

# LEAST-SQUARES FINITE ELEMENT METHODS FOR OPTIMALITY SYSTEMS ARISING IN OPTIMIZATION AND CONTROL PROBLEMS

PAVEL BOCHEV\* AND MAX D. GUNZBURGER†

**Abstract.** The approximate solution of optimization and optimal control problems for systems governed by linear, elliptic partial differential equations is considered. Such problems are most often solved using methods based on applying the Lagrange multiplier rule to obtain an optimality system consisting of the state system, an adjoint-state system, and optimality conditions. Galerkin methods applied to this system result in indefinite matrix problems. Here, we consider using modern least-squares finite element methods for the solution of the optimality systems. The matrix equations resulting from this approach are symmetric and positive definite, and are readily amenable to uncoupling strategies. This is an important advantage of least-squares principles as they allows for a more efficient computational solution of the optimization problem. We develop an abstract theory that includes optimal error estimates for least-squares finite element methods applied to optimality systems. We then provide an application of the theory to optimization problems for the Stokes equations.

**Key words.** Optimal control, optimization, least-squares finite element methods, optimality systems, Lagrange multipliers

**AMS subject classifications.** 65N30, 65N22, 49J20, 49K20

**1. Introduction.** Optimization and control problems for systems governed by partial differential equations arise in many applications. Experimental studies of such problems go back 100 years [22] and computational approaches have been applied since the advent of the computer age. Most of the efforts in the latter direction have employed elementary optimization strategies but, more recently, there has been considerable practical and theoretical interest in the application of sophisticated local and global optimization strategies, e.g., Lagrange multiplier methods, sensitivity or adjoint-based gradient methods, quasi-Newton methods, evolutionary algorithms, etc.

The optimal control or optimization problems we consider consist of

- *state variables*, i.e., variables that describe the system being modeled;
- *control variables* or *design parameters*, i.e., variables at our disposal that can be used to affect the state variables;
- a *state system*, i.e., partial differential equations relating the state and control variables; and
- a *functional* of the state and control variables whose minimization is the goal.

Then, the problems we consider consist of finding state and control variables that minimize the given functional subject to the state system being satisfied. Here, we restrict attention to linear, elliptic state systems and to quadratic functionals.

The Lagrange multiplier rule is a standard approach for solving finite-dimensional, constrained optimization problems. It is not surprising then that several popular approaches to solving optimization and control problems constrained by partial dif-

---

\*Computational Mathematics and Algorithms Department, Sandia National Laboratories, Albuquerque NM 87185-1110 ([pbochoe@sandia.gov](mailto:pbochoe@sandia.gov)). Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

†School of Computational Science and Information Technology, Florida State University, Tallahassee FL 32306-4120 ([gunzburg@csit.fsu.edu](mailto:gunzburg@csit.fsu.edu)). Supported in part by CSRI, Sandia National Laboratories under contract 18407.

ferential equations are also based on solving optimality systems deduced from the application of the Lagrange multiplier rule. The optimality system consists of:

- the *state system*, i.e., the given partial differential equations that relate the unknown state and control variables;
- an *adjoint* or *co-state system* which are also partial differential equations involving the adjoint operator of the state system; and
- an *optimality condition* that reflects the fact that the gradient of the functional vanishes for optimal values of the state and control variables.

The three components of the optimality system are coupled. In the linear constraints/quadratic functional context we consider in this paper, the optimality system, viewed as a coupled system, is a symmetric and weakly coercive linear system in the state, adjoint-state, and control variables.

In the context of finite element methods, optimality systems are usually discretized using Galerkin methods, resulting in typical saddle-point type matrix problems that are symmetric and indefinite. In many if not most practical situations, the coupled optimality system is a formidable system to solve; compared to solving direct problems involving the state system alone, discrete optimality systems typically involve at least double the number of unknowns. For this reason, many approaches have been proposed for decoupling, through iterative processes, the different components of the optimality system. An extensive discussion of several such strategies in both an abstract setting and for fluid flow problems can be found in [18].

In this paper, we discuss the use of modern least-squares finite element methods for finding approximate solutions of the optimality system. The resulting matrix problems are symmetric and *positive definite*. Moreover, their diagonal blocks are also symmetric and positive definite, thus opening up better possibilities for devising efficient uncoupling methods than is the case for Galerkin discretizations. In order to develop a basic theory for least-squares finite element methods for optimization and control problems, we focus on treating the optimality system as a fully coupled system and only briefly discuss the application of decoupling strategies. Nevertheless, the reader should keep in mind that amenability to efficient uncoupling strategies is, perhaps, the chief reason to consider the application of least-squares principles to optimization problems. The application of least-squares principles to optimality systems was previously discussed, in a concrete setting, in [19].

The approach we have described for finding approximate solutions of optimal control and optimization problems for partial differential equations is of the *differentiate-then-discretize* or *optimize-then-discretize* type. One first applies, at the continuous partial differential equations level, the first-order necessary conditions for finding saddle points of a Lagrangian functional and then one uses a finite element method, be it of Galerkin or of least-square type, to discretize the resulting optimality system. For the alternative *discretize-then-differentiate* or *discretize-then-optimize* type approach, one reverses the steps: one first discretizes the optimization or control problem by some means and then applies the Lagrange multiplier rule to the resulting discrete optimization problem. The two steps do not, in general, commute so that the discrete systems determined by the two approaches are not the same. A discussion of the relative merits of the two approaches can be found, e.g., in [18]. Here, we focus on the differentiate-then-discretize approach.

Instead of using the Lagrange multiplier rule for solving constrained optimization problems, one may use a penalty method. Penalty/least-squares finite element methods are the subject of the companion paper [9]; see also [20]. Other applica-

tions of least-squares finite element methods to optimization problems may be found in [2, 3, 5, 8].

The paper is organized as follows. In §3, we study, in an abstract setting, Lagrange multiplier methods for quadratic optimization and control problems constrained by linear, elliptic partial differential equations. In §4, we study least-squares finite element methods for the approximate solution of the optimality system resulting from the application of the Lagrange multiplier rule. In §5, we provide concrete examples that illustrate the theory of §§3–4. In passing, we briefly remark on several related topics, including decoupling strategies for the solution of the discretized optimality system. Before we embark, however, in §2, we present mostly well-known results about general constrained optimization problems and their solution via Lagrange multiplier methods. These results serve as the foundation for the considerations of §§3–4. We remark that in several inequalities appearing in the throughout the paper,  $C$  denotes a positive constant whose value changes with context but that is independent of any of the data or solution functions appearing in the inequalities.

**2. Linearly constrained quadratic minimization problems in Hilbert spaces.** In this section, we review the now classical theory (see [12] and also [13, 16]) for finite element methods for constrained quadratic minimization problems. The optimization and control problems that are the subject of this paper can be profitably viewed as special cases of the types of problems treated by the classic theory.

Given Hilbert spaces  $V$  and  $S$  along with their dual spaces  $V^*$  and  $S^*$ , respectively, the symmetric bilinear form  $a(\cdot, \cdot)$  on  $V \times V$ , the bilinear form  $b(\cdot, \cdot)$  on  $V \times S$ , the functions  $f \in V^*$  and  $g \in S^*$ , and the real number  $t$ , we define the *quadratic functional*<sup>1</sup>

$$\mathcal{J}(u) = \frac{1}{2}a(u, u) - \langle f, u \rangle_{V^*, V} + t \quad \forall u \in V, \quad (2.1)$$

the *linear constraint equation*

$$b(u, q) = \langle g, q \rangle_{S^*, S} \quad \forall q \in S, \quad (2.2)$$

where  $\langle \cdot, \cdot \rangle$  denotes an appropriate duality pairing, and the *constrained minimization problem*<sup>2</sup>.

$$\min_{u \in V} \mathcal{J}(u) \quad \text{subject to (2.2)}. \quad (2.4)$$

---

<sup>1</sup>The value of  $t$  does not affect the minimizer of  $\mathcal{J}(\cdot)$ . We include it in the definition of  $\mathcal{J}(u)$  only to facilitate, in later sections, the identification of concrete functionals with the abstract functional (2.1).

<sup>2</sup>Such problems arise in many applications. A classical example is provided by the functional

$$\mathcal{J}(\mathbf{v}) = \frac{1}{2} \int_{\Omega} |\mathbf{v}|^2 d\Omega,$$

the linear constraint  $\nabla \cdot \mathbf{v} = f$ , and the minimization problem

$$\min J(\mathbf{v}) \quad \text{subject to } \nabla \cdot \mathbf{v} = f, \quad (2.3)$$

where the minimization is effected over suitable function space. For example, in fluid mechanics, (2.3) is known as the *Kelvin principle* and, in structural mechanics (where  $\mathbf{v}$  is a tensor), as the *complementary energy principle*. For the Kelvin principle  $A$  is the identity operator,  $B$  is the divergence,  $S$  is the space  $L^2(\Omega)$  of all square integrable functions, and  $V$  is the space  $H(\text{div}, \Omega)$  of all square integrable vector fields whose divergencies are also square integrable.

The bilinear forms serve to define the associated operators

$$A : V \rightarrow V^*, \quad B : V \rightarrow S^*, \quad \text{and} \quad B^* : S \rightarrow V^* \quad (2.5)$$

through the relations

$$\begin{cases} a(u, v) = \langle Au, v \rangle_{V^*, V} & \forall u, v \in V \\ b(v, q) = \langle Bv, q \rangle_{S^*, S} = \langle B^*q, v \rangle_{V^*, V} & \forall v \in V, q \in S. \end{cases}$$

The minimization problem (2.4) can then be given the form

$$\min_{u \in V} \mathcal{J}(u) \quad \text{subject to} \quad Bu = g,$$

where the constraint equations  $Bu = g$  holds in  $S^*$ . We define the subspace

$$Z = \{v \in V : b(v, q) = 0 \quad \forall q \in S\} \quad (2.6)$$

and make the following assumptions about the bilinear forms:

$$\begin{cases} a(u, v) \leq C_a \|u\|_V \|v\|_V & \forall u, v \in V \\ b(u, q) \leq C_b \|u\|_V \|q\|_S & \forall u \in V, q \in S \\ a(u, u) \geq 0 & \forall u \in V \\ a(u, u) \geq K_a \|u\|_V^2 & \forall u \in Z \\ \sup_{v \in V, v \neq 0} \frac{b(v, q)}{\|v\|_V} \geq K_b \|q\|_S & \forall q \in S, \end{cases} \quad (2.7)$$

where  $C_a, C_b, K_a,$  and  $K_b$  are all positive constants.

**2.1. Existence of solutions.** The following result is well known; see, e.g. [21].

PROPOSITION 2.1. *Let the assumptions (2.7) hold. Then, the constrained minimization problem (2.4) has a unique solution  $u \in V$ .  $\square$*

**2.2. Solution via Lagrange multipliers.** For all  $v \in V$  and  $q \in S$ , we introduce the Lagrangian functional

$$\mathcal{L}(v, q) = \mathcal{J}(v) + b(v, q) - \langle g, q \rangle_{S^*, S} = \frac{1}{2}a(v, v) + b(v, q) - \langle f, v \rangle_{V^*, V} - \langle g, q \rangle_{S^*, S} + t.$$

Then, the constrained minimization problem (2.4) is equivalent to the unconstrained optimization problem of finding saddle points  $(u, p) \in V \times S$  of the Lagrangian functional. These saddle points may be found by solving the optimality system

$$\begin{cases} a(u, v) + b(v, p) & = \langle f, v \rangle_{V^*, V} & \forall v \in V \\ b(u, q) & = \langle g, q \rangle_{S^*, S} & \forall q \in S. \end{cases} \quad (2.8)$$

The following result is also well known; see, e.g., [12].

PROPOSITION 2.2. *Let the assumptions (2.7) hold. Then, the system (2.8) has a unique solution  $(u, p) \in V \times S$ . Moreover,*

$$\|u\|_V + \|p\|_S \leq C(\|f\|_{V^*} + \|g\|_{S^*}) \quad (2.9)$$

and  $u \in V$  is the unique solution of the constrained minimization problem (2.4).  $\square$

In terms of the operators introduced in (2.5), the system (2.8) takes the form

$$\begin{cases} Au + B^*p &= f & \text{in } V^* \\ Bu &= g & \text{in } S^*. \end{cases}$$

REMARK 2.3. The unique solvability of (2.8) and the estimate (2.9) do not require that the bilinear form  $a(\cdot, \cdot)$  be symmetric or that it satisfy the third condition in (2.7). Also, the fourth condition in (2.7) may be weakened to a weak coercivity condition. However, these conditions are required to make the connection between (2.8) and the constrained minimization problem (2.4). So, throughout, we will assume that all the conditions in (2.7) hold.

**2.3. Galerkin approximations of the optimality system.** We choose (conforming) finite dimensional subspaces  $V^h \subset V$  and  $S^h \subset S$ , and then restrict (2.8) to these subspaces, i.e., we seek  $u^h \in V^h$  and  $p^h \in S^h$  that satisfy

$$\begin{cases} a(u^h, v^h) + b(v^h, p^h) &= \langle f, v^h \rangle_{V^*, V} & \forall v^h \in V^h \\ b(u^h, q^h) &= \langle g, q^h \rangle_{S^*, S} & \forall q^h \in S^h. \end{cases} \quad (2.10)$$

This is also the optimality system for the minimization of the functional  $\mathcal{J}(\cdot)$  over  $V^h$  subject to  $b(u^h, q^h) = \langle g, q^h \rangle_{S^*, S}$  for all  $q^h \in S^h$ . Let

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

In general,  $Z^h \not\subset Z$  even though  $V^h \subset V$  and  $S^h \subset S$ , and so the last two assumptions in (2.7) may not be satisfied with respect to the subspaces. If  $V^h$  and  $S^h$  are such that the last two assumptions hold, then one obtains the following well-known result; see, e.g., [12].

PROPOSITION 2.4. *Let the hypotheses of Proposition 2.1 hold and assume that*

$$a(u^h, u^h) \geq K_a^h \|u^h\|_V^2 \quad \forall u^h \in Z^h \quad (2.11)$$

and

$$\sup_{v^h \in V^h, v^h \neq 0} \frac{b(v^h, q^h)}{\|v^h\|_V} \geq K_b^h \|q^h\|_S \quad \forall q^h \in S^h, \quad (2.12)$$

where  $K_a^h$  and  $K_b^h$  are positive constants independent of  $h$ . Then, the discrete system (2.10) has a unique solution  $(u^h, p^h) \in V^h \times S^h$  and moreover

$$\|u^h\|_V + \|p^h\|_S \leq C(\|f\|_{V^*} + \|g\|_{S^*}).$$

Furthermore, if  $(u, p) \in V \times S$  denotes the unique solution of (2.8), then

$$\|u - u^h\|_V + \|p - p^h\|_S \leq C \left( \inf_{v^h \in V^h} \|u - v^h\|_V + \inf_{q^h \in S^h} \|p - q^h\|_S \right). \quad (2.13)$$

The discrete problem (2.10) is equivalent to a linear system. Indeed, let  $\{U_i\}_{i=1}^n$  and  $\{P_i\}_{i=1}^m$ , where  $n = \dim(V^h)$  and  $m = \dim(S^h)$ , denote bases for  $V^h$  and  $S^h$ , respectively, and let  $\bar{\mathbf{u}} = (u_1, \dots, u_n)^T$  and  $\bar{\mathbf{p}} = (p_1, \dots, p_m)^T$  denote the vectors of

coefficients in the expansions of  $u^h$  and  $p^h$  in terms of the respective bases. Furthermore, let  $f_i = \langle f, U_i \rangle_{V^*, V}$  for  $i = 1, \dots, n$ ,  $g_i = \langle g, P_i \rangle_{S^*, S}$  for  $i = 1, \dots, m$ ,  $\vec{f} = (f_1, \dots, f_n)^T$ , and  $\vec{g} = (g_1, \dots, g_m)^T$  and define the elements of the  $n \times n$  matrix  $\mathbb{A}$  and the  $m \times n$  matrix  $\mathbb{B}$  by  $\mathbb{A}_{ij} = a(U_i, U_j)$  for  $i, j = 1, \dots, n$  and  $\mathbb{B}_{ij} = b(U_j, P_i)$  for  $i = 1, \dots, m, j = 1, \dots, n$ , respectively. Then, (2.10) is equivalent to the linear system

$$\begin{pmatrix} \mathbb{A} & \mathbb{B}^T \\ \mathbb{B} & 0 \end{pmatrix} \begin{pmatrix} \vec{u} \\ \vec{p} \end{pmatrix} = \begin{pmatrix} \vec{f} \\ \vec{g} \end{pmatrix}. \quad (2.14)$$

REMARK 2.5. The coefficient matrix in (2.14) is symmetric and indefinite. This is universal for discretizations of saddle-point problems arising from the use of the Lagrange multiplier rule for constrained optimization problems.

