

Least-squares finite element methods for optimization and control problems for the Stokes equations

Pavel Bochev*

Max D. Gunzburger†

Dedicated to the memory of our friend George J. Fix

Abstract

The approximate solution of optimization and control problems for systems governed by the Stokes equations is considered. Modern computational techniques for such problems are predominantly based on the application of the Lagrange multiplier rule, while penalty formulations, even though widely used in other settings, have not enjoyed the same level of popularity for this class of problems. A discussion is provided that explains why naively defined penalty methods may not be practical. Then, practical penalty methods are defined using methodologies associated with modern least-squares finite element methods. The advantages, with respect to efficiency, of penalty/least-squares methods for optimal control problems compared to methods based on Lagrange multipliers are highlighted. A tracking problem for the Stokes system is used for illustrative purposes.

1 Introduction

In many applications and for many years, optimization and control problems for systems governed by partial differential equations (PDE's) have been a subject of interest to experimentalists. For example, the control of boundary layers in fluid flows was studied by Prandtl as early as 1904 [?]. These problems have also been a subject of theoretical interest and, for almost as long as computers have been around, of computational interest as well. Most of the efforts in the latter directions have employed elementary optimization strategies. For a historical perspective of such efforts in the fluid mechanics setting, see [?]; experiences in other settings, e.g., electromagnetics, heat transfer, structural mechanics, etc., are very similar.

More recently, mathematicians, scientists, and engineers have turned to the application of sophisticated optimization strategies, e.g., Lagrange multiplier methods, sensitivity or adjoint-based gradient methods, quasi-Newton methods, evolutionary algorithms, etc., for solving optimization and control problems for systems governed by PDE's. On the mathematical side, one may credit J.-L. Lions and D. Russell for helping popularize and foment these trends.

Several popular approaches to solving optimization and control problems constrained by PDE's are based, one way or another, on optimality systems deduced from the application of the Lagrange

*Computational Mathematics and Algorithms Department, Sandia National Laboratories, Albuquerque NM 87185-1110 (pbboche@sandia.gov). Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed-Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC-94AL85000.

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

multiplier rule. This is not surprising since the Lagrange multiplier rule is a standard approach to solving finite-dimensional optimization problems. What is perhaps surprising is that penalty methods, another popular approach for the latter setting, have not engendered anywhere near as much interest for the infinite-dimensional problems that are of interest here. The main advantage of taking the penalty approach over Lagrange multiplier or sensitivity equations based methods is that the former involves fewer unknowns. The main disadvantage of the penalty approach is the relative poor conditioning of the linear algebraic systems that one must solve; this results in inefficiencies in the application of iterative methods for the solution of the linear systems.

The poor conditioning problem arises from the need to choose small values for the penalty parameter so that the error due to penalization does not dominate the discretization error. This problem can be ameliorated by invoking an iterative penalty approach (see, e.g., [?, ?]) that, at the price of having to solve multiple linear systems, can render the error due to penalization as small as one wants even for relatively large values of the penalty parameter. However, one now is faced with the possible inefficiency of having to solve more than one linear system, so that anything that can be done to reduce the difficulty and cost of effecting that solution is crucial to the success of the overall penalty-based optimization algorithm.

In this paper, we expand on these observations to show why naively defined penalty methods may not be practical. We then show how, by incorporating modern least-squares finite element methodologies, the penalty approach can be rehabilitated to yield practical and efficient algorithms for optimal control problems. These algorithms, referred hereafter as *penalty/least-squares methods*, use least-squares variational principles to treat the PDE constraints. This type of penalty methods offers certain efficiency related advantages compared to methods based on the application of Lagrange multiplier techniques and the solution of the resulting optimality system by either Galerkin or least-squares finite element methods.

Penalty/least-squares methods were pioneered by G. Fix, et al. [?] for an optimal shape design problem. This was followed in [?, ?] with a study of penalty/least-squares methods for the Dirichlet control of the Navier-Stokes system and, in [?], with like methods for optimal control problems constrained by first-order elliptic systems. In [?], an alternate approach was developed wherein least-squares principles are applied to the optimality system that results from the application of the Lagrange multiplier rule.

The paper is organized as follows. In §§?, we present mostly well-known results about general constrained optimization problems and their solution via Lagrange multiplier and penalty approaches. Then, in §§??-??, we apply the framework of §§? to optimization problems for the Stokes equations. In particular, in §§?, we consider a straightforward penalization approach and show why the resulting method is not totally practical and, in §§?, we develop a least-squares finite element approach that realizes all the potential advantages of penalty-based formulations without compromising efficiency. Finally, in §§?, we consider some practical issues that arise in the efficient implementation of the methodologies presented in §§?. Although our discussion is in the context of the Stokes equations, most of what we say applies to more general quadratic optimization problems with linear, elliptic PDE constraints.

2 Constrained minimization problems in Hilbert spaces

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 *functional*

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

the *constraint equation*

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

where $\langle \cdot, \cdot \rangle$ denotes the duality pairing, and the *constrained minimization problem*¹

$$\min_{u \in V} \mathcal{J}(u) \quad \text{subject to } (??). \quad (2.3)$$

The bilinear forms serve to define associated operators

$$A : V \rightarrow V^*, \quad B : V \rightarrow S^*, \quad \text{and} \quad B^* : S \rightarrow V^* \quad (2.4)$$

through the relations

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

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

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

where the constraint equation $Bu = g$ holds in S^* . We define the space

$$Z = \{v \in V : b(v, \xi) = 0 \quad \forall \xi \in S\} \quad (2.5)$$

and make the following assumptions about the bilinear forms:

$$\left\{ \begin{array}{ll} a(u, v) \leq C_a \|u\|_V \|v\|_V & \forall u, v \in V \\ b(u, \xi) \leq C_b \|u\|_V \|\xi\|_S & \forall u \in V, \xi \in S \\ a(u, u) \geq 0 & \forall u \in V \\ a(u, u) \geq K_a \|u\|_V^2 & \forall u \in Z \\ \sup_{v \in V, v \neq 0} \frac{b(v, \xi)}{\|v\|_V} \geq K_b \|\xi\|_S & \forall \xi \in S, \end{array} \right. \quad (2.6)$$

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. [?].

Proposition 2.1 *Let the assumptions (??) hold. Then, the constrained minimization problem (??) has a unique solution $u \in V$.*

¹The 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 the identification of functionals to be encountered in later sections with the functional (??).

2.2 Solution via Lagrange multipliers

For all $v \in V$ and $\xi \in S$, we introduce the Lagrangian functional

$$\mathcal{L}(v, \xi) = \mathcal{J}(v) + b(v, \xi) - \langle g, \xi \rangle_{S^*, S} = \frac{1}{2}a(v, v) + b(v, \xi) - \langle f, v \rangle_{V^*, V} - \langle g, \xi \rangle_{S^*, S} + t. \quad (2.7)$$

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

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

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

Proposition 2.2 *Let the assumptions (??) hold. Then, the system (??) has a unique solution $(u, \lambda) \in V \times S$ and moreover*

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

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

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

2.2.1 Approximation of the Lagrange multiplier optimality system

We choose (conforming) finite dimensional subspaces $V^h \subset V$ and $S^h \subset S$, and then restrict (??) to the subspaces, i.e., we seek $u^h \in V^h$ and $\lambda^h \in S^h$ that satisfy

$$\begin{cases} a(u^h, v^h) + b(v^h, \lambda^h) &= \langle f, v^h \rangle_{V^*, V} & \forall v^h \in V^h \\ b(u^h, \xi^h) &= \langle g, \xi^h \rangle_{S^*, S} & \forall \xi^h \in S^h. \end{cases} \quad (2.9)$$

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

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

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

Proposition 2.3 *Let the hypotheses of Proposition ?? hold and assume that*

$$a(u^h, u^h) \geq K_a^h \|u^h\|_V^2 \quad \forall u^h \in Z^h \quad (2.10)$$

and²

$$\sup_{v^h \in V^h, v^h \neq 0} \frac{b(v^h, \xi^h)}{\|v^h\|_V} \geq K_b^h \|\xi^h\|_S \quad \forall \xi^h \in S^h, \quad (2.11)$$

²The assumption (??) is commonly referred to as the (discrete) *inf-sup condition* owing to the equivalent form

$$\inf_{\xi^h \in S^h, \xi^h \neq 0} \sup_{v^h \in V^h, v^h \neq 0} \frac{b(v^h, \xi^h)}{\|v^h\|_V \|\xi^h\|_S} \geq K_b^h.$$

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

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

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

$$\|u - u^h\|_V + \|\lambda - \lambda^h\|_S \leq C \left(\inf_{v^h \in V^h} \|u - v^h\|_V + \inf_{\xi^h \in S^h} \|\lambda - \xi^h\|_S \right). \quad (2.12)$$

The discrete problem (??) is equivalent to a linear system. Indeed, let $\{U_i\}_{i=1}^n$ and $\{\Lambda_i\}_{i=1}^m$, where $n = \dim V^h$, $m = \dim S^h$, denote bases for V^h and S^h , respectively, and let $\vec{u} = (u_1, \dots, u_n)^T$ and $\vec{\lambda} = (\lambda_1, \dots, \lambda_m)^T$ denote the coefficients in the expansion of u^h and λ^h in terms of the bases. Furthermore, let $f_i = \langle f, U_i \rangle_{V^*, V}$ for $i = 1, \dots, n$, $g_i = \langle g, \Lambda_i \rangle_{S^*, S}$ 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 $\mathbb{B}_{ij} = b(U_j, \Lambda_i)$ for $i = 1, \dots, m$, $j = 1, \dots, n$, respectively. Then, (??) is equivalent to the linear system

$$\begin{pmatrix} \mathbb{A} & \mathbb{B}^T \\ \mathbb{B} & 0 \end{pmatrix} \begin{pmatrix} \vec{u} \\ \vec{\lambda} \end{pmatrix} = \begin{pmatrix} \vec{f} \\ \vec{g} \end{pmatrix}. \quad (2.13)$$

Remark 2.4 The coefficient matrix in (??) is symmetric and indefinite. This is universal for discretization of the saddle-point problems arising from the use of the Lagrange multiplier rule for constrained optimization problems.

Remark 2.5 The assumptions (??) and (??) guarantee that the $(m+n) \times (m+n)$ coefficient matrix in (??) is uniformly invertible with respect to h .

2.3 Solution via penalization

Let $d(\cdot, \cdot)$ denote a symmetric bilinear form on $S \times S$ satisfying the assumptions

$$\begin{cases} d(\lambda, \xi) \leq C_d \|\lambda\|_S \|\xi\|_S & \forall \lambda, \xi \in S \\ d(\lambda, \lambda) \geq K_d \|\lambda\|_S^2 & \forall \lambda \in S, \end{cases} \quad (2.14)$$

where both C_d and K_d are positive constants. The bilinear form $d(\cdot, \cdot)$ serves to define an invertible operator $D : S \rightarrow S^*$ through

$$d(\lambda, \xi) = \langle D\lambda, \xi \rangle_{S^*, S}.$$

Proposition 2.6 Assume that (??) and (??) hold. Then,

$$a(v, v) + \langle Bv, D^{-1}Bv \rangle_{S^*, S} \geq K \|v\|_V^2 \quad \forall v \in V, \quad (2.15)$$

where $K = \min\{K_a, K_d K_b^2 / C_d^2\}$.

Proof: For all $v \in Z^\perp$, we have from (??) and (??) that

$$a(v, v) + \langle Bv, D^{-1}Bv \rangle_{S^*, S} \geq \langle Bv, D^{-1}Bv \rangle_{S^*, S} \geq \frac{K_d}{C_d^2} \|Bv\|_{S^*}^2 \geq \frac{K_d K_b^2}{C_d^2} \|v\|_V^2.$$

Furthermore, the assumptions on $b(\cdot, \cdot)$ contained in (??) imply that the subspace $Z \subset V$ is closed and that $\|Bv\|_{S^*} \geq K_b \|v\|_V$ for all $v \in Z^\perp$; see, e.g., [?]. Then,

$$a(v, v) + \langle Bv, D^{-1}Bv \rangle_{S^*, S} = a(v, v) \geq K_a \|v\|_V^2 \quad \forall v \in Z.$$

Thus, (??) is proved. \square

Let $\epsilon > 0$ be a parameter that tends to zero and consider the *penalized functional*

$$\begin{aligned}\mathcal{J}_\epsilon(v) &= \mathcal{J}(v) + \frac{1}{2\epsilon} \left\langle Bv - g, D^{-1}(Bv - g) \right\rangle_{S^*, S} \\ &= \frac{1}{2} a(u, u) - \langle f, u \rangle_{V^*, V} + t + \frac{1}{2\epsilon} \left\langle Bv - g, D^{-1}(Bv - g) \right\rangle_{S^*, S}\end{aligned}\quad (2.16)$$

that is defined for all $v \in V$. Then, consider the unconstrained minimization problem

$$\min_{u \in V} \mathcal{J}_\epsilon(u). \quad (2.17)$$

For any fixed $\epsilon > 0$, the minimizer u_ϵ can be found by solving the optimality equation

$$a(u_\epsilon, v) + \frac{1}{\epsilon} \langle Bv, D^{-1}Bu_\epsilon \rangle_{S^*, S} = \langle f, v \rangle_{V^*, V} + \frac{1}{\epsilon} \langle Bv, D^{-1}g \rangle_{S^*, S} \quad \forall v \in V. \quad (2.18)$$

The following result follows easily from (??) and is well known; see, e.g., [?].

