Electronic Transactions on Numerical Analysis.
Volume 22, pp. 97-113, 2006.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

REGULARIZATION AND STABILIZATION OF DISCRETE SADDLE-POINT VARIATIONAL PROBLEMS*

P. B. BOCHEV AND R. B. LEHOUCQ†

Abstract. Our paper considers parameterized families of saddle-point systems arising in the finite element solution of PDEs. Such saddle point systems are ubiquitous in science and engineering. Our motivation is to explain how these saddle-point systems can be modified to avoid onerous stability conditions and to obtain linear systems that are amenable to iterative methods of solution. In particular, the algebraic and variational perspectives of regularization and stabilization are explained.

Key words. constrained minimization, saddle point systems, mixed finite elements, regularization, stabilization, penalty, Stokes problem, Darcy flow problem

AMS subject classifications. 65F15, 65N25, 65N30, 65N22, 65M60, 65N55, 65M55

1. Introduction.

Saddle-point equations arise when a constrained optimization problem is solved by Lagrange multipliers. For example, in many physical models, admissible states can be characterized as constrained minimizers of a quadratic energy functional [34]. A computational science example are the FETI domain decomposition methods (see [38] for an overview) used in computational mechanics. There, Lagrange multipliers are used to connect solutions among the subdomains. A related example is non-conforming finite element methods [12] where Lagrange multipliers are applied to enforce inter-element continuity of the finite element solution. Lagrange multipliers can also be used to enforce Dirichlet boundary conditions [3] giving another instance of a saddle-point problem. An important example that also leads to saddle-point problems is constrained interpolation that is required in discretized multicomponent systems on different grids. To enable communication between the components, fields must be converted from one representation to another without spurious effects [14] such as artificial energy or mass dissipation or accumulation.

In this paper, we consider saddle-point systems of equations resulting from the approximate numerical solution of PDEs by mixed finite elements. Many important PDEs, such as the incompressible Stokes and Darcy flow problems, can be derived from an optimality system (Euler-Lagrange equations) associated with a constrained minimization problem. Their finite element discretization leads to mixed Galerkin methods and a family of algebraic saddle-point problems, parameterized by some measure h of the grid size. This is in contrast to discrete optimization, where the problem size remains fixed. This poses some significant challenges in both formulation and solution of such saddle-point problems. Most notably, the operators in this family must be invertible stably and uniformly in h. For PDEs related to constrained minimization, this cannot be accomplished by merely choosing conforming finite element spaces, spaces that are finite dimensional subspaces of the respective continuum spaces. In addition to conformity, these spaces must satisfy the onerous inf-sup or LBB condition [9, 12]. Among other things, this condition precludes the use of equal order finite element spaces defined with respect to the same partition of the computational domain in finite elements. Besides the programming inconvenience, using different spaces for the different variables requires more complicated data structures. Moreover, the resulting

*Received March 3, 2005. Accepted for publication November 17, 2005. Recommended by E. de Sturler.

†Sandia National Laboratories, P.O. Box 5800, MS 1110, Albuquerque, NM 87185-1110 ({pbboche, rblehou}@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. This work was partially funded by the Applied Mathematical Sciences program, U.S. Department of Energy, Office of Energy Research.

ETNA
Kent State University
etna@mcs.kent.edu

saddle-point problems have symmetric indefinite matrices. These systems typically have a 2 by 2 block structure with the (2,2) block being identically zero. This makes their numerical solution by preconditioned iterative methods challenging. The reader is referred to [5, 17] for further information and references on preconditioned iterative methods for saddle point systems.

For these and other reasons, two alternative finite element formulations for PDEs related to constrained optimization have been pursued ever since the inception of mixed Galerkin methods. The first, *stabilization* circumvents the onerous stability conditions inherent in mixed methods. The second, *regularization*[1], does not circumvent stability conditions, but renders the solution of the saddle point system with preconditioned iterative methods less difficult. At the discrete level these two fundamentally different approaches may appear indistinguishable. One of our goals will be to expound the distinctions between the two approaches and how they are propagated to the algebraic level. For this purpose we will always treat modifications as occurring at the variational level even though in many cases they can be introduced directly into the algebraic problem. This will also help to clarify the subtleties that arise with an algebraic view of the discrete problem that does not account for their origin.

Our paper is organized as follows. Section 2 presents a discussion of abstract variational problems that leads to saddle point systems. The abstract problem is placed into context with two examples in Section 3. The abstract algebraic problem is the subject of 4. We then modify the constrained minimization problem in sections 6–7 with regularization and stabilization, respectively. Section 8 concludes with a brief summary.

2. Abstract setting. Let $\mathcal{V}$ and $\mathcal{S}$ denote two Hilbert spaces with inner products $(\cdot,\cdot)_{\mathcal{V}}$ and $(\cdot,\cdot)_{\mathcal{S}}$, respectively, and $\langle\cdot,\cdot\rangle$ denote the duality pairing between $\mathcal{V}$, $\mathcal{S}$ and their dual spaces $\mathcal{V}'$, $\mathcal{S}'$ respectively. Let

$$a(\cdot,\cdot):\mathcal{V}\times\mathcal{V}\mapsto\mathbb{R},\quad b(\cdot,\cdot):\mathcal{V}\times\mathcal{S}\mapsto\mathbb{R} \tag{2.1}$$

be two continuous bilinear forms, such that

$$a(\mathbf{v},\mathbf{v})\geq 0\quad\text{for all } \mathbf{v}\in\mathcal{V} \tag{2.2}$$

and let $f\in\mathcal{V}'$, $g\in\mathcal{S}'$ be given data functions. We consider the optimization problem:

$$\left\{\begin{array}{l}\text{find } \mathbf{u}\in\mathcal{V} \text{ that minimizes } \; J(\mathbf{v})=\dfrac{1}{2}a(\mathbf{v},\mathbf{v})-\langle f,\mathbf{v}\rangle\\[2ex] \text{subject to } \; b(\mathbf{v},\mu)=\langle g,\mu\rangle \quad \text{for all } \mu\in\mathcal{S}.\end{array}\right. \tag{2.3}$$

By introducing a Lagrange multiplier $\lambda\in\mathcal{S}$ this constrained optimization problem is transformed into the unconstrained problem of finding the saddle-point $(\mathbf{u},\lambda)\in\mathcal{V}\times\mathcal{S}$ of the Lagrangian functional

$$L(\mathbf{v},\mu)=J(\mathbf{v})+b(\mathbf{v},\mu)-\langle g,\mu\rangle\,. \tag{2.4}$$

Using standard calculus of variations techniques we obtain the saddle-point problem:

$$\left\{\begin{array}{l}\text{find } (\mathbf{u},\lambda)\in\mathcal{V}\times\mathcal{S} \text{ such that}\\[1ex] \begin{array}{lll} a(\mathbf{u},\mathbf{v})+b(\mathbf{v},\lambda) & =\langle f,\mathbf{v}\rangle & \forall\mathbf{v}\in\mathcal{V},\\ b(\mathbf{u},\mu) & =\langle g,\mu\rangle & \forall\mu\in\mathcal{S},\end{array}\end{array}\right. \tag{2.5}$$

[1] We call attention to the non-standard use of the term "regularization" in the paper. Typically, regularization refers to a process wherein an ill-posed problem is replaced by a well-posed problem. Here, regularization denotes a technique to replace one well-posed problem by another well-posed problem that is "easier" to solve.

