can be cast into an abstract canonical form given by

$$F(\lambda, U) \equiv U + T \cdot G(\lambda, U) = 0\,, \tag{5.1}$$

where $\lambda = Re$, T corresponds to a least-squares solution operator for the associated Stokes problem, and G is a nonlinear operator. Similarly, the corresponding discrete nonlinear problem can be identified with an abstract equation of the form

$$F^h(\lambda, U^h) \equiv U^h + T^h \cdot G(\lambda, U^h) = 0\,, \tag{5.2}$$

where T^h is a discrete counterpart of T. The importance of this abstract form is signified by the fact that discretization in (5.2) is introduced solely by means of an approximation to the *linear* operator T in (5.1). As a result, under some assumptions, we can show that the error in the nonlinear approximation defined by (5.2) is of the same order as the error in the least-squares solution of the linear Stokes problem.

5.2. Least-squares methods for compressible, potential flow. We now consider another nonlinear example, namely potential flow over a body $\widehat{\Omega}$ contained in a box Ω; see [61] and [64]. The relevant state equations are given by (3.12)–(3.13) in the domain $\Omega/\widehat{\Omega}$ along with (3.15). On the boundary $\partial\widehat{\Omega}$ of the body we consider the boundary condition (3.14), i.e., $\mathbf{u} \cdot \mathbf{n} = 0$, and on the boundary of the flow domain $\partial\Omega$, a condition of the form $\mathbf{u} \cdot \mathbf{n} = g$ is assumed.

In view of (3.12), we also have that

$$\operatorname{curl} \mathbf{u} = 0 \quad \text{in } \Omega/\widehat{\Omega}\,, \tag{5.3}$$

so that together with (3.13) and the boundary conditions, we arrive at a div-curl–type system. The system (3.13) and (5.3) is known to be effective for subsonic flows (smooth density ρ and velocity field $\mathbf{u}$; see [65]), and nonlinear least-squares finite element methods based on the functional

$$\mathcal{J}(\mathbf{u}) = \frac{1}{2}\big(\|\mathrm{div}\,(\rho\mathbf{u})\|_0^2 + \|\mathrm{curl}\,\mathbf{u}\|_0^2\big) \tag{5.4}$$

are well suited for the approximation of such flows.

However, the use of (5.4) for the approximation of transonic flows containing shock waves is problematic. Indeed, although the mass flow $(\rho\mathbf{u}\cdot\mathbf{n})$ and the tangential velocity are continuous across the shock, the terms $\mathrm{div}\,(\rho\mathbf{u})$ and $\mathrm{curl}\,\mathbf{u}$ may not necessarily belong to $L^2(\Omega)$, i.e., a basic L^2-functional would be meaningless for such flows. To remedy this situation, the residual of the mass balance equation can be measured in a weighted L^2-norm given by

$$\|\eta\mathrm{div}\,(\rho\mathbf{u})\|_0^2\,,$$

where the weight function η is subject to the condition

$$\int_{\Omega/\widehat{\Omega}} \eta|\mathrm{div}\,(\rho\mathbf{u})|^2\,d\Omega < \infty\,.$$

This condition essentially means that η vanishes on the shock, i.e., we are dealing with a degenerate L^2-norm. The second term in (5.4) can be dealt with in a similar manner.

In [64] it has been suggested that a least-squares method for transonic flows should be based instead on the *mass-flow-potential* variables, i.e., instead of (5.4), we consider a weighted functional of the form

$$\mathcal{J}(\mathbf{v},\phi) = \frac{1}{2}\Big(\|\eta\mathrm{div}\,\mathbf{v}\|_0^2 + \|\frac{\mathbf{v}}{\rho} - \mathrm{grad}\,\phi\|_0^2\Big)\,. \tag{5.5}$$

In (5.5), the density ρ is a function of $\mathrm{grad}\,\phi$ through (1.15). For an analysis of resulting nonlinear least-squares finite element method, the reader can consult [61].

6. Other least-squares methods. So far the least-squares finite element methods we have studied fit into the framework of section 2. We now examine three types of methods that do not fit into that framework. The first is represented by collocation least-squares methods. Here we consider examples of point and subdomain collocation methods. The second includes a method that combines least-squares ideas with the technique of Lagrange multipliers in order to enhance mass conservation. Lastly, the third method casts the original boundary value problem into the framework of an optimal control or optimization problem with a least-squares functional serving the role of the cost or objective functional.