Proposition 2.7 *Let the assumptions (??) and (??) hold. Then, for any fixed $0 < \epsilon \leq 1$, there exists a unique $u_\epsilon \in V$ satisfying (??).*

In terms of the operators A , B , and D , (??) takes the form

$$Au_\epsilon + \frac{1}{\epsilon} B^* D^{-1} Bu_\epsilon = f + \frac{1}{\epsilon} B^* D^{-1} g \quad \text{in } V^*. \quad (2.19)$$

For any fixed $\epsilon > 0$ and for given $u_\epsilon \in V$, define $\lambda_\epsilon \in S$ through

$$\epsilon d(\lambda_\epsilon, \xi) = b(u_\epsilon, \xi) - \langle g, \xi \rangle_{S^*, S} \quad \forall \xi \in S. \quad (2.20)$$

Again, the following result is well known; see, e.g., [?].

Proposition 2.8 *Let the assumptions (??) and (??) hold. Then, for any fixed $0 < \epsilon \leq 1$ and for any given $u_\epsilon \in V$, there exists a unique $\lambda_\epsilon \in S$ satisfying (??).*

In terms of the operators B and D , we have that $\epsilon D\lambda_\epsilon = Bu_\epsilon - g$ in S^* or $\lambda_\epsilon = \frac{1}{\epsilon} D^{-1} Bu_\epsilon - \frac{1}{\epsilon} D^{-1} g$ in S . We may then write (??) in the form

$$a(u_\epsilon, v) + \langle Bv, \lambda_\epsilon \rangle_{S^*, S} = \langle f, v \rangle_{V^*, V} \quad \forall v \in V. \quad (2.21)$$

Consequently, $(u_\epsilon, \lambda_\epsilon) \in V \times S$ is the unique solution of the *regularized system*

$$\begin{cases} a(u_\epsilon, v) + b(v, \lambda_\epsilon) &= \langle f, v \rangle_{V^*, V} & \forall v \in V \\ b(u_\epsilon, \xi) - \epsilon d(\lambda_\epsilon, \xi) &= \langle g, \xi \rangle_{S^*, S} & \forall \xi \in S \end{cases} \quad (2.22)$$

or, in terms of the operators A , B , and D ,

$$\begin{cases} Au_\epsilon + B^* \lambda_\epsilon &= f & \text{in } V^* \\ Bu_\epsilon - \epsilon D\lambda_\epsilon &= g & \text{in } S^*. \end{cases} \quad (2.23)$$

Remark 2.9 Problems (??) and (??) are completely equivalent. If $(u_\epsilon, \lambda_\epsilon) \in V \times S$ solves (??), then u_ϵ solves (??). On the other hand, if $u_\epsilon \in V$ is a solution of (??), then u_ϵ and $\lambda_\epsilon \in S$, where the latter is the solution of (??), is a solution of (??). Alternately, we could have stated these equivalences using (??) and (??). Thus, penalization of $\mathcal{J}(v)$ and regularization of the Lagrange multiplier optimality system are equivalent. Note that (??) is the optimality system corresponding to the regularized Lagrangian functional

$$\mathcal{L}_\epsilon(v, \xi) = \left(\frac{1}{2}a(v, v) + b(v, \xi) - \langle f, v \rangle_{V^*, V} - \langle g, \xi \rangle_{S^*, S} + t \right) - \frac{\epsilon}{2}d(\xi, \xi) = \mathcal{L}(v, \xi) - \frac{\epsilon}{2}d(\xi, \xi)$$

which may be viewed as a regularization of the functional (??).

Once again, the following result is well known; see, e.g., [?].

Proposition 2.10 *Let the assumptions (??) and (??) hold. Let $(u, \lambda) \in V \times S$ denote the unique solution of (??) or, equivalently, of the optimization problem (??), and, for each fixed $0 < \epsilon \leq 1$, let $(u_\epsilon, \lambda_\epsilon) \in V \times S$ denote the unique solution of (??) or, equivalently, of (??) and (??). Then,*

$$\|u - u_\epsilon\|_V + \|\lambda - \lambda_\epsilon\|_S \leq \epsilon C(\|f\|_{V^*} + \|g\|_{S^*}) \quad (2.24)$$

so that $u_\epsilon \rightarrow u$ and $\lambda_\epsilon \rightarrow \lambda$ as $\epsilon \rightarrow 0$.

Remark 2.11 An iterative process may defined through which a sequence of solutions of penalty systems can be sequentially determined and for which the differences between the members of the sequence and the solution of the constrained minimization problem (??) are of $O(\epsilon^k)$, where k is the index of the sequence. In this way, at the cost of an iteration, the penalty solutions can be made as accurate as desired. See, e.g., [?, ?] for details.

2.3.1 Approximation of penalty solution

To approximate (u_ϵ, p_ϵ) , we can start with either (??) or (??). Although these two problems are equivalent, they do not engender the same discrete equations. This point will be clarified shortly.

Discretization of the optimality system. First, let us start with (??). We choose (conforming) finite dimensional subspaces $V^h \subset V$ and $S^h \subset S$ and pose (??) over the subspaces, i.e., we seek $(u_\epsilon^h, \lambda_\epsilon^h) \in V^h \times S^h$ that satisfies

$$\begin{cases} a(u_\epsilon^h, v^h) + b(v^h, \lambda_\epsilon^h) &= \langle f, v^h \rangle_{V^*, V} & \forall v^h \in V^h \\ b(u_\epsilon^h, \xi^h) - \epsilon d(\lambda_\epsilon^h, \xi^h) &= \langle g, \xi^h \rangle_{S^*, S} & \forall \xi^h \in S^h. \end{cases} \quad (2.25)$$

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

Proposition 2.12 *Let the assumptions (??), (??), (??), and (??) hold. Then, for any fixed $0 < \epsilon \leq 1$, (??) has a unique solution $(u_\epsilon^h, \lambda_\epsilon^h) \in V^h \times S^h$ and, moreover, that solution satisfies*

$$\|u^h - u_\epsilon^h\|_V + \|\lambda^h - \lambda_\epsilon^h\|_S \leq \epsilon C(\|f\|_{V^*} + \|g\|_{S^*}), \quad (2.26)$$

where $\{u^h, \lambda^h\}$ denotes the unique solution of (??). Combining with (??), we obtain

$$\begin{aligned} \|u - u_\epsilon^h\|_V + \|\lambda - \lambda_\epsilon^h\|_S &\leq C \left(\epsilon (\|f\|_{V^*} + \|g\|_{S^*}) \right. \\ &\quad \left. + \inf_{v^h \in V^h} \|u - v^h\|_V + \inf_{\xi^h \in S^h} \|\lambda - \xi^h\|_S \right), \end{aligned} \quad (2.27)$$

where $\{u, \lambda\}$ denotes the unique solution of (??) or, equivalently, of (??).

In addition to the matrices \mathbb{A} and \mathbb{B} previously introduced, we define the $m \times m$ matrix \mathbb{D} by $\mathbb{D}_{ij} = d(\Lambda_i, \Lambda_j)$. Then, (??) is equivalent to the linear system

$$\begin{pmatrix} \mathbb{A} & \mathbb{B}^T \\ \mathbb{B} & -\epsilon \mathbb{D} \end{pmatrix} \begin{pmatrix} \vec{\mathbf{u}}_\epsilon \\ \vec{\boldsymbol{\lambda}}_\epsilon \end{pmatrix} = \begin{pmatrix} \vec{\mathbf{f}} \\ \vec{\mathbf{g}} \end{pmatrix}. \quad (2.28)$$

It is now easy to see how one can eliminate λ^h from (??) or, equivalently, $\vec{\boldsymbol{\lambda}}$ from (??). Assumptions (??) imply that the matrix \mathbb{D} is symmetric and positive definite, and therefore invertible. Then, one easily deduces from (??) that $\vec{\mathbf{u}}_\epsilon$ solves

$$\left(\mathbb{A} + \frac{1}{\epsilon} \mathbb{B}^T \mathbb{D}^{-1} \mathbb{B} \right) \vec{\mathbf{u}}_\epsilon = \vec{\mathbf{f}} + \frac{1}{\epsilon} \mathbb{B}^T \mathbb{D}^{-1} \vec{\mathbf{g}}. \quad (2.29)$$

Note that (??) only involves the approximation $u_\epsilon^h \in V^h$ of $u \in V$.

Proposition 2.13 *Let the assumptions (??), (??), (??), and (??) hold. Then, for any fixed $0 < \epsilon \leq 1$, (??) has a unique solution $\vec{\mathbf{u}}_\epsilon$.*

Proof: The assumptions imply that \mathbb{A} is symmetric and positive definite on the kernel of \mathbb{B} , that \mathbb{B} is of full row-rank, and that \mathbb{D} is positive definite. Thus, $\mathbb{A} + \frac{1}{\epsilon} \mathbb{B}^T \mathbb{D}^{-1} \mathbb{B}$ is symmetric and positive definite. \square

Once $\vec{\mathbf{u}}_\epsilon$ is determined from (??), $\vec{\boldsymbol{\lambda}}_\epsilon$ may be determined from

$$\epsilon \mathbb{D} \vec{\boldsymbol{\lambda}}_\epsilon = \mathbb{B} \vec{\mathbf{u}}_\epsilon - \vec{\mathbf{g}}. \quad (2.30)$$

Remark 2.14 The system (??) is a regular perturbation of (??). Thus, if the assumptions of Proposition ?? hold, then, since the coefficient matrix in (??) is uniformly invertible (with respect to h), so is the coefficient matrix in (??) (with respect to both h and ϵ). On the other hand, if those assumptions are not satisfied, and in particular if the condition (??) is not satisfied, then neither of the systems (??) or (??) are stably invertible. Thus, there is no advantage to solving (??) as opposed to (??). On the other hand, as we see in the next remark, there are advantages to solving (??) instead of (??).

Remark 2.15 One obvious advantage of penalty methods over Lagrange multiplier methods for problems such as (??) is that they involve less unknowns. Comparing (??) with (??), we see that the addition of the penalty term allows for the elimination of $\vec{\boldsymbol{\lambda}}_\epsilon$ to obtain (??). Thus, we may solve for $\vec{\mathbf{u}}_\epsilon$ directly from (??) which involves less equations and unknowns than does (??), and subsequently, if desired, solve for $\vec{\boldsymbol{\lambda}}_\epsilon$ from (??). Furthermore, the coefficient matrix in the system (??) is symmetric and positive definite (provided the assumptions of Proposition ?? are satisfied) while the one for the system (??) is indefinite. On the other hand, for small values of ϵ , the coefficient matrix in the penalized system (??) may be ill conditioned. Small values of ϵ need to be employed in order to balance the errors arising from penalization, i.e., the terms involving ϵ in (??), and the errors arising from approximation, i.e., the remaining terms in that estimate. By using an iterative penalty approach (see Remark ??), the estimate (??) can be replaced by one that involves terms proportional to ϵ^k (instead of ϵ), where k is the number of iterations applied. In this manner, the two types of terms in the error estimate can be balanced even if ϵ is not so small so that the conditioning of the matrix in (??) is compromised. Of course, one has to pay the price of having to solve k linear systems. Note that all these systems involve the same coefficient matrix.

Remark 2.16 At first glance it seems that (??) does not involve the space S^h ; however, S^h does enter into (??) through the definitions of the matrices \mathbb{B} and \mathbb{D} . If V^h and S^h are chosen so that (??), (??), (??), and, in particular, the discrete inf-sup condition (??) are satisfied, then we saw by Propositions ?? and ?? that (??), or equivalently, (??), is uniquely solvable and the error estimates (??) and (??) hold. If (??) is not satisfied, then, as noted above, (??) or (??) may not be stably invertible. Because (??) is still required for these problems, one should view (??) or (??) as a solution method for the discrete system (??). See Remark ??.