ETNA
Kent State University
etna@mcs.kent.edu

that expresses a first-order optimality condition for (2.4). Using the operators $A : \mathcal{V} \longmapsto \mathcal{V}'$ and $B : \mathcal{V} \longmapsto \mathcal{S}'$ defined by

$$\langle A\mathbf{u}, \mathbf{v}\rangle = a(\mathbf{u}, \mathbf{v}) \quad \text{and} \quad \langle B\mathbf{u}, \mu\rangle = b(\mathbf{u}, \mu)$$

the saddle-point problem (2.5) is recast as

$$\left\{ \begin{array}{lll} A\mathbf{u} + B'\lambda & = f & \text{in } \mathcal{V}' \\ B\mathbf{u} & = g & \text{in } \mathcal{S}' \end{array} \right. \tag{2.6}$$

To discuss well-posedeness of (2.5) or (2.6) we need the kernel space

$$\mathcal{Z} = \{\mathbf{v} \in \mathcal{V} \,|\, b(\mathbf{v}, \mu) = 0 \quad \forall \mu \in \mathcal{S}\}\,. \tag{2.7}$$

In his seminal paper [9] Brezzi developed conditions for the unique and stable solution of (2.5) in terms of the two bilinear forms.

THEOREM 2.1. *The saddle-point problem (2.5) defines an isomorphism $\mathcal{V}\times\mathcal{S} \mapsto \mathcal{V}'\times\mathcal{S}'$ if and only if $a(\cdot,\cdot)$ is coercive on the kernel $\mathcal{Z}$, i.e.,*

$$a(\mathbf{v}, \mathbf{v}) \geq \alpha \|\mathbf{v}\|^2_{\mathcal{V}} \quad \forall \mathbf{v} \in \mathcal{Z}, \quad \alpha > 0 \tag{2.8}$$

and $b(\cdot,\cdot)$ satisfies the inf-sup condition

$$\inf_{\mu\in\mathcal{S}} \sup_{\mathbf{v}\in\mathcal{V}} \frac{b(\mathbf{v}, \mu)}{\|\mathbf{v}\|_{\mathcal{V}}\|\mu\|_{\mathcal{S}}} \geq \beta > 0 \tag{2.9}$$

If $g = 0$, (2.3) is equivalent to the reduced problem

$$\text{find } \mathbf{u} \in \mathcal{Z} \text{ that minimizes} \quad J(\mathbf{v})\,. \tag{2.10}$$

This minimizer solves the variational equation: find $\mathbf{u} \in \mathcal{Z}$ such that

$$a(\mathbf{u}, \mathbf{v}) = \langle f, \mathbf{v}\rangle \quad \forall \mathbf{v} \in \mathcal{Z}\,. \tag{2.11}$$

According to (2.8), Theorem 2.1, this problem has a unique solution.

3. Examples. Stokes equations and the Darcy equations are examples of PDEs associated with constrained minimization principles. We briefly examine how they fit in the abstract setting of Section 2 and discuss some of their differences. To describe the two problems, we assume that Ω is a bounded open domain in $\mathbb{R}^n$, $n = 2, 3$ with Lipschitz continuous boundary Γ. We recall the space $L^2(\Omega)$ of all square integrable functions and its subspace $L^2_0(\Omega)$ of zero mean functions. We will also need the space $\mathbf{H}^1_0(\Omega)$ of all square integrable vector functions $\mathbf{u}$ that vanish on Γ and whose first derivatives are also square integrable. Lastly, we recall the space $H_N(\Omega, \mathrm{div}\,)$ of all square integrable vector fields $\mathbf{u}$ with square integrable divergence and whose normal components vanish on Γ. Because our focus is on finite element methods, we assume that $\mathcal{T}_h$ is a partition of Ω into finite elements K. In two dimensions K can be triangles or quadrilaterals, in three dimensions the elements are tetrahedra, hexahedra, prisms or pyramids. Finite element spaces are defined by combining local polynomial spaces defined on each element. The so-called nodal or Lagrangian finite elements are subspaces of $H^1(\Omega)$ and contain C^0 piecewise polynomial functions. Other finite element spaces provide only normal continuity across element faces and form subspaces of $H(\Omega, \mathrm{div}\,)$. For more details about these spaces and their construction we refer the reader to [12, 23].

ETNA
Kent State University
etna@mcs.kent.edu

3.1. Stokes equations. We consider the constrained optimization problem

$$\left\{ \begin{array}{c} \text{find } \mathbf{u} \in \mathbf{H}_0^1(\Omega) \text{ that minimizes} \\ J(\mathbf{v}) = \dfrac{1}{2}\displaystyle\int_\Omega \nabla \mathbf{v} \cdot \nabla \mathbf{v}\, d\Omega - \int_\Omega \mathbf{f} \cdot \mathbf{v}\, d\Omega \quad \text{subject to } \nabla \cdot \mathbf{v} = 0 \end{array} \right. . \tag{3.1}$$

With the identifications $\mathcal{V} = \mathbf{H}_0^1(\Omega)$, $\mathcal{S} = L_0^2(\Omega)$,

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

problem (3.1) is of the form (2.3). The null-space (2.7) is given by the subspace

$$\mathcal{Z} = \{\mathbf{v} \in \mathbf{H}_0^1(\Omega) \,|\, b(\mathbf{v}, \mu) = 0 \quad \forall \mu \in L_0^2(\Omega)\}$$

of weakly solenoidal vector fields in $\mathbf{H}_0^1(\Omega)$. A classical result [9, 23] shows that (2.9) holds for the Stokes problem. The bilinear form $a(\cdot, \cdot)$ is coercive on $\mathbf{H}_0^1(\Omega) \times \mathbf{H}_0^1(\Omega)$ and so (2.8) is trivially satisfied.