6.1. Least-squares collocation methods. In this section, we briefly review a class of least-squares methods in which the discretization step is taken prior to the least-squares step. Such methods are commonly known as *least-squares collocation*, *point least-squares, point-matching,* or *overdetermined collocation* methods; see [69]. The main idea is as follows. Consider again the linear boundary value problem (2.1)–(2.2). We assume that an approximate solution is sought in the form

$$U(x) \approx U_N(\mathbf{a}, x)\,,$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_N)$ is a vector of unknown coefficients. Let $R^j_{\mathcal{L}}(\mathbf{a}, x)$, $j = 1, \ldots, K$, and $R^j_{\mathcal{R}}(\mathbf{a}, x)$, $j = 1, \ldots, L$ denote residuals of the equations in (2.1) and (2.2), respectively. To define a least-squares collocation method, we choose a finite set of points $\{x_i\}_{i=1}^{M_1}$ in Ω, and another set of points $\{x_i\}_{i=M_1+1}^{M}$ on Γ. Then, a least-squares functional is defined by summing the weighted squares of the residuals evaluated at the points x_i, as follows:

$$\mathcal{J}(\mathbf{a}) = \sum_{j=1}^{K} \sum_{i=1}^{M_1} \alpha_{ji} \big(R^j_{\mathcal{L}}(\mathbf{a}, x_i)\big)^2 + \sum_{j=1}^{L} \sum_{i=M_1+1}^{M} \beta_{ji} \big(R^j_{\mathcal{R}}(\mathbf{a}, x_i)\big)^2 . \tag{6.1}$$

The weights α_{ji} and β_{ji} may depend on both the particular equation and collocation point. Minimization of (6.1) with respect to the parameters in $\mathbf{a}$ leads to a (usually overdetermined) algebraic system of the form $A\mathbf{a} = \mathbf{b}$, where A is an M by N matrix. Then, a discrete solution is determined by solving the normal equations $A^T A\mathbf{a} = A^T\mathbf{b}$. Methods formulated along these lines have been used for the numerical solution of the Navier–Stokes equations (see [105]) and hyperbolic problems, including the shallow water equations (see [119], [120], [100], and [101]). For numerous other applications of collocation least-squares, see [69].

Evidently, when the number of collocation points M equals the number of degrees of freedom N in $U_N(\mathbf{a}, x)$, the above methods reduce to a standard collocation procedure. Similarly, if $U_N(\mathbf{a}, x)$ is defined using a finite element space and the collocation points and weights correspond to a quadrature rule, then collocation is equivalent to a finite element least-squares method in which integration has been replaced by quadrature. Collocation least-squares methods offer some specific advantages. For example, since only a finite set of points x_i in the domain Ω need be specified, collocation least-squares are attractive for problems posed on irregularly shaped domains; see [100]. On the other hand, since the normal equations tend to become ill conditioned, such methods require additional techniques, like scaling or orthonormalization, in order to obtain a reliable solution; see [69].

Standard collocation, as well as collocation least-squares methods, use point-by-point matching criteria to define the discrete problem. Instead of a set of points we can also consider collocation over a set of subdomains of Ω. In such a case, the discrete problems are obtained by averaging differential equations over each subdomain. Here, for an illustration of this approach, we consider the subdomain Galerkin least-squares method of [56]. Let (2.1)–(2.2) correspond to a planar first-order elliptic boundary value problem of Petrovsky type with $C = 0$, i.e., $\mathcal{L}(U) = AU_x + BU_y$, $\mathcal{R}U = RU$ where R is again a full-rank $n \times 2n$ matrix. To define the subdomain Galerkin least-squares method for (2.1)–(2.2), we consider a finite element space X^h consisting of continuous piecewise-linear functions defined on a regular triangulation $\mathcal{T}_h$ of the domain Ω into triangles Ω_k. These triangles will also serve as collocation subdomains. We let K and N denote the number of triangles and vertices, respectively, in $\mathcal{T}_h$. For simplicity, we shall assume that the finite element functions in X^h satisfy the essential boundary conditions (2.2). Then, a set of discrete equations is formed by averaging separately the components of the differential system (2.1)–(2.2) over each of the triangles $\Omega_k \in \mathcal{T}_h$, as follows:

$$\int_{\Omega_k} \big(\mathcal{L}U^h\big)_j \, d\Omega = \int_{\Omega_k} (f)_j \, d\Omega \quad \text{for} \quad k = 1, \ldots, K \quad \text{and} \quad j = 1, \ldots, 2n\,. \tag{6.2}$$

Once a basis for X^h is chosen, it is not difficult to see that (6.2) is equivalent to a rectangular linear algebraic system of the form $CU = F$, which consists of $2nK$ equa-

tions in approximately $2nN$ unknowns, i.e., there are about twice as many equations as unknowns. The subdomain Galerkin least-squares method of [56] consists per se of forming the matrix C and subsequently solving the above linear system by a discrete least-squares technique. If the data F are sufficiently smooth, we can show (see [56]) that the resulting method is optimal in the sense that

$$\|U - U^h\|_1 \le C_1 h \|F\|_1 \quad \text{and} \quad \|U - U^h\|_0 \le C_0 h^2 \|F\|_1 .$$

We note that the discretization step in (6.2) can also be interpreted as an application of a nonstandard Galerkin method to the system (2.1)–(2.2) in which the test space consists of piecewise-constant test functions with respect to $\mathcal{T}_h$.

Similar subdomain collocation least-squares methods have also been developed for the numerical solution of Maxwell's equations; see [46].