Discretization of the penalized problem. Instead of discretizing (??), we can instead discretize the penalized optimization problem (??). To this end, choose a conforming subspace $\tilde{V}^h \subset V$ and consider the optimization problem

$$\min_{v^h \in \tilde{V}^h} \mathcal{J}_\epsilon(v^h). \quad (2.31)$$

It is easy to see that the problem (??) is equivalent to seeking $\tilde{u}_\epsilon^h \in \tilde{V}^h$ such that

$$a(\tilde{u}_\epsilon^h, v^h) + \frac{1}{\epsilon} \langle Bv^h, D^{-1}B\tilde{u}_\epsilon^h \rangle_{S^*, S} = \langle f, v^h \rangle_{V^*, V} + \frac{1}{\epsilon} \langle Bv^h, D^{-1}g \rangle_{S^*, S} \quad (2.32)$$

for all $v^h \in \tilde{V}^h$. Obviously, (??) can be obtained by restricting (??) to the subspace $\tilde{V}^h \subset V$. It is also easily seen that (??) is equivalent to the linear system

$$\left(\tilde{\mathbb{A}} + \frac{1}{\epsilon} \tilde{\mathbb{B}} \right) \tilde{\mathbf{u}}_\epsilon = \tilde{\mathbf{f}} + \frac{1}{\epsilon} \tilde{\mathbf{g}}, \quad (2.33)$$

where $\tilde{\mathbb{A}}_{ij} = a(U_i, U_j)$, $\tilde{\mathbb{B}}_{ij} = \langle BU_i, D^{-1}BU_j \rangle_{S^*, S}$, $\tilde{\mathbf{f}}_i = \langle f, U_i \rangle_{V^*, V}$, and $\tilde{\mathbf{g}}_i = \langle BU_i, D^{-1}g \rangle_{S^*, S}$ for $i, j = 1, \dots, n$.

Proposition 2.17 *Let the assumptions (??), (??), and (??) hold. Assume further that*

$$a(v^h, v^h) + \langle Bv^h, D^{-1}Bv^h \rangle_{S^*, S} \geq K^h \|v^h\|_V^2 \quad \forall v^h \in V^h \quad (2.34)$$

for some positive constant K^h . Then, for any fixed $0 < \epsilon \leq 1$, the coefficient matrix in (??) is positive definite as well as symmetric so that that equation has a unique solution $\tilde{\mathbf{u}}_\epsilon$, or equivalently, (??) has a unique solution \tilde{u}_ϵ^h .

Proof: Assumption (??) easily implies that, for $0 < \epsilon \leq 1$, the coefficient matrix in (??) is positive definite. \square

Proposition 2.18 *Let the assumptions (??), (??), (??), and (??) hold. Then, (??) holds with $K^h = \min\{K_a^h, K_d(K_b^h)^2/C_d^2\}$ so that, for any fixed $0 < \epsilon \leq 1$, the matrix in (??) is symmetric and positive definite and (??) has a unique solution $\tilde{\mathbf{u}}_\epsilon$, or equivalently, (??) has a unique solution \tilde{u}_ϵ^h .*

Proof: The proof follows the same lines as that for Proposition ??.

Remark 2.19 Proposition ?? shows that the discrete penalty equation will have a unique solution as long as one can verify the coercivity assumption (??). From Proposition ??, it is clear that (??) is a sufficient condition for (??) to hold. It is also clear that (??) is not a necessary condition for (??) and so the discrete penalty equation may have a unique solution even in cases when the discrete inf-sup condition does not hold. However, (??) is necessary to prove that the discrete penalty solution converges to the correct solution as $\epsilon \rightarrow 0$; see Remarks ?? and ??.

Remark 2.20 As it stands, (??) does not, in general define a practical method. The need to invert the operator D in order to determine both $\tilde{\mathbb{B}}$ and $\tilde{\mathbf{g}}$ is usually not possible except in the case of $S = S^*$ and D the identity operator. In other cases, one can replace $\tilde{\mathbb{B}}$ and $\tilde{\mathbf{g}}$ by

$$\tilde{\mathbb{B}}_{ij}^h = (U_i, U_j)_{*,h} \quad \forall i, j = 1, \dots, n \quad \text{and} \quad (\tilde{\mathbf{g}})_i^h = \langle U_i, g \rangle_{*,h} \quad \forall i = 1, \dots, n,$$

where $(\cdot, \cdot)_{*,h}$ and $\langle \cdot, \cdot \rangle_{*,h}$ are a mesh-dependent inner product and a mesh-dependent duality pairing whose definitions require the definition of a discrete approximation to the operator D^{-1} and may also require a discrete approximation to the operator B . We will return to this issue in §???. Note that, on the other hand, (??) involves the inverse of *the discrete operator* \mathbb{D} so that it can be implemented for any bilinear form $d(\cdot, \cdot)$ that satisfies the assumptions (??).

Remark 2.21 The advantages of penalty methods over Lagrange multiplier methods for problems such as (??) that are discussed in Remark ?? for the system (??) also apply to (??). Comparing (??) with (??), the former involves less equations and unknowns and has a coefficient matrix that is symmetric and positive definite.

Remark 2.22 Clearly, $\tilde{\mathbb{A}} + \frac{1}{\epsilon} \tilde{\mathbb{B}} \neq \mathbb{A} + \frac{1}{\epsilon} \mathbb{B}^T \mathbb{D}^{-1} \mathbb{B}$ and $\tilde{\mathbf{f}} + \frac{1}{\epsilon} \tilde{\mathbf{g}} \neq \vec{\mathbf{f}} + \frac{1}{\epsilon} \mathbb{B}^T \mathbb{D}^{-1} \vec{\mathbf{g}}$ so that (??) and (??) are not the same even though the parent infinite-dimensional problems (??) and (??), respectively, are equivalent. Note that (??) is obtained by first discretizing the infinite-dimensional regularized optimality system (??) to obtain (??) and then eliminating the discrete Lagrange multiplier $\vec{\lambda}_\epsilon$ from the latter. On the other hand, (??) can be viewed as being obtained by first eliminating the Lagrange multiplier λ_ϵ from (??) to obtain (??) and then discretizing the latter. Clearly, in general, the two steps do not commute. Thus, discretizations of the regularized optimality system and the penalized optimization problem do not yield the same approximations to the solution of (??), i.e., in general, $\tilde{\mathbf{u}}_\epsilon \neq \vec{\mathbf{u}}_\epsilon$.

Remark 2.23 It is clear that (??) is defined without needing to choose a subspace $\tilde{S}^h \subset S$. (This should be contrasted with (??) for which the subspace S^h explicitly enters into the definition of the matrices \mathbb{B} and \mathbb{D} .) Note, however, that in some sense we are implicitly defining a subspace $\tilde{S}^h = D^{-1} B \tilde{V}^h \subset S$ for the Lagrange multiplier. This subspace, when paired with \tilde{V}^h , may not satisfy the discrete inf-sup condition (??) which shows that approximations obtained through (??) with an arbitrary choice for \tilde{V}^h may not yield stable approximations. See the next remark.

Remark 2.24 The *locking effect*, in the context of penalty methods, describes the phenomena in which the finite element solution approaches zero as $\epsilon \rightarrow 0$. The locking effect is caused by the overconstraining of the discrete solution. If $b(\cdot, \cdot)$, V^h , and S^h satisfy the discrete inf-sup condition (??), then the system (??), or equivalently (??), does not suffer from locking. However, the system (??) can suffer locking with an improper choice for \tilde{V}^h . For a discussion of the locking effect and ways to ameliorate it, see, e.g., [?].

2.4 Examples

We now provide some very brief illustrations of constrained optimization problems of the type (??). In the examples, Ω is an open, bounded domain in \mathcal{R}^s , $s = 2$ or 3 , with boundary Γ . We recall the space $L^2(\Omega)$ of all square integrable functions with norm $\|\cdot\|$, the space $L_0^2(\Omega) \equiv \{\xi \in L^2(\Omega) : \int_\Omega \xi d\Omega = 0\}$, 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 $\|v\|_1 \equiv (\|\nabla v\|^2 + \|v\|^2)^{1/2}$.

2.4.1 The Stokes problem

Let $V = [H_0^1(\Omega)]^s$, $S = L_0^2(\Omega)$, $g = 0$, $t = 0$,

$$a(\mathbf{u}, \mathbf{v}) = \int_{\Omega} \nabla \mathbf{u} : \nabla \mathbf{v} \, d\Omega, \quad \text{and} \quad b(\mathbf{v}, \xi) = - \int_{\Omega} \xi \nabla \cdot \mathbf{v} \, d\Omega.$$

Then,

$$Z = \left\{ \mathbf{v} \in [H_0^1(\Omega)]^s : \int_{\Omega} \xi \nabla \cdot \mathbf{v} \, d\Omega = 0 \quad \forall \xi \in L_0^2(\Omega) \right\} \quad (2.35)$$

is the subspace of all divergence free functions in V . For this setting, all the assumptions in (??) are satisfied. In fact, we have that

$$a(\mathbf{u}, \mathbf{u}) \geq K_a \|\mathbf{u}\|_1^2 \quad \forall \mathbf{u} \in [H_0^1(\Omega)]^s$$

and not just for the subspace Z . Thus, for any conforming choices of subspaces $V^h \subset [H_0^1(\Omega)]^s$ and $S^h \subset L_0^2(\Omega)$, the assumptions in (??) are all satisfied except for the inf-sup condition

$$\inf_{\mathbf{v}^h \in V^h, \mathbf{v}^h \neq \mathbf{0}} \frac{\int_{\Omega} \xi^h \nabla \cdot \mathbf{v}^h \, d\Omega}{\|\mathbf{v}^h\|_1} \geq C_b^h \|\xi^h\|_0 \quad \forall \xi^h \in S^h. \quad (2.36)$$

The inclusions $V^h \subset [H_0^1(\Omega)]^s$ and $S^h \subset L_0^2(\Omega)$ are not sufficient for (??) to hold. Thus, stable approximations of the Stokes problem require that the finite element spaces additionally satisfy (??); see, e.g., [?, ?] for details.

The constrained minimization problem (??) is equivalent to the Stokes system for the velocity \mathbf{u} and the pressure λ :

$$\begin{cases} -\Delta \mathbf{u} + \nabla \lambda = \mathbf{f} & \text{in } \Omega \\ \nabla \cdot \mathbf{u} = 0 & \text{in } \Omega \\ \mathbf{u} = \mathbf{0} & \text{on } \Gamma \end{cases} \quad \text{and} \quad \int_{\Omega} \lambda \, d\Omega = 0. \quad (2.37)$$

With the choice

$$d(\lambda, \xi) = \int_{\Omega} \lambda \xi \, d\Omega \quad (2.38)$$

(so that the operator D is the identity operator), the penalized optimization problem corresponding to (??) obtained by the elimination of λ is given by

$$\begin{cases} -\Delta \mathbf{u}_\epsilon - \frac{1}{\epsilon} \nabla (\nabla \cdot \mathbf{u}_\epsilon) = \mathbf{f} & \text{in } \Omega \\ \mathbf{u}_\epsilon = \mathbf{0} & \text{on } \Gamma. \end{cases} \quad (2.39)$$

Remark 2.25 To highlight the difference between (??) and (??), we note that in the current context (??) involves the matrix $\mathbb{A} + \frac{1}{\epsilon} \mathbb{B}^T \mathbb{B}$, where

$$\begin{aligned} \mathbb{A}_{ij} &= \int_{\Omega} \nabla \mathbf{U}_i : \nabla \mathbf{U}_j \, d\Omega & \text{for } i, j = 1, \dots, n \\ \mathbb{B}_{ij} &= \int_{\Omega} \Lambda_i \nabla \cdot \mathbf{U}_j \, d\Omega & \text{for } i = 1, \dots, m, j = 1, \dots, n, \end{aligned}$$

and where $\{\mathbf{U}_i\}_{i=1}^n$ and $\{\Lambda_i\}_{i=1}^m$ denote bases for V^h and S^h , respectively. On the other hand, (??) involves the matrix $\tilde{\mathbb{A}} + \frac{1}{\epsilon}\tilde{\mathbb{B}}$, where

$$\tilde{\mathbb{A}}_{ij} = \int_{\Omega} \nabla \mathbf{U}_i : \nabla \mathbf{U}_j \, d\Omega \quad \text{and} \quad \tilde{\mathbb{B}}_{ij} = \int_{\Omega} (\nabla \cdot \mathbf{U}_i)(\nabla \cdot \mathbf{U}_j) \, d\Omega \quad \text{for } i, j = 1, \dots, n.$$

Clearly, $\tilde{\mathbb{A}} + \frac{1}{\epsilon}\tilde{\mathbb{B}} \neq \mathbb{A} + \frac{1}{\epsilon}\mathbb{B}^T\mathbb{B}$.