REMARK 2.6. The assumptions (2.11) and (2.12) guarantee that the  $(m+n) \times (m+n)$  coefficient matrix in (2.14) is invertible and that the norms of its inverse are bounded from above independently of  $m$  and  $n$ , i.e., independently of the grid size  $h$ .

REMARK 2.7. The observations made in Remark 2.3 about the bilinear form  $a(\cdot, \cdot)$  and (2.8) also apply to (2.10).

**3. Quadratic optimization and control problems in Hilbert spaces with linear constraints.** In this section, we specialize the results of §2 to the type of optimization and control problems described in §1. We identify the variable  $u$  of §2 with the pair  $(\phi, \theta)$ , where  $\phi$  and  $\theta$  are the state and control variables, respectively, of the control problem.

We begin with four given Hilbert spaces  $\Theta$ ,  $\Phi$ ,  $\widehat{\Phi}$ , and  $\widetilde{\Phi}$  along with their dual spaces denoted by  $(\cdot)^*$ . We assume that  $\Phi \subseteq \widehat{\Phi} \subseteq \widetilde{\Phi}$  with continuous embeddings and that  $\widetilde{\Phi}$  acts as the pivot space for both the pair  $\{\Phi^*, \Phi\}$  and the pair  $\{\widehat{\Phi}^*, \widehat{\Phi}\}$  so that we not only have that  $\Phi \subseteq \widehat{\Phi} \subseteq \widetilde{\Phi} \subseteq \widehat{\Phi}^* \subseteq \Phi^*$ , but also

$$\langle \psi, \phi \rangle_{\Phi^*, \Phi} = \langle \psi, \phi \rangle_{\widehat{\Phi}^*, \widehat{\Phi}} = (\psi, \phi)_{\widetilde{\Phi}} \quad \forall \psi \in \widehat{\Phi}^* \subseteq \Phi^* \quad \text{and} \quad \forall \phi \in \Phi \subseteq \widehat{\Phi}, \quad (3.1)$$

where  $(\cdot, \cdot)_{\widetilde{\Phi}}$  denotes the inner product on  $\widetilde{\Phi}$ .

Next, we define the *quadratic functional*

$$\mathcal{J}(\phi, \theta) = \frac{1}{2} a_1(\phi - \widehat{\phi}, \phi - \widehat{\phi}) + \frac{1}{2} a_2(\theta, \theta) \quad \forall \phi \in \Phi, \theta \in \Theta, \quad (3.2)$$

where  $a_1(\cdot, \cdot)$  and  $a_2(\cdot, \cdot)$  are symmetric bilinear forms on  $\widehat{\Phi} \times \widehat{\Phi}$  and  $\Theta \times \Theta$ , respectively, and  $\widehat{\phi} \in \widehat{\Phi}$  is a given function. In the language of control theory,  $\Phi$  is called the *state space*,  $\phi$  the *state variable*,  $\Theta$  the *control space*, and  $\theta$  the *control variable*. In many applications, the control space is finite dimensional in which case  $\theta$  is often referred to as the vector of *design variables*. We note that often  $\Theta$  is chosen to be a bounded set in a Hilbert space but, for our purposes, we consider the less general situation of  $\Theta$  itself being a Hilbert space. We make the following assumptions about the bilinear forms  $a_1(\cdot, \cdot)$  and  $a_2(\cdot, \cdot)$ :

$$\begin{cases} a_1(\phi, \mu) \leq C_1 \|\phi\|_{\widehat{\Phi}} \|\mu\|_{\widehat{\Phi}} & \forall \phi, \mu \in \widehat{\Phi} \\ a_2(\theta, \nu) \leq C_2 \|\theta\|_{\Theta} \|\nu\|_{\Theta} & \forall \theta, \nu \in \Theta \\ a_1(\phi, \phi) \geq 0 & \forall \phi \in \widehat{\Phi} \\ a_2(\theta, \theta) \geq K_2 \|\theta\|_{\Theta}^2 & \forall \theta \in \Theta, \end{cases} \quad (3.3)$$

where  $C_1$ ,  $C_2$ , and  $K_2$  are all positive constants. The second term in the functional (3.2) can be interpreted as a penalty term which limits the size of the control  $\theta$ .

Given another Hilbert space  $\Lambda$ , the additional bilinear forms  $b_1(\cdot, \cdot)$  on  $\Phi \times \Lambda$  and  $b_2(\cdot, \cdot)$  on  $\Theta \times \Lambda$ , and the function  $g \in \Lambda^*$ , we define the *linear constraint equation*

$$b_1(\phi, \psi) + b_2(\theta, \psi) = \langle g, \psi \rangle_{\Lambda^*, \Lambda} \quad \forall \psi \in \Lambda. \quad (3.4)$$

We make the following assumptions about the bilinear forms  $b_1(\cdot, \cdot)$  and  $b_2(\cdot, \cdot)$ :

$$\left\{ \begin{array}{ll} b_1(\phi, \psi) \leq c_1 \|\phi\|_{\Phi} \|\psi\|_{\Lambda} & \forall \phi \in \Phi, \psi \in \Lambda \\ b_2(\theta, \psi) \leq c_2 \|\psi\|_{\Phi} \|\theta\|_{\Theta} & \forall \theta \in \Theta, \psi \in \Lambda \\ \sup_{\psi \in \Lambda, \psi \neq 0} \frac{b_1(\phi, \psi)}{\|\psi\|_{\Lambda}} \geq k_1 \|\phi\|_{\Phi} & \forall \phi \in \Phi \\ \sup_{\phi \in \Phi, \phi \neq 0} \frac{b_1(\phi, \psi)}{\|\phi\|_{\Phi}} > 0 & \forall \psi \in \Lambda, \end{array} \right. \quad (3.5)$$

where  $c_1$ ,  $c_2$ , and  $k_1$  are all positive constants. These assumptions suffice to guarantee that, given any  $\theta \in \Theta$ , the constraint equation (3.4) is uniquely solvable for  $\phi \in \Phi$ ; this observation easily follows from [1] and (3.5).

We consider the *optimal control problem*<sup>3</sup>

$$\min_{(\phi, \theta) \in \Phi \times \Theta} \mathcal{J}(\phi, \theta) \quad \text{subject to} \quad b_1(\phi, \psi) + b_2(\theta, \psi) = \langle g, \psi \rangle_{\Lambda^*, \Lambda} \quad \forall \psi \in \Lambda. \quad (3.6)$$

It is easy to verify that the problem (3.6) falls into the framework of §2. To this end, we let  $V \equiv \Phi \times \Theta$ ,  $S \equiv \Lambda$ ,  $\|\{\phi, \theta\}\|_V = \sqrt{\|\phi\|_{\Phi}^2 + \|\theta\|_{\Theta}^2}$  for all  $\{\phi, \theta\} \in V$ ,

$$\left\{ \begin{array}{ll} a(\{\phi, \theta\}, \{\mu, \nu\}) \equiv a_1(\phi, \mu) + a_2(\theta, \nu) & \forall \phi, \mu \in \Phi, \theta, \nu \in \Theta, \\ b(\{\phi, \theta\}, \{\psi\}) \equiv b_1(\phi, \psi) + b_2(\theta, \psi) & \forall \phi \in \Phi, \theta \in \Theta, \psi \in \Lambda, \\ \langle f, \{\mu, \nu\} \rangle_{V^*, V} \equiv a_1(\mu, \hat{\phi}) & \forall \mu \in \Phi, \nu \in \Theta, \\ t = \frac{1}{2} a_1(\hat{\phi}, \hat{\phi}). \end{array} \right. \quad (3.7)$$

From (3.3), it follows that  $t \leq (C_1/2) \|\hat{\phi}\|_{\hat{\Phi}}^2$  and, also using the continuous embedding  $\Phi \subseteq \hat{\Phi}$ ,

$$\frac{\langle f, \{\mu, \nu\} \rangle_{V^*, V}}{\|\{\mu, \nu\}\|_V} \leq \frac{\langle f, \{\mu, \nu\} \rangle_{V^*, V}}{\|\mu\|_{\Phi}} = \frac{a_1(\mu, \hat{\phi})}{\|\mu\|_{\Phi}} \leq C_1 \|\hat{\phi}\|_{\hat{\Phi}} \quad \forall \{\mu, \nu\} \in \Phi \times \Theta = V,$$

so that  $\|f\|_{V^*} \leq C_1 \|\hat{\phi}\|_{\hat{\Phi}}$ , i.e.,  $f$  does indeed belong to  $V^*$ . Then, with the obvious identifications  $u = \{\phi, \theta\}$ ,  $v = \{\mu, \nu\}$ , and  $q = \{\psi\}$ , the functionals (2.1) and

<sup>3</sup>In section 5 we will consider an example where the linear constraint will be the weak form of the Stokes equations of incompressible viscous flows. We draw attention to the fact that these equations themselves are another example of a problem that fits the abstract setting of section 2 with

$$\mathcal{J}(\mathbf{v}; \mathbf{f}) = \frac{1}{2} \int_{\Omega} |\nabla \mathbf{v}|^2 d\Omega - \int_{\Omega} \mathbf{f} \cdot \mathbf{v} d\Omega,$$

$$b(\mathbf{v}, q) = \int_{\Omega} q \nabla \cdot \mathbf{v} d\Omega \quad \text{nd} \quad g = 0.$$

(3.2) are equivalent as are the constraint equations (2.2) and (3.4). The constrained optimization problem (2.4) and the optimal control problem (3.6) are also equivalent.

We will use the framework and results established in §2 to study the optimal control problem (3.6). Many of the results we discuss are well known, but we repeat them here to establish a context for later discussions.

**3.1. Existence of optimal states and controls.** We begin with the following preliminary result.

LEMMA 3.1. *Let the assumptions (3.3) and (3.5) hold. Then, the spaces  $V \equiv \Phi \times \Theta$  and  $S \equiv \Lambda$  and the bilinear forms  $a(\cdot, \cdot)$  and  $b(\cdot, \cdot)$  defined in (3.7) satisfy the assumptions (2.7).*

*Proof.* Using (3.3) and the continuous embedding  $\Phi \subset \widehat{\Phi}$ , we have that

$$\begin{aligned} a(\{\phi, \theta\}, \{\mu, \nu\}) &= a_1(\phi, \mu) + a_2(\theta, \nu) \leq C_1 \|\phi\|_{\widehat{\Phi}} \|\mu\|_{\widehat{\Phi}} + C_2 \|\theta\|_{\Theta} \|\nu\|_{\Theta} \\ &\leq C_1 \|\phi\|_{\Phi} \|\mu\|_{\Phi} + C_2 \|\theta\|_{\Theta} \|\nu\|_{\Theta} \leq \max\{C_1, C_2\} \sqrt{\|\phi\|_{\Phi}^2 + \|\theta\|_{\Theta}^2} \sqrt{\|\mu\|_{\Phi}^2 + \|\nu\|_{\Theta}^2} \end{aligned}$$

for all  $\phi, \mu \in \Phi$ ,  $\theta, \nu \in \Theta$  so that  $a(u, v) \leq C_a \|u\|_V \|v\|_V$  for all  $u, v \in V$  with  $C_a = \max\{C_1, C_2\}$ . Similarly, we have using (3.5) that  $b(u, q) \leq C_b \|u\|_V \|q\|_S$  for all  $u \in V$ ,  $q \in S$  with  $C_b = \max\{c_1, c_2\}$ . Next, from (3.3) we have that

$$a(\{\phi, \theta\}, \{\phi, \theta\}) = a_1(\phi, \phi) + a_2(\theta, \theta) \geq a_2(\theta, \theta) \geq K_2 \|\theta\|_{\Theta}^2 \quad \forall \phi \in \Phi, \theta \in \Theta$$

so that  $a(u, u) \geq 0$  for all  $u \in V$ . We next define the subspace  $Z \subset \Phi \times \Theta$  by

$$Z = \left\{ \{\phi, \theta\} \in \Phi \times \Theta : b_1(\phi, \psi) + b_2(\theta, \psi) = 0 \quad \forall \psi \in \Lambda \right\}. \quad (3.8)$$

The assumptions (3.5) imply that, given any  $\theta \in \Theta$ , the problem

$$b_1(\phi, \psi) = -b_2(\theta, \psi) \quad \forall \psi \in \Lambda \quad (3.9)$$

has a unique solution  $\phi_\theta$  and, moreover,

$$\|\phi_\theta\|_{\Phi} \leq \frac{c_2}{k_1} \|\theta\|_{\Theta}; \quad (3.10)$$

see, e.g., [1]. Thus,  $Z$  can be completely characterized by  $(\phi_\theta, \theta) \in \Phi \times \Theta$ , where, for arbitrary  $\theta \in \Theta$ ,  $\phi_\theta$  is the solution of (3.9). Then, (3.10) and (3.3) imply that

$$\begin{aligned} a(\{\phi_\theta, \theta\}, \{\phi_\theta, \theta\}) &= a_1(\phi_\theta, \phi_\theta) + a_2(\theta, \theta) \geq a_2(\theta, \theta) \geq K_2 \|\theta\|^2 \\ &\geq \frac{K_2}{2} \min \left\{ 1, \frac{k_1^2}{c_2^2} \right\} \|\{\phi_\theta, \theta\}\|_V^2 \quad \forall \{\phi_\theta, \theta\} \in Z. \end{aligned}$$

As a result,  $a(u, u) \geq K_a \|u\|_V^2$  for all  $u \in Z$  with  $K_a = \frac{1}{2} K_2 \min \left\{ 1, \frac{k_1^2}{c_2^2} \right\}$  so that the third assumption in (2.7) is also satisfied.

To verify the last assumption in (2.7), note that

$$\sup_{\phi \in \Phi, \phi \neq 0} \frac{b_1(\phi, \psi)}{\|\phi\|_{\Phi}} \geq k_1 \|\psi\|_{\Lambda} \quad \forall \psi \in \Lambda. \quad (3.11)$$

Indeed, assumptions (3.5) imply that (see [1]), for any  $\psi \in \Lambda$ , the problem

$$b_1(\phi, \mu) = (\psi, \mu)_{\Lambda} \quad \forall \mu \in \Lambda \quad (3.12)$$

has a unique solution  $\phi_\psi$  and, moreover,

$$\|\phi_\psi\|_\Phi \leq \frac{1}{k_1} \|\psi\|_\Lambda. \quad (3.13)$$

Using (3.12)–(3.13), it is easy to see that

$$\frac{b_1(\phi_\psi, \psi)}{\|\phi_\psi\|_\Phi} = \frac{\|\psi\|_\Lambda^2}{\|\phi_\psi\|_\Phi} \geq k_1 \|\psi\|_\Lambda \quad \forall \psi \in \Lambda$$

which immediately implies (3.11). Finally, using (3.11),

$$\sup_{(\phi, \theta) \in \Phi \times \Theta, (\phi, \theta) \neq (0, 0)} \frac{b(\{\phi, \theta\}, \{\psi\})}{\sqrt{\|\phi\|_\Phi^2 + \|\theta\|_\Theta^2}} \geq \sup_{\phi \in \Phi, \phi \neq 0} \frac{b_1(\phi, \psi)}{\|\phi\|_\Phi} \geq k_1 \|\psi\|_\Lambda \quad \forall \psi \in \Lambda$$

so that

$$\sup_{u \in V, u \neq 0} \frac{b(u, q)}{\|u\|_V} \geq K_b \|q\|_S \quad \forall q \in S$$

with  $K_b = k_1 \square$

Having verified the assumptions (2.7) for the optimal control problem (3.6), we immediately have the following result.