6.2. Restricted least-squares methods. In general, when a least-squares method is used for the numerical solution of incompressible flow problems, computed velocity fields do not exactly satisfy the continuity equation. As a result, least-squares methods conserve mass only in an approximate manner and usually we can show that $\|\operatorname{div} \mathbf{u}^h\|_0 = O(h^r)$, where $r > 0$ depends on the particular finite element space employed. One way to enhance mass conservation involves the use of local mesh-dependent weights along with special weights for the continuity equation. For example, the weighted functional (4.14) can be modified as follows (see [68]):

(6.3)

$$J_K(\boldsymbol{\omega}, p, \mathbf{u}) = \frac{1}{2}\Big(\|\nu \operatorname{curl} \boldsymbol{\omega} + \operatorname{grad} p - \mathbf{f}\|_0^2 + \sum_j^J h_j^{-2} \big(W \|\operatorname{div} \mathbf{u}\|_{0,\Omega_j}^2 + \|\operatorname{curl} \mathbf{u} - \boldsymbol{\omega}\|_{0,\Omega_j}^2 \big) \Big) ,$$

where Ω_j, $j = 1, \ldots, J$, denotes the jth finite element, h_j denotes the diameter of Ω_j, and W is a weight for the continuity equation. Computational results with the corresponding finite element method reported in [68] indicate very good mass conservation properties with a moderate continuity equation weight ($W = 10$). Note that finite element methods based on the functional (6.3) do fit into the framework of section 2.

Another approach, suggested in [59], which does not fit into the framework of section 2, combines least-squares and Lagrange multiplier techniques into a method called *restricted least-squares*. The main idea of this method is to consider the continuity equation as a constraint that is enforced on each finite element via Lagrange multipliers. In this sense, the restricted least-squares method has some similarities with the methods of [5] and [7], where Lagrange multipliers are employed to enforce boundary conditions. To state the method of [59], let $\mathcal{T}_h$ denote a triangulation of Ω with n finite elements, $\mathcal{L}$ denote a first-order Stokes differential operator, and X^h denote a suitable finite element space defined over $\mathcal{T}_h$. The variational problem associated with the restricted least-squares method for the Stokes equations is then given by

seek $U^h \in X^h$ and $\lambda_j \in \mathbb{R}$, $j = 1, \ldots, J$, such that

$$\int_\Omega \mathcal{L}U^h \cdot \mathcal{L}V^h d\Omega + \sum_j^J \Big(\lambda_j \int_{\Omega_j} \operatorname{div} \mathbf{v}^h d\Omega + \mu_j \int_{\Omega_j} \operatorname{div} \mathbf{u}^h d\Omega \Big) = \int_\Omega \mathcal{L}V^h \cdot F \quad \forall\, V^h \in X^h, \quad \mu_j \in \mathbb{R}, \quad j = 1, \ldots, J .$$

Although computational results obtained with the restricted method are very satisfactory, the method also has some shortcomings. The use of Lagrange multipliers leads to a linear algebraic system with a symmetric but indefinite matrix that has a structure very similar to the matrices arising in mixed methods. Likewise, the size of the discrete problem increases by the number of additional constraints. Thus, at present it remains unclear whether the advantages of the restricted method outweigh the problems associated with imposing constraints on the velocity approximation. In particular, the loss of positive definiteness negates the main advantage of the least-squares formalism.

6.3. Least-squares/optimization methods. The main idea of least-squares/optimization methods is to transform the original boundary value problem into an optimal control or optimization problem for which a cost functional is given by a least-squares–type functional. To describe the method consider the following nonlinear Dirichlet problem (see [31]):

$$-\triangle\phi - G(\phi) = 0 \quad \text{in } \Omega \tag{6.4}$$

along with the boundary condition (1.6). Then, an H^{-1} least-squares functional for (6.4) is given by

$$\mathcal{J}(\phi) = \|\triangle\phi + G(\phi)\|^2_{-1}\,, \tag{6.5}$$

where $\|\cdot\|_{-1}$ denotes the negative norm of subsection 2.3. Minimization of (6.5) over $H^1_0(\Omega)$ would lead to a least-squares principle that is similar to the principles of section 2.

The least-squares/optimization approach, however, considers minimization of

$$\mathcal{K}(\phi,\xi) = \|\triangle(\phi-\xi)\|^2_{-1}\,, \tag{6.6}$$

where $\xi \in H^1_0(\Omega)$ is a solution of

$$-\triangle\xi = G(\phi) \quad \text{in } \Omega \quad \text{and} \quad \xi = 0 \quad \text{on } \Gamma\,. \tag{6.7}$$

In the context of optimal control problems, we can identify ϕ with the control vector, ξ with the state variable, (6.7) with the state equation, and (6.6) with the cost functional. Furthermore, using the identity

$$\|\triangle\phi\|_{-1} = \|\operatorname{grad}\phi\|_0 \quad \forall\ \phi \in H^1_0(\Omega)\,,$$

we can replace (6.6) with the more easily computable (and therefore practical) cost functional

$$\mathcal{K}(\phi,\xi) = \|\operatorname{grad}(\phi-\xi)\|^2_0\,. \tag{6.8}$$

To summarize, the least-squares/optimization method for (6.4) can be stated as follows:

minimize $\mathcal{K}(\phi,\xi)$ given by (6.8) over $\phi \in H^1_0(\Omega)$, subject to the state equation (6.7).