2.4.2 A curl-curl formulation of the Stokes problem

Let $V = [H_0^1(\Omega)]^s$, $S = L_0^2(\Omega)$. We keep $g = 0$, $t = 0$, and $b(\cdot, \cdot)$ as defined in §?? but change $a(\cdot, \cdot)$ to

$$a(\mathbf{u}, \mathbf{v}) = \int_{\Omega} (\nabla \times \mathbf{u}) \cdot (\nabla \times \mathbf{v}) \, d\Omega.$$

Then, Z is given by (??) and it is not hard to see that

$$a(\mathbf{u}, \mathbf{u}) \geq K_a \|\mathbf{u}\|_1^2 \quad \forall \mathbf{u} \in Z$$

but that this inequality does not hold on all of V . In this case, we must verify that (??) is satisfied on the approximating subspaces. The constrained minimization problem (??) is now equivalent to the system

$$\begin{cases} \nabla \times \nabla \times \mathbf{u} + \nabla \lambda &= \mathbf{f} & \text{in } \Omega \\ \nabla \cdot \mathbf{u} &= 0 & \text{in } \Omega \\ \mathbf{u} = \mathbf{0} && \text{on } \Gamma \end{cases} \quad \text{and} \quad \int_{\Omega} \lambda \, d\Omega = 0. \quad (2.40)$$

The corresponding penalized optimization problem, with the bilinear form $d(\cdot, \cdot)$ chosen as in (??), is equivalent to

$$\begin{cases} \nabla \times \nabla \times \mathbf{u} - \frac{1}{\epsilon} \nabla(\nabla \cdot \mathbf{u}) = \mathbf{f} & \text{in } \Omega \\ \mathbf{u} = \mathbf{0} & \text{on } \Gamma. \end{cases}$$

In this system, the term arising from the penalization is crucial to the coercivity of the operator on the left-hand side, i.e., to the validity of (??), while for the penalized Stokes system (??) the first term on the left-hand side by itself was sufficient to guarantee the validity of that result.

Remark 2.26 Note that $g = 0$ implies that $\nabla \cdot \mathbf{u} = 0$ in Ω . Then, since $-\Delta \mathbf{u} = \nabla \times \nabla \times \mathbf{u} - \nabla(\nabla \cdot \mathbf{u})$, (??) is equivalent to the Stokes problem (??). This illustrates the point that different formulations of the same problem may result in considerably different properties of the corresponding penalized problems.

3 Quadratic optimization problems with Stokes equations constraints

We now apply the results of §?? to quadratic optimization problems constrained by the Stokes system. We identify $\{\mathbf{u}, p; \boldsymbol{\theta}\}$ with the variable u of §??, where \mathbf{u} denotes the velocity, p the pressure, and $\boldsymbol{\theta}$ the body force which acts as the control.

We consider the control problem consisting of minimizing the quadratic functional

$$\mathcal{J}(\mathbf{u}, p, \boldsymbol{\theta}) = \frac{1}{2} \int_{\Omega} |\mathbf{u} - \hat{\mathbf{u}}|^2 \, d\Omega + \frac{\delta}{2} \int_{\Omega} |\boldsymbol{\theta}|^2 \, d\Omega \quad (3.1)$$

subject to the Stokes system

$$\begin{cases} -\Delta \mathbf{u} + \nabla p - \boldsymbol{\theta} = \mathbf{0} & \text{in } \Omega \\ \nabla \cdot \mathbf{u} = 0 & \text{in } \Omega \\ \mathbf{u} = \mathbf{0} & \text{on } \Gamma \end{cases} \quad \text{and} \quad \int_{\Omega} p \, d\Omega = 0 \quad (3.2)$$

being satisfied, where $\delta > 0$ is a given constant and $\widehat{\mathbf{u}} \in [L^2(\Omega)]^s$ a given function. This optimal control problem may be interpreted as follows: we are trying to find a velocity \mathbf{u} and a control function $\boldsymbol{\theta}$ such that \mathbf{u} matches as well as possible, in an $L^2(\Omega)$ sense, a given velocity field $\widehat{\mathbf{u}}$ and such that the Stokes system is satisfied. The matching is done by the first term in the functional; the second term is used to limit the size of the control function $\boldsymbol{\theta}$. This optimization problem is often referred to as the *velocity tracking problem with distributed controls* for the Stokes system.

Remark 3.1 Note that in §??, the pressure was denoted by λ while we now use p for that purpose. This is done to achieve consistency with the notation of §?? where λ is used to denote the Lagrange multiplier used to enforce the constraints in a constrained optimization problem. In §??, the pressure acts as such a Lagrange multiplier, while in this section, it is a state variable. Its role as an “inner” Lagrange multiplier is within the constraint equations (??), not in the “outer” sense of the optimization problem at hand.

Let $V = [H_0^1(\Omega)]^s \times L_0^2(\Omega) \times [L^2(\Omega)]^s$, $V^* = [H^{-1}(\Omega)]^s \times L_0^2(\Omega) \times [L^2(\Omega)]^s$,

$$a(\{\mathbf{u}, p; \boldsymbol{\theta}\}, \{\mathbf{v}, q; \boldsymbol{\sigma}\}) = \int_{\Omega} \mathbf{u} \cdot \mathbf{v} \, d\Omega + \delta \int_{\Omega} \boldsymbol{\theta} \cdot \boldsymbol{\sigma} \, d\Omega \quad \forall \{\mathbf{u}, p; \boldsymbol{\theta}\}, \{\mathbf{v}, q; \boldsymbol{\sigma}\} \in V \times V, \quad (3.3)$$

$$\left\langle \{\widehat{\mathbf{u}}, 0; \mathbf{0}\}, \{\mathbf{v}, q; \boldsymbol{\sigma}\} \right\rangle_{V^*, V} = \int_{\Omega} \widehat{\mathbf{u}} \cdot \mathbf{v} \, d\Omega \quad \forall \{\mathbf{v}, q; \boldsymbol{\sigma}\} \in V,$$

and $t = \int_{\Omega} |\widehat{\mathbf{u}}|^2 \, d\Omega$. Then, using the correspondences $\{\mathbf{u}, p; \boldsymbol{\theta}\} \leftrightarrow u$ and $\{\widehat{\mathbf{u}}, 0; \mathbf{0}\} \leftrightarrow f$, it is clear that the functional (??) is of the form (??). Next, let $\Theta = [L^2(\Omega)]^s$, $S = [H_0^1(\Omega)]^s \times L_0^2(\Omega)$, $S^* = [H^{-1}(\Omega)]^s \times L_0^2(\Omega)$, and consider the form

$$b_1(\{\mathbf{u}, p\}, \{\boldsymbol{\xi}, \nu\}) = \left\langle -\Delta \mathbf{u} + \nabla p, \boldsymbol{\xi} \right\rangle_{H^{-1}(\Omega), H_0^1(\Omega)} + \int_{\Omega} \nu \nabla \cdot \mathbf{u} \, d\Omega \quad (3.4)$$

defined on $S \times S$, the form

$$b_2(\{\boldsymbol{\theta}\}, \{\boldsymbol{\xi}, \nu\}) = -\left\langle \boldsymbol{\theta}, \boldsymbol{\xi} \right\rangle_{H^{-1}(\Omega), H_0^1(\Omega)}$$

defined on $\Theta \times S$, and the form

$$b(\{\mathbf{u}, p; \boldsymbol{\theta}\}, \{\boldsymbol{\xi}, \nu\}) = b_1(\{\mathbf{u}, p\}, \{\boldsymbol{\xi}, \nu\}) + b_2(\{\boldsymbol{\theta}\}, \{\boldsymbol{\xi}, \nu\}) \quad (3.5)$$

defined for all $\{\mathbf{u}, p; \boldsymbol{\theta}\} \in V$ and $\{\boldsymbol{\xi}, \nu\} \in S$. Then, with the additional correspondences $\{\boldsymbol{\xi}, \nu\} \leftrightarrow \xi$ and $\{\mathbf{0}, 0\} \leftrightarrow g$, the Stokes system (??) is equivalent to (??), i.e., to

$$b(\{\mathbf{u}, p; \boldsymbol{\theta}\}, \{\boldsymbol{\xi}, \nu\}) = 0 \quad \forall \{\boldsymbol{\xi}, \nu\} \in S = [H_0^1(\Omega)]^s \times L_0^2(\Omega). \quad (3.6)$$

Thus, the problem of minimizing the functional (??) subject to (??) is equivalent to the problem (??).

Suppose for the moment that $\boldsymbol{\theta} \in [L^2(\Omega)]^s$ is given. Then, consider the following least-squares functional for the Stokes system (??):

$$\mathcal{K}(\{\mathbf{u}, p; \boldsymbol{\theta}\}) = \frac{1}{2} \left(\| -\Delta \mathbf{u} + \nabla p - \boldsymbol{\theta} \|_{-1}^2 + \| \nabla \cdot \mathbf{u} \|_0^2 \right). \quad (3.7)$$

The choice of norms in which to measure the residuals of the Stokes system, i.e., a negative norm for the momentum equation and an L^2 norm for the continuity equation, is dictated by the a priori estimate

$$\| -\Delta \mathbf{u} + \nabla p \|_{-1} + \| \nabla \cdot \mathbf{u} \|_0 \geq C (\| \mathbf{u} \|_1 + \| p \|_0) \quad (3.8)$$

that holds for all $\{\mathbf{u}, p\} \in [H_0^1(\Omega)]^s \times L_0^2(\Omega)$ and for some constant $C > 0$; see [?]. This choice makes the least-squares functional (??) *norm-equivalent* and is sufficient to guarantee that, for every $\boldsymbol{\theta} \in [L^2(\Omega)]^s$, the least-squares optimization problem

$$\min_{\{\mathbf{v}, q\} \in [H_0^1(\Omega)]^s \times L_0^2(\Omega)} \mathcal{K}(\{\mathbf{v}, q; \boldsymbol{\theta}\}) \quad (3.9)$$

has a unique minimizer $\{\mathbf{u}, p\}$ out of $[H_0^1(\Omega)]^s \times L_0^2(\Omega)$; see [?, ?].

Remark 3.2 Using the least-squares minimization problem (??) as a basis for finding finite element approximations of solutions of the Stokes problem offers the advantage of circumventing the need to satisfy any inf-sup conditions that arise in mixed Galerkin-based discretizations. In addition, the least-squares-based method results in symmetric, positive definite linear systems instead of the indefinite linear systems that arise in mixed Galerkin-based methods.

We will consider two different ways of using the least-squares functional (??) for the solution of the velocity tracking problem. First, in §??, the cost functional (??) is simply penalized by the least-squares functional (??). Subsequently, approximate solutions can be determined from either the discretized optimality system corresponding to the penalized functional (the eliminate and then discretize approach discussed in Remark ??) or by eliminating the Lagrange multiplier in the discretized optimality system corresponding to a regularized Lagrangian functional (the discretize and then eliminate approach discussed in Remark ??). While it is true that one obtains symmetric, positive definite systems through these approaches, we will see that one still needs to apply inf-sup type conditions in order to guarantee the stability and convergence of the approximations of the penalized optimization problem. Thus, one of the great advantages of least-squares finite element methods for the Stokes problem is negated.

In §??, a second way is introduced for using the least-squares functional (??) for the solution of the velocity tracking problem. Instead of using the least-squares functional to *penalize* the functional, we will use them to *replace* the original PDE constraint by a least-squares formulation. This will allow us from the very beginning to define a setting that is guaranteed to satisfy a discrete inf-sup condition *for any choice of conforming discrete subspaces* so that elimination of the discrete Lagrange multiplier will be guaranteed to give a symmetric and positive definite linear system that is uniformly invertible with respect to both the grid size h and the penalty parameter ϵ .