If solutions $(\mathbf{u}, \lambda)$ of the optimality system (2.5) are sufficiently smooth, integration by parts can be used to show that $(\mathbf{u}, \lambda)$ satisfy the Stokes equations

$$\left\{ \begin{array}{lll} -\triangle \mathbf{u} + \nabla \lambda & = f & \text{in } \Omega \\ \nabla \cdot \mathbf{u} & = 0 & \text{in } \Omega \end{array} \right. \tag{3.2}$$

augmented with the boundary condition $\mathbf{u} = 0$ on $\partial\Omega$. With the obvious identifications $A = -\triangle$, $B = -\nabla \cdot$ and $B' = \nabla$ these equations provide the operator form (2.6) of the saddle-point problem. For the Stokes equations $\mathbf{u}$ is the velocity field and the Lagrange multiplier λ turns out to be the pressure.

3.2. Darcy problem. We consider the constrained optimization problem

$$\left\{ \begin{array}{c} \text{find } \mathbf{u} \in H_N(\Omega, \mathrm{div}\,) \text{ that minimizes} \\ J(\mathbf{v}) = \dfrac{1}{2}\displaystyle\int_\Omega |\mathbf{v}|^2\, d\Omega \quad \text{subject to } \nabla \cdot \mathbf{v} = f \end{array} \right. . \tag{3.3}$$

For this problem $\mathcal{V} = H_N(\Omega, \mathrm{div}\,)$, $\mathcal{S} = L_0^2(\Omega)$,

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

The null-space of the Darcy problem is given by

$$\mathcal{Z} = \{\mathbf{v} \in H_N(\Omega, \mathrm{div}\,) \,|\, b(\mathbf{v}, \mu) = 0 \quad \forall \mu \in L_0^2(\Omega)\}\,,$$

and is a proper subspace of $H_N(\Omega, \mathrm{div}\,)$. Again, classical results [12] show that both (2.8) and (2.9) hold for $a(\cdot, \cdot)$ and $b(\cdot, \dot{)}$.

If solutions $(\mathbf{u}, \lambda)$ of the associated optimality system (2.5) are sufficiently smooth then $(\mathbf{u}, \lambda)$ satisfy the *Darcy* equations

$$\left\{ \begin{array}{ll} \mathbf{u} + \nabla \lambda & = 0 \\ \nabla \cdot \mathbf{u} & = f \end{array} \right. \tag{3.4}$$

ETNA
Kent State University
etna@mcs.kent.edu

augmented with the boundary condition $\mathbf{u} \cdot \mathbf{n} = 0$ on $\partial\Omega$. With the identifications $A = I$, $B = -\nabla\cdot$ and $B' = \nabla$ these equations take the operator form (2.6). For the Darcy equations $\mathbf{u}$ is the velocity and λ is the pressure.

The bilinear form $b(\cdot,\cdot)$ and associated operators B and B' are the same for the Stokes and Darcy problems. However, the form $a(\cdot,\cdot)$ is fundamentally different for the two problems. For the Stokes problem this form defines an inner product on $\mathbf{H}_0^1(\Omega) \times \mathbf{H}_0^1(\Omega)$, while for the Darcy equations $a(\cdot,\cdot)$ is merely the $L^2(\Omega)$ inner product. As a result, for the Stokes equations, $a(\cdot,\cdot)$ is coercive on all of $\mathcal{V}$, including the subspace $\mathcal{Z}$. In contrast, for the Darcy problem coercivity of $a(\cdot,\cdot)$ is restricted to the null-space $\mathcal{Z}$.

4. Approximation of saddle-point problems. This section briefly reviews the algebraic interpretation associated with the abstract saddle-point problem (2.5). The reader is referred to [12, pp.73–80], and [10, 11, 20, 26] for further information on an algebraic treatment.

The saddle-point problem (2.5) and the reduced equation (2.11) are completely equivalent. However, their numerical approximation leads to methods that are not equivalent. Moreover, deriving a conforming approximation of the nullspace (2.7) is difficult, which leaves the saddle-point problem (2.5) as the preferred setting for the constrained minimization.

To discretize (2.5) we need finite dimensional subspaces $\mathcal{V}_h \subset \mathcal{V}$ and $\mathcal{S}_h \subset \mathcal{S}$, parameterized by h. For the examples of interest to us (see Section 3) $\mathcal{V}_h$ and $\mathcal{S}_h$ will be finite element spaces and h is some measure of the element size. The discrete version of (2.5) is:

$$\left\{ \begin{array}{l} \text{find } (\mathbf{u}_h, \lambda_h) \in \mathcal{V}_h \times \mathcal{S}_h \text{ such that} \\[2mm] \begin{array}{lll} a(\mathbf{u}_h, \mathbf{v}_h) + b(\mathbf{v}_h, \lambda_h) & = \langle f, \mathbf{v}_h \rangle & \forall \mathbf{v}_h \in \mathcal{V}_h \\ b(\mathbf{u}_h, \mu_h) & = \langle g, \mu_h \rangle & \forall \mu_h \in \mathcal{S}_h \end{array} \end{array} \right. \tag{4.1}$$

In finite elements this formulation is known as the *mixed Galerkin* method, and gives rise to a parameterized family of linear algebraic equations

$$\begin{bmatrix} \mathbf{A}_h & \mathbf{B}_h^T \\ \mathbf{B}_h & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_h \\ \lambda_h \end{bmatrix} = \begin{bmatrix} \mathbf{f}_h \\ \mathbf{g}_h \end{bmatrix} \tag{4.2}$$

where, with some abuse of notation, $\mathbf{u}_h$ and λ_h are used to denote the coefficients of the finite element functions $\mathbf{u}_h$ and λ_h in terms of bases of $\mathcal{V}_h$ and $\mathcal{S}_h$. The discrete data functions $\mathbf{f}_h$ and $\mathbf{g}_h$ are the representations of the functions f and g. The linear system (4.2) is the first-order necessary condition for the discrete optimization problem

$$\left\{ \begin{array}{c} \min\limits_{\mathbf{x}_h} Q_h(\mathbf{x}_h) \stackrel{\text{def}}{=} \dfrac{1}{2}\mathbf{x}_h^T \mathbf{A}_h \mathbf{x}_h - \mathbf{x}_h^T \mathbf{f}_h \\ \text{subject to } \mathbf{B}_h \mathbf{x}_h = \mathbf{g}_h \end{array} \right. . \tag{4.3}$$

This problem is an equality constrained quadratic program (QP) [33, p.444], and the matrix in (4.2) is called the Karush-Kuhn-Tucker (KKT) matrix. Therefore, approximation of the constrained optimization problem (4.3) gives rise to a sequence of equality constrained QPs parameterized by the mesh size h.