To solve the above optimization problem we can use an abstract version of the conjugate gradient method; see [31]. At each iteration, this method would require solution of two Dirichlet problems (6.7) for the computation of the descent direction; see [31].

This class of methods has been developed for nonlinear flow problems, including compressible flows (see [31], [32], and [81]) and the Navier–Stokes equations (see [31] and [80]). For example, to derive the least-squares/optimization method for the Navier–Stokes equations (1.25)–(1.28), let

$$\mathbf{Z} = \{\mathbf{u} \in \mathbf{H}_0^1(\Omega) \mid \operatorname{div} \mathbf{u} = 0 \text{ in } \Omega\}$$

and

$$\mathcal{K}(\mathbf{u}, \boldsymbol{\xi}) = \frac{\nu}{2} \|\triangle(\boldsymbol{\xi} - \mathbf{u})\|_{-1}^2 = \frac{\nu}{2} \int_\Omega |\operatorname{grad}(\boldsymbol{\xi} - \mathbf{u})|^2 dx, \tag{6.9}$$

and consider the Stokes problem

$$\begin{aligned} -\nu \triangle \boldsymbol{\xi} + \operatorname{grad} q &= -\mathbf{u} \cdot \operatorname{grad} \mathbf{u} \qquad \text{in } \Omega, \\ \operatorname{div} \boldsymbol{\xi} &= 0 \qquad \text{in } \Omega, \\ \boldsymbol{\xi} &= \mathbf{0} \qquad \text{on } \Gamma. \end{aligned} \tag{6.10}$$

Then, the least-squares/optimization method for (1.25)–(1.28) is given by

minimize $\mathcal{K}(\mathbf{u}, \boldsymbol{\xi})$ given by (6.9) over $\mathbf{u} \in \mathbf{Z}$, subject to the state equation (6.10).

To solve the above optimal control problem, we can again use an abstract conjugate gradient process. Now, computation of the descent direction at each iteration involves the solution of several Stokes problems; see [31] and [80].

7. Concluding remarks. In this paper, we have focused primarily on least-squares methods for elliptic boundary value problems. Among the various applications of least-squares principles, these methods appear currently to be at the most advanced theoretical and practical stage. The mathematical framework for the analysis of such methods is well developed, and their computational performance is well documented in the literature. The limited space did not allow us to consider many other important areas, such as hyperbolic problems, time-dependent problems, and time-space least-squares. For further details on such applications, we refer interested readers to [4], [9], [39], [40], [41], [62], [63], [88], [89], [110], [111], [112], and [113], among others.

REFERENCES

[1] R. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.

[2] S. Agmon, A. Douglis, and L. Nirenberg, *Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions* II, Comm. Pure Appl. Math., 17 (1964), pp. 35–92.

[3] A. Aziz, R. Kellog, and A. Stephens, *Least-squares methods for elliptic systems*, Math. Comput., 10, (1985) pp. 53–70.

[4] A.K. Aziz and J.-L. Liu, *A weighted least squares method for the backward-forward heat equation*, SIAM J. Numer. Anal., 28 (1991), pp. 156–167.

[5] I. Babuska, *The finite element method with Lagrange multipliers*, Numer. Math., 20 (1973), pp. 179–192.

[6] G. Baker, *Simplified proofs or error estimates for the least-squares method for Dirichlet's problem*, Math. Comput., 27, (1973), pp. 229–235.

[7] H. Barbosa and T. Hughes, *The finite element method with Lagrange multipliers on the boundary: Circumventing the Babuska-Brezzi condition*, Comput. Methods Appl. Mech. Engrg., 85 (1991), pp. 109–128.

[8] M. Behr, L. Franca, and T. Hughes, *Stabilized finite element methods for the velocity-pressure-stress formulation of incompressible flows*, Comput. Methods Appl. Mech. Engrg., 104 (1993), pp. 31–48.

[9] B. Bell and K. Surana, *A space time coupled p-version least-squares finite element formulation for unsteady fluid dynamics problems*, Internat. J. Numer. Methods Engrg., 37 (1994), pp. 3545–3569.

[10] M. Bercovier, *Perturbation of mixed variational problems. Application to mixed finite element methods*, RAIRO Anal. Numer., 12 (1978), pp. 211–236.

[11] P. Bochev, *Least-squares finite element methods for the Stokes and Navier-Stokes equations*, Ph.D. thesis, Virginia Tech, Blacksburg, VA, 1994.

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

[13] P. Bochev, Z. Cai, T. Manteuffel, and S. McCormick, *First order least squares for the Navier-Stokes equations*, Proc. Seventh Copper Mountain Multigrid Conference, NASA Conference Publication 3339, Part 1 (1996), pp. 41–55.

[14] P. Bochev, Z. Cai, T.A. Manteuffel, and S.F. McCormick, *Analysis of velocity-flux first-order system least-squares principles for the Navier-Stokes equations: Part* I, SIAM J. Numer. Anal., 35 (1998), pp. 990–1009.

[15] P. Bochev and M. Gunzburger, *A least-squares finite element method for the Navier-Stokes equations*, Appl. Math. Lett., 6 (1993), pp. 27–30.