**THEOREM 3.2.** *Let the assumptions (3.3) and (3.5) hold. Then, the optimal control problem (3.6) has a unique solution  $(\phi, \theta) \in \Phi \times \Theta$ .*

*Proof.* The result immediately follows from Proposition 2.1 and Lemma 3.1.  $\square$

It is instructive to rewrite the functional (3.2), the constraint (3.4), and the optimal control problem (3.6) in operator notation. To this end, we note that the bilinear forms serve to define operators

$$\begin{aligned} A_1 : \widehat{\Phi} &\rightarrow \widehat{\Phi}^*, & A_2 : \Theta &\rightarrow \Theta^*, & B_1 : \Phi &\rightarrow \Lambda^*, \\ B_1^* : \Lambda &\rightarrow \Phi^*, & B_2 : \Theta &\rightarrow \Lambda^*, & B_2^* : \Lambda &\rightarrow \Theta^* \end{aligned}$$

through the following relations:

$$\begin{cases} a_1(\phi, \mu) = \langle A_1 \phi, \mu \rangle_{\widehat{\Phi}^*, \widehat{\Phi}} & \forall \phi, \mu \in \widehat{\Phi} \\ a_2(\theta, \nu) = \langle A_2 \theta, \nu \rangle_{\Theta^*, \Theta} & \forall \theta, \nu \in \Theta \\ b_1(\phi, \psi) = \langle B_1 \phi, \psi \rangle_{\Lambda^*, \Lambda} = \langle B_1^* \psi, \phi \rangle_{\Phi^*, \Phi} & \forall \phi \in \Phi, \psi \in \Lambda \\ b_2(\psi, \theta) = \langle B_2 \theta, \psi \rangle_{\Lambda^*, \Lambda} = \langle B_2^* \psi, \theta \rangle_{\Theta^*, \Theta} & \forall \theta \in \Theta, \psi \in \Lambda. \end{cases} \quad (3.14)$$

Then, the functional (3.2) and the constraint (3.4) respectively take the forms

$$\mathcal{J}(\phi, \theta) = \frac{1}{2} \langle A_1(\phi - \widehat{\phi}), (\phi - \widehat{\phi}) \rangle_{\widehat{\Phi}^*, \widehat{\Phi}} + \frac{1}{2} \langle A_2 \theta, \theta \rangle_{\Theta^*, \Theta} \quad \forall \phi \in \Phi, \theta \in \Theta \quad (3.15)$$

and

$$B_1 \phi + B_2 \theta = g \quad \text{in } \Lambda^* \quad (3.16)$$

and the optimal control problem (3.6) takes the form

$$\min_{(\phi, \theta) \in \Phi \times \Theta} \mathcal{J}(\phi, \theta) \quad \text{subject to (3.16)}. \quad (3.17)$$

Assumptions (3.3) and (3.5) imply that  $A_1, A_2, B_1, B_2, B_1^*$ , and  $B_2^*$  are bounded with

$$\begin{aligned} \|A_1\|_{\widehat{\Phi} \rightarrow \widehat{\Phi}^*} &\leq C_1, & \|A_2\|_{\Theta \rightarrow \Theta^*} &\leq C_2, & \|B_1\|_{\Phi \rightarrow \Phi^*} &\leq c_1, \\ \|B_1^*\|_{\Phi \rightarrow \Phi^*} &\leq c_1, & \|B_2\|_{\Phi \rightarrow \Theta^*} &\leq c_2, & \|B_2^*\|_{\Theta \rightarrow \Phi^*} &\leq c_2 \end{aligned}$$

and that the operator  $B_1$  is invertible with  $\|B_1^{-1}\|_{\Lambda^* \rightarrow \Phi} \leq 1/k_1$ . Note also that the subspace  $Z \subset V = \Phi \times \Theta$  can be defined by

$$Z = \left\{ \{\phi, \theta\} \in \Phi \times \Theta : \phi = -B_1^{-1}B_2\theta \quad \forall \theta \in \Theta \right\}.$$

**3.2. Solution via Lagrange multipliers and the optimality system.** For all  $\{\mu, \nu\} \in V = \Phi \times \Theta$  and  $\psi \in S = \Lambda$ , we introduce the Lagrangian functional

$$\begin{aligned} \mathcal{L}(\{\mu, \nu\}, \{\psi\}) &= \mathcal{J}(\{\mu, \nu\}) + b(\{\mu, \nu\}, \{\psi\}) - \langle g, \psi \rangle_{\Lambda^*, \Lambda} \\ &= \frac{1}{2}a_1(\mu - \widehat{\phi}, \mu - \widehat{\phi}) + \frac{1}{2}a_2(\nu, \nu) + b_1(\mu, \psi) + b_2(\nu, \psi) - \langle g, \psi \rangle_{\Lambda^*, \Lambda}. \end{aligned}$$

Then, (3.6) is equivalent to the unconstrained optimization problem of finding saddle points  $(\{\phi, \theta\}, \{\lambda\})$  in  $V \times S$  of the Lagrangian functional. These saddle points may be found by solving the *optimality system*

$$\begin{cases} a_1(\phi, \mu) & + b_1(\mu, \lambda) & = a_1(\widehat{\phi}, \mu) & \forall \mu \in \Phi \\ & a_2(\theta, \nu) + b_2(\nu, \lambda) & = 0 & \forall \nu \in \Theta \\ b_1(\phi, \psi) + b_2(\theta, \psi) & & = \langle g, \psi \rangle_{\Lambda^*, \Lambda} & \forall \psi \in \Lambda. \end{cases} \quad (3.18)$$

The third equation in the optimality system (3.18) is simply the constraint equation. The first equation is commonly referred to as the *adjoint* or *co-state equation* and the Lagrange multiplier  $\lambda$  is referred as the *adjoint* or *co-state* variable. The second equation in (3.18) is referred to as the *optimality condition* since it is merely a statement that the gradient of the functional  $\mathcal{J}(\cdot, \cdot)$  defined in (3.2) vanishes at the optimum.

Using the framework of §2.2, the following result is immediate.

**THEOREM 3.3.** *Let the assumptions (3.3) and (3.5) hold. Then, the optimality system (3.18) has a unique solution  $(\phi, \theta, \lambda) \in \Phi \times \Theta \times \Lambda$ . Moreover*

$$\|\phi\|_{\Phi} + \|\theta\|_{\Theta} + \|\lambda\|_{\Lambda} \leq C(\|g\|_{\Lambda^*} + \|\widehat{\phi}\|_{\widehat{\Phi}})$$

and  $(\phi, \theta) \in \Phi \times \Theta$  is the unique solution of the optimal control problem (3.6).

*Proof.* With the associations  $V = \Phi \times \Theta$ ,  $S = \Lambda$ ,  $u = \{\phi, \theta\}$ , and  $p = \{\lambda\}$ , the results immediately follow from Lemma 3.1 and Proposition 2.2.  $\square$

Using the operators introduced in (3.14) and (3.1), the optimality system (3.18) takes the form

$$\begin{cases} A_1\phi & + B_1^*\lambda & = A_1\widehat{\phi} & \text{in } \Phi^* \\ & A_2\theta + B_2^*\lambda & = 0 & \text{in } \Theta^* \\ B_1\phi + B_2\theta & & = g & \text{in } \Lambda^*. \end{cases} \quad (3.19)$$

**3.3. Galerkin approximation of the optimality system.** We choose (conforming) finite dimensional subspaces  $\Phi^h \subset \Phi$ ,  $\Theta^h \subset \Theta$ , and  $\Lambda^h \subset \Lambda$  and then restrict (3.18) to the subspaces, i.e., we seek  $(\phi^h, \theta^h, \lambda^h) \in \Phi^h \times \Theta^h \times \Lambda^h$  that satisfies

$$\begin{cases} a_1(\phi^h, \mu^h) & + b_1(\mu^h, \lambda^h) & = a_1(\widehat{\phi}, \mu^h) & \forall \mu^h \in \Phi^h \\ & a_2(\theta^h, \nu^h) & + b_2(\nu^h, \lambda^h) & = 0 & \forall \nu^h \in \Theta^h \\ b_1(\phi^h, \psi^h) & + b_2(\theta^h, \psi^h) & & = \langle g, \psi^h \rangle_{\Lambda^*, \Lambda} & \forall \psi^h \in \Lambda^h. \end{cases} \quad (3.20)$$

This is also the optimality system for the minimization of (3.2) over  $\Phi^h \times \Theta^h$  subject to the constraint  $b_1(\phi^h, \psi^h) + b_2(\psi^h, \theta^h) = \langle g, \psi^h \rangle_{\Lambda^*, \Lambda}$  for all  $\psi^h \in \Lambda^h$ .

We next define the subspace  $Z^h \subset \Phi^h \times \Theta^h$  by

$$Z^h = \left\{ \{\phi^h, \theta^h\} \in \Phi^h \times \Theta^h : b_1(\phi^h, \psi^h) + b_2(\theta^h, \psi^h) = 0 \quad \forall \psi^h \in \Lambda^h \right\}. \quad (3.21)$$

Note that, in general,  $Z^h \not\subset Z$  even though  $\Phi^h \subset \Phi$ ,  $\Theta^h \subset \Theta$ , and  $\Lambda^h \subset \Lambda$ . Thus, we make the following additional assumptions about  $b_1(\cdot, \cdot)$  and  $\Phi^h$ :

$$\begin{cases} \sup_{\psi^h \in \Lambda^h, \psi^h \neq 0} \frac{b_1(\phi^h, \psi^h)}{\|\psi^h\|_{\Lambda}} \geq k_1^h \|\phi^h\|_{\Phi} & \forall \phi^h \in \Phi^h \\ \sup_{\phi^h \in \Phi^h, \phi^h \neq 0} \frac{b_1(\phi^h, \psi^h)}{\|\phi^h\|_V} > 0 & \forall \psi^h \in \Lambda^h, \end{cases} \quad (3.22)$$

where  $k_1^h$  is a positive constant whose value is independent of  $h$ . Analogous to Lemma 3.1, we have the following result.

**LEMMA 3.4.** *Let the assumptions (3.3), (3.5), and (3.22) hold. Then, the spaces  $V^h = \Phi^h \times \Theta^h$  and  $S^h = \Phi^h$  and the bilinear forms  $a(\cdot, \cdot)$  and  $b(\cdot, \cdot)$  defined in (3.7) satisfy the assumptions (2.11) and (2.12).*

*Proof.* The proof proceeds exactly as that for for Lemma 3.1; the constants in (2.11) are given by  $K_a^h = \frac{1}{2} K_2 \min \left\{ 1, \frac{(k_1^h)^2}{c_2^2} \right\}$  and  $K_b^h = k_1^h$ .  $\square$

We then easily obtain the following results.

**THEOREM 3.5.** *Let the assumptions (3.3), (3.5), and (3.22) hold. Then, the discrete optimality system (3.20) has a unique solution  $(\phi^h, \theta^h, \lambda^h) \in \Phi^h \times \Theta^h \times \Lambda^h$  and moreover*

$$\|\phi^h\|_{\Phi} + \|\theta^h\|_{\Theta} + \|\lambda^h\|_{\Lambda} \leq C(\|g\|_{\Lambda^*} + \|\widehat{\phi}\|_{\widehat{\Phi}}).$$

Furthermore, let  $(\phi, \theta, \lambda) \in \Phi \times \Theta \times \Lambda$  denote the unique solution of the optimality system (3.18), or, equivalently, of the optimal control problem (3.6). Then,

$$\begin{aligned} & \|\phi - \phi^h\|_{\Phi} + \|\theta - \theta^h\|_{\Theta} + \|\lambda - \lambda^h\|_{\Lambda} \\ & \leq C \left( \inf_{\mu^h \in \Phi^h} \|\phi - \mu^h\|_{\Phi} + \inf_{\xi^h \in \Theta^h} \|\theta - \xi^h\|_{\Theta} + \inf_{\psi^h \in \Lambda^h} \|\lambda - \psi^h\|_{\Lambda} \right). \end{aligned} \quad (3.23)$$

*Proof.* The results immediately follow from Proposition 2.4 and Lemma 3.4.  $\square$

The discrete optimality system (3.20) is equivalent to the linear system

$$\begin{pmatrix} \mathbb{A}_1 & 0 & \mathbb{B}_1^T \\ 0 & \mathbb{A}_2 & \mathbb{B}_2^T \\ \mathbb{B}_1 & \mathbb{B}_2 & 0 \end{pmatrix} \begin{pmatrix} \vec{\phi} \\ \vec{\theta} \\ \vec{\lambda} \end{pmatrix} = \begin{pmatrix} \vec{\mathbf{f}} \\ \vec{\mathbf{0}} \\ \vec{\mathbf{g}} \end{pmatrix}, \quad (3.24)$$

where  $\vec{\mathbf{f}}$  and  $\vec{\mathbf{g}}$  are defined using  $a_1(\widehat{\phi}, \mu^h)$  and  $\langle g, \psi^h \rangle_{\Lambda^*, \Lambda}$ , respectively, and  $\mathbb{A}_k$  and  $\mathbb{B}_k$  are defined in the standard manner from the bilinear forms  $a_k(\cdot, \cdot)$  and  $b_k(\cdot, \cdot)$ ,  $k = 1, 2$ , respectively.

REMARK 3.6. There are two sets of inf-sup conditions associated with the problems (3.18) and (3.20). First, we have the “inner” conditions (3.5) and (3.22) that involve only the state variable and that guarantee the unique solvability of the state equation and the discrete state equation, respectively, i.e., of the third equations in (3.18) and (3.20). Second, we have the “outer” conditions (2.7) and (2.12) involving the bilinear form  $b(\cdot, \cdot)$  defined in (3.7) and that involve both the state and control variables. These latter conditions help guarantee the unique solvability of the optimality system (3.18) and the discrete optimality system (3.20), respectively. Note that the outer conditions and the related saddle point nature of the optimality systems occur regardless of the the nature of the inner problem, i.e., the state equations. For example, even if the state equations involve a strongly coercive bilinear form  $b_1(\cdot, \cdot)$  so that the last two inequalities in (3.5) can be replaced by  $b_1(\phi, \phi) \geq k_1 \|\phi\|_{\Phi}^2$  for all  $\phi \in \Phi$ , we would still have the inf-sup condition in the form of the last equation in (2.7).

REMARK 3.7. As mentioned in Remark 3.6, the assumptions in (3.22) guarantee the unique solvability of the discrete state equation (the third equation in (3.20)) for the discrete state variable  $\phi^h \in \Phi^h$ . Thus, if the constraint equation (3.4) is a partial differential equation problem, then (3.22) are the general assumptions on the associated bilinear form and the approximating space that are made to guarantee the stability and convergence of Galerkin finite element discretizations; see, e.g., [1]. Furthermore, because of the nature of the assumptions (3.3) and (3.5), the inf-sup condition on the bilinear form  $b(\cdot, \cdot)$  is satisfied merely by assuming that (3.22) holds. Thus, by merely guaranteeing that the discrete constraint equations within the discretized optimal control problem are uniquely solvable for any given discrete control, i.e., assuming that the “inner” inf-sup conditions holds, we have that the “outer” inf-sup condition on the bilinear form  $b(\cdot, \cdot)$  holds. The latter, of course, is crucial to the stability and convergence of finite element approximations to any saddle point problem, including the optimality systems we consider here.