4.1. Well-posedeness of parameterized QPs. For every $h > 0$ the KKT, or saddle-point, matrix is nonsingular provided the following two conditions [33, p.445] are met:

1. The reduced Hessian matrix $\mathbf{Z}_h^T \mathbf{A}_h \mathbf{Z}_h$, where the columns of $\mathbf{Z}$ are a basis for $\ker(\mathbf{B}_h)$, is positive definite,

2. The matrix $\mathbf{B}$ has full row rank.

However, when dealing with families of QPs that approximate (2.3) these two conditions are not enough to ensure convergence to the exact solution of (2.3). The reason is that solvability of (4.2) for a fixed value of h does not imply that solutions remain well-behaved, e.g., depend continuously on the data, when $h \to 0$. Consequently, the key to developing a useful stability criterion for (4.2) is to treat this problem as an instance of the abstract system (2.6) rather than as a linear system of algebraic equations. Then, the well-posedness of (4.2) is subject to the conditions in Section 2, including Theorem 2.1, restricted to $\mathcal{V}_h$ and $\mathcal{S}_h$. To obtain these conditions in terms of matrices we note that

$$\mathcal{Z}_h = \{\mathbf{v}_h \in \mathcal{V}_h \,|\, b(\mathbf{v}_h, \lambda_h) = 0 \quad \forall \lambda_h \in \mathcal{S}_h\},$$

is the null-space of $\mathbf{B}_h$ and so is spanned by the columns of $\mathbf{Z}_h$. There also exist symmetric and positive definite matrices $\mathbf{V}_h$ and $\mathbf{S}_h$, such that

$$\|\mathbf{v}_h\|_{\mathcal{V}}^2 = \mathbf{v}_h^T \mathbf{V}_h \mathbf{v}_h \quad \text{and} \quad \|\mu_h\|_{\mathcal{S}}^2 = \mu_h^T \mathbf{S}_h \mu_h\,, \tag{4.4}$$

respectively. In terms of these matrices, continuity of $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ implies

$$\mathbf{v}_h^T \mathbf{A}_h \mathbf{u}_h \le \kappa_a \left(\mathbf{v}_h^T \mathbf{V}_h \mathbf{v}_h\right)^{1/2} \left(\mathbf{u}_h^T \mathbf{V}_h \mathbf{u}_h\right)^{1/2} \tag{4.5}$$

$$\mu_h^T \mathbf{B}_h \mathbf{u}_h \le \kappa_b \left(\mu_h^T \mathbf{S}_h \mu_h\right)^{1/2} \left(\mathbf{u}_h^T \mathbf{V}_h \mathbf{u}_h\right)^{1/2} \tag{4.6}$$

where κ_a and κ_b are positive real constants independent of h. Coercivity on the kernel condition (2.8) specializes to

$$\mathbf{v}_h^T \mathbf{A}_h \mathbf{v}_h \ge \alpha^h \mathbf{v}_h^T \mathbf{V}_h \mathbf{v}_h \quad \forall \mathbf{v}_h \in \mathcal{Z}_h\,, \tag{4.7}$$

while the inf-sup condition (2.9) can be expressed as

$$\inf_{\mu_h \in \mathcal{S}_h} \sup_{\mathbf{v}_h \in \mathcal{V}_h} \frac{\mu_h^T \mathbf{B}_h \mathbf{v}_h}{\left(\mathbf{v}_h^T \mathbf{V}_h \mathbf{v}_h\right)^{1/2} \left(\mu_h^T \mathbf{S}_h \mu_h\right)^{1/2}} \ge \beta^h\,, \tag{4.8}$$

where $0 < \hat{\alpha} \le \alpha^h$ and $0 < \hat{\beta} \le \beta^h$ as $h \to 0$. We remark that $\hat{\beta}$ is a lower bound on the smallest generalized singular value of $\mathbf{B}_h$ (see [12, pp.76–77]).

Let us compare (4.7) and (4.8) with the two conditions implying that the KKT matrix is nonsingular. Because $\mathbf{v}_h \in \mathcal{Z}_h$ if and only if there exists $\mathbf{y}$ such that $\mathbf{v}_h = \mathbf{Z}_h\mathbf{y}$, condition (4.7) can be restated as

$$(\mathbf{Z}_h\mathbf{y})^T \mathbf{A}_h \mathbf{Z}_h \mathbf{y} \ge \alpha^h (\mathbf{Z}_h\mathbf{y})^T \mathbf{V}_h \mathbf{Z}_h \mathbf{y} \qquad \forall \mathbf{y}\,. \tag{4.9}$$

While this condition implies the first condition for the nonsingularity of the KKT matrix, (4.9) is clearly more restrictive and is equivalent to requirement that the smallest eigenvalue of the pencil $(\mathbf{Z}_h^T \mathbf{A}_h \mathbf{Z}_h, \mathbf{Z}_h^T \mathbf{V}_h \mathbf{Z}_h)$ remains bounded away from zero *independently* of h. Similarly, the second condition for the nonsingularity of the KKT matrix follows from (4.8), but does not imply (4.8) that requires the smallest generalized singular value of $\mathbf{B}_h$ to be bounded away from zero independently of h.

4.2. Difficulties of parameterized QPs. Designing a pair of finite dimensional subspaces $(\mathcal{V}_h, \mathcal{S}_h)$ so that the restriction of (2.3) gives rise to an invertible KKT, or saddle point, matrix for a fixed value of h is not difficult. Considerably more difficult is to find a sequence of subspaces $(\mathcal{V}_h, \mathcal{S}_h)_{h>0}$ such that restriction of (2.3) to these spaces gives rise to a uniformly invertible saddle point matrix (satisfies the stability conditions (4.5)–(4.8)).

ETNA
Kent State University
etna@mcs.kent.edu

Clearly, (4.5) and (4.6) are implied by the continuity of the bilinear forms in (2.1) and the inclusions $\mathcal{V}_h \subset \mathcal{V}$ and $\mathcal{S}_h \subset \mathcal{S}$, i.e., conformity is sufficient for continuity. Unfortunately, conformity is not sufficient for (4.7) and (4.8) and so these two conditions do not follow from (2.8) and (2.9) and the inclusions $\mathcal{V}_h \subset \mathcal{V}$ and $\mathcal{S}_h \subset \mathcal{S}$. First, even if these inclusions hold, the null space $\mathcal{Z}_h$ is not necessarily a subspace of $\mathcal{Z}$ and so (4.7) does not follow from (2.8). Second, because $\mathcal{S}_h \subset \mathcal{S}$, the inf-sup condition (2.9) implies that for every $\lambda_h \in \mathcal{S}_h$ there is a $\mathbf{v}(\lambda_h) \in \mathcal{V}$ such that

$$b(\mathbf{v}(\lambda_h), \lambda_h) \geq \beta \|\lambda_h\|_{\mathcal{S}}.$$

However, existence of $\mathbf{v}(\lambda_h)$ is only guaranteed in the infinite dimensional space $\mathcal{V}$, while for (4.8) to hold, $\mathbf{v}(\lambda_h)$ must belong to $\mathcal{V}_h$.