[16] P. Bochev and M. Gunzburger, *Accuracy of least-squares methods for the Navier-Stokes equations,* Comput. & Fluids, 22 (1993), pp. 549–563

[17] P. Bochev and M. Gunzburger, *Analysis of least-squares finite element methods for the Stokes equations,* Math. Comput., 63 (1994), pp. 479–506.

[18] P. Bochev and M. Gunzburger, *Analysis of weighted least-squares finite element method for the Navier-Stokes equations*, Proc. 14th IMACS World Congress, Georgia Tech, Atlanta, GA, 1994, pp. 584–587.

[19] P. Bochev and M. Gunzburger, *Least-squares for the velocity-pressure-stress formulation of the Stokes equations*, Comput. Methods Appl. Mech. Engrg., 126 (1995), pp. 267–287.

[20] J.M. Boland and R.A. Nicolaides, *Stability of finite elements under divergence constraints*, SIAM J. Numer. Anal., 20 (1983), pp. 722–731.

[21] J. Boland and R. Nicolaides, *On the stability of bilinear-constant velocity pressure finite elements*, Numer. Math., 44 (1984), pp. 219–222.

[22] J.M. Boland and R.A. Nicolaides, *Stable and semistable low order finite elements for viscous flows*, SIAM J. Numer. Anal., 22 (1985), pp. 474–492.

[23] J. Bramble, R. Lazarov, and J. Pasciak, *A Least Squares Approach Based on a Discrete Minus One Inner Product for First Order Systems*, Technical report 94-32, Mathematical Science Institute, Cornell University, Ithaca, NY, 1994.

[24] J.H. Bramble and J.A. Nitsche, *A generalized Ritz-least-squares method for Dirichlet problems*, SIAM J. Numer. Anal., 10 (1973), pp. 81–93.

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

[26] J. Bramble and A. Schatz, *Least-squares methods for* $2m$*th order elliptic boundary value problems*, Math. Comput., 25 (1971), pp. 1–32.

[27] F. Brezzi, *On the existence, uniqueness, and approximation of saddle-point problems arising from Lagrange multipliers*, RAIRO Anal. Numer., 8 (1974), pp. 129–151.

[28] F. Brezzi and J. Douglas, *Stabilized mixed methods for the Stokes problem*, Numer. Math., 53 (1988), pp. 225–235.

[29] F. Brezzi and J. Pitkaranta, *On the stabilization of finite element approximations of the Stokes problem*, in Efficient Solutions of Elliptic Systems, W. Hackbush, ed., Vieweg, Braunschweig, 1984, pp. 11–19.

[30] F. Brezzi, J. Rappaz, and P. Raviart, *Finite-dimensional approximation of nonlinear problems, Part* I*: Branches of nonsingular solutions*, Numer. Math., 36 (1980), pp. 1–25.

[31] M. Bristeau, O. Pironneau, R. Glowinski, J. Periaux, P. Perrier, and G. Poirier, *Application of optimal control and finite element methods to the calculation of transonic flows*, in Numerical Methods in Applied Fluid Dynamics, B. Hunt, ed., Academic Press, New York, 1980, pp. 203–312.

[32] M. Bristeau, O. Pironneau, R. Glowinski, J. Periaux, P. Perrier, and G. Poirier, *Finite element methods for transonic flow calculations*, in Recent Advances in Numerical Methods in Fluids, Vol. 4, W. G. Habashi, ed., Pineridge Press, Swansea, 1985, pp. 703–731.

[33] Z. Cai, R. Lazarov, T.A. Manteuffel, and S.F. McCormick, *First-order system least-squares for second-order partial differential equations: Part* 1, SIAM J. Numer. Anal., 31 (1994), pp. 1785–1799.

[34] Z. Cai, T. Manteuffel, and S. McCormick, *First-order system least-squares for velocity-vorticity-pressure form of the Stokes equations, with application to linear elasticity*, ETNA, 3 (1995), pp. 150–159.

[35] 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.

[36] 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.

[37] Z. Cai, T.A. Manteuffel, S.F. McCormick, and S.V. Parter, *First-order system least squares (FOSLS) for planar linear elasticity: Pure traction problem*, SIAM J. Numer. Anal., 35 (1998), pp. 320–335.

[38] Y. Cao and M.D. Gunzburger, *Least-square finite element approximations to solutions of interface problems*, SIAM J. Numer. Anal., 35 (1998), pp. 393–405.

[39] G. Carey and B. Jiang, *Least-squares finite element method and preconditioned conjugate gradient solution*, Int. J. Num. Meth. Engng., 24 (1987), pp. 1283–1296.

[40] G. Carey and B. Jiang, *Least-squares finite elements for first-order hyperbolic systems*, Int. J. Numer. Methods Engrg., 26 (1988), pp. 81–93.

[41] G. Carey, B. Jiang, and R. Showalter, *A regularization-stabilization technique for nonlinear conservation equations computations*, Numer. Methods Partial Differential Equations, 4 (1988), pp. 165–171.