REMARK 3.8. The discrete optimality system (3.20) or its matrix equivalent (3.24), have the typical saddle point structure and thus, the stability and convergence of the approximations they define depend on the bilinear form  $b(\cdot, \cdot) = b_1(\cdot, \cdot) + b_2(\cdot, \cdot)$  satisfying the discrete inf-sup condition (2.12) with respect to  $V^h = \Phi^h \times \Theta^h$  and  $S^h = \Lambda^h$ . In the current context, this assumption is satisfied (see Remark 3.7) merely by assuming that (3.22) holds for the bilinear form  $b_1(\cdot, \cdot)$  and the spaces  $\Phi^h$  and  $\Lambda_h$ . Thus, as discussed in Remark 3.7, the stability and convergence of solutions of (3.20) or (3.24) depends solely on the ability to stably solve, given any discrete control variable, the discrete state equation for a discrete state variable. On the other hand, if (3.22) does not hold, then there exists a  $\phi_0^h \in \Phi^h$  such that  $\phi_0^h \neq 0$  and  $b_1(\phi_0^h, \psi^h) = 0$  for all  $\psi^h \in \Lambda^h$ . Then,  $b(\{\phi_0^h, 0\}, \{\psi^h\}) = b_1(\phi_0^h, \psi^h) = 0$  for all  $\psi^h \in \Lambda^h$  so that

$$\sup_{\psi^h \in \Lambda^h, \psi^h \neq 0} \frac{b(\{\phi_0^h, 0\}, \{\psi^h\})}{\|\psi^h\|_{\Lambda}} = 0.$$

It can be shown that this implies that the discrete inf-sup condition (2.12) does not hold so that (3.20) or its matrix equivalent (3.24) may not be solvable, i.e., the coefficient matrix in (3.24) may not be invertible. In fact, the assumptions (3.22) imply that  $\mathbb{B}_1$  is uniformly invertible. This, and the facts (which follow from (3.3))

that the symmetric matrices  $\mathbb{A}_1$  and  $\mathbb{A}_2$  are positive semi-definite and positive definite, respectively, is enough to guarantee that the coefficient matrix in (3.24) is invertible. On the other hand, if (3.22) does not hold so that the matrix  $\mathbb{B}_1$  has a nontrivial null space, then, under the other assumptions that have been made, one cannot guarantee the invertibility of the coefficient matrix in (3.24).

REMARK 3.9. Solving the discrete optimality system (3.20), or equivalently, the linear system (3.24), is often a formidable task. If the constraint equations (3.4) are a system of partial differential equations, then the last (block) row of (3.24) represents a Galerkin finite element discretization of that system. The discrete adjoint equations, i.e., the first row in (3.24), are also a discretization of a system of partial differential equations. Moreover, the dimension of the discrete adjoint vector  $\vec{\lambda}$  is essentially the same as that of discrete state vector  $\vec{\phi}$ . Thus, (3.24) is at least twice the size (we have yet to account for the discrete control variables in  $\vec{\theta}$ ) of the discrete system corresponding to the discretization of the partial differential equation constraints. Thus, if these equations are difficult to approximate, the discrete optimality system will be even more difficult to deal with. For this reason, there have been many approaches suggested for uncoupling the three components of discrete optimality systems such as (3.20), or equivalently, (3.24). See, e.g., [18], for a discussion of several of these approaches. We note that these approaches rely on the invertibility of the matrices  $\mathbb{B}_1$  and  $\mathbb{A}_2$ , properties that follow from (3.22) and (3.3), respectively.

#### 4. Least-squares finite element methods for the optimality system.

Even if the state equation (3.4) (or (3.16)) involves a symmetric, positive definite operator  $B_1$ , i.e., even if the bilinear form  $b_1(\cdot, \cdot)$  is symmetric and strongly coercive, the discrete optimality system (3.20) (or (3.24)) obtained through a Galerkin discretization is indefinite. For example, if  $B_1 = -\Delta$  with zero boundary conditions, then  $\mathbb{B}_1$  is a symmetric, positive definite matrix, but the coefficient matrix in (3.24) is indefinite. In order to obtain a discrete optimality system that is symmetric and positive definite, we will apply a least-squares finite element discretization. In fact, these desirable properties for the discrete system will remain in place even if the state system bilinear form  $b_1(\cdot, \cdot)$  is only weakly coercive, i.e., even if the operator  $B_1$  is merely invertible and not necessarily positive definite.

Given a system of partial differential equations, there are many ways to define least-squares finite element methods for determining approximate solutions. Practicality issues can be used to select the “best” methods from among the many choices available. See, e.g., [6] for a discussion of what factors enter into the choice of a particular least-squares finite element method for a given problem. Here, we will consider the most straightforward means for defining a least-squares finite element method. When, in §5, we consider a specific example, we will return to a discussion of practicality issues in the choice of a least-squares finite element formulation.

**4.1. A least-squares finite element method for a generalization of the optimality system.** We start with the generalized form of the optimality system (3.19) written in operator form, i.e.,

$$\begin{cases} A_1\phi & + B_1^*\lambda = f & \text{in } \Phi^* \\ & A_2\theta + B_2^*\lambda = s & \text{in } \Theta^* \\ B_1\phi + B_2\theta & = g & \text{in } \Lambda^*, \end{cases} \quad (4.1)$$

where  $(f, s, g) \in \Phi^* \times \Theta^* \times \Lambda^*$  is a general data triple and  $(\phi, \theta, \lambda) \in \Phi \times \Theta \times \Lambda$  is the corresponding solution triple. In the same way that Theorem 3.3 was proved, we

have the following result.

PROPOSITION 4.1. *Let the assumptions (3.3) and (3.5) hold. Then, for any  $(f, s, g) \in \Phi^* \times \Theta^* \times \Lambda^*$ , the generalized optimality system (4.1) has a unique solution  $(\phi, \theta, \lambda) \in \Phi \times \Theta \times \Lambda$ . Moreover,*

$$\|\phi\|_{\Phi} + \|\theta\|_{\Theta} + \|\lambda\|_{\Lambda} \leq C(\|f\|_{\Phi^*} + \|s\|_{\Theta^*} + \|g\|_{\Lambda^*}). \quad (4.2)$$

A least-squares functional can be defined by summing the squares of the norms of the residuals of the three equations in (4.1) to obtain

$$\mathcal{K}(\phi, \theta, \lambda; f, s, g) = \|A_1\phi + B_1^*\lambda - f\|_{\Phi^*}^2 + \|A_2\theta + B_2^*\lambda - s\|_{\Theta^*}^2 + \|B_1\phi + B_2\theta - g\|_{\Lambda^*}^2. \quad (4.3)$$

Clearly, the unique solution of (4.1) is also the solution of the problem

$$\min_{(\phi, \theta, \lambda) \in \Phi \times \Theta \times \Lambda} \mathcal{K}(\phi, \theta, \lambda; f, s, g). \quad (4.4)$$

The first-order necessary conditions corresponding to (4.4) are easily found to be

$$B((\phi, \theta, \lambda), (\mu, \nu, \psi)) = F((\mu, \nu, \psi); (f, s, g)) \quad \forall (\mu, \nu, \psi) \in \Phi \times \Theta \times \Lambda, \quad (4.5)$$

where

$$\begin{aligned} B((\phi, \theta, \lambda), (\mu, \nu, \psi)) &= (A_1\mu + B_1^*\psi, A_1\phi + B_1^*\lambda)_{\Phi^*} \\ &\quad + (A_2\nu + B_2^*\psi, A_2\theta + B_2^*\lambda)_{\Theta^*} + (B_1\mu + B_2\nu, B_1\phi + B_2\theta)_{\Lambda^*} \end{aligned} \quad (4.6)$$

$$\forall (\phi, \theta, \lambda), (\mu, \nu, \psi) \in \Phi \times \Theta \times \Lambda$$

and

$$\begin{aligned} F((\mu, \nu, \psi); (f, s, g)) &= (A_1\mu + B_1^*\psi, f)_{\Phi^*} + (A_2\nu + B_2^*\psi, s)_{\Theta^*} \\ &\quad + (B_1\mu + B_2\nu, g)_{\Lambda^*} \quad \forall (\mu, \nu, \psi) \in \Phi \times \Theta \times \Lambda. \end{aligned} \quad (4.7)$$

LEMMA 4.2. *Let the assumptions (3.3) and (3.5) hold. Then, the bilinear form  $B(\cdot, \cdot)$  is symmetric and continuous on  $(\Phi \times \Theta \times \Lambda) \times (\Phi \times \Theta \times \Lambda)$  and the linear functional  $F(\cdot)$  is continuous on  $(\Phi \times \Theta \times \Lambda)$ . Moreover, the bilinear form  $B(\cdot, \cdot)$  is coercive on  $(\Phi \times \Theta \times \Lambda)$ , i.e.,*

$$B((\phi, \theta, \lambda), (\phi, \theta, \lambda)) \geq C(\|\phi\|_{\Phi}^2 + \|\theta\|_{\Theta}^2 + \|\lambda\|_{\Lambda}^2) \quad \forall (\phi, \theta, \lambda) \in \Phi \times \Theta \times \Lambda. \quad (4.8)$$

*Proof.* The symmetry and continuity of the form  $B(\cdot, \cdot)$  and the continuity of the form  $F(\cdot)$  are clear. From (4.6), we have that

$$B((\phi, \theta, \lambda), (\phi, \theta, \lambda)) = \|A_1\phi + B_1^*\lambda\|_{\Phi^*}^2 + \|A_2\theta + B_2^*\lambda\|_{\Theta^*}^2 + \|B_1\phi + B_2\theta\|_{\Lambda^*}^2. \quad (4.9)$$

Clearly, for any  $(\phi, \theta, \lambda) \in \Phi \times \Theta \times \Lambda$ , there exists  $(f, s, g) \in \Phi^* \times \Theta^* \times \Lambda^*$  such that  $(\phi, \theta, \lambda)$  is a solution of (4.1). This observation and Proposition 4.1 then yield that

$$\begin{aligned} \|\phi\|_{\Phi}^2 + \|\theta\|_{\Theta}^2 + \|\lambda\|_{\Lambda}^2 &\leq C(\|f\|_{\Phi^*}^2 + \|s\|_{\Theta^*}^2 + \|g\|_{\Lambda^*}^2) \\ &= C(\|A_1\phi + B_1^*\lambda\|_{\Phi^*}^2 + \|A_2\theta + B_2^*\lambda\|_{\Theta^*}^2 + \|B_1\phi + B_2\theta\|_{\Lambda^*}^2) \end{aligned} \quad (4.10)$$

$$\forall (\phi, \theta, \lambda) \in \Phi \times \Theta \times \Lambda.$$

Combining (4.9) and (4.10) then easily yields (4.8).  $\square$

REMARK 4.3. Since

$$\begin{aligned} \mathcal{K}(\phi, \theta, \lambda; 0, 0, 0) &= \|A_1\phi + B_1^*\lambda\|_{\Phi^*}^2 + \|A_2\theta + B_2^*\lambda\|_{\Theta^*}^2 + \|B_1\phi + B_2\theta\|_{\Lambda^*}^2 \\ &= B((\phi, \theta, \lambda), (\phi, \theta, \lambda)), \end{aligned}$$

the coercivity and continuity of the bilinear form  $B(\cdot, \cdot)$  are equivalent to stating that the functional  $\mathcal{K}(\phi, \theta, \lambda; 0, 0, 0)$  is norm-equivalent, i.e., that there exist constants  $\gamma_1 > 0$  and  $\gamma_2 > 0$  such that

$$\gamma_1(\|\phi\|_{\Phi}^2 + \|\theta\|_{\Theta}^2 + \|\lambda\|_{\Lambda}^2) \leq \mathcal{K}(\phi, \theta, \lambda; 0, 0, 0) \leq \gamma_2(\|\phi\|_{\Phi}^2 + \|\theta\|_{\Theta}^2 + \|\lambda\|_{\Lambda}^2) \quad (4.11)$$

for all  $(\phi, \theta, \lambda) \in \Phi \times \Theta \times \Lambda$ .

PROPOSITION 4.4. *Let the assumptions (3.3) and (3.5) hold. Then, for any  $(f, s, g) \in \Phi^* \times \Theta^* \times \Lambda^*$ , the problem (4.5) has a unique solution  $(\phi, \theta, \lambda) \in \Phi \times \Theta \times \Lambda$ . Moreover, this solution coincides with the solution of the problems (4.1) and (4.4) and satisfies the estimate (4.2).*

*Proof.* The results follow from Lemma 4.2 and the Lax-Milgram lemma.  $\square$

We define a finite element discretization of (4.1) or, equivalently, of (4.5), by choosing conforming finite element subspaces  $\Phi^h \subset \Phi$ ,  $\Theta^h \subset \Theta$ , and  $\Lambda^h \subset \Lambda$  and then requiring that  $(\phi^h, \theta^h, \lambda^h) \in \Phi^h \times \Theta^h \times \Lambda^h$  satisfy

$$\begin{aligned} B((\phi^h, \theta^h, \lambda^h), (\mu^h, \nu^h, \psi^h)) &= F((\mu^h, \nu^h, \psi^h); (f, s, g)) \\ \forall (\mu^h, \nu^h, \psi^h) &\in \Phi^h \times \Theta^h \times \Lambda^h. \end{aligned} \quad (4.12)$$

Note that  $(\phi^h, \theta^h, \lambda^h)$  can also be characterized as the solution of the problem

$$\min_{(\phi^h, \theta^h, \lambda^h) \in \Phi^h \times \Theta^h \times \Lambda^h} \mathcal{K}(\phi^h, \theta^h, \lambda^h; f, s, g).$$

PROPOSITION 4.5. *Let the assumptions (3.3) and (3.5) hold. Then, for any  $(f, h, g) \in \Phi^* \times \Theta^* \times \Lambda^*$ , the problem (4.12) has a unique solution  $(\phi^h, \theta^h, \lambda^h) \in \Phi^h \times \Theta^h \times \Lambda^h$ . Moreover, we have the optimal error estimate*

$$\begin{aligned} &\|\phi - \phi^h\|_{\Phi} + \|\theta - \theta^h\|_{\Theta} + \|\lambda - \lambda^h\|_{\Lambda} \\ &\leq C \left( \inf_{\tilde{\phi}^h \in \Phi^h} \|\phi - \tilde{\phi}^h\|_{\Phi} + \inf_{\tilde{\theta}^h \in \Theta^h} \|\theta - \tilde{\theta}^h\|_{\Theta} + \inf_{\tilde{\lambda}^h \in \Lambda^h} \|\lambda - \tilde{\lambda}^h\|_{\Lambda} \right), \end{aligned} \quad (4.13)$$

where  $(\phi, \theta, \lambda) \in \Phi \times \Theta \times \Lambda$  is the unique solution of the problem (4.5), or equivalently, of the problems (4.1) or (4.4).

*Proof.* The results follow from Lemma 4.2 and standard finite element analyses.  $\square$

#### 4.2. A least-squares finite element method for the optimality system.

The results of §4.1 easily specialize to the optimality system (3.19). Indeed, letting  $f = A_1\hat{\phi} \in \hat{\Phi}^* \subset \Phi^*$  and  $s = 0$ , we have that (4.1) reduces to (3.19). We now have the least-squares functional,

$$\mathcal{K}(\phi, \theta, \lambda; \hat{\phi}, g) = \|A_1\phi + B_1^*\lambda - A_1\hat{\phi}\|_{\Phi^*}^2 + \|A_2\theta + B_2^*\lambda\|_{\Theta^*}^2 + \|B_1\phi + B_2\theta - g\|_{\Lambda^*}^2, \quad (4.14)$$

the minimization problem

$$\min_{(\phi, \theta, \lambda) \in \Phi \times \Theta \times \Lambda} \mathcal{K}(\phi, \theta, \lambda; \hat{\phi}, g), \quad (4.15)$$

the first-order necessary conditions

$$B((\phi, \theta, \lambda), (\mu, \nu, \psi)) = F((\mu, \nu, \psi); (A_1 \widehat{\phi}, 0, g)) \quad \forall (\mu, \nu, \psi) \in \Phi \times \Theta \times \Lambda, \quad (4.16)$$

where  $B(\cdot, \cdot)$  and  $F(\cdot)$  are defined as in (4.6) and (4.7), respectively.