4 Direct penalization by the least-squares functional

We consider using the least-squares functional (??) to directly penalize the cost functional (??) of the optimization problem. To this end, consider the penalized functional

$$\begin{aligned}\mathcal{J}_\epsilon(\{\mathbf{u}, p; \boldsymbol{\theta}\}) &= \mathcal{J}(\{\mathbf{u}, p; \boldsymbol{\theta}\}) + \frac{1}{2\epsilon}\mathcal{K}(\{\mathbf{u}, p; \boldsymbol{\theta}\}) \\ &= \frac{1}{2} \int_{\Omega} |\mathbf{u} - \hat{\mathbf{u}}|^2 d\Omega + \frac{\delta}{2} \int_{\Omega} |\boldsymbol{\theta}|^2 d\Omega + \frac{1}{2\epsilon} \left(\|\cdot - \Delta \mathbf{u} + \nabla p - \boldsymbol{\theta}\|_{-1}^2 + \|\nabla \cdot \mathbf{u}\|_0^2 \right).\end{aligned}\quad (4.1)$$

The optimality system corresponding to the minimization of (??) is given by: seek $\{\mathbf{u}_\epsilon, p_\epsilon, \boldsymbol{\theta}_\epsilon\}$ in $V \equiv [H_0^1(\Omega)]^s \times L_0^2(\Omega) \times [L^2(\Omega)]^s$ such that

$$\begin{aligned}\int_{\Omega} \mathbf{u}_\epsilon \cdot \mathbf{v} d\Omega + \delta \int_{\Omega} \boldsymbol{\theta}_\epsilon \cdot \boldsymbol{\sigma} d\Omega + \frac{1}{\epsilon} \int_{\Omega} (\nabla \cdot \mathbf{u}_\epsilon) (\nabla \cdot \mathbf{v}) d\Omega \\ + \frac{1}{\epsilon} \left(-\Delta \mathbf{u}_\epsilon + \nabla p_\epsilon - \boldsymbol{\theta}_\epsilon, -\Delta \mathbf{v} + \nabla q - \boldsymbol{\sigma} \right)_{-1} = \int_{\Omega} \hat{\mathbf{u}} \cdot \mathbf{v} d\Omega, \quad \forall \{\mathbf{v}, q; \boldsymbol{\sigma}\} \in V.\end{aligned}\quad (4.2)$$

Using (??), one can show that the bilinear form in (??) is continuous and coercive on $V \times V$ so the problem (??) has a unique solution. However, to demonstrate $O(\epsilon)$ convergence of this solution to the exact solution of the velocity tracking problem, it is necessary to show that the associated regularized Lagrange multiplier optimality system is well posed. For this purpose, we need to identify a form $d(\cdot, \cdot)$ that would allow us to obtain (??) by eliminating a set of yet unknown Lagrange multipliers. To this end, we note the well-known identity [?]:

$$\|\mathbf{v}\|_{-1}^2 \equiv \int_{\Omega} \mathbf{v} \cdot (-\Delta)^{-1} \mathbf{v} d\Omega = \langle \mathbf{v}, (-\Delta)^{-1} \mathbf{v} \rangle_{H^{-1}(\Omega), H_0^1(\Omega)}, \quad (4.3)$$

where $(\Delta)^{-1} : [H^{-1}(\Omega)]^s \rightarrow [H_0^1(\Omega)]^s$ denotes the inverse of the (vector) Laplace operator with zero Dirichlet boundary conditions. Using (??) and the definition of the form $a(\cdot, \cdot)$ from (??) the optimality system (??) can be expressed as

$$\begin{aligned}a(\{\mathbf{u}_\epsilon, p_\epsilon, \boldsymbol{\theta}_\epsilon\}, \{\mathbf{v}, q, \boldsymbol{\sigma}\}) + \frac{1}{\epsilon} \int_{\Omega} (\nabla \cdot \mathbf{v}) (\nabla \cdot \mathbf{u}_\epsilon) d\Omega \\ + \frac{1}{\epsilon} \left\langle -\Delta \mathbf{v} + \nabla q - \boldsymbol{\sigma}, (-\Delta)^{-1} (-\Delta \mathbf{u}_\epsilon + \nabla p_\epsilon - \boldsymbol{\theta}_\epsilon) \right\rangle_{H^{-1}(\Omega), H_0^1(\Omega)} = \int_{\Omega} \hat{\mathbf{u}} \cdot \mathbf{v} d\Omega\end{aligned}\quad (4.4)$$

for all $\{\mathbf{v}, q; \boldsymbol{\sigma}\} \in V$. It is now clear that

$$\boldsymbol{\lambda}_\epsilon = \frac{1}{\epsilon} (-\Delta)^{-1} (-\Delta \mathbf{u}_\epsilon + \nabla p_\epsilon - \boldsymbol{\theta}_\epsilon) \quad \text{and} \quad \mu_\epsilon = \frac{1}{\epsilon} \nabla \cdot \mathbf{u}_\epsilon \quad (4.5)$$

are the ‘‘missing’’ Lagrange multipliers, and that (??) can be rewritten as

$$a(\{\mathbf{u}_\epsilon, p_\epsilon, \boldsymbol{\theta}_\epsilon\}, \{\mathbf{v}, q, \boldsymbol{\sigma}\}) + b(\{\mathbf{v}, q, \boldsymbol{\sigma}\}, \{\boldsymbol{\lambda}_\epsilon, \mu_\epsilon\}) = \int_{\Omega} \hat{\mathbf{u}} \cdot \mathbf{v} d\Omega \quad \forall \{\mathbf{v}, q; \boldsymbol{\sigma}\} \in V, \quad (4.6)$$

where $b(\cdot, \cdot)$ is the form defined in (??). Next, recall that the operator $-\Delta : [H_0^1(\Omega)]^s \rightarrow [H^{-1}(\Omega)]^s \equiv ([H_0^1(\Omega)]^s)^*$ can be defined through

$$\langle (-\Delta) \mathbf{u}, \mathbf{v} \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} = \int_{\Omega} \nabla \mathbf{u} : \nabla \mathbf{v} d\Omega \quad \forall \mathbf{u}, \mathbf{v} \in [H_0^1(\Omega)]^s. \quad (4.7)$$

Then, (??) can be recast as

$$\epsilon \int_{\Omega} (\nabla \boldsymbol{\lambda}_{\epsilon} : \nabla \boldsymbol{\xi} + \mu_{\epsilon} \nu) d\Omega = \langle -\Delta \mathbf{u}_{\epsilon} + \nabla p_{\epsilon} - \boldsymbol{\theta}_{\epsilon}, \boldsymbol{\xi} \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} + \int_{\Omega} \nu \nabla \cdot \mathbf{u}_{\epsilon} d\Omega \quad (4.8)$$

for all $\{\boldsymbol{\xi}, \nu\} \in [H_0^1(\Omega)]^s \times L_0^2(\Omega)$. If we define the bilinear form

$$d(\{\boldsymbol{\lambda}, \mu\}, \{\boldsymbol{\xi}, \nu\}) = \int_{\Omega} (\nabla \boldsymbol{\lambda} : \nabla \boldsymbol{\xi} + \mu \nu) d\Omega \quad \forall \{\boldsymbol{\lambda}, \mu\}, \{\boldsymbol{\xi}, \nu\} \in [H_0^1(\Omega)]^s \times L_0^2(\Omega), \quad (4.9)$$

then (??) can be rewritten as

$$b(\{\mathbf{u}_{\epsilon}, p_{\epsilon}, \boldsymbol{\theta}_{\epsilon}\}, \{\boldsymbol{\xi}, \nu\}) - \epsilon d(\{\boldsymbol{\lambda}_{\epsilon}, \mu_{\epsilon}\}, \{\boldsymbol{\xi}, \nu\}) = 0 \quad (4.10)$$

for all $\{\boldsymbol{\xi}, \nu\} \in [H_0^1(\Omega)]^s \times L_0^2(\Omega)$.

Using the correspondences $\{\mathbf{u}, p; \boldsymbol{\theta}\} \leftrightarrow u$, $\{\mathbf{v}, q; \boldsymbol{\sigma}\} \leftrightarrow v$, $\{\boldsymbol{\lambda}, \mu\} \leftrightarrow \lambda$, $\{\boldsymbol{\xi}, \nu\} \leftrightarrow \xi$, $\{\widehat{\mathbf{u}}, 0; \mathbf{0}\} \leftrightarrow f$, and $\{\mathbf{0}, 0\} \leftrightarrow g$ along with the definitions $V = [H_0^1(\Omega)]^s \times L_0^2(\Omega) \times [L^2(\Omega)]^s$ and $S = [H_0^1(\Omega)]^s \times L_0^2(\Omega)$, it is clear that (??) and (??) are equivalent to (??). It is also easy to see, using the above correspondences, that (??) is equivalent to (??). Thus, we are position to invoke the results of §??, provided that we can verify the assumptions (??) and (??).

Proposition 4.1 *The assumptions (??) are valid for the forms $a(\cdot, \cdot)$ and $b(\cdot, \cdot)$ defined in (??) and (??), respectively, and the assumptions (??) are valid for the bilinear form $d(\cdot, \cdot)$ defined in (??).*

Proof: For the sake of brevity, we only demonstrate that $a(\cdot, \cdot)$ is coercive on the kernel space

$$Z = \left\{ \{\mathbf{u}, p; \boldsymbol{\theta}\} \in [H_0^1(\Omega)]^s \times L_0^2(\Omega) \times [L^2(\Omega)]^s : \right.$$

$$\left. b(\{\mathbf{u}, p; \boldsymbol{\theta}\}, \{\boldsymbol{\xi}, \nu\}) = 0 \quad \forall \{\boldsymbol{\xi}, \nu\} \in [H_0^1(\Omega)]^s \times L_0^2(\Omega) \right\}$$

and that $b(\cdot, \cdot)$ satisfies the last condition in (??). The remaining assumptions in (??) can be easily verified and, obviously, the assumptions (??) are satisfied with $C_d = K_d = 1$.

Note that $\{\mathbf{u}, p; \boldsymbol{\theta}\} \in Z$ if and only if the pair $\{\mathbf{u}, p\}$ solves the Stokes system (??). Therefore, from (??), we obtain that

$$\|\mathbf{u}\|_1 + \|p\|_0 \leq C \|\boldsymbol{\theta}\|_{-1} \quad \forall \{\mathbf{u}, p; \boldsymbol{\theta}\} \in Z.$$

From the definition of $a(\cdot, \cdot)$, and the fact that $\|\boldsymbol{\theta}\|_{-1} \leq \|\boldsymbol{\theta}\|_0$, it follows that

$$a(\{\mathbf{u}, p; \boldsymbol{\theta}\}, \{\mathbf{u}, p; \boldsymbol{\theta}\}) = \|\mathbf{u}\|_0^2 + \delta \|\boldsymbol{\theta}\|_0^2 \geq \delta \|\boldsymbol{\theta}\|_0^2 \geq \frac{\delta}{2} \min(1, 1/C) \left(\|\mathbf{u}\|_1^2 + \|p\|_0^2 + \|\boldsymbol{\theta}\|_0^2 \right)$$

for all $\{\mathbf{u}, p; \boldsymbol{\theta}\} \in Z$. To prove the last assumption in (??), let $\{\boldsymbol{\xi}, \nu\}$ be an arbitrary pair in $[H_0^1(\Omega)]^s \times L_0^2(\Omega)$ and consider the Stokes system

$$\begin{cases} -\Delta \mathbf{u} + \nabla p & = -\Delta \boldsymbol{\xi} & \text{in } \Omega \\ \nabla \cdot \mathbf{u} & = \nu & \text{in } \Omega \\ \mathbf{u} = \mathbf{0} & & \text{on } \Gamma \end{cases} \quad \text{and} \quad \int_{\Omega} p d\Omega = 0.$$

We recall (see [?, p.299]) that for every $\boldsymbol{\xi} \in [H_0^1(\Omega)]^s$ and $\nu \in L_0^2(\Omega)$, there exists a unique solution $\{\mathbf{u}, p\} \in [H_0^1(\Omega)]^s \times L_0^2(\Omega)$ of this system and, moreover,

$$\|\mathbf{u}\|_1 + \|p\|_0 \leq C(\|\Delta \boldsymbol{\xi}\|_{-1} + \|\nu\|_0) \leq C(\|\boldsymbol{\xi}\|_1 + \|\nu\|_0),$$

where the last inequality follows from (??). Then,

$$b(\{\mathbf{u}, p; \boldsymbol{\theta}\}, \{\boldsymbol{\xi}, \nu\}) = \|\nabla \boldsymbol{\xi}\|_0^2 + \|\nu\|_0^2 \geq \frac{1}{2C}(\|\boldsymbol{\xi}\|_1 + \|\nu\|_0)(\|\mathbf{u}\|_1 + \|p\|_0)$$

from which the last assumption in (??) easily follows. \square

With this proposition we have verified all the hypotheses of Propositions ??, ??, and ?? and thus we have the following result.