[42] G. Carey, A. Pehlivanov, and Y. Shen, *Least-squares mixed finite elements*, in Finite Element Methods—Proc. Conf. Fifty Years of the Courant Element, M. Kryzek, ed., Jyväskylä, Findland, 1993, pp. 105–107.

[43] G.F. Carey, A.I. Pehlivanov, and P.S. Vassilevski, *Least-squares mixed finite element methods for non-selfadjoint problems:* II. *Performance of block-ILU factorization methods*, SIAM J. Sci. Comput., 16 (1995), pp. 1126–1136.

[44] G. Carey, A. Pehlivanov, and P. Vassilevski, *Least-squares mixed finite element methods for non-selfadjoint elliptic problems:* I. *Error estimates*, Numer. Math., 72 (1996), pp. 501–522.

[45] G. Carey and Y. Shen, *Convergence studies of least-squares finite elements for first order systems*, Comm. Appl. Numer. Methods, 5 (1989), pp. 427–434.

[46] C. Chang, *Finite element method for the solution of Maxwell's equations in multiple media*, Appl. Math. Comput., 25 (1988), pp. 89–99.

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

[48] C. Chang, *A least squares finite element method for the Helmholtz equation*, Comput. Methods Appl. Mech. Engrg., 83 (1990), pp. 1–7.

[49] C.L. Chang, *Finite element approximation for grad-div type systems in the plane*, SIAM J. Numer. Anal., 29 (1992), pp. 452–461.

[50] C. Chang, *An error estimate of the least squares finite element method for the Stokes problem in three dimensions*, Math. Comput., 63 (1994), pp. 41–50.

[51] C. Chang, *An Error Analysis of Least Squares Finite Element Method of Velocity-Pressure-Vorticity Formulation for Stokes Problem: Correction*, Mathematics Research report 95-53, Department of Mathematics, Cleveland State University, Cleveland, OH, 1995.

[52] C. Chang, *Least squares finite element for second order boundary value problems*, Appl. Math. Comput., to appear.

[53] C. Chang, *A least squares finite element method for incompressible flow in stress-velocity-pressure version*, Comput. Methods Appl. Mech. Engrg., to appear.

[54] C. Chang, *Piecewise linear approximations to the Stokes problem with velocity boundary conditions*, submitted.

[55] C. Chang and M. Gunzburger, *A finite element method for first order elliptic systems in three dimensions*, Appl. Math. Comput., 23 (1987), pp. 171–184.

[56] C.-L. Chang and M.D. Gunzburger, *A subdomain-Galerkin/least squares method for first-order elliptic systems in the plane*, SIAM J. Numer. Anal., 27 (1990), pp. 1197–1211.

[57] C. Chang and B. Jiang, *An error analysis of least squares finite element method of velocity-pressure-vorticity formulation for Stokes problem*, Comput. Methods Appl. Mech. Engrg., 84 (1990), pp. 247–255.

[58] C. Chang, J. Li, X. Xiang, Y. Yu, and W. Ni, *Least squares finite element method for electromagnetic fields in* 2-*D*, Appl. Math. Comput., 58 (1993), pp. 143–167.

[59] C.L. Chang and J.J. Nelson, *Least-squares finite element method for the Stokes problem with zero residual of mass conservation*, SIAM J. Numer. Anal., 34 (1997), pp. 480–489.

[60] C. Chang, X. Xiang, and J. Li, *An analysis of the eddy current problem by the least squares finite element in* 2-*D*, Appl. Math. Comput., 60 (1994), pp. 179–191.

[61] T. Chen, *On least-squares approximations to compressible flow problems*, Numer. Methods Partial Differential Equations, 2 (1986), pp. 207–228.