We define a finite element discretization of (4.16) by again choosing conforming finite element subspaces  $\Phi^h \subset \Phi$ ,  $\Theta^h \subset \Theta$ , and  $\Lambda^h \subset \Lambda$  and then requiring that  $(\phi^h, \theta^h, \lambda^h) \in \Phi^h \times \Theta^h \times \Lambda^h$  satisfy

$$\begin{aligned} B((\phi^h, \theta^h, \lambda^h), (\mu^h, \nu^h, \psi^h)) &= F((\mu^h, \nu^h, \psi^h); (A_1 \widehat{\phi}, 0, g)) \\ \forall (\mu^h, \nu^h, \psi^h) &\in \Phi^h \times \Theta^h \times \Lambda^h. \end{aligned} \quad (4.17)$$

Then, Proposition 4.5 takes the following form.

**THEOREM 4.6.** *Let the assumptions (3.3) and (3.5) hold. Then, for any  $(\widehat{\phi}, g) \in \widehat{\Phi}^* \times \Lambda^*$ , the problem (4.17) has a unique solution  $(\phi^h, \theta^h, \lambda^h) \in \Phi^h \times \Theta^h \times \Lambda^h$ . Moreover, we have the optimal error estimate: there exists a constant  $C > 0$  whose value is independent of  $h$ , such that*

$$\begin{aligned} &\|\phi - \phi^h\|_{\Phi} + \|\theta - \theta^h\|_{\Theta} + \|\lambda - \lambda^h\|_{\Lambda} \\ &\leq C \left( \inf_{\tilde{\phi}^h \in \Phi^h} \|\phi - \tilde{\phi}^h\|_{\Phi} + \inf_{\tilde{\theta}^h \in \Theta^h} \|\theta - \tilde{\theta}^h\|_{\Theta} + \inf_{\tilde{\lambda}^h \in \Lambda^h} \|\lambda - \tilde{\lambda}^h\|_{\Lambda} \right), \end{aligned} \quad (4.18)$$

where  $(\phi, \theta, \lambda) \in \Phi \times \Theta \times \Lambda$  is the unique solution of the problem (4.16) or, equivalently, of the problems (3.19) or (3.18). Note also that  $(\phi, \theta) \in \Phi \times \Theta$  is the unique solution of the problem (3.6).

**REMARK 4.7.** The discrete problem (4.17) is equivalent to the linear algebraic system

$$\begin{pmatrix} \mathbb{K}_1 & \mathbb{C}_1^T & \mathbb{C}_2^T \\ \mathbb{C}_1 & \mathbb{K}_2 & \mathbb{C}_3^T \\ \mathbb{C}_2 & \mathbb{C}_3 & \mathbb{K}_3 \end{pmatrix} \begin{pmatrix} \vec{\phi} \\ \vec{\theta} \\ \vec{\lambda} \end{pmatrix} = \begin{pmatrix} \vec{\mathbf{f}} \\ \vec{\mathbf{h}} \\ \vec{\mathbf{g}} \end{pmatrix}. \quad (4.19)$$

Indeed, if one chooses bases  $\{\mu_j^h(\mathbf{x})\}_{j=1}^J$ ,  $\{\nu_k^h(\mathbf{x})\}_{k=1}^K$ , and  $\{\psi_\ell^h(\mathbf{x})\}_{\ell=1}^L$  for  $\Phi^h$ ,  $\Theta^h$ , and  $\Lambda^h$ , respectively, we then have  $\phi^h = \sum_{j=1}^J \phi_j \mu_j^h$ ,  $\theta^h = \sum_{k=1}^K \theta_k \nu_k^h$ , and  $\lambda^h = \sum_{\ell=1}^L \lambda_\ell \psi_\ell^h$  for some sets of coefficients  $\{\phi_j\}_{j=1}^J$ ,  $\{\theta_k\}_{k=1}^K$ , and  $\{\lambda_\ell\}_{\ell=1}^L$  that are determined by solving (4.19). In (4.19), we have that  $\vec{\phi} = (\phi_1, \dots, \phi_J)^T$ ,  $\vec{\theta} = (\theta_1, \dots, \theta_K)^T$ ,

$$\begin{aligned}
\vec{\lambda} &= (\lambda_1, \dots, \lambda_L)^T, \\
(\mathbb{K}_1)_{ij} &= (A_1\mu_i, A_1\mu_j)_{\Phi^*} + (B_1\mu_i, B_1\mu_j)_{\Lambda^*} && \text{for } i, j = 1, \dots, J, \\
(\mathbb{K}_2)_{ik} &= (A_2\nu_i, A_1\nu_k)_{\Theta^*} + (B_2\nu_i, B_2\nu_k)_{\Lambda^*} && \text{for } i, k = 1, \dots, K, \\
(\mathbb{K}_3)_{i\ell} &= (B_1^*\psi_i, B_1^*\psi_\ell)_{\Phi^*} + (B_2\psi_i, B_2\psi_\ell)_{\Theta^*} && \text{for } i, \ell = 1, \dots, L, \\
(\mathbb{C}_1)_{ij} &= (B_2\nu_i, B_1\mu_j)_{\Lambda^*} && \text{for } i = 1, \dots, K, j = 1, \dots, J, \\
(\mathbb{C}_2)_{ij} &= (B_1^*\psi_i, A_1\nu_j)_{\Phi^*} && \text{for } i = 1, \dots, L, j = 1, \dots, J, \\
(\mathbb{C}_3)_{ik} &= (B_2^*\psi_i, A_2\nu_k)_{\Theta^*} && \text{for } i = 1, \dots, L, k = 1, \dots, K, \\
(\vec{\mathbf{f}})_i &= (A_1\mu_i, A_1\hat{\phi})_{\Phi^*} + (B_1\mu_i, g)_{\Lambda^*} && \text{for } i = 1, \dots, J, \\
(\vec{\mathbf{h}})_i &= (B_2\nu_i, g)_{\Lambda^*} && \text{for } i = 1, \dots, K, \\
(\vec{\mathbf{g}})_i &= (B_1^*\psi_i, A_1\hat{\phi})_{\Phi^*} && \text{for } i = 1, \dots, L.
\end{aligned}$$

REMARK 4.8. It easily follows from Lemma 4.2 that the coefficient matrix of (4.19) is *symmetric and positive definite*. This should be compared to the linear system (3.24) that results from a Galerkin finite element discretization of the optimality system (3.18) for which the coefficient matrix is symmetric and *indefinite*.

REMARK 4.9. The stability of the discrete problem (4.17), the convergence and optimal accuracy of the approximate solution  $(\phi^h, \theta^h, \lambda^h)$ , and the symmetry and positive definiteness of the discrete system (4.19) obtained by the least-squares finite element method follow from the assumptions (3.3) and (3.5) that guarantee the well posedness of the infinite-dimensional optimization problem (3.6) and its corresponding optimality system (3.18). It is important to note that all of these desirable properties of the least-squares finite element method do not require that the bilinear form  $b_1(\cdot, \cdot)$  and the finite element spaces  $\Phi^h$  and  $\Lambda^h$  satisfy the inner (see Remark 3.6) inf-sup conditions (3.22) that are necessary for the well posedness of the Galerkin finite element discretization (3.20) of the optimality system (3.18). In fact, this is why least-squares finite element methods are often an attractive alternative to Galerkin discretizations; see, e.g., [6].

REMARK 4.10. The observations made in Remark 3.9 about the possible need to uncouple the equations in (3.24) hold as well for the linear system (4.19). Uncoupling approaches for (3.24) rely on the invertibility of the matrices  $\mathbb{B}_1$  and  $\mathbb{A}_2$ ; the first of these is, in general, non-symmetric and indefinite, even when the necessary discrete inf-sup conditions in (3.22) are satisfied. For (4.19), uncoupling strategies would rely on the invertibility of the matrices  $\mathbb{K}_1$ ,  $\mathbb{K}_2$ , and  $\mathbb{K}_3$ ; all three of these matrices are symmetric and positive definite even when (3.22) is not satisfied. An example of a simple uncoupling strategy is to apply a block-Gauss Seidel method to (4.19), which would proceed as follows.

Start with initial guesses  $\vec{\phi}^{(0)}$  and  $\vec{\theta}^{(0)}$  for the discretized state and control; then, for  $k = 1, 2, \dots$ , successively solve the linear systems

$$\begin{aligned}
\mathbb{K}_3 \vec{\lambda}^{(k+1)} &= \vec{\mathbf{g}} - \mathbb{C}_2 \vec{\phi}^{(k)} - \mathbb{C}_3 \vec{\theta}^{(k)} \\
\mathbb{K}_1 \vec{\phi}^{(k+1)} &= \vec{\mathbf{f}} - \mathbb{C}_1^T \vec{\theta}^{(k)} - \mathbb{C}_2^T \vec{\lambda}^{(k+1)} \\
\mathbb{K}_2 \vec{\theta}^{(k+1)} &= \vec{\mathbf{h}} - \mathbb{C}_1 \vec{\phi}^{(k+1)} - \mathbb{C}_3^T \vec{\lambda}^{(k+1)}
\end{aligned} \tag{4.20}$$

until satisfactory convergence is achieved, e.g., until some norm of the difference between successive iterates is less than some prescribed tolerance.

Since the coefficient matrix in (4.19) is symmetric and positive definite, this iteration will converge. Moreover, all three coefficient matrices  $\mathbb{K}_3$ ,  $\mathbb{K}_1$ , and  $\mathbb{K}_2$  of the linear systems in (4.20) are themselves symmetric and positive definite so that very efficient solution methodologies, including parallel ones, can be applied for their solution. We also note that, in order to obtain faster convergence rates, better uncoupling iterative methods, e.g., over-relaxation schemes or a conjugate gradient method, can be applied instead of the Gauss-Seidel iteration of (4.20).

REMARK 4.11. The discrete problem (4.17) (or equivalently, (4.19)) resulting from the least-squares method for the optimality system (3.19) can be viewed as a Galerkin discretization of the system

$$\begin{aligned} (A_1^*A_1 + B_1^*B_1)\phi + (B_1^*B_2)\theta + (A_1^*B_1^*)\lambda &= (A_1^*A_1)\widehat{\phi} + (B_1^*)g && \text{in } \Phi \\ (A_2^*A_2 + B_2^*B_2)\theta + (A_2^*B_2^*)\lambda + (B_2^*B_1)\phi &= (B_2^*)g && \text{in } \Theta \\ (B_1B_1^* + B_2B_2^*)\lambda + (B_1A_1)\phi + (B_2A_2)\theta &= (B_1A_1)\widehat{\phi} && \text{in } \Lambda. \end{aligned} \quad (4.21)$$

The first equation of this system is the sum of  $A_1^*$  applied to the first equation of the optimality system (3.19) and  $B_1^*$  applied to the third equation of that system. The other equations of (4.21) are related to the equations of (3.19) in a similar manner. The system (4.21) shows that the discrete system (4.19) essentially involves the discretization of “squares” of operators, e.g.,  $A_1^*A_1$ ,  $B_1^*B_1$ , etc. This observation has a profound effect in how one chooses the form of the constraint equation in (3.6), i.e., the form of (3.16). We will return to this point in the next section when we consider a concrete example.

**5. Example: Optimization problems for the Stokes system.** Let  $\Omega$  denote an open, bounded domain in  $\mathcal{R}^s$ ,  $s = 2$  or  $3$ , with boundary  $\Gamma$ . Let  $\mathbf{u}$  and  $p$  denote the velocity and pressure fields, respectively, and let  $\boldsymbol{\theta}$  denote a distributed control. Then, consider the Stokes system

$$\begin{cases} -\Delta \mathbf{u} + \nabla p + \boldsymbol{\theta} &= \mathbf{g} \\ \nabla \cdot \mathbf{u} &= 0 \end{cases} \quad \text{in } \Omega, \quad \mathbf{u} = \mathbf{0} \quad \text{on } \Gamma, \quad \int_{\Omega} p \, d\Omega = 0 \quad (5.1)$$

and the functionals

$$\text{Case I:} \quad \mathcal{J}_1(\mathbf{u}, \boldsymbol{\theta}) = \frac{1}{2} \int_{\Omega} |\nabla \times \mathbf{u}|^2 \, d\Omega + \frac{\delta}{2} \int_{\Omega} |\boldsymbol{\theta}|^2 \, d\Omega \quad (5.2)$$

$$\text{Case II:} \quad \mathcal{J}_2(\mathbf{u}, \boldsymbol{\theta}; \widehat{\mathbf{u}}) = \frac{1}{2} \int_{\Omega} |\mathbf{u} - \widehat{\mathbf{u}}|^2 \, d\Omega + \frac{\delta}{2} \int_{\Omega} |\boldsymbol{\theta}|^2 \, d\Omega, \quad (5.3)$$

where  $\mathbf{g}$  and  $\widehat{\mathbf{u}}$  are given functions. We study the two problems of finding  $(\mathbf{u}, p, \boldsymbol{\theta})$  that minimizes either the functional in (5.2) or (5.3), subject to the Stokes system (5.1) being satisfied. In the first case, i.e., for the functional (5.2), the problem we study is to find a distributed control function  $\boldsymbol{\theta}$  that minimizes, in the  $\mathbf{L}^2(\Omega)$  sense, the vorticity over the flow domain  $\Omega$ . In the second case, i.e., for the functional (5.3), the problem we study is to find a distributed control function  $\boldsymbol{\theta}$  such that flow velocity  $\mathbf{u}$  matches as well as possible, in the  $\mathbf{L}^2(\Omega)$  sense, a given velocity field  $\widehat{\mathbf{u}}$ .

In Remark 4.11, it was noted that least-squares finite element methods for optimization problems result in the “squaring” of the constraint operator, in this case, of the Stokes system (5.1). This results in biharmonic type terms appearing in the system corresponding to (4.21), or, equivalently, in (4.9). A conforming finite element discretization would then require the use of continuously differentiable approximation spaces. In order to overcome this impracticality, it has become a standard procedure in least-squares finite element methods to write the state system in an equivalent first-order formulation; see, e.g., [6], for a detailed discussion of this issue.

There are many ways to rewrite the Stokes system (5.1) as a first-order system of partial differential equations. Here, we choose the *velocity-vorticity-pressure formulation* that is the most commonly used system for this purpose. Let  $\boldsymbol{\omega} = \nabla \times \mathbf{u}$  denote the vorticity. Then, using the well-known vector identity  $-\Delta \mathbf{u} = \nabla \times \nabla \times \mathbf{u} - \nabla(\nabla \cdot \mathbf{u}) = \nabla \times \boldsymbol{\omega} - \nabla(\nabla \cdot \mathbf{u})$ , the Stokes system (5.1) can be expressed as

$$\begin{cases} \nabla \times \boldsymbol{\omega} + \nabla p + \boldsymbol{\theta} &= \mathbf{g} \\ \nabla \cdot \mathbf{u} &= 0 \quad \text{in } \Omega, \quad \mathbf{u} = \mathbf{0} \quad \text{on } \Gamma, \quad \int_{\Omega} p \, d\Omega = 0. \\ \nabla \times \mathbf{u} - \boldsymbol{\omega} &= \mathbf{0} \end{cases} \quad (5.4)$$

Note that the functional (5.2) can now be written as

$$\text{Case I:} \quad \mathcal{J}_1(\boldsymbol{\omega}, \boldsymbol{\theta}) = \frac{1}{2} \int_{\Omega} |\boldsymbol{\omega}|^2 \, d\Omega + \frac{\delta}{2} \int_{\Omega} |\boldsymbol{\theta}|^2 \, d\Omega. \quad (5.5)$$

Thus, the optimization problems we study are to find  $(\mathbf{u}, \boldsymbol{\omega}, p, \boldsymbol{\theta})$  that minimizes either the functional in (5.3) or (5.5), subject to the Stokes system in the form (5.4) being satisfied.