There exist simple tests that can rule out some poor choices of $\mathcal{V}_h$ and $\mathcal{S}_h$. For example, if $\mathbf{B}_h$ is full rank and has more rows than columns, then $\mathcal{Z}_h = \{0\}$ and so $\mathbf{A}_h$ is not positive definite on the nullspace of $\mathbf{B}_h$. Calculating the dimension of $\mathbf{B}_h$ is the basis of the popular "counting test" in the engineering literature. The problem with this and other similar tests is that they cannot be used to show that (4.7)-(4.8) hold with mesh-independent $\hat{\alpha}$ and $\hat{\beta}$. When $\hat{\alpha}$ and $\hat{\beta}$ are mesh-dependent, the saddle point matrix may be invertible, but the quality of the solution sequence as $h \to 0$ may degenerate.

Let us examine how the difference between the Stokes and Darcy equations impacts construction of well-posed parameterized QPs for these equations. Recall that for the Stokes equations $a(\cdot,\cdot)$ is coercive on $\mathcal{V}$. The single most-important consequence of this fact is that the first discrete condition (4.7) is satisfied whenever $\mathcal{V}_h \subset \mathcal{V}$, or equivalently, by the conformity of $\mathcal{V}_h$. Therefore, the choice of stable conforming pairs $(\mathcal{V}_h, \mathcal{S}_h)$ for the Stokes equations is subject *only* to the discrete inf-sup condition (4.8). This and the fact that (3.2) is one of the studied settings for the mixed Galerkin method, is the reason why the first condition in Theorem 2.1 is often overlooked.

For the Darcy problem $a(\cdot,\cdot)$ is not coercive on $\mathcal{V}$ but only on the proper subspace $\mathcal{Z}$. As a result, for this problem the discrete condition (4.7) is not implied by the inclusion $\mathcal{V}_h \subset \mathcal{V}$. Therefore, the choice of conforming pairs $(\mathcal{V}_h, \mathcal{S}_h)$ for (3.4) is subject to *both* discrete conditions (4.7) and (4.8). One practical consequence is that stable pairs for the Stokes equations are unstable for the Darcy problem, while stable pairs [13, 31] for the latter are not conforming for the former. For instance, stable discretization for the Darcy equations can be obtained using the lowest order Raviart-Thomas spaces [12] RT_0 for $\mathcal{V}_h$ and piecewise constant approximation for $\mathcal{S}_h$. However, RT_0 contains piecewise polynomial functions that are only continuous in the normal component and do not form a proper subspace of $\mathbf{H}^1(\Omega)$—they cannot be used in a conforming method for the Stokes equations.

In either case stable pairs $(\mathcal{V}_h, \mathcal{S}_h)$ cannot be obtained by using equal order finite element spaces defined with respect to the same triangulation of $\mathcal{T}_h$ of Ω into finite elements. For example, the equal order $P1 - P1$ (piecewise linear spaces on simplices) and $Q1 - Q1$ (bilinear or trilinear spaces on quads or hexes) pairs are unstable for both the Stokes and the Darcy equations. See [24, 12] for more examples of stable and unstable spaces for Stokes and Darcy flow.

5. Modification of saddle-point problems. The approximate numerical solution of the constrained optimization problem (2.3) faces two main difficulties. To obtain a well-posed parameterized QPs it is necessary to find finite dimensional pairs $(\mathcal{V}_h, \mathcal{S}_h)$ that satisfy the restrictive stability conditions (4.7)-(4.8). These conditions may be difficult to verify [8]. Then, there are PDEs, such as mixed elasticity, where stable pairs have been found only recently and only in 2D [2]. These stable elements also tend to have a rather complicated nature. In the context of finite elements, compatibility with (4.7)-(4.8) requires the use of unequal order approximations and/or different grids. The attendant heterogeneity of the data structures increases code complexity.

Provided stable pairs are available, we still face the task of solving a sequence saddle point linear systems (4.2). Because systems with millions of unknowns are common in PDE discretizations, iterative solution methods are typically employed, however, indefinite systems are still a challenge for iterative solvers. The stable mixed method may also need special preconditioners [1]. We refer the reader to [5, 17] and the references listed therein for further information.

The difficulties in obtaining uniformly invertible saddle point matrices and solving the resulting linear set of equations has prompted numerous techniques that aim to circumvent stability conditions, improve the efficiency of iterative methods used for the solution, or address both issues simultaneously. Without an exception, these techniques can be related to modifications of either the constrained minimization problem (2.3) or the optimality system (2.5). In this paper we refer to modifications that attempt to improve solver efficiency and require spaces that are *compatible* with (4.7)-(4.8), as *regularized methods*. We reserve the term *stabilized methods* for modifications that circumvent the two compatibility conditions (4.7)-(4.8), allow the use of arbitrary discrete pairs $(\mathcal{V}_h, \mathcal{S}_h)$ to derive the algebraic problems, and retain asymptotic convergence to the solution of (2.5) as $h \to 0$.

Regularization and stabilization often result in nearly identical saddle point matrices. For this reason, our discussion always begins with a statement of the modified variational problem and then proceeds to develop the associated algebraic system. For simplicity we consider only homogeneous constraints. Having $g \neq 0$ will only change the right hand side of the modified equations while the structure of the modified bilinear form, responsible for its properties, will remain the same.

6. Regularized methods. We first discuss penalty methods followed by augmented Lagrangian methods.

6.1. Penalty. Penalty methods are among the earliest examples of modified variational formulations. A penalty formulation can be obtained by modifying the quadratic energy functional $J(\mathbf{v})$, the Lagrangian functional $L(\mathbf{v}, \mu)$ or by direct manipulation of the optimality system (2.5). In the first case the modified problem is to find the unconstrained minimizer $\mathbf{u}_r \in \mathcal{V}$ of the penalized energy functional

$$J_r(\mathbf{v}) = J(\mathbf{v}) + \frac{r}{2}\|B\mathbf{v}\|^2_{\mathcal{S}'} \tag{6.1}$$

When $r \to \infty$, the penalty term in (6.1) enforces the constraint without using a Lagrange multiplier. In practice, only finite values of r can be used in a numerical computation and so a solution of (6.1) is subject to a $O(r^{-1})$ penalty error.

In the second case the modified problem is to find the saddle-point $(\mathbf{u}_r, \lambda_r) \in \mathcal{V} \times \mathcal{S}$ of the penalized Lagrangian

$$L_r(\mathbf{v}, \mu) = L(\mathbf{v}, \mu) - \frac{2}{r}\|\mu\|^2_{\mathcal{S}} \tag{6.2}$$

The penalty term in (6.2) improves the properties of the variational equation (2.5) and allows us to eliminate the Lagrange multiplier from the optimality system.