[62] T. Chen, *Semidiscrete least-squares methods for linear convection-diffusion problem*, Comput. Math. Appl., 24 (1992), pp. 29–44.
[63] T. Chen, *Semidiscrete least-squares methods for linear hyperbolic systems,* Numer. Methods Partial Differential Equations, 8 (1992), pp. 423–442.
[64] T. Chen and G. Fix, *Least Squares Finite Element Simulation of Transonic Flows*, ICASE report 86-27, NASA Langley Research Center, Hampton, VA, 1986.
[65] P. Ciarlet, *Finite Element Method for Elliptic Problems*, North–Holland, Amsterdam, 1978.
[66] C. Cox and G. Fix, *On the accuracy of least-squares methods in the presence of corner singularities*, Comput. Math. Appl., 10 (1984), pp. 463–476.
[67] C. Cox, G. Fix, and M. Gunzburger, *A least-squares finite element scheme for transonic flow around harmonically oscillating wings*, J. Comput. Phys., 51 (1983), pp. 387–403.
[68] J. Deang and M. Gunzburger, *Issues related to least squares finite element methods for the Stokes equations*, SIAM J. Sci. Comput., 20 (1998), pp. 878–906.
[69] E. Eason, *A review of least-squares methods for solving partial differential equations*, Internat. J. Numer. Methods Engrg., 10 (1976), pp. 1021–1046
[70] R. Falk, *An analysis of the penalty method and extrapolation for the stationary Stokes equations*, in Advances in Computer Methods for Partial Differential Equations, R. Vichnevetsky, ed., AICA, 1975, pp. 66–69.
[71] G. Fix and M. Gunzburger, *On least squares approximations to indefinite problems of the mixed type*, Internat. J. Numer. Methods Engng., 12 (1978), pp. 453–469.
[72] G. Fix, M. Gunzburger, and R. Nicolaides, *On finite element methods of the least-squares type*, Comput. Math. Appl., 5 (1979), pp. 87–98.
[73] G. Fix, M. Gunzburger and R. Nicolaides, *On mixed finite element methods for first-order elliptic systems*, Numer. Math., 37 (1981), pp. 29–48.
[74] G.J. Fix and M.E. Rose, *A comparative study of finite element and finite difference methods for Cauchy–Riemann type equations,* SIAM J. Numer. Anal., 22 (1985), pp. 250–261.
[75] G. Fix and E. Stephan, *Finite Element Methods of Least-Squares Type for Regions with Corners*, ICASE report 81-41, NASA Langley Research Center, Hampton, VA, 1981.
[76] G. Fix and E. Stephan, *On the finite element least squares approximation to higher order elliptic systems,* Arch. Rational Mech. Anal., 91 (1986), pp. 137–151.
[77] L. Franca, S. Frey, and T. Hughes, *Stabilized finite element methods:* I. *Application to the advective-diffusive model*, Comput. Methods Appl. Mech. Engrg., 95 (1992), pp. 253–276.
[78] L.P. Franca and R. Steberg, *Error analysis of some Galerkin least squares methods for the elasticity equations*, SIAM J. Numer. Anal., 28 (1991), pp. 1680–1697.
[79] V. Girault and P. Raviart, *Finite Element Methods for Navier-Stokes Equations*, Springer-Verlag, Berlin, 1986.
[80] R. Glowinski, B. Mantel, J. Periaux, P. Perrier, and O. Pironneau, *On an efficient new preconditioned conjugate gradient method. Application to the in-core solution of the Navier-Stokes equations via non-linear least-squares and finite element methods*, in Finite Elements in Fluids, Volume 4, R. H. Gallagher, ed., John Wiley, New York, 1982, pp. 365–401.
[81] R. Glowinski, J. Periaux, and O. Pironneau, *Transonic flow simulation by the finite element method via optimal control*, in Finite Elements in Fluids, Volume 3, R. H. Gallagher, ed., John Wiley, New York, 1978, pp. 205–217.
[82] M. Gunzburger, *Finite Element Methods for Viscous Incompressible Flows*, Academic Press, Boston, 1989.
[83] C. Hirsch, *Numerical Computation of Internal and External Flows*, John Wiley, Chichester, 1990.
[84] T. Hughes, L. Franca, and M. Ballestra, *A new finite element formulation for computational fluid dynamics:* V. *Circumventing the Babuska-Brezzi condition: A stable Petrov-Galerkin formulation of the Stokes problem accommodating equal-order interpolations*, Comput. Methods Appl. Mech. Engrg., 59 (1986), pp. 85–99.
[85] T. Hughes, W. Liu, and A. Brooks, *Finite element analysis of incompressible viscous flows by the penalty function formulation*, J. Comput. Phys., 30 (1979), pp. 1–60.
[86] D. Jespersen, *A least-squares decomposition method for solving elliptic equations*, Math. Comput., 31 (1977), pp. 873–880.
[87] B. Jiang, *A least-squares finite element method for incompressible Navier-Stokes problems*, Internat. J. Numer. Methods Fluids, 14 (1992), pp. 843–859.
[88] B. Jiang and G. Carey, *Adaptive refinement for least-squares finite elements with element by element conjugate gradient solution*, Internat. J. Numer. Methods Engrg., 24 (1987), pp. 569–580.