**5.1. Precise statement of optimization problems.** We recall the space  $L^2(\Omega)$  of all square integrable functions with norm  $\|\cdot\|_0$  and inner product  $(\cdot, \cdot)$ , the space  $L_0^2(\Omega) \equiv \{q \in L^2(\Omega) : \int_{\Omega} p \, d\Omega = 0\}$ , the space  $H^1(\Omega) \equiv \{v \in L^2(\Omega) : \nabla v \in [L^2(\Omega)]^s\}$ , and the space  $H_0^1(\Omega) \equiv \{v \in H^1(\Omega) : v = 0 \text{ on } \Gamma\}$ . A norm for functions  $v \in H^1(\Omega)$  is given by  $\|v\|_1 \equiv (\|\nabla v\|^2 + \|v\|_0^2)^{1/2}$ . The dual space of  $H_0^1(\Omega)$  is denoted by  $H^{-1}(\Omega)$ . The corresponding spaces of vector-valued functions are denoted in bold face, e.g.,  $\mathbf{H}^1(\Omega) = [H^1(\Omega)]^s$  is the space of vector-valued functions each of whose components belongs to  $H^1(\Omega)$ . We note the following equivalence of norms [16]:

$$\tilde{C}_1 \|\mathbf{v}\|_1^2 \leq \|\nabla \times \mathbf{v}\|_0^2 + \|\nabla \cdot \mathbf{v}\|_0^2 \leq \tilde{C}_2 \|\mathbf{v}\|_1^2 \quad \forall \mathbf{v} \in \mathbf{H}_0^1(\Omega) \quad (5.6)$$

for some constants  $\tilde{C}_1 > 0$  and  $\tilde{C}_2 > 0$ .

Let  $\Phi = \Lambda = \mathbf{H}_0^1(\Omega) \times \mathbf{L}^2(\Omega) \times L_0^2(\Omega)$  and  $\Theta = \mathbf{L}^2(\Omega)$  so that  $\Phi^* = \Lambda^* = \mathbf{H}^{-1}(\Omega) \times \mathbf{L}^2(\Omega) \times L_0^2(\Omega)$  and  $\Theta^* = \mathbf{L}^2(\Omega)$ . Let  $\hat{\Phi} = \tilde{\Phi} = \mathbf{L}^2(\Omega) \times \mathbf{L}^2(\Omega) \times L_0^2(\Omega)$ . Then,  $\Phi \subset \hat{\Phi} = \tilde{\Phi} = \hat{\Phi}^* \subset \Phi^*$ . For  $\phi = \{\mathbf{u}, \boldsymbol{\omega}, p\} \in \Phi$ , we define the norm

$$\|\phi\|_{\Phi} = (\|\mathbf{u}\|_1^2 + \|\boldsymbol{\omega}\|_0^2 + \|p\|_0^2)^{1/2}$$

and likewise for the other product spaces.

We make the associations of

$$\begin{array}{lll} \text{trial functions:} & \phi = \{\mathbf{u}, \boldsymbol{\omega}, p\} \in \Phi, & \theta = \{\boldsymbol{\theta}\} \in \Theta, \quad \lambda = \{\mathbf{v}, \boldsymbol{\sigma}, q\} \in \Lambda, \\ \text{test functions:} & \mu = \{\tilde{\mathbf{u}}, \tilde{\boldsymbol{\omega}}, \tilde{p}\} \in \Phi, & \nu = \{\tilde{\boldsymbol{\theta}}\} \in \Theta, \quad \psi = \{\tilde{\mathbf{v}}, \tilde{\boldsymbol{\sigma}}, \tilde{r}\} \in \Lambda, \\ \text{data:} & g = \{\mathbf{g}, \mathbf{0}, 0\} \in \Lambda^*, & \hat{\phi} = \{\hat{\mathbf{u}}, \mathbf{0}, 0\} \in \hat{\Phi}. \end{array}$$

We next define the bilinear forms

$$\begin{aligned}
a_1(\phi, \mu) &= \begin{cases} (\tilde{\boldsymbol{\omega}}, \boldsymbol{\omega}) & \text{for Case I} \\ (\tilde{\mathbf{u}}, \mathbf{u}) & \text{for Case II} \end{cases} \quad \forall \phi = \{\mathbf{u}, \boldsymbol{\omega}, p\} \in \widehat{\Phi}, \quad \mu = \{\tilde{\mathbf{u}}, \tilde{\boldsymbol{\omega}}, \tilde{p}\} \in \widehat{\Phi}, \\
a_2(\theta, \nu) &= \delta(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) \quad \forall \theta = \{\boldsymbol{\theta}\} \in \Theta, \quad \nu = \{\tilde{\boldsymbol{\theta}}\} \in \Theta, \\
b_1(\phi, \psi) &= (\boldsymbol{\omega}, \nabla \times \tilde{\mathbf{v}}) - (p, \nabla \cdot \tilde{\mathbf{v}}) + (\nabla \times \mathbf{u} - \boldsymbol{\omega}, \tilde{\boldsymbol{\sigma}}) - (\nabla \cdot \mathbf{u}, \tilde{r}) \\
&\quad \forall \phi = \{\mathbf{u}, \boldsymbol{\omega}, p\} \in \Phi, \quad \psi = \{\tilde{\mathbf{v}}, \tilde{\boldsymbol{\sigma}}, \tilde{r}\} \in \Lambda, \\
b_2(\theta, \psi) &= (\boldsymbol{\theta}, \tilde{\mathbf{v}}) \quad \forall \theta = \{\boldsymbol{\theta}\} \in \Theta, \quad \psi = \{\tilde{\mathbf{v}}, \tilde{\boldsymbol{\sigma}}, \tilde{r}\} \in \Lambda.
\end{aligned}$$

For  $\mathbf{g} \in \mathbf{H}^{-1}(\Omega)$ , we also define the linear functional

$$\langle g, \psi \rangle_{\Lambda^*, \Lambda} = \langle \mathbf{g}, \tilde{\mathbf{v}} \rangle_{\mathbf{H}^{-1}(\Omega), \mathbf{H}_0^1(\Omega)} \quad \forall \psi = \{\tilde{\mathbf{v}}, \tilde{\boldsymbol{\sigma}}, \tilde{r}\} \in \Lambda.$$

The operators associated with the bilinear forms are then

$$\begin{aligned}
A_1 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{for Case I,} \quad A_1 = \begin{pmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{for Case II,} \\
A_2 &= \delta I, \quad B_1 = \begin{pmatrix} 0 & \nabla \times & \nabla \\ \nabla \times & -I & 0 \\ -\nabla \cdot & 0 & 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix}.
\end{aligned} \tag{5.7}$$

It is now easily seen that the functionals  $\mathcal{J}_1(\cdot, \cdot)$  and  $\mathcal{J}_2(\cdot, \cdot; \cdot)$  defined in (5.5) and (5.3), respectively, can be written in the form (3.2). Likewise, the Stokes system (5.4) can be written in the form (3.4). Thus, the two optimization problems for the Stokes system can both be written in the form (3.6), with  $\mathcal{J}(\cdot, \cdot)$  being either  $\mathcal{J}_1(\cdot, \cdot)$  or  $\mathcal{J}_2(\cdot, \cdot)$  as appropriate. Thus, if the assumptions (3.3) and (3.5) can be verified in the context of the two optimization problems for the Stokes system, then all the results of §4 will apply to those systems.

**PROPOSITION 5.1.** *Let the spaces  $\Phi$ ,  $\widehat{\Phi}$ ,  $\Theta$ , and  $\Lambda$  and the bilinear forms  $a_1(\cdot, \cdot)$ ,  $a_2(\cdot, \cdot)$ ,  $b_1(\cdot, \cdot)$ , and  $b_2(\cdot, \cdot)$  be defined as in §5.1. Then, the assumptions (3.3) and (3.5) are satisfied.*

*Proof.* The four inequalities in (3.3) and the first two inequalities in (3.5) are easily verified with  $C_1 = 1$ ,  $C_2 = \delta$ ,  $K_2 = \delta$ ,  $c_1 = 3$ , and  $c_2 = 1$ . The third inequality in (3.5) is verified if, for any  $\phi = \{\mathbf{u}, \boldsymbol{\omega}, p\} \in \Phi$ , one can find a  $\tilde{\psi} = \{\tilde{\mathbf{v}}, \tilde{\boldsymbol{\sigma}}, \tilde{r}\} \in \Lambda$  such that

$$\begin{aligned}
b_1(\phi, \tilde{\psi}) &= (\boldsymbol{\omega}, \nabla \times \tilde{\mathbf{v}}) - (p, \nabla \cdot \tilde{\mathbf{v}}) + (\nabla \times \mathbf{u} - \boldsymbol{\omega}, \tilde{\boldsymbol{\sigma}}) - (\nabla \cdot \mathbf{u}, \tilde{r}) \\
&\geq k_1 \|\phi\|_{\Phi} \|\tilde{\psi}\|_{\Lambda} = k_1 (\|\mathbf{u}\|_1^2 + \|\boldsymbol{\omega}\|_0^2 + \|p\|_0^2)^{1/2} (\|\tilde{\mathbf{v}}\|_1^2 + \|\tilde{\boldsymbol{\sigma}}\|_0^2 + \|\tilde{r}\|_0^2)^{1/2}
\end{aligned}$$

for some constant  $k_1 > 0$ . To this end, for any  $\phi = \{\mathbf{u}, \boldsymbol{\omega}, p\} \in \Phi$ , let  $\tilde{\psi} = \{\tilde{\mathbf{v}}, \tilde{\boldsymbol{\sigma}}, \tilde{r}\} \in \Lambda$  satisfy the system

$$\nabla \times \tilde{\mathbf{v}} = \boldsymbol{\omega}, \quad \nabla \cdot \tilde{\mathbf{v}} = -p, \quad \tilde{\boldsymbol{\sigma}} = \nabla \times \mathbf{u}, \quad \text{and} \quad \tilde{r} = -\nabla \cdot \mathbf{u}$$

in  $\Omega$ . Clearly, from the last two equations, we have that

$$\|\tilde{\boldsymbol{\sigma}}\|_0^2 + \|\tilde{r}\|_0^2 = \|\nabla \times \mathbf{u}\|_0^2 + \|\nabla \cdot \mathbf{u}\|_0^2 \leq \tilde{C}_2 \|\mathbf{u}\|_1^2.$$

Also, since  $\tilde{\mathbf{v}} \in \mathbf{H}_0^1(\Omega)$ , we have from the first two equations and (5.6) that

$$\tilde{C}_1 \|\tilde{\mathbf{v}}\|_1^2 \leq (\|\nabla \times \tilde{\mathbf{v}}\|_0^2 + \|\nabla \cdot \tilde{\mathbf{v}}\|_0^2) = (\|\boldsymbol{\omega}\|_0^2 + \|p\|_0^2).$$

Combining the last two results yields that

$$\left( \|\tilde{\mathbf{v}}\|_1^2 + \|\tilde{\boldsymbol{\sigma}}\|_0^2 + \|\tilde{r}\|_0^2 \right) \leq \max \left\{ \frac{1}{\tilde{C}_1}, \tilde{C}_2 \right\} \left( \|\mathbf{u}\|_1^2 + \|\boldsymbol{\omega}\|_0^2 + \|p\|_0^2 \right).$$

Then, with  $\phi = \{\mathbf{u}, \boldsymbol{\omega}, p\} \in \Phi$  and  $\tilde{\psi} = \{\tilde{\mathbf{v}}, \tilde{\boldsymbol{\sigma}}, \tilde{r}\} \in \Lambda$ , we have that

$$\begin{aligned} b_1(\phi, \tilde{\psi}) &= \|\boldsymbol{\omega}\|_0^2 + \|p\|_0^2 + \|\nabla \times \mathbf{u}\|_0^2 + \|\nabla \cdot \mathbf{u}\|_0^2 - (\boldsymbol{\omega}, \nabla \times \mathbf{u}) \\ &\geq \|\boldsymbol{\omega}\|_0^2 + \|p\|_0^2 + \|\nabla \times \mathbf{u}\|_0^2 + \|\nabla \cdot \mathbf{u}\|_0^2 - \|\boldsymbol{\omega}\|_0 \|\nabla \times \mathbf{u}\|_0 \\ &\geq \frac{1}{2} \|\boldsymbol{\omega}\|_0^2 + \|p\|_0^2 + \frac{1}{2} \|\nabla \times \mathbf{u}\|_0^2 + \|\nabla \cdot \mathbf{u}\|_0^2 \\ &\geq \frac{1}{2} \min\{1, \tilde{C}_1\} \left( \|\mathbf{u}\|_1^2 + \|\boldsymbol{\omega}\|_0^2 + \|p\|_0^2 \right) \\ &\geq \frac{\min\{1, \tilde{C}_1\}}{2\sqrt{\max\{\frac{1}{\tilde{C}_1}, \tilde{C}_2\}}} \left( \|\mathbf{u}\|_1^2 + \|\boldsymbol{\omega}\|_0^2 + \|p\|_0^2 \right)^{1/2} \left( \|\tilde{\mathbf{v}}\|_1^2 + \|\tilde{\boldsymbol{\sigma}}\|_0^2 + \|\tilde{r}\|_0^2 \right)^{1/2} \\ &= \frac{\min\{1, \tilde{C}_1\}}{2\sqrt{\max\{\frac{1}{\tilde{C}_1}, \tilde{C}_2\}}} \|\phi\|_{\Phi} \|\tilde{\psi}\|_{\Lambda}. \end{aligned}$$

Thus, with  $k_1 = \min\{1, \tilde{C}_1\} / \left( 2\sqrt{\max\{\frac{1}{\tilde{C}_1}, \tilde{C}_2\}} \right)$ , the third inequality in (3.5) is verified. Note that  $k_1$  depends only on the comparability constants in (5.6).  $\square$

REMARK 5.2. We have now verified the assumptions (3.3) and (3.5) for the two optimization problems of finding  $(\mathbf{u}, \boldsymbol{\omega}, p, \boldsymbol{\theta})$  that minimizes either the functional in (5.3) or (5.5), subject to the Stokes system in the form (5.4) being satisfied. Thus, all the results of Sections 3.1 and 3.2 hold. In particular, with the associations already defined between spaces, operators, etc., we could use the Lagrange multiplier rule to characterize the solutions of the optimization problems as solutions of the optimality system (3.19).

REMARK 5.3. We could apply, as in §3.3, a Galerkin finite element method for determining approximate solutions of the optimality system (3.19). Such an approach, unlike least-squares finite element discretizations, does not involve the “squaring” of operators so that there is no need to transform the Stokes system (5.1) into an equivalent first-order form as in (5.4); one would then also use the form (5.2) for the functional  $\mathcal{J}_1$  instead of the form (5.5). We then would have  $\phi = \{\mathbf{u}, p\}$ ,  $\theta = \{\boldsymbol{\theta}\}$ , etc., and use, instead of the operators defined in (5.7), the operators

$$\begin{aligned} A_1 &= \begin{pmatrix} \nabla \times & 0 \\ 0 & 0 \end{pmatrix} \quad \text{for Case I,} & A_1 &= \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \quad \text{for Case II,} \\ A_2 &= \delta I, & B_1 &= \begin{pmatrix} -\Delta & \nabla \\ -\nabla \cdot & 0 \end{pmatrix}, & B_2 &= \begin{pmatrix} I \\ 0 \end{pmatrix}. \end{aligned} \tag{5.8}$$

The assumptions (3.3) and (3.5) can also be verified for the bilinear forms associated with these operators.

REMARK 5.4. As noted in §3.3, a Galerkin discretization of the optimality system (3.19) using either of the forms (5.7) or (5.8) for the operators, requires that the

assumptions in (3.22) hold. If one uses (5.8), one can easily show that the finite element spaces for the velocity and pressure approximations have to satisfy the inf-sup condition [12, 13, 16, 17]

$$\inf_{q^h \in S^h, q^h \neq 0} \sup_{\mathbf{v}^h \in \mathbf{V}^h, \mathbf{v}^h \neq \mathbf{0}} \frac{\int_{\Omega} q^h \nabla \cdot \mathbf{v}^h d\Omega}{\|q^h\|_0 \|\mathbf{v}^h\|_1} \geq \gamma \quad (5.9)$$

for some constant  $\gamma > 0$ . This condition guarantees the unique solvability of the discrete Stokes system and restricts the choice of finite element spaces used for the velocity and pressure approximations; see [12, 13, 17] for details. In particular, one cannot use piecewise polynomial spaces of the same order and defined with respect to the same grid for the velocity and pressure approximations. If one instead uses (5.7), an even more onerous inf-sup condition is required of the finite element spaces for the velocity, vorticity, and pressure.