In the last case, we seek solutions $(\mathbf{u}_r, \lambda_r) \in \mathcal{V} \times \mathcal{S}$ of the modified first-order optimality system

$$\left\{ \begin{array}{ll} a(\mathbf{u}, \mathbf{v}) + b(\mathbf{v}, \lambda) & = \langle f, \mathbf{v}\rangle \quad \forall \mathbf{v} \in \mathcal{V} \\ b(\mathbf{u}, \mu) & = \frac{1}{r}(\lambda, \mu)_{\mathcal{S}} \quad \forall \mu \in \mathcal{S} \end{array} \right. \tag{6.3}$$

ETNA
Kent State University
etna@mcs.kent.edu

An important practical case is when $\mathcal{S}$ coincides with its dual $\mathcal{S}'$. Then, the three modified problems (6.1)-(6.3) are equivalent in the sense that for a given value of r they all have the same solution $\mathbf{u}_r$. To show this equivalence it suffices to demonstrate that the necessary optimality conditions for (6.1) and (6.2) are given by (6.3). This equivalence is obvious for (6.2) and so we proceed with (6.1). Let $\mathbf{u}_r \in \mathcal{V}$ denote the solution of

$$a(\mathbf{u}_r, \mathbf{v}) + r\langle B\mathbf{u}_r, B\mathbf{v}\rangle = \langle f, \mathbf{v}\rangle \quad \forall \mathbf{v} \in \mathcal{V}\,, \tag{6.4}$$

that is a necessary optimality condition for (6.1). Let $\lambda_r = rB\mathbf{u}_r$. By assumption $\lambda_r \in \mathcal{S}$ and so,

$$(\lambda_r, \mu)_{\mathcal{S}} = \langle B\mathbf{u}_r, \mu\rangle = b(\mathbf{u}_r, \mu) \quad \forall \mu \in \mathcal{S}\,.$$

This equation and (6.4) with $rB\mathbf{u}_r$ substituted by λ_r give the modified problem (6.3).

Discretization of the penalized Lagrangian (6.2) and the modified optimality system (6.3) by the same pair of spaces $(\mathcal{V}_h, \mathcal{S}_h)$ give the same parameterized modified saddle point system

$$\begin{bmatrix} \mathbf{A}_h & \mathbf{B}_h^T \\ \mathbf{B}_h & -r^{-1}\mathbf{S}_h \end{bmatrix} \begin{bmatrix} \mathbf{u}_h \\ \lambda_h \end{bmatrix} = \begin{bmatrix} \mathbf{f}_h \\ 0 \end{bmatrix} \tag{6.5}$$

where $\mathbf{S}_h$ is the symmetric and positive definite matrix from (4.4). As a result, λ_h can be eliminated from (6.5) to obtain the reduced parameterized system

$$\left(\mathbf{A}_h + r\mathbf{B}_h^T\mathbf{S}_h^{-1}\mathbf{B}_h\right)\mathbf{u}_h = \mathbf{f}_h\,. \tag{6.6}$$

Discretization of (6.1) requires only a space $\mathcal{V}_h$ for $\mathbf{u}_h$. The parameterized linear system is

$$(\mathbf{A}_h + r\mathbf{G}_h)\,\mathbf{u}_h = \mathbf{f}_h\,, \tag{6.7}$$

where $\mathbf{G}_h$ is a symmetric semi-definite matrix obtained in the usual manner from the second term in (6.4). In general, $\mathbf{G}_h \neq \mathbf{B}_h^T\mathbf{S}_h^{-1}\mathbf{B}_h$ and so (6.7) and (6.6) are not equivalent, even though their continuous prototypes are. This lack of equivalence can be explained by comparing the order of the discretization and elimination steps in the two problems. The system (6.6) is obtained by discretization of the saddle-point problem (6.3) followed by elimination of the *discrete* Lagrange multiplier λ_h. In contrast, (6.7) is a discretization of (6.1) that can be obtained from (6.3) by elimination of the Lagrange multiplier λ, i.e., in this problem the elimination step precedes the discretization step. Changing the order of discretization and elimination leads to different discrete problems with distinct solutions.

Finally, let us discuss well-posedeness of (6.5)-(6.7). For large values of the penalty parameter r the system (6.5) approaches the unmodified KKT problem (4.2). As a result, despite the presence of the symmetric positive definite matrix $\mathbf{S}_h$, stability of (6.5) and (6.6) remains subject to (4.7)-(4.8). Because (6.7) is defined by a single discrete space, it seems logical to expect that this problem will be well-posed regardless of $\mathcal{V}_h$. Unfortunately, this is not true because the penalty term defines an implicit discrete Lagrange multiplier. Indeed, if $\mathbf{G}_h$ is positive definite, (6.7) can be converted to

$$\begin{bmatrix} \mathbf{A}_h & \tilde{\mathbf{B}}_h^T \\ \tilde{\mathbf{B}}_h & -r^{-1}\mathbf{I}_h \end{bmatrix} \begin{bmatrix} \mathbf{u}_h \\ \lambda_h \end{bmatrix} = \begin{bmatrix} \mathbf{f}_h \\ 0 \end{bmatrix} \tag{6.8}$$

where we assume the existence of a $\tilde{\mathbf{B}}_h$ such that $\tilde{\mathbf{B}}_h^T\tilde{\mathbf{B}}_h = \mathbf{G}_h$ and $\lambda_h = r\tilde{\mathbf{B}}_h\mathbf{u}_h$ is a discrete Lagrange multiplier. Therefore, (6.7) remains associated with a parameterized saddle point matrix in which the space $\mathcal{S}_h$ [30] is implicitly defined from $\mathcal{V}_h$ by $\mathcal{S}_h = \tilde{\mathbf{B}}_h\mathcal{V}_h$. Consequently, stability of (6.8), and by extension, of (6.7) is contingent upon the compatibility of $\mathcal{V}_h$ and the implicit Lagrange multiplier space $\mathcal{S}_h$. In some extreme cases the nullspace $\mathcal{Z}_h$ implied by the implicit space $\mathcal{S}_h$ can be empty, causing the finite element method to *lock*. A classical example of locking is when (6.1) is used to solve the Stokes equations by piecewise linear elements. Then, for large penalty, the piecewise linear minimizer of (6.1) is identically zero. To eliminate locking and achieve compatibility, the penalty term in (6.1) is evaluated using a reduced integration order; see [18, 29, 35]. To minimize the attendant penalty error, penalty methods can be applied in an iterative manner [25].

In summary, our discussion shows that modifications in (6.1)-(6.3) cannot be used to circumvent the stability conditions—in all forms the penalty approach does not lead to *stabilized* formulations. A proper interpretation of a penalty method is as a solution technique for the parameterized saddle point system (4.2) that allows us to eliminate the Lagrange multiplier and to replace the indefinite saddle point matrix with a symmetric and positive definite matrix.