Theorem 4.2 *The velocity tracking problem consisting of minimizing the functional (??) subject to the Stokes system (??) and the penalized form of this problem consisting of minimizing the penalized functional (??) both have unique solutions. Moreover, the solution of the latter problem, i.e., the solution of (??) and (??), converges to the solution of the former problem with an error that is $O(\epsilon)$.*

4.1 Approximation of the penalized optimization problem

To approximate the penalty solution, we have the choice of discretizing either the penalized problem (??) or the associated regularized optimality system (??) and (??). These approaches respectively corresponds to the *eliminate and then discretize* and the *discretize and then eliminate* approaches discussed in Remark ??.

If we choose to first discretize and then eliminate, it turns out that finite element spaces for the velocity and the pressure cannot be chosen independently *even though we have penalized the cost functional by a well-posed least-squares formulation for the Stokes system* for which such restrictions do not exist! To see this, it suffices to inspect the definition of the discrete space Z^h . Given finite element subspaces W^h , P^h , and Θ^h of $[H_0^1(\Omega)]^s$, $L_0^2(\Omega)$, and $[L^2(\Omega)]^s$, respectively, it is not hard to see that $\{\mathbf{u}^h, p^h, \boldsymbol{\theta}^h\} \in Z^h$ if and only if

$$\begin{aligned} \int_{\Omega} \nabla \mathbf{u}^h \cdot \nabla \mathbf{v}^h d\Omega - \int_{\Omega} p^h \nabla \cdot \mathbf{v}^h &= \int_{\Omega} \boldsymbol{\theta}^h \cdot \mathbf{v}^h \quad \forall \mathbf{v}^h \in W^h \\ - \int_{\Omega} q^h \nabla \cdot \mathbf{u}^h &= 0 \quad \forall q^h \in P^h. \end{aligned} \tag{4.11}$$

This problem is a mixed Galerkin discretization of the Stokes system and as such it is subject to the inf-sup condition [?, ?]. Therefore, we conclude that the proper definition of the discrete kernel space Z^h requires a stable pair of velocity and pressure subspaces. In particular, this excludes the possibility of using equal order interpolation spaces defined with respect to the same triangulation of the domain Ω into finite elements; see [?, ?].

With respect to the last assumption in (??), the inf-sup condition on the state variables is an *inner* stability condition required to ensure well-posedness of the discrete constraint equation, i.e., the mixed Stokes problem (??). Without this condition, the *outer* inf-sup condition in (??) will fail as well.

If we choose to first eliminate and then discretize, the finite element approximation of (??) is easily defined by restricting this problem to a finite element subspace of $V = [H_0^1(\Omega)]^s \times L_0^2(\Omega) \times [L^2(\Omega)]^s$. In the usual manner, one can show that the ensuing problem is a linear system whose

solution defines a finite element approximation $(\mathbf{u}_\epsilon^h, p_\epsilon^h, \boldsymbol{\theta}_\epsilon^h)$ to the penalty solution. As stated in Remark ??, this approach does not require an explicit choice of finite element subspaces for the Lagrange multipliers, neither does it seem to require a special choice of velocity and pressure subspaces. Indeed, since the bilinear form in (??) is coercive on $V \times V$, it is easy to see that all assumptions of Proposition ?? will hold for any conforming subspace V^h of V and so the discrete penalty problem will have a unique solution for any *fixed* value of the penalty parameter ϵ . Nevertheless, stability and convergence of this method as $\epsilon \rightarrow 0$ will depend on whether or not the implicitly defined multiplier space satisfies a discrete inf-sup condition when paired with the spaces used to discretize (??); see Remark ?. As a result, discretization of the penalty system (??) is not guaranteed to work for all possible choices of conforming finite element subspaces, in particular, the penalty finite element solution is not guaranteed to be free of locking as $\epsilon \rightarrow 0$; see Remark ?.

To summarize, we have seen that the direct approach of penalizing an objective functional with a well-posed least-squares functional for the constraint equations does not in general lead to a computational method that takes advantage of some of the most desirable features of least-square finite element methods. In particular, the need to have the discrete constraint system to be stably solvable for the state variable for any choice of control variables can, e.g., for the Stokes system, negate the advantages of least-square finite element methods in optimal control settings. In the next section, we consider an alternate method that circumvents this problem and leads to formulations of the velocity tracking problem that can be discretized using any choice of conforming finite element subspaces for the state and control variables.

5 Constraining by the least-squares functional

In the last section, we saw that direct penalization of (??) by the least-squares functional (??) led to a regularized Lagrange multiplier system that still required an *internal* inf-sup stability condition. This, of course, is caused by the fact that re-introducing the Lagrange multipliers (??) back into (??) recovers the mixed form of the Stokes system, which, as we know from §??, is a saddle-point problem in its own right.

One of the chief reasons for the widespread use of least-squares principles has been their ability to circumvent saddle-point stability conditions; see [?]. Since the main disadvantage of direct least-squares penalization is the reappearance of the mixed form of the constraint equation, it is natural to seek a solution of this problem by replacing the original constraint equation (??) by an equivalent least-squares formulation so that the new bilinear form $b_1(\cdot, \cdot)$ appearing in the constraint equation is symmetric and coercive. Then, the solution of the optimization problem will not require any *internal* discrete stability conditions

In sum, we propose to use the least-squares functional $\mathcal{K}(\mathbf{u}, p; \boldsymbol{\theta})$, defined in (??), to constrain rather than to penalize the functional (??). Thus, *least-squares-constrained* formulation of the velocity tracking problem is given by the optimization problem

$$\min_{(\mathbf{u}, \boldsymbol{\theta}) \in [H_0^1(\Omega)]^s \times [L^2(\Omega)]^s} \mathcal{J}(\mathbf{u}, \boldsymbol{\theta}) \quad \text{subject to} \quad \min_{(\mathbf{u}, p) \in [H_0^1(\Omega)]^s \times L_0^2(\Omega)} \mathcal{K}(\mathbf{u}, p; \boldsymbol{\theta}). \quad (5.1)$$

Standard techniques from calculus of variations can be used to show that, for any given $\boldsymbol{\theta} \in [L^2(\Omega)]^s$, the minimizer $(\mathbf{u}, p) \in [H_0^1(\Omega)]^s \times L_0^2(\Omega)$ of $\mathcal{K}(\mathbf{u}, p; \boldsymbol{\theta})$ solves the variational equation

$$\left(-\Delta \mathbf{u} + \nabla p, -\Delta \mathbf{v} + \nabla q \right)_{-1} + \int_{\Omega} \nabla \cdot \mathbf{u} \nabla \cdot \mathbf{v} \, d\Omega = \left(\boldsymbol{\theta}, -\Delta \mathbf{v} + \nabla q \right)_{-1} \quad (5.2)$$

for all $(\mathbf{v}, q) \in [H_0^1(\Omega)]^s \times L_0^2(\Omega)$. To cast the least-squares constraint equation (??) into the form

of (??), let $S = [H_0^1(\Omega)]^s \times L_0^2(\Omega)$, $\Theta = [L^2(\Omega)]^s$, and $V = S \times \Theta$. Then, consider the bilinear form

$$b_1(\{\mathbf{u}, p\}, \{\mathbf{v}, q\}) = \left(-\Delta \mathbf{u} + \nabla p, -\Delta \mathbf{v} + \nabla q \right)_{-1} + \int_{\Omega} \nabla \cdot \mathbf{u} \nabla \cdot \mathbf{v} \, d\Omega \quad (5.3)$$

defined on $S \times S$, the bilinear form

$$b_2(\{\boldsymbol{\theta}\}, \{\mathbf{v}, q\}) = -\left(\boldsymbol{\theta}, -\Delta \mathbf{v} + \nabla q \right)_{-1}$$

defined on $\Theta \times S$, and the form

$$b(\{\mathbf{u}, p, \boldsymbol{\theta}\}, \{\mathbf{v}, q\}) = b_1(\{\mathbf{u}, p\}, \{\mathbf{v}, q\}) + b_2(\{\boldsymbol{\theta}\}, \{\mathbf{v}, q\})$$

defined on $V \times S$. With the additional correspondence $\{\mathbf{0}, 0\} \leftrightarrow g$, the least-squares Stokes constraint (??) is equivalent to (??), i.e., to

$$b(\{\mathbf{u}, p, \boldsymbol{\theta}\}, \{\mathbf{v}, q\}) = 0 \quad \forall \{\mathbf{v}, q\} \in S = [H_0^1(\Omega)]^s \times L^2(\Omega).$$

The bilinear form $b_1(\cdot, \cdot)$ serves to define a self-adjoint operator $B_1 : S \rightarrow S^*$. Using (??), one can show that this form is continuous and coercive on $S \times S$ so that the operator B_1 is invertible with a bounded inverse.

Remark 5.1 It is instructive to compare the operator engendered by $b_1(\cdot, \cdot)$ as defined in (??) and used in used in §?? with that corresponding to (??). In the first case, $b_1(\cdot, \cdot)$ gives rise to the Stokes operator

$$B_1 = \begin{bmatrix} -\Delta & \nabla \\ \nabla \cdot & 0 \end{bmatrix} \quad (5.4)$$

plus some suitable boundary conditions while in the second case, (??) can be used to show that $b_1(\cdot, \cdot)$ leads to the operator

$$B_1 = \begin{bmatrix} -\Delta + \nabla \nabla \cdot & \nabla \\ \nabla \cdot & \nabla \cdot (-\Delta)^{-1} \nabla \end{bmatrix} \quad (5.5)$$

plus some suitable boundary conditions. Note that B_1 as defined in (??) is merely the standard symmetric but indefinite Stokes operator and the weak formulation involving the corresponding bilinear form (??) is subject to an inf-sup condition on the velocity-pressure spaces. On the other hand, B_1 as defined in (??) is a symmetric and positive definite operator and the weak formulation involving the corresponding bilinear form (??) does not require inf-sup conditions for stability. This is the main difference between the two ways of using a least-squares functional for optimization problems. Thus, in this sense, the least-squares constraint (??) can be viewed as a regularized form of the original Stokes constraint (??).

For $\{\mathbf{u}, p, \boldsymbol{\theta}\} \in V = [H_0^1(\Omega)]^s \times L_0^2(\Omega) \times [L^2(\Omega)]^s$ and $\{\mathbf{v}, q\} \in S = \mathbf{H}_0^1(\Omega) \times L_0^2(\Omega)$, we introduce the Lagrangian functional

$$\begin{aligned} \mathcal{L}(\{\mathbf{u}, p, \boldsymbol{\theta}\}, \{\mathbf{v}, q\}) &= \mathcal{J}(\mathbf{u}, \boldsymbol{\theta}) + b(\{\mathbf{u}, p, \boldsymbol{\theta}\}, \{\mathbf{v}, q\}) \\ &= \frac{1}{2} \int_{\Omega} (\mathbf{u} - \hat{\mathbf{u}})^2 \, d\Omega + \frac{\delta}{2} \int_{\Omega} |\boldsymbol{\theta}|^2 \, d\Omega + \left(-\Delta \mathbf{u} + \nabla p - \boldsymbol{\theta}, -\Delta \mathbf{v} + \nabla q \right)_{-1} \\ &\quad + \int_{\Omega} (\nabla \cdot \mathbf{u}) (\nabla \cdot \mathbf{v}) \, d\Omega. \end{aligned} \quad (5.6)$$

Then, the constrained optimization problem (??) is equivalent to the unconstrained optimization problem of finding the saddle points $(\{\mathbf{u}, p, \boldsymbol{\theta}\}, \{\mathbf{v}, q\})$ of (??). The saddle points may be found by solving the optimality system

$$\left\{ \begin{array}{l} \int_{\Omega} \mathbf{u} \cdot \mathbf{z} d\Omega + \int_{\Omega} (\nabla \cdot \mathbf{v})(\nabla \cdot \mathbf{z}) d\Omega + \left(-\Delta \mathbf{v} + \nabla q, -\Delta \mathbf{z} \right)_{-1} = \int_{\Omega} \widehat{\mathbf{u}} \cdot \mathbf{z} d\Omega \\ \left(-\Delta \mathbf{v} + \nabla q, \nabla r \right)_{-1} = 0 \\ \int_{\Omega} \boldsymbol{\theta} \cdot \boldsymbol{\sigma} d\Omega - \left(-\Delta \mathbf{v} + \nabla q, \boldsymbol{\sigma} \right)_{-1} = 0 \\ \left(-\Delta \mathbf{u} + \nabla p, -\Delta \mathbf{w} \right)_{-1} + \int_{\Omega} (\nabla \cdot \mathbf{u})(\nabla \cdot \mathbf{w}) d\Omega - \left(\boldsymbol{\theta}, -\Delta \mathbf{w} \right)_{-1} = 0 \\ \left(-\Delta \mathbf{u} + \nabla p, \nabla s \right)_{-1} - \left(\boldsymbol{\theta}, \nabla s \right)_{-1} = 0 \end{array} \right. \quad (5.7)$$

for all $\mathbf{z} \in [H_0^1(\Omega)]^s$, $r \in L_0^2(\Omega)$, $\boldsymbol{\sigma} \in [L^2(\Omega)]^s$, $\mathbf{w} \in [H_0^1(\Omega)]^s$, and $s \in L_0^2(\Omega)$.