REMARK 5.5. Note that (5.9) is a *third* level of inf-sup conditions that we have encountered in our deliberations: (5.9) is necessary and sufficient to guarantee that the inf-sup condition (3.22) holds; the latter is necessary and sufficient to guarantee that the inf-sup condition (2.12) holds.

**5.2. Least-squares finite element methods for the two optimization problems.** Using the associations of spaces and variables defined in §5.1 and the operators defined in (5.7), it is easy to see that the least-squares functional (4.14) is given by, for the example problems we are considering,

$$\begin{aligned} & \mathcal{K}(\{\mathbf{u}, \boldsymbol{\omega}, p\}, \boldsymbol{\theta}, \{\mathbf{v}, \boldsymbol{\sigma}, q\}; \hat{\mathbf{u}}, \mathbf{g}) \\ &= \|\nabla \times \boldsymbol{\sigma} + \nabla q + \delta_2(\mathbf{u} - \hat{\mathbf{u}})\|_{-1}^2 + \|\nabla \times \mathbf{v} - \boldsymbol{\sigma} + \delta_1 \boldsymbol{\omega}\|_0^2 + \|\nabla \cdot \mathbf{v}\|_0^2 \\ & \quad + \|\delta \boldsymbol{\theta} + \mathbf{v}\|_0^2 \\ & \quad + \|\nabla \times \boldsymbol{\omega} + \nabla p + \boldsymbol{\theta} - \mathbf{g}\|_{-1}^2 + \|\nabla \times \mathbf{u} - \boldsymbol{\omega}\|_0^2 + \|\nabla \cdot \mathbf{u}\|_0^2, \end{aligned} \quad (5.10)$$

where

$$\delta_1 = \begin{cases} 1 & \text{for Case I} \\ 0 & \text{for Case II} \end{cases} \quad \text{and} \quad \delta_2 = \begin{cases} 0 & \text{for Case I} \\ 1 & \text{for Case II.} \end{cases}$$

We also have the bilinear form

$$\begin{aligned} & B(\{\mathbf{u}, \boldsymbol{\omega}, p\}, \boldsymbol{\theta}, \{\mathbf{v}, \boldsymbol{\sigma}, q\}; \{\tilde{\mathbf{u}}, \tilde{\boldsymbol{\omega}}, \tilde{p}\}, \tilde{\boldsymbol{\theta}}, \{\tilde{\mathbf{v}}, \tilde{\boldsymbol{\sigma}}, \tilde{q}\}) \\ &= \left( \nabla \times \boldsymbol{\sigma} + \nabla q + \delta_2 \mathbf{u}, \nabla \times \tilde{\boldsymbol{\sigma}} + \nabla \tilde{q} + \delta_2 \tilde{\mathbf{u}} \right)_{-1} \\ & \quad + \left( \nabla \times \mathbf{v} - \boldsymbol{\sigma} + \delta_1 \boldsymbol{\omega}, \nabla \times \tilde{\mathbf{v}} - \tilde{\boldsymbol{\sigma}} + \delta_1 \tilde{\boldsymbol{\omega}} \right) + \left( \nabla \cdot \mathbf{v}, \nabla \cdot \tilde{\mathbf{v}} \right) \\ & \quad + \left( \delta \boldsymbol{\theta} + \mathbf{v}, \delta \tilde{\boldsymbol{\theta}} + \tilde{\mathbf{v}} \right) \\ & \quad + \left( \nabla \times \boldsymbol{\omega} + \nabla p + \boldsymbol{\theta}, \nabla \times \tilde{\boldsymbol{\omega}} + \nabla \tilde{p} + \tilde{\boldsymbol{\theta}} \right)_{-1} \\ & \quad + \left( \nabla \times \mathbf{u} - \boldsymbol{\omega}, \nabla \times \tilde{\mathbf{u}} - \tilde{\boldsymbol{\omega}} \right) + \left( \nabla \cdot \mathbf{u}, \nabla \cdot \tilde{\mathbf{u}} \right), \end{aligned} \quad (5.11)$$

where  $(\cdot, \cdot)_{-1}$  denotes the inner product in  $\mathbf{H}^{-1}(\Omega)$ , and the linear functional

$$\begin{aligned} F(\{\tilde{\mathbf{u}}, \tilde{\boldsymbol{\omega}}, \tilde{p}\}, \tilde{\boldsymbol{\theta}}, \{\tilde{\mathbf{v}}, \tilde{\boldsymbol{\sigma}}, \tilde{q}\}; \hat{\mathbf{u}}, \mathbf{g}) \\ = \left( \delta_2 \hat{\mathbf{u}}, \nabla \times \tilde{\boldsymbol{\sigma}} + \nabla \tilde{q} + \delta_2 \tilde{\mathbf{u}} \right)_{-1} + \left( \mathbf{g}, \nabla \times \tilde{\boldsymbol{\omega}} + \nabla \tilde{p} + \tilde{\boldsymbol{\theta}} \right)_{-1}. \end{aligned} \quad (5.12)$$

Then, as in (4.16), we have that the unique minimizer of the least-squares functional (5.10) can be characterized as being the solution of the problem: find  $\{\mathbf{u}, \boldsymbol{\omega}, p\} \in \mathbf{H}_0^1(\Omega) \times \mathbf{L}^2(\Omega) \times L_0^2(\Omega)$ ,  $\boldsymbol{\theta} \in \mathbf{L}^2(\Omega)$ , and  $\{\mathbf{v}, \boldsymbol{\sigma}, q\} \in \mathbf{H}_0^1(\Omega) \times \mathbf{L}^2(\Omega) \times L_0^2(\Omega)$  such that

$$\begin{aligned} B(\{\mathbf{u}, \boldsymbol{\omega}, p\}, \boldsymbol{\theta}, \{\mathbf{v}, \boldsymbol{\sigma}, q\}; \{\tilde{\mathbf{u}}, \tilde{\boldsymbol{\omega}}, \tilde{p}\}, \tilde{\boldsymbol{\theta}}, \{\tilde{\mathbf{v}}, \tilde{\boldsymbol{\sigma}}, \tilde{q}\}; \hat{\mathbf{u}}, \mathbf{g}) \\ \forall \{\tilde{\mathbf{u}}, \tilde{\boldsymbol{\omega}}, \tilde{p}\} \in \mathbf{H}_0^1(\Omega) \times \mathbf{L}^2(\Omega) \times L_0^2(\Omega), \tilde{\boldsymbol{\theta}} \in \mathbf{L}^2(\Omega), \\ \{\tilde{\mathbf{v}}, \tilde{\boldsymbol{\sigma}}, \tilde{q}\} \in \mathbf{H}_0^1(\Omega) \times \mathbf{L}^2(\Omega) \times L_0^2(\Omega). \end{aligned} \quad (5.13)$$

To define least-squares finite element approximations of the optimization problems, we first choose conforming finite element subspaces  $\mathbf{V}^h \subset \mathbf{H}_0^1(\Omega)$ ,  $\mathbf{W}^h \subset \mathbf{L}^2(\Omega)$ ,  $S^h \subset L_0^2(\Omega)$ , and  $\mathbf{T}^h \subset \mathbf{L}^2(\Omega)$ . We then minimize the functional in (5.10) over the subspaces, or equivalently, solve the problem: find  $\{\mathbf{u}^h, \boldsymbol{\omega}^h, p^h\} \in \mathbf{V}^h \times \mathbf{W}^h \times S^h$ ,  $\boldsymbol{\theta}^h \in \mathbf{T}^h$ , and  $\{\mathbf{v}^h, \boldsymbol{\sigma}^h, q^h\} \in \mathbf{V}^h \times \mathbf{W}^h \times S^h$  such that

$$\begin{aligned} B(\{\mathbf{u}^h, \boldsymbol{\omega}^h, p^h\}, \boldsymbol{\theta}^h, \{\mathbf{v}^h, \boldsymbol{\sigma}^h, q^h\}; \{\tilde{\mathbf{u}}^h, \tilde{\boldsymbol{\omega}}^h, \tilde{p}^h\}, \tilde{\boldsymbol{\theta}}^h, \{\tilde{\mathbf{v}}^h, \tilde{\boldsymbol{\sigma}}^h, \tilde{q}^h\}) \\ = F(\{\tilde{\mathbf{u}}^h, \tilde{\boldsymbol{\omega}}^h, \tilde{p}^h\}, \tilde{\boldsymbol{\theta}}^h, \{\tilde{\mathbf{v}}^h, \tilde{\boldsymbol{\sigma}}^h, \tilde{q}^h\}; \hat{\mathbf{u}}, \mathbf{g}) \\ \forall \{\tilde{\mathbf{u}}^h, \tilde{\boldsymbol{\omega}}^h, \tilde{p}^h\} \in \mathbf{V}^h \times \mathbf{W}^h \times S^h, \tilde{\boldsymbol{\theta}}^h \in \mathbf{T}^h, \\ \{\tilde{\mathbf{v}}^h, \tilde{\boldsymbol{\sigma}}^h, \tilde{q}^h\} \in \mathbf{V}^h \times \mathbf{W}^h \times S^h. \end{aligned} \quad (5.14)$$

Proposition (5.1) and the results of §4 allow us to prove the following results.

**THEOREM 5.6.** *Let  $\Phi = \Lambda = \mathbf{H}_0^1(\Omega) \times \mathbf{L}^2(\Omega) \times L_0^2(\Omega)$  and let  $\Theta = \mathbf{L}^2(\Omega)$ . Then,*

*i) the bilinear form  $B(\cdot; \cdot)$  defined in (5.11) is symmetric, continuous, and coercive on  $\{\Phi \times \Theta \times \Lambda\} \times \{\Phi \times \Theta \times \Lambda\}$ .*

*Let  $\hat{\mathbf{u}} \in \mathbf{L}^2(\Omega)$  and  $\mathbf{g} \in \mathbf{H}^{-1}(\Omega)$  be given. Then,*

*ii) the linear functional  $F(\cdot)$  defined in (5.12) is continuous on  $\{\Phi \times \Theta \times \Lambda\}$ ;*

*iii) the problem (5.13) has a unique solution  $(\{\mathbf{u}, \boldsymbol{\omega}, p\}, \boldsymbol{\theta}, \{\mathbf{v}, \boldsymbol{\sigma}, q\}) \in \Phi \times \Theta \times \Lambda$ .*

*Let  $\mathbf{V}^h \subset \mathbf{H}_0^1(\Omega)$ ,  $\mathbf{W}^h \subset \mathbf{L}^2(\Omega)$ ,  $S^h \subset L_0^2(\Omega)$ , and  $\mathbf{T}^h \subset \mathbf{L}^2(\Omega)$  and let  $\Phi^h = \Lambda^h = \mathbf{V}^h \times \mathbf{W}^h \times S^h$  and  $\Theta^h = \mathbf{T}^h$ . Then,*

*iv) the discrete problem (5.13) has a unique solution  $(\{\mathbf{u}^h, \boldsymbol{\omega}^h, p^h\}, \boldsymbol{\theta}^h, \{\mathbf{v}^h, \boldsymbol{\sigma}^h, q^h\}) \in \Phi^h \times \Theta^h \times \Lambda^h$ ;*

*v) we have the error estimate:*

$$\begin{aligned}
& \|\mathbf{u} - \mathbf{u}^h\|_1 + \|\boldsymbol{\omega} - \boldsymbol{\omega}^h\|_0 + \|p - p^h\|_0 + \|\boldsymbol{\theta} - \boldsymbol{\theta}^h\|_0 \\
& \quad + \|\mathbf{v} - \mathbf{v}^h\|_1 + \|\boldsymbol{\sigma} - \boldsymbol{\sigma}^h\|_0 + \|q - q^h\|_0 \\
& \leq C \left( \inf_{\tilde{\mathbf{u}}^h \in \mathbf{V}^h} \|\mathbf{u} - \tilde{\mathbf{u}}^h\|_1 + \inf_{\tilde{\boldsymbol{\omega}}^h \in \mathbf{W}^h} \|\boldsymbol{\omega} - \tilde{\boldsymbol{\omega}}^h\|_0 + \inf_{\tilde{p}^h \in S^h} \|p - \tilde{p}^h\|_0 \right. \\
& \quad + \inf_{\tilde{\boldsymbol{\theta}}^h \in \mathbf{T}^h} \|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}^h\|_0 + \inf_{\tilde{\mathbf{v}}^h \in \mathbf{V}^h} \|\mathbf{v} - \tilde{\mathbf{v}}^h\|_1 \\
& \quad \left. + \inf_{\tilde{\boldsymbol{\sigma}}^h \in \mathbf{W}^h} \|\boldsymbol{\sigma} - \tilde{\boldsymbol{\sigma}}^h\|_0 + \inf_{\tilde{q}^h \in S^h} \|q - \tilde{q}^h\|_0 \right). \tag{5.15}
\end{aligned}$$

*Proof.* The results follow in a straightforward manner from Proposition 5.1 along with Lemma 4.2, Proposition 4.4, and Theorem 4.6.  $\square$

REMARK 5.7. Following Remark 4.8, the discrete problem (5.13) is equivalent to a linear algebraic system having a symmetric, positive definite coefficient matrix. In the case of a Galerkin discretization of the optimality system, the coefficient matrix is indefinite.

REMARK 5.8. Following Remark 4.9, the results in Theorem (5.6) about the solution of the discrete problem (5.13) follow merely from the conformity of the finite element subspaces, i.e., merely from the inclusions  $\mathbf{V}^h \subset \mathbf{H}_0^1(\Omega)$ ,  $\mathbf{W}^h \subset \mathbf{L}^2(\Omega)$ ,  $S^h \subset L_0^2(\Omega)$ , and  $\mathbf{T}^h \subset \mathbf{L}^2(\Omega)$ . In particular, unlike the case of Galerkin finite element discretizations of the optimality system, they do not require that the finite element spaces satisfy additional conditions such as (5.9); see Remark 5.4. In particular, in (5.13), one can choose the same degree piecewise polynomials defined with respect to the same grid for all variables.

REMARK 5.9. The discrete problem (5.13) is a rather formidable one in that it involves many unknowns, i.e., 10 scalar fields in two dimensions and 17 scalar fields in three dimensions. However, following Remark 4.10, the discrete problem (5.13) can be efficiently uncoupled, more so than is the case for Galerkin finite element discretizations of optimality systems.

REMARK 5.10. A practical Galerkin finite element discretization of the optimality system can use a formulation in terms of the operators defined in (5.8) while the least-squares based discretization employs a formulation in terms of the operators defined in (5.7). Thus, the latter approach involves more unknowns compared to the former that involves 8 scalar fields in two dimensions and 11 scalar fields in three dimensions. This apparent disadvantage of the least-squares approach should be balanced against the advantages discussed in Remarks 5.7, 5.8, and 5.9.