6.2. Augmented Lagrangian methods. The principal drawback of penalty methods are the contradictory demands on r placed by accuracy and efficiency requirements, respectively. On one hand, r must be large enough so that the penalty error is comparable to the discretization error. On the other hand, r must be small enough so that (6.6) has a smaller condition number.

Augmented Lagrangian methods [19] are an alternative modification technique that combines the features of (6.1) and (6.2). Like (6.2) they penalize the Lagrangian but use the same penalty term as in (6.1). The modified problem is to find the saddle-point $(\mathbf{u}_r, \lambda_r) \in \mathcal{V}_h \times \mathcal{S}_h$ of the following Lagrangian:

$$L_r(\mathbf{v}, \mu) = L(\mathbf{v}, \mu) + \frac{r}{2}\|B\mathbf{v}\|^2_{\mathcal{S}'}. \tag{6.9}$$

The optimality system of (6.9) is

$$\begin{cases} \tilde{a}(\mathbf{u}_r, \mathbf{v}) + b(\mathbf{v}, \lambda_r) & = \langle f, \mathbf{v}\rangle \quad \forall \mathbf{v} \in \mathcal{V} \\ b(\mathbf{u}_r, \mu) & = 0 \qquad\quad \forall \mu \in \mathcal{S} \end{cases} \tag{6.10}$$

where $\tilde{a}(\mathbf{u}, \mathbf{v}) = a(\mathbf{u}, \mathbf{v}) + r\langle B\mathbf{u}, B\mathbf{v}\rangle$. Discretization of (6.3) is accomplished by a pair of subspaces $\mathcal{V}_h \subset \mathcal{V}$ and $\mathcal{S}_h \subset \mathcal{S}$. The resulting discrete problem

$$\begin{bmatrix} \mathbf{A}_h + r\mathbf{G}_h & \mathbf{B}_h^T \\ \mathbf{B}_h & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_h \\ \lambda_h \end{bmatrix} = \begin{bmatrix} \mathbf{f}_h \\ 0 \end{bmatrix} \tag{6.11}$$

is a modification of the parameterized saddle point system (4.2). In contrast to (6.5), the modification now affects the first equation in (4.2), where the (1,1) block is replaced by the same matrix as in (6.7), and leaves the second equation unchanged. As a result, solutions of (6.11) satisfy the discrete constraint equation and are not subject to a penalty error. Consequently, large values of r are unnecessary for accuracy and the value of this parameter can be optimized for solver efficiency. Stability of (6.11) remains subject of the two conditions (4.7) and (4.8) and so (6.11) is not a *stabilized* method.

Compared with the penalty formulations, augmented Lagrangian methods improve the efficiency of iterative solvers by decreasing the condition number of the $\mathbf{A}_h + r\mathbf{G}_h$ block. This is of particular interest for problems where $a(\cdot, \cdot)$ is coercive strictly on the kernel $\mathcal{Z}$ but $\mathbf{A}_h$ is not invertible. The drawback is that (6.11) is an indefinite linear system and the Lagrange multiplier cannot be eliminated.

As a final remark, we note that all methods considered in this section can be obtained from a modification of an optimization problem. Thus, they retain the symmetry of the original unmodified equations.

7. Stabilized methods. To discuss stabilized methods we rewrite the unmodified optimality system (2.5) as

$$Q(\{\mathbf{u},\lambda\},\{\mathbf{v},\mu\}) = F(\{\mathbf{v},\mu\}) \tag{7.1}$$

where

$$Q(\{\mathbf{u},\lambda\},\{\mathbf{v},\mu\}) = a(\mathbf{u},\mathbf{v}) + b(\mathbf{v},\lambda) + b(\mathbf{u},\mu) \quad \text{and} \quad F(\{\mathbf{v},\mu\}) = \langle f,\mathbf{v}\rangle\,.$$

The goal of stabilization is to replace (7.1) by a modified problem

$$Q_M(\{\mathbf{u},\lambda\},\{\mathbf{v},\mu\}) = F_M(\{\mathbf{v},\mu\}) \tag{7.2}$$

that is weakly coercive over a wider range of discrete spaces than (7.1) and gives rise to a sequence of problems whose solutions converge to a solution of (7.1). Specifically, we seek formulations such that

$$\sup_{(\mathbf{v}_h,\mu_h)\in\mathcal{V}_h\times\mathcal{S}_h} \frac{Q_M(\{\mathbf{u}_h,\lambda_h\},\{\mathbf{v}_h,\mu_h\})}{\|\{\mathbf{v}_h,\mu_h\}\|_{\mathcal{V}_h\times\mathcal{S}_h}} \geq \gamma\|\{\mathbf{u}_h,\lambda_h\}\|_{\mathcal{V}_h\times\mathcal{S}_h},\ \forall\{\mathbf{u}_h,\lambda_h\}\in\mathcal{V}_h\times\mathcal{S}_h \tag{7.3}$$

$$\sup_{(\mathbf{u}_h,\lambda_h)\in\mathcal{V}_h\times\mathcal{S}_h} \frac{Q_M(\{\mathbf{u}_h,\lambda_h\},\{\mathbf{v}_h,\mu_h\})}{\|\{\mathbf{u}_h,\lambda_h\}\|_{\mathcal{V}_h\times\mathcal{S}_h}} > 0,\ \forall\{\mathbf{v}_h,\mu_h\}\in\mathcal{V}_h\times\mathcal{S}_h \tag{7.4}$$

with $\gamma > 0$ independent of h, for pairs $(\mathcal{V}_h,\mathcal{S}_h)$ that are not subject to (4.7)-(4.8). There are two basic approaches that lead to such modified equations. The first is based on the observation that (7.3)-(7.4) can be achieved by adding a properly weighted term $B'\mu$ to the bilinear form Q. The second approach uses projection operators acting on the discrete Lagrange multiplier space. These operators are seen as filters that remove unwanted discrete modes from $\mathcal{S}_h$.

If $\mathbf{K}_h$ is the matrix obtained from the modified bilinear form, and $\mathbf{N}_h = (\mathbf{V},\mathbf{S})$ is the matrix that generates the norm on $\mathcal{V}_h\times\mathcal{S}_h$, then (7.3) is equivalent to

$$\inf_{\mathbf{y}}\sup_{\mathbf{x}} \frac{\mathbf{y}^T\mathbf{K}_h\mathbf{x}}{\mathbf{y}^T\mathbf{N}_h\mathbf{x}} \geq \gamma > 0\,.$$

Therefore, algebraically, weak coercivity is equivalent [4] to having the smallest generalized singular value of $\mathbf{K}_h$ bounded away from zero independently of h.