Theorem 5.2 *The system (??) has a unique solution.*

Proof: Note that $\{\mathbf{u}, p, \boldsymbol{\theta}\}$ is in the kernel space Z if and only if, given a $\boldsymbol{\theta} \in [L^2(\Omega)]^s$, the pair $\{\mathbf{u}, p\}$ solves the least-squares variational equation (??). Because the form $b_1(\cdot, \cdot)$ is continuous and coercive, this equation has a unique solution and this solution depends continuously on the data. Therefore, for any $\{\mathbf{u}, p, \boldsymbol{\theta}\} \in Z$, it holds that

$$\|\mathbf{u}\|_1 + \|p\|_0 \leq C\|\boldsymbol{\theta}\|_{-1}.$$

Note that the form $a(\cdot, \cdot)$ is the same as in Theorem ?? and so its coercivity on Z follows in exactly the same manner.

It now remains to verify that the last assumption in (??) holds for the form $b(\cdot, \cdot)$, i.e., that there exists a constant K_b such that

$$\sup_{\{\mathbf{u}, p, \boldsymbol{\theta}\} \in V} \frac{b(\{\mathbf{u}, p, \boldsymbol{\theta}\}, \{\mathbf{v}, q\})}{\|\{\mathbf{u}, p, \boldsymbol{\theta}\}\|_V} \geq K_b \|\{\mathbf{v}, q\}\|_S.$$

Let $\{\mathbf{v}, q\} \in S$ be an arbitrary but fixed function and take $\{\mathbf{u}, p, \boldsymbol{\theta}\} \equiv \{\mathbf{v}, p, \mathbf{0}\}$. Then,

$$b(\{\mathbf{u}, p, \boldsymbol{\theta}\}, \{\mathbf{v}, q\}) = b_1(\{\mathbf{v}, q\}, \{\mathbf{v}, q\}) \geq C\|\{\mathbf{v}, q\}\|_S^2,$$

where the last inequality follows from (??). This shows that the last assumption in (??) holds.

□

Finite element discretizations of (??) are defined in the usual manner. We choose conforming subspaces $W^h \subset [H_0^1(\Omega)]^s$, $P^h \subset L^2(\Omega)$, and $\Theta^h \subset [L^2(\Omega)]^s$ and then seek $\{\mathbf{u}^h, p^h, \boldsymbol{\theta}^h\} \in W^h \times$

$P^h \times \Theta^h$ that solves

$$\left\{ \begin{array}{l} \int_{\Omega} \mathbf{u}^h \cdot \mathbf{z}^h d\Omega + \left(-\Delta \mathbf{v}^h + \nabla q^h, -\Delta \mathbf{z}^h \right)_{-1} + \int_{\Omega} (\nabla \cdot \mathbf{v}^h)(\nabla \cdot \mathbf{z}^h) d\Omega = \int_{\Omega} \hat{\mathbf{u}}^h \cdot \mathbf{z}^h d\Omega \\ \left(-\Delta \mathbf{v}^h + \nabla q^h, \nabla r^h \right)_{-1} = 0 \\ \int_{\Omega} \boldsymbol{\theta} \cdot \boldsymbol{\sigma}^h d\Omega - \left(-\Delta \mathbf{v}^h + \nabla q^h, \boldsymbol{\sigma}^h \right)_{-1} = 0 \\ \left(-\Delta \mathbf{u}^h + \nabla p^h, -\Delta \mathbf{w}^h \right)_{-1} + \int_{\Omega} (\nabla \cdot \mathbf{u}^h)(\nabla \cdot \mathbf{w}^h) d\Omega - \left(\boldsymbol{\theta}^h, -\Delta \mathbf{w}^h \right)_{-1} = 0 \\ \left(-\Delta \mathbf{u}^h + \nabla p^h, \nabla s^h \right)_{-1} - \left(\mathbf{f}^h, \nabla s^h \right)_{-1} = 0 \end{array} \right. \quad (5.8)$$

for all $\{\mathbf{z}^h, r^h, \boldsymbol{\sigma}^h\} \in W^h \times P^h \times \Theta^h$ and $\{\mathbf{w}^h, s^h\} \in W^h \times P^h$.

Theorem 5.3 *The system (??) has a unique solution for any conforming choice of the finite element spaces W^h , P^h , and Θ^h .*

Proof: Since $b_1(\cdot, \cdot)$ is coercive and continuous on $[H_0^1(\Omega)]^s \times L_0^2(\Omega)^2$ it will remain coercive for any choice of conforming subspaces W^h and P^h and so the proof of the discrete inf-sup condition follows in exactly the same way as in Theorem ???. To show that $a(\cdot, \cdot)$ as defined in (??) is coercive on the discrete kernel space Z^h , note that $\{\mathbf{u}^h, p^h, \boldsymbol{\theta}^h\} \in Z^h$ if and only if $\{\mathbf{u}^h, p^h\}$ solves the problem

$$b_1(\{\mathbf{u}^h, p^h\}, \{\mathbf{v}^h, q^h\}) + b_2(\boldsymbol{\theta}^h, \{\mathbf{v}^h, q^h\}) \quad \forall \{\mathbf{v}^h, q^h\} \in W^h \times P^h.$$

Clearly, this problem has a unique solution $\{\mathbf{u}^h, p^h\} \in W^h \times P^h$ and moreover,

$$\|\mathbf{u}^h\|_1 + \|p^h\|_0 \leq C \|\boldsymbol{\theta}^h\|_{-1}.$$

Now (??) easily follows by noting that

$$a(\{\mathbf{u}^h, p^h, \boldsymbol{\theta}^h\}, \{\mathbf{u}^h, p^h, \boldsymbol{\theta}^h\}) = \|\mathbf{u}^h\|_0^2 + \frac{\delta}{2} \|\boldsymbol{\theta}^h\|_0^2$$

and that $\|\boldsymbol{\theta}^h\|_0 \geq \|\boldsymbol{\theta}^h\|_{-1}$. Thus, we have established that all assumptions of Proposition ?? hold for any choice of conforming finite element subspaces of $[H_0^1(\Omega)]^s$, $L_0^2(\Omega)$, and $L^2(\Omega)$. \square

Let $\vec{\phi}$, $\vec{\theta}$, and $\vec{\lambda}$ denote the coefficient vectors of the state variables $\{\mathbf{u}^h, p^h\}$, the control $\boldsymbol{\theta}^h$, and the adjoint variables $\{\mathbf{v}^h, q^h\}$, respectively. The discretized optimality system (??) is equivalent to the linear system

$$\begin{pmatrix} \mathbb{M}_1 & 0 & \mathbb{K} \\ 0 & \mathbb{M}_2 & \mathbb{Q}^T \\ \mathbb{K} & \mathbb{Q} & 0 \end{pmatrix} \begin{pmatrix} \vec{\phi} \\ \vec{\theta} \\ \vec{\lambda} \end{pmatrix} = \begin{pmatrix} \vec{\mathbf{f}} \\ \vec{\mathbf{0}} \\ \vec{\mathbf{0}} \end{pmatrix}. \quad (5.9)$$

Here, \mathbb{M}_1 and \mathbb{M}_2 are matrices corresponding to the bilinear form $a(\cdot, \cdot)$ and the bases chosen for the subspaces V^h and Θ^h , respectively, \mathbb{K} is the stiffness matrix corresponding to the bilinear form $b_1(\cdot, \cdot)$ for the basis chosen for the subspace $V^h \times P^h$, and \mathbb{Q} is the rectangular matrix generated by the bilinear form $b_2(\cdot, \cdot)$ with respect to the bases chosen for the subspaces $V^h \times P^h$ and Θ^h . Theorem ?? implies that (??) is uniformly invertible for any choice for the finite element subspaces.

Remark 5.4 Here we have chosen the same finite element subspaces $W^h \times P^h$ to approximate both the state $\{\mathbf{u}, p\}$ and the adjoint $\{\mathbf{v}, q\}$ variables. If these variables are approximated using two different subspaces of $[H_0^1(\Omega)]^s \times L^2(\Omega)$, then the two matrices \mathbb{K} appearing in (??) are not the same and in fact, they are the transpose of each other. Even though nothing prevents us from using different subspaces, this would clearly complicate the exposition and so we will not pursue this approach.

Of course, (??) is a formidable system to solve; it is at least twice the size of the least-squares problem for the Stokes system $\mathbb{K}\vec{\phi} + \mathbb{Q}\vec{\theta} = \vec{\mathbf{0}}$, not counting the size of the control variable. To reduce its size, we proceed to eliminate the adjoint variables from the discrete problem (??), i.e., we first discretize and then eliminate; see Remark ???. Elimination of the adjoint variables requires a form $d(\cdot, \cdot)$ that satisfies assumptions (??). Here we will use the form defined in (??) and the *relaxed* constraint equation

$$\begin{aligned} \epsilon d(\{\mathbf{u}_\epsilon^h, p_\epsilon^h\}, \{\mathbf{w}^h, s^h\}) &= b(\{\mathbf{u}_\epsilon^h, p_\epsilon^h, \boldsymbol{\theta}_\epsilon^h\}, \{\mathbf{w}^h, s^h\}) \\ &= \left(-\Delta \mathbf{u}_\epsilon^h + \nabla q_\epsilon^h - \boldsymbol{\theta}_\epsilon^h, -\Delta \mathbf{w}^h + \nabla s^h \right)_{-1} + \int_{\Omega} (\nabla \cdot \mathbf{u}_\epsilon^h)(\nabla \cdot \mathbf{w}^h) d\Omega \end{aligned} \quad (5.10)$$

to effect the regularization. After the original constraint in (??) is replaced by (??), the linear system (??) takes the form

$$\begin{pmatrix} \mathbb{M}_1 & 0 & \mathbb{K} \\ 0 & \mathbb{M}_2 & \mathbb{Q}^T \\ \mathbb{K} & \mathbb{Q} & -\epsilon \mathbb{D} \end{pmatrix} \begin{pmatrix} \vec{\phi}_\epsilon \\ \vec{\theta}_\epsilon \\ \vec{\lambda}_\epsilon \end{pmatrix} = \begin{pmatrix} \vec{\mathbf{f}} \\ \vec{\mathbf{0}} \\ \vec{\mathbf{0}} \end{pmatrix}. \quad (5.11)$$

Here \mathbb{D} is a matrix corresponding to the form $d(\cdot, \cdot)$ and the bases chosen for the subspaces W^h and P^h (recall that we restrict attention to approximation of the state and adjoint variables by the same finite element spaces). It is easy to see that

$$\mathbb{D} = \begin{pmatrix} \mathbb{K}_2 & 0 \\ 0 & \mathbb{M} \end{pmatrix},$$

where \mathbb{K}_2 and \mathbb{M} are matrices corresponding to

$$\int_{\Omega} \nabla \mathbf{v} : \nabla \mathbf{w} d\Omega \quad \text{and} \quad \int_{\Omega} q s d\Omega$$

and the bases chosen for W^h and P^h , respectively. One can easily eliminate the discrete adjoint vector $\vec{\lambda}_\epsilon$ from (??) to obtain

$$\begin{pmatrix} \mathbb{M}_1 + \frac{1}{\epsilon} \mathbb{K} \mathbb{D}^{-1} \mathbb{K} & \frac{1}{\epsilon} \mathbb{K} \mathbb{D}^{-1} \mathbb{Q} \\ \frac{1}{\epsilon} \mathbb{Q}^T \mathbb{D}^{-1} \mathbb{K} & \mathbb{M}_2 + \frac{1}{\epsilon} \mathbb{Q}^T \mathbb{D}^{-1} \mathbb{Q} \end{pmatrix} \begin{pmatrix} \vec{\phi}_\epsilon \\ \vec{\theta}_\epsilon \end{pmatrix} = \begin{pmatrix} \vec{\mathbf{f}} \\ \vec{\mathbf{0}} \end{pmatrix}. \quad (5.12)$$

The coefficient matrix of this linear system is symmetric and positive definite. Note the appearance of \mathbb{D}^{-1} in the system (??). Formally, computation of \mathbb{D}^{-1} requires a solution of a vector Poisson equation to invert \mathbb{K}_2 and inversion of a consistent mass matrix. Even though \mathbb{K}_2^{-1} can be computed fairly quickly by multilevel methods, it turns out that it is possible to improve the efficiency of the penalized formulation even more. We will consider this issue in the next section.