REMARK 5.11. Suppose one chooses continuous, piecewise polynomial finite element spaces of degree  $r$  for the approximation of all variables; this is permissible for least-squares finite element methods; see Remark 5.8. Suppose also that the solution of the optimality system satisfies  $\mathbf{u} \in \mathbf{H}^{r+1}(\Omega) \cap \mathbf{H}_0^1(\Omega)$ ,  $\boldsymbol{\omega} \in \mathbf{H}^r(\Omega)$ ,  $p \in H^r(\Omega) \cap L_0^2(\Omega)$ ,  $\boldsymbol{\theta} \in \mathbf{H}^r(\Omega)$ ,  $\mathbf{v} \in \mathbf{H}^{r+1}(\Omega) \cap \mathbf{H}_0^1(\Omega)$ ,  $\boldsymbol{\sigma} \in \mathbf{H}^r(\Omega)$ , and  $q \in H^r(\Omega) \cap L_0^2(\Omega)$ . Then, the error estimate (5.15) implies that

$$\begin{aligned}
& \|\mathbf{u} - \mathbf{u}^h\|_1 + \|\boldsymbol{\omega} - \boldsymbol{\omega}^h\|_0 + \|p - p^h\|_0 + \|\boldsymbol{\theta} - \boldsymbol{\theta}^h\|_0 \\
& \quad + \|\mathbf{v} - \mathbf{v}^h\|_1 + \|\boldsymbol{\sigma} - \boldsymbol{\sigma}^h\|_0 + \|q - q^h\|_0 = O(h^r), \tag{5.16}
\end{aligned}$$

where  $h$  is a measure of the grid size.

**5.2.1. Circumventing the use of negative norms.** The least-squares functional (5.10) makes use of the  $\mathbf{H}^{-1}(\Omega)$  norm. As a result, both the bilinear form  $B(\cdot; \cdot)$

and the linear functional  $F(\cdot)$  appearing in least-squares finite element discretization (5.14) of the optimality system involve the  $H^{-1}(\Omega)$  inner product  $(\cdot, \cdot)_{-1}$ . Computing the  $H^{-1}(\Omega)$  inner product of two functions essentially requires the solution of a Poisson problem, i.e., for two functions  $\omega, \sigma \in H^{-1}(\Omega)$ , we can write

$$(\omega, \sigma)_{-1} = \int_{\Omega} \omega v \, d\Omega, \quad \text{where} \quad -\Delta v = \sigma \quad \text{in } \Omega \quad \text{and} \quad v = 0 \quad \text{on } \Gamma.$$

Having to solve a Poisson problem every time one has to evaluate the  $H^{-1}(\Omega)$  inner product of two functions renders impractical the implementation of (5.14).

There is substantial temptation to avoid the appearance of negative norms in the least-square finite element formulation by simply replacing the negative norm in (5.10) with the  $\mathbf{L}^2(\Omega)$  norm, i.e., to base a least-squares finite element method on the functional

$$\begin{aligned} \tilde{\mathcal{K}}(\{\mathbf{u}, \boldsymbol{\omega}, p\}, \boldsymbol{\theta}, \{\mathbf{v}, \boldsymbol{\sigma}, q\}; \hat{\mathbf{u}}, \mathbf{g}) \\ = \|\nabla \times \boldsymbol{\sigma} + \nabla q + \delta_2(\mathbf{u} - \hat{\mathbf{u}})\|_0^2 + \|\nabla \times \mathbf{v} - \boldsymbol{\sigma} + \delta_1 \boldsymbol{\omega}\|_0^2 + \|\nabla \cdot \mathbf{v}\|_0^2 \\ + \|\delta \boldsymbol{\theta} + \mathbf{v}\|_0^2 + \|\nabla \times \boldsymbol{\omega} + \nabla p + \boldsymbol{\theta} - \mathbf{g}\|_0^2 + \|\nabla \times \mathbf{u} - \boldsymbol{\omega}\|_0^2 + \|\nabla \cdot \mathbf{u}\|_0^2 \end{aligned}$$

instead of the functional (5.10); note that now we would have to choose  $\Phi = \Lambda = \mathbf{H}_0^1(\Omega) \times \mathbf{H}^1(\Omega) \times L_0^2(\Omega)$ . Doing this would indeed lead to a discrete problem involving only easily implementable  $L^2(\Omega)$  inner products. However, in this case, the norm-equivalence relation does not hold (see [6, 7]) so that the resulting bilinear form associated with the minimization of the functional  $\tilde{\mathcal{K}}$  is not coercive. As a result, the discrete problem will not have a (uniformly, as  $h \rightarrow 0$ ) positive definite coefficient matrix and the least-squares finite element approximations may not be stable and will certainly not be optimally accurate.

Another approach for avoiding the use of  $H^{-1}(\Omega)$  inner products is to replace the velocity-vorticity-pressure formulation (5.4) of the Stokes problem with another first-order formulation whose residuals, when measured in  $L^2(\Omega)$  norms, do result in a norm equivalent functional. Such formulations, involving additional unknowns and redundant equations, were developed in [14, 15]. For example, if we use the velocity-velocity gradient-pressure formulation due to [14], we would employ the least-squares functional

$$\begin{aligned} \hat{\mathcal{K}}(\{\mathbf{u}, \mathbf{U}, p\}, \boldsymbol{\theta}, \{\mathbf{v}, \mathbf{V}, q\}; \hat{\mathbf{u}}, \mathbf{g}) \\ = \|\nabla \cdot \mathbf{V} + \nabla q + \delta_2(\mathbf{u} - \hat{\mathbf{u}})\|_0^2 + \|(\nabla \mathbf{v})^T - \mathbf{V} + \delta_1 \boldsymbol{\omega}\|_0^2 + \|\nabla \cdot \mathbf{v}\|_0^2 \\ + \|\delta \boldsymbol{\theta} + \mathbf{v}\|_0^2 + \|\nabla \cdot \mathbf{U} + \nabla p + \boldsymbol{\theta} - \mathbf{g}\|_0^2 + \|(\nabla \mathbf{u})^T - \mathbf{U}\|_0^2 + \|\nabla \cdot \mathbf{u}\|_0^2 \\ + \|\nabla(\text{Tr} \mathbf{V})\|_0^2 + \|\nabla \times \mathbf{V}\|_0^2 + \|\nabla(\text{Tr} \mathbf{U})\|_0^2 + \|\nabla \times \mathbf{U}\|_0^2, \end{aligned}$$

where  $(\cdot)^T$  and  $\text{Tr}(\cdot)$  denote the transpose and the trace of a tensor, and where the components of  $\boldsymbol{\omega}$  can be easily expressed as linear combinations of the off-diagonal elements of the tensor  $\mathbf{U}$ . Instead of the vorticity  $\boldsymbol{\omega} = \nabla \times \mathbf{u}$  and adjoint vorticity  $\boldsymbol{\sigma} = \nabla \times \mathbf{v}$ , the new variables introduced to effect the first-order formulation are  $\mathbf{U} = (\nabla \mathbf{u})^T$  and  $\mathbf{V} = (\nabla \mathbf{v})^T$ . Also, now we have that  $\Phi = \Lambda = \mathbf{H}_0^1(\Omega) \times Q \times L_0^2(\Omega)$ , where  $Q = \{\mathbf{V} \in [H^1(\Omega)]^9 \mid \mathbf{V} \times \mathbf{n} = 0\}$ . The equations whose residuals appear in the last line of the definition of  $\hat{\mathcal{K}}$  are all redundant in the sense that they are all already

implied by the other equations. Note that now we have even more unknowns than that for the velocity-vorticity-pressure; e.g., in three dimensions, the least-squares discrete problem resulting from minimizing the functional  $\tilde{\mathcal{K}}$  would now involve 27 scalar fields. Furthermore, the addition of redundant equations requires more regular data and solutions and preclude the use of the least-squares methodology in, e.g., non-convex polygonal domains.

A third and more practical approach to avoiding the use of  $H^{-1}(\Omega)$  inner products, is to replace the functional (5.10) by the mesh-weighted functional

$$\begin{aligned} \tilde{\mathcal{K}}_h(\{\mathbf{u}, \boldsymbol{\omega}, p\}, \boldsymbol{\theta}, \{\mathbf{v}, \boldsymbol{\sigma}, q\}; \hat{\mathbf{u}}, \mathbf{g}) \\ = h^2 \|\nabla \times \boldsymbol{\sigma} + \nabla q + \delta_2(\mathbf{u} - \hat{\mathbf{u}})\|_0^2 + \|\nabla \times \mathbf{v} - \boldsymbol{\sigma} + \delta_1 \boldsymbol{\omega}\|_0^2 + \|\nabla \cdot \mathbf{v}\|_0^2 \\ + \|\delta \boldsymbol{\theta} + \mathbf{v}\|_0^2 + h^2 \|\nabla \times \boldsymbol{\omega} + \nabla p + \boldsymbol{\theta} - \mathbf{g}\|_0^2 + \|\nabla \times \mathbf{u} - \boldsymbol{\omega}\|_0^2 + \|\nabla \cdot \mathbf{u}\|_0^2. \end{aligned}$$

This approach is motivated by the finite element inverse inequality  $C\|\omega^h\|_0 \leq h^{-1}\|\omega^h\|_{-1}$  which leads to the norm “equivalence”  $Ch\|\omega^h\|_0 \leq \|\omega^h\|_{-1} \leq \|\omega^h\|_0$  between the  $H^{-1}(\Omega)$  and  $L^2(\Omega)$  norms of finite element functions. One can then show, using the analyses developed in [6], that one obtains an optimal convergence for the functional  $\tilde{\mathcal{K}}_h$ , even though this functional is not norm-equivalent. One possible drawback of this approach is that the condition number of the resulting matrix may be too large for the practical use of some iterative solution techniques.

Perhaps the most practical approach to avoiding the use of  $H^{-1}(\Omega)$  inner products is to replace the  $H^{-1}(\Omega)$  norm terms in the functional (5.10) by more sophisticated “equivalent” discrete norms that involve only  $L^2(\Omega)$  norms. Such ideas have been widely used in the least-squares finite element literature; see, e.g., [4, 10, 11]. As noted above, the computation of negative norms requires inversion of a Laplacian operator (with zero boundary conditions). It was shown in [10] that for finite element functions, it is equivalent to use the discrete minus one inner product

$$(\omega^h, \sigma^h)_h = ((L^h + h^2 I)\omega^h, \sigma^h)_0,$$

where  $L^h$  is a discrete inverse Laplace operator (with zero boundary conditions) that is spectrally equivalent to inverse Laplace operator itself. In practice, the computation of  $L^h \omega^h$  for any finite element function  $\omega^h$  is often implemented by using a few multigrid cycles which makes its computation very efficient. The application of this approach in our context results in the minimization of the functional

$$\begin{aligned} \bar{\mathcal{K}}_h(\{\mathbf{u}^h, \boldsymbol{\omega}^h, p^h\}, \boldsymbol{\theta}^h, \{\mathbf{v}^h, \boldsymbol{\sigma}^h, q^h\}; \hat{\mathbf{u}}, \mathbf{g}) \\ = \|\nabla \times \boldsymbol{\sigma}^h + \nabla q^h + \delta_2(\mathbf{u}^h - \hat{\mathbf{u}})\|_h^2 + \|\nabla \times \mathbf{v}^h - \boldsymbol{\sigma}^h + \delta_1 \boldsymbol{\omega}^h\|_0^2 + \|\nabla \cdot \mathbf{v}^h\|_0^2 \\ + \|\delta \boldsymbol{\theta}^h + \mathbf{v}^h\|_0^2 + \|\nabla \times \boldsymbol{\omega}^h + \nabla p^h + \boldsymbol{\theta}^h - \mathbf{g}\|_h^2 + \|\nabla \times \mathbf{u}^h - \boldsymbol{\omega}^h\|_0^2 + \|\nabla \cdot \mathbf{u}^h\|_0^2, \end{aligned}$$

where  $\|\omega^h\|_h^2 = ((L^h + h^2 I)\omega^h, \omega^h)_0$ . Using the techniques of [4], it can be shown that this functional leads to a practical least-squares finite element method yielding positive definite coefficient matrices and an error estimate such as (5.16).

**6. Conclusions.** Optimization and control problems governed by PDEs are most often solved by Lagrange multiplier techniques which lead to variational equations consisting of the state system, an adjoint-state system, and optimality conditions. Galerkin discretization of these equations results in discrete problems that are

not only formidable in size but are indefinite as well, which makes their iterative solution difficult.

In this paper we formulated a new approach for the finite element discretization of the optimality system, based on the application of least-squares principles. The main advantage of this formulation is seen in the better possibilities that it affords for the uncoupling of the discrete optimality equations and their efficient iterative solution. Least-squares principles result in symmetric and positive definite algebraic systems. Moreover, for the optimization and control problems considered in this paper, these linear systems have a  $3 \times 3$  block structure where the diagonal blocks themselves are symmetric and positive definite. As an example of a simple, but convergent uncoupling strategy, we considered a block-Gauss Seidel method. To illustrate the issues involved in the formulation of effective and practical least-squares methods we considered two optimization problems for the Stokes equations.

#### REFERENCES

- [1] I. BABUSKA AND A. AZIZ, *Survey lectures on the mathematical foundations of the finite element method*, in *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations*, Academic, New York, 1972, pp. 1–359.
- [2] D. BEDIVAN AND G. FIX, *Least-squares methods for optimal shape design problems*, *Computers Math. Applic.* **30** 1995, pp. 17–25.
- [3] P. BOCHEV, *Least-squares methods for optimal control*, *Nonl. Anal.*, **30** 1997, pp. 1875–1885.
- [4] P. BOCHEV, *Negative norm least-squares methods for the velocity-vorticity-pressure Navier-Stokes equations*, *Num. Meth. PDE's* **15** 1999, pp. 237–256.
- [5] P. BOCHEV AND D. BEDIVAN, *Least-squares methods for Navier-Stokes boundary control problems*, *Int. J. Comp. Fluid Dyn.* **9** 1997, pp. 43–58.
- [6] P. BOCHEV AND M. GUNZBURGER, *Least-squares finite element methods for elliptic equations*, *SIAM Review* **40** 1998, pp. 789–837.
- [7] P. BOCHEV AND M. GUNZBURGER, *Analysis of least-squares finite element methods for the Stokes equations*, *Math. Comp.* **63** 1994, pp. 479–506.
- [8] P. BOCHEV AND M. GUNZBURGER, *Least-squares finite element methods for optimization and control problems for the Stokes equations*, *Comp. Math. Appl.*, Vol. 48, No.7, 2004, pp. 1035–1057.
- [9] P. BOCHEV AND M. GUNZBURGER, *Least-squares/penalty finite element methods for optimization and control problems*, to appear.
- [10] 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, 1994.
- [11] J. BRAMBLE AND J. PASCIAK, *Least-squares methods for Stokes equations based on a discrete minus one inner product*, *J. Comp. App. Math.* **74** 1996, pp. 155–173.
- [12] F. BREZZI, *On the existence, uniqueness and approximation of saddle-point problems arising from Lagrange multipliers*, *RAIRO Anal. Numer.* R2 ?? 1974, pp. 129–151.
- [13] F. BREZZI AND M. FORTIN, *Mixed and Hybrid Finite Element Methods*, Springer, New York, 1991.
- [14] CAI, T. MANTEUFFEL, AND S. MCCORMICK, *First-order system least-squares for the Stokes equations, with application to linear elasticity*, *SIAM J. Num. Anal.* **34** 1997, pp.1727–1741.
- [15] C.-L. CHANG, *A mixed finite element method for the Stokes problem: an acceleration pressure formulation*, *Appl. Math. Comp.* **36** 1990, pp. 135–146.
- [16] V. GIRAULT AND P.-A. RAVIART, *Finite Element Methods for Navier-Stokes Equations*, Springer, Berlin, 1986.
- [17] M. GUNZBURGER, *Finite Element Methods for Viscous Incompressible Flows*, Academic, Boston, 1989.
- [18] M. GUNZBURGER, *Perspectives in Flow Control and Optimization*, SIAM, Philadelphia, 2003.
- [19] M. GUNZBURGER AND H.C. LEE, *Analysis and approximation of optimal control problems for first-order elliptic systems in three dimensions*, *Appl. Math. Comp.* **100** 1999, pp. 49–70.
- [20] M. GUNZBURGER AND H.C. LEE, *A penalty/least-squares method for optimal control problems for first-order elliptic systems*, *Appl. Math. Comp.* **107** 2000, pp. 57–75.

- [21] J. LIONS, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer, New York, 1971.
- [22] H. SCHLICHTING AND K. GERSTEN, *Boundary Layer Theory*, Springer, Berlin, 2000.