[89] B. Jiang and G. Carey, *A stable least-squares finite element method for nonlinear hyperbolic problems*, Internat. J. Numer. Methods Fluids, 8 (1988), pp. 933–942.
[90] B. Jiang and C. Chang, *Least-squares finite elements for the Stokes problem*, Comput. Methods Appl. Mech. Engrg., 78 (1990), pp. 297–311.
[91] B. Jiang, T. Lin, and L. Povinelli, *A Least-Squares Finite Element Method for* 3*D Incompressible Navier-Stokes Equations*, AIAA report 93-0338, AIAA, New York, 1993.
[92] B. Jiang, T. Lin, and L. Povinelli, *Large-scale computation of incompressible viscous flow by least-squares finite element method*, Comput. Methods Appl. Mech. Engrg., 114 (1994), pp. 213–231.
[93] B. Jiang and L. Povinelli, *Theoretical Study of the Incompressible Navier-Stokes Equations by the Least Squares Method*, NASA TM 106535, ICOMP-94-04, NASA, Cleveland, OH, 1994.
[94] B. Jiang and L. Povinelli, *Least-squares finite element method for fluid dynamics*, Comput. Methods Appl. Mech. Engrg., 81 (1990), pp. 13–37.
[95] B. Jiang and L. Povinelli, *Optimal least-squares finite element methods for elliptic problems*, Comput. Methods Appl. Mech. Engrg., 102 (1993), pp. 199–212.
[96] B. Jiang and V. Sonad, *Least-Squares Solution of Incompressible Navier-Stokes Equations with the p-version of Finite Elements*, NASA TM 105203 (ICOMP report 91-14), NASA, Cleveland, OH, 1991.
[97] B. Jiang, J. Wu, and L. Povinelli, *The Origin of Spurious Solutions in Computational Electromagnetics*, NASA Technical memorandum TM-106921 ICOMP-95-8, NASA, Cleveland, OH, 1994.
[98] S. Krein and Y. Petunin, *Scales of Banach spaces*, Russian Math. Surveys, 21 (1966), pp. 85–160.
[99] D. Lefebvre, J. Peraire, and K. Morgan, *Least-squares finite element solution of compressible and incompressible flows*, Internat. J. Numer. Methods Heat Fluid Flow, 2 (1992), pp. 99–113.
[100] J. Liable and G. Pinder, *Least-squares collocation solution of differential equations on irregularly shaped domains using orthogonal meshes*, Numer. Methods Partial Differential Equations, 5 (1989), pp. 347–361.
[101] J. Liable and G. Pinder, *Solution of shallow water equations by least squares collocation*, Water Resources Research, 29 (1993), pp. 445–456.
[102] J. Lions and E. Magenes, *Nonhomogeneous Elliptic Boundary Value Problems and Applications*, Vol. I, Springer-Verlag, Berlin, 1972.
[103] P. Lynn and S. Arya, *Use of least-squares criterion in the finite element formulation*, Internat. J. Numer. Methods Engrg., 6 (1973), pp. 75–88.
[104] T. Manteuffel, S. McCormick, and G. Starke, *First-order system least-squares for second-order elliptic problems with discontinuous coefficients*, Proc. 7th Copper Mountain Conference on Multigrid Methods, NASA Pubs., Hampton, 1995.
[105] J. Milthorpe and G. Steven, *On a least squares approach to the integration of the Navier-Stokes equations*, in Finite Elements in Fluids, Volume 3, R. H. Gallagher, ed., John Wiley, New York, 1978, pp. 89–103.
[106] A. Pehlivanov and G. Carey, *Error estimates for least-squares mixed finite elements*, Math. Mod. Numer. Anal., 28 (1994), pp. 499–516
[107] A.I. Pehlivanov, G.F. Carey, and R.D. Lazarov, *Least-squares mixed finite element methods for second-order elliptic problems*, SIAM J. Numer. Anal., 31 (1994), pp. 1368–1377.
[108] J. Reddy and J. Oden, *Mixed finite element approximations of linear boundary value problems*, Quart. Appl. Math., 33 (1975), pp. 255–280.
[109] Y. Roitberg, *A theorem about the complete set of isomorphisms for systems elliptic in the sense of Douglis and Nirenberg*, Ukrain. Mat. Zh., 27 (1975), pp. 447–450.
[110] L. Tang, T. Cheng, and T. Tsang, *Transient solutions for three-dimensional lid-driven cavity flows by a least-squares finite element method*, Internat. J. Numer. Methods Fluids, 21 (1995), pp. 413–432
[111] L. Tang and T. Tsang, *Transient solutions by a least-squares finite element method and Jacobi conjugate gradient technique*, Numer. Heat Trans. Part B, 28 (1995), pp. 183–198.
[112] L. Tang and T. Tsang, *A least-squares finite element method for time dependent incompressible flows with thermal convection*, Internat. J. Numer. Methods Fluids, 17 (1993), pp. 271–289.
[113] L. Tang and T. Tsang, *Temporal, spatial and thermal features of* 3*-D Rayleigh-Benard convection by least-squares finite element method*, Comput. Methods Appl. Mech. Engrg., 140 (1997), pp. 201–219.
[114] R. Temam, *Navier-Stokes Equations*, North–Holland, Amsterdam, 1979.

[115] W. Wedland, *Elliptic Systems in the Plane*, Pitman, London, 1979.
[116] J. Wu and B. Jiang, *A Least-Squares Finite Element Method for Electromagnetic Scattering Problems*, Report ICOMP-96-12, NASA Lewis Research Center, Cleveland, OH, 1996.
[117] J. Wu, S. Yu, and B. Jiang, *Simulation of Two-Fluid Flows by the Least-Squares Finite Element Method Using a Continuum Surface Tension Model*, Report ICOMP-96-13, NASA Lewis Research Center, Cleveland, OH, 1996.
[118] S. Yu, B. Jiang, J. Wu, and N. Liu, *A div-curl-grad formulation for compressible buoyant flows solved by the least-squares finite element method*, Comput. Methods Appl. Mech. Engrg., 137 (1996), pp. 59–88
[119] D. Zeitoun, J. Liable, and G. Pinder, *A weighted least-squares method for first order hyperbolic systems*, Internat. J. Numer. Methods Fluids, 20 (1995), pp. 191–212.
[120] D. Zeitoun and G. Pinder, *A least squares approach for solving remediation problems of contaminated aquifers*, in Computational Methods in Water Resources, Vol. 1, Computational Mechanics Publications, Elsevier Applied Science, New York, 1992, pp. 329–336.