6 Further practicality considerations

In this section, we briefly discuss important issues related to the implementation of least-squares constrained methods. In particular, we use several popular techniques from least-squares finite element methodologies to demonstrate the formulation of practical and efficient computational algorithms based on the ideas from the last section.

6.1 Discrete norms

The choice of $d(\cdot, \cdot)$ is guided by the assumptions in (??) which require this form to be symmetric, continuous and coercive, i.e., inner-product equivalent. These assumptions also guarantee that the matrix \mathbb{D} engendered by $d(\cdot, \cdot)$ is invertible for any conforming choice of finite element subspaces for the adjoint variables. In the present context, inversion of \mathbb{D} includes inversion of \mathbb{K}_2 , i.e., a solution of a discrete Poisson equation on the same mesh on which we discretize our primary problem. Thus, it would be advantageous to find a cheaper alternative. Since the only relevant assumption on $d(\cdot, \cdot)$ is its inner-product equivalence, it is clear that we can replace \mathbb{D} by an arbitrary symmetric and positive definite matrix as long as it remains spectrally equivalent to \mathbb{D} . This idea has been widely used in the least-squares community in the implementation of *negative norm* least-squares methods; see [?, ?, ?], among others.

The computation of negative norms requires the inversion of the Laplace operator; see (??). Because this operation is not practical, least-squares methods that require negative norms have relied on computable discrete equivalents to replace the actual negative norm. It can be shown (see [?]) that for finite element functions such an equivalence can be defined using the discrete minus one inner product

$$(\phi, \psi)_h = \left((\mathbb{B}^h + h^2\mathbb{I})\phi, \psi \right)_0, \quad (6.1)$$

where \mathbb{B}^h is a preconditioner for the Laplace equation that is spectrally equivalent to \mathbb{K}_2^{-1} . Thus, \mathbb{D}^{-1} can be replaced by the matrix

$$\mathbb{C} = \begin{pmatrix} \mathbb{B}^h + h^2\mathbb{I} & \mathbf{0} \\ \mathbf{0} & h^{-2}\mathbb{I} \end{pmatrix}$$

In practice, \mathbb{B}^h is often implemented by using several multigrid cycles which makes its computation very efficient compared to the evaluation of \mathbb{K}_2^{-1} .

Of course, the negative norms and inner products in (??), (??), (??), and (??) also must be replaced by computable equivalents before we can actually use the methods in computations. Likewise, it is preferable to replace the Laplace operator $-\Delta$ by a discrete equivalent Δ^h so as to allow the use of standard C^0 finite element subspaces in the discrete problem. For instance, we can define $\Delta^h : [H^{-1}(\Omega)]^s \mapsto W^h$ by $-\Delta^h \mathbf{u} = \mathbf{v}^h$ if and only if

$$\left(\nabla \mathbf{v}^h, \nabla \mathbf{z}^h \right)_0 = \left(\mathbf{u}, \mathbf{z}^h \right)_0 \quad \forall \mathbf{z}^h \in W^h.$$

It can be shown (see [?]) that the use of Δ^h does not lead to loss of accuracy in the discrete problem.

Thus, using (??) in lieu of the minus one inner product and Δ^h in lieu of Δ gives the computable

alternative of (??):

$$\left\{ \begin{array}{l} \int_{\Omega} \mathbf{u}^h \cdot \mathbf{z}^h d\Omega + \left(-\Delta^h \mathbf{v}^h + \nabla q^h, -\Delta^h \mathbf{z}^h \right)_h + \int_{\Omega} \nabla \cdot \mathbf{v}^h \nabla \cdot \mathbf{z}^h d\Omega = \int_{\Omega} \hat{\mathbf{u}}^h \cdot \mathbf{z}^h d\Omega \\ \left(-\Delta^h \mathbf{v}^h + \nabla q^h, \nabla r^h \right)_h = 0 \\ \int_{\Omega} \boldsymbol{\theta} \cdot \boldsymbol{\sigma}^h d\Omega - \left(-\Delta^h \mathbf{v}^h + \nabla q^h, \boldsymbol{\sigma}^h \right)_h = 0 \\ \left(-\Delta^h \mathbf{u}^h + \nabla p^h, -\Delta \mathbf{w}^h \right)_h + \int_{\Omega} \nabla \cdot \mathbf{u}^h \nabla \cdot \mathbf{w}^h d\Omega - \left(\boldsymbol{\theta}^h, -\Delta \mathbf{w}^h \right)_h = 0 \\ \left(-\Delta^h \mathbf{u}^h + \nabla p^h, \nabla s^h \right)_h - \left(\boldsymbol{\theta}^h, \nabla s^h \right)_h = 0. \end{array} \right. \quad (6.2)$$

The linear system obtained from (??) by using \mathbb{C} to eliminate the adjoint variables is given by

$$\begin{pmatrix} \mathbb{M}_1 + \frac{1}{\epsilon} \mathbb{K}^h \mathbb{C} \mathbb{K}^h & \frac{1}{\epsilon} \mathbb{K}^h \mathbb{C} \mathbb{Q} \\ \frac{1}{\epsilon} \mathbb{Q}^T \mathbb{C} \mathbb{K}^h & \mathbb{M}_2 + \frac{1}{\epsilon} \mathbb{Q}^T \mathbb{C} \mathbb{Q} \end{pmatrix} \begin{pmatrix} \vec{\phi}_{\epsilon} \\ \vec{\theta}_{\epsilon} \end{pmatrix} = \begin{pmatrix} \vec{\mathbf{f}} \\ \vec{\mathbf{0}} \end{pmatrix}, \quad (6.3)$$

where \mathbb{K}^h is the analogue of \mathbb{K} obtained with the discrete minus one inner product. All matrices in (??) are computable and the system can be solved by, e.g., preconditioned conjugate gradient methods so that (??), when coupled to an iterative penalty method, defines a truly practical algorithm for the solution of the velocity tracking problem.

6.2 First-order formulations

In (??), we constrained the optimization problem by a least-squares functional based on the second-order Stokes system. A popular and widely used practice in least-squares finite element methods is to apply least-squares principles to equivalent *first-order* formulations of the PDE problem. This reduces the continuity requirements on the finite element spaces, but also increases the number of dependent variables. However, this reformulation may well be worth the effort, especially when the optimization problem involves physically important variables such as vorticity or stress.

To illustrate this idea, consider the functional

$$\mathcal{J}(\mathbf{u}, \boldsymbol{\theta}) = \frac{1}{2} \int_{\Omega} |\nabla \times \mathbf{u}|^2 d\Omega + \frac{\delta}{2} \int_{\Omega} |\boldsymbol{\theta}|^2 d\Omega \quad (6.4)$$

and the optimization problem

$$\min_{\{\mathbf{v}, \boldsymbol{\theta}\} \in [H_0^1(\Omega)]^s \times [L^2]^s(\Omega)} \mathcal{J}(\mathbf{v}, \boldsymbol{\theta})$$

subject to the Stokes system (??). This optimization problem calls for finding a distributed control $\boldsymbol{\theta}$ that minimizes the total flow vorticity.

Using the vorticity

$$\boldsymbol{\omega} = \nabla \times \mathbf{u}$$

in (??) allows us to write that functional as

$$\mathcal{J}(\boldsymbol{\omega}, \boldsymbol{\theta}) = \frac{1}{2} \int_{\Omega} |\boldsymbol{\omega}|^2 d\Omega + \frac{\delta}{2} \int_{\Omega} |\boldsymbol{\theta}|^2 d\Omega. \quad (6.5)$$

We can also consider the vorticity as a new dependent variable in the Stokes system. Using the well-known vector identity

$$-\Delta \mathbf{u} = \nabla \times \nabla \times \mathbf{u} - \nabla(\nabla \cdot \mathbf{u}),$$

the Stokes system (??) can be expressed as

$$\begin{aligned} \nabla \times \boldsymbol{\omega} + \nabla p &= \boldsymbol{\theta} && \text{in } \Omega \\ \nabla \cdot \mathbf{u} &= 0 && \text{in } \Omega \\ \nabla \times \mathbf{u} - \boldsymbol{\omega} &= 0 && \text{in } \Omega \\ \mathbf{u} &= 0 && \text{on } \Gamma. \end{aligned} \tag{6.6}$$

The problem (??) is known as the *velocity-vorticity-pressure* formulation of the Stokes problem. In [?], it was shown that a norm-equivalent functional for this system is given by

$$\mathcal{K}(\mathbf{u}, \boldsymbol{\omega}, p; \boldsymbol{\theta}) = \frac{1}{2} \left(\|\nabla \times \boldsymbol{\omega} + \nabla p - \boldsymbol{\theta}\|_{-1}^2 + \|\nabla \cdot \mathbf{u}\|_0^2 + \|\nabla \times \mathbf{u} - \boldsymbol{\omega}\|_0^2 \right).$$

Now, the vorticity minimization problem can be restated as

$$\min_{(\boldsymbol{\omega}, \boldsymbol{\theta}) \in [L^2(\Omega)]^s \times [L^2(\Omega)]^s} \mathcal{J}_\epsilon(\boldsymbol{\omega}, \boldsymbol{\theta}) \quad \text{subject to} \quad \min_{(\mathbf{u}, \boldsymbol{\omega}, p) \in [H_0^1(\Omega)]^s \times [L^2(\Omega)]^s \times L_0^2(\Omega)} \mathcal{K}(\mathbf{u}, \boldsymbol{\omega}, p; \boldsymbol{\theta}). \tag{6.7}$$

One advantage of (??) is that its optimality system will involve at most first-order derivatives of the dependent variables which makes it easier to discretize by standard C^0 finite element subspaces.

References

- [1] D. BEDIVAN AND G. FIX, Least-squares methods for optimal shape design problems, *Comput. Math. Appl.* **30**, 1995, pp. 7-25.
- [2] P. BOCHEV, Least-squares methods for optimal control, *Nonlin. Anal. Theo. Meth. Appl.* **30**, 1997, pp. 1875-1885.
- [3] P. BOCHEV, Negative norm least-squares methods for the velocity-vorticity-pressure Navier-Stokes equations, *Num. Meth. PDE's* **15**, 1999, pp. 237-256.
- [4] P. BOCHEV AND D. BEDIVAN, Least-squares methods for Navier-Stokes boundary control problems, *Int. J. Comp. Fluid Dyn* **9**, 1997, pp. 43-58.
- [5] P. BOCHEV AND M. GUNZBURGER, Least-squares finite element methods for elliptic equations, *SIAM Rev.* **40**, 1998 pp. 789-837.
- [6] P. BOCHEV AND M. GUNZBURGER, Analysis of least-squares finite element methods for the Stokes equations, *Math. Comp.* **63**, 1994, pp. 479-506.
- [7] D. BRAESS, *Finite Elements*, Cambridge, Cambridge, 1997.
- [8] J. BRAMBLE, R. LAZAROV, AND J. PASCIAK, A least squares approach based on a discrete minus one inner product for first order systems, *Technical Report 94-32*, Mathematical Science Institute, Cornell University, Ithaca, 1994.

- [9] 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.
- [10] F. BREZZI, On the existence, uniqueness and approximation of saddle-point problems arising from Lagrange multipliers, *RAIRO Anal. Numer. R2*, 1974, pp. 129–151.
- [11] V. GIRAULT AND P.-A. RAVIART, *Finite Element Methods for Navier-Stokes Equations*, Springer, Berlin, 1986.
- [12] M. GUNZBURGER, *Perspectives in Flow Control and Optimization*, SIAM, Philadelphia, 2002.
- [13] M. GUNZBURGER, *Finite Element Methods for Viscous Incompressible Flows*, Academic, Boston, 1989.
- [14] 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.
- [15] 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.
- [16] J. LIONS, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer, New York, 1971.
- [17] H. SCHLICHTING AND K. GERSTEN, *Boundary Layer Theory*, Springer, Berlin, 2000.