7.1. Residual stabilization. Given an arbitrary pair $(\mathbf{v},\mu)\in\mathcal{V}\times\mathcal{S}$ the residual of (2.6) is the function

$$\mathcal{R}(\mathbf{v},\mu) = \begin{pmatrix} A\mathbf{v} + B'\mu - f \\ B\mathbf{v}\end{pmatrix} \in \mathcal{V}'\times\mathcal{S}'\,.$$

Residual stabilization takes advantage of the fact that the term $B'\mu$ can be introduced through the residual in a consistent manner. Indeed, if $(\mathbf{u},\lambda)$ is a solution of (2.6), then $\mathcal{R}(\mathbf{u},\lambda) = 0$, and any terms that were added to (2.5) will vanish. As a result, $(\mathbf{u},\lambda)$ will also be a solution of (7.2). Such consistency is the chief appeal of residual stabilization because a definition of high-order stabilized methods is automatic.

ETNA
Kent State University
etna@mcs.kent.edu

One approach to residual stabilization is to use a least-squares form of the residual in the Lagrangian functional. In this case we have a choice of the full least-squares modification

$$L_{FLS}(\mathbf{v},\mu) = L(\mathbf{v},\mu) - \frac{\tau_1}{2}\|A\mathbf{v} + B'\mu - f\|^2_{\mathcal{V}'} + \frac{\tau_2}{2}\|B\mathbf{v}\|^2_{\mathcal{S}'} \tag{7.5}$$

or the partial least-squares modification

$$L_{PLS}(\mathbf{v},\mu) = L(\mathbf{v},\mu) - \frac{\tau_1}{2}\|A\mathbf{v} + B'\mu - f\|^2_{\mathcal{V}'}\,, \tag{7.6}$$

where τ_1 and τ_2 are real stabilization parameters. The augmented Lagrangian functional (6.9) uses the second component of the residual and so is another variant of the partial least-squares modification. However, this term lacks the critical operator $B'\mu$ needed for (7.3)-(7.4).

The residual terms in (7.5)-(7.6) change (7.1) to a problem (7.2) with

$$F_M(\{\mathbf{v},\mu\}) = F(\{\mathbf{v},\mu\}) - \tau_1\langle f, A\mathbf{v} + B'\mu\rangle$$

and

$$Q_M(\{\mathbf{u},\lambda\},\{\mathbf{v},\mu\}) =$$
$$Q(\{\mathbf{u},\lambda\},\{\mathbf{v},\mu\}) - \begin{cases} \tau_1\langle A\mathbf{u} + B'\lambda, A\mathbf{v} + B'\mu\rangle - \tau_2\langle B\mathbf{u}, B\mathbf{v}\rangle & \text{for (7.5)} \\ \tau_1\langle A\mathbf{u} + B'\lambda, A\mathbf{v} + B'\mu\rangle & \text{for (7.6)} \end{cases}$$

A second approach to residual stabilization is to derive the modified variational equation (7.2) directly by adding weighted residual terms to (2.5):

$$\begin{aligned} &Q_M(\{\mathbf{u},\lambda\},\{\mathbf{v},\mu\}) = \\ &\qquad Q(\{\mathbf{u},\lambda\},\{\mathbf{v},\mu\}) - \tau_1\langle A\mathbf{u} + B'\lambda, W(\{\mathbf{v},\mu\})\rangle + \tau_2\langle B\mathbf{u}, V(\{\mathbf{v},\mu\})\rangle\,; \\ &F_M(\{\mathbf{v},\mu\}) = F(\{\mathbf{v},\mu\}) - \langle f, W(\{\mathbf{v},\mu\})\rangle\,. \end{aligned}$$

The type of stabilized formulation depends on the weight operators W and V. The second operator is often set to zero and methods are defined by a choice of W. Following the taxonomy introduced in [4], three important cases are

$$W(\{\mathbf{v},\mu\}) = \begin{cases} A\mathbf{v} + B'\mu & \text{Least-squares (LS)} \\ -A\mathbf{v} + B'\mu & \text{Reflected least-squares (RLS)} \\ B'\mu & \text{Simplified least-squares (SLS)}\,. \end{cases} \tag{7.7}$$

The least-squares weight operator gives the same formulation as (7.6). The other two weight operators result in variational equations that cannot be derived from optimization problems, or in other words are not first-order optimality conditions of some Lagrangian. As a result, the symmetry of the original variational problem is lost under such modifications. If (7.3)-(7.4) hold for all values of τ_1 and/or τ_2, the stabilized form is called *absolutely* stable. If this form is weakly coercive for a range of stabilizing parameters, the form is called *conditionally* stable.

The main challenge in the implementation of residual stabilization is the term $\langle A\mathbf{u} + B'\lambda, W(\{\mathbf{v},\mu\})$. For the Stokes problem $\mathcal{V}'$ is the dual space $H^{-1}(\Omega)$ and so $\langle\mathbf{u},\mathbf{v}\rangle_{-1} = \langle(-\triangle)^{-1}\mathbf{u},\mathbf{v}\rangle_0$. Hence the computation of the consistent residual stabilization term $\langle -\triangle\mathbf{u}+$

$\nabla\lambda, W(\{\mathbf{v},\mu\})\rangle_{-1}$ impractical and is often replaced by a sum of weighted L^2 norms computed element by element:

$$\sum_K \tau_K \int_K (-\triangle \mathbf{u}_h + \nabla\lambda_h)\, W(\{\mathbf{v}_h,\mu_h\})dx\,, \tag{7.8}$$

where τ_K is stabilization parameter. Typically $\tau_K = \delta h_K^2$ for some positive real δ. The use of a *broken* L^2 norm in (7.8) is required because standard finite element functions are not continuously differentiable and $-\triangle$ cannot be applied globally in Ω. The first choice in (7.7) leads to the original *Galerkin Least-Squares* [27] method. The last choice is the so-called *Pressure-Poisson* stabilized Galerkin method [28]. These two methods are conditionally stable [21] because (7.3)-(7.4) hold for $0 < \delta \le \delta_{\max}$, where $\delta_{\max}$ is a positive real number that may depend on the shape of Ω, the Poincare inequality constant and the inverse inequality constant. The range of δ in these methods is limited by the need to balance terms that provide stability with terms that fulfill the consistency requirement. The second choice in (7.7) gives the absolutely stable *Douglas-Wang* stabilized Galerkin method [16]. For a further discussion and taxonomy of these methods and their extensions to linear elasticity we refer the reader to [21, 4, 22] respectively.

We remark that $-\triangle$ may be replaced with a discrete operator $-\triangle^h$ that is meaningful for C^0 finite element functions [7]. The resulting method is slightly inconsistent in the sense that the modified residual term does not vanish for the exact solution. However, the inconsistency is within the approximation order and so convergence does not suffer.

Residual stabilization of the Darcy problem does not experience this kind of problems because the relevant residual $\mathbf{u} + \nabla\lambda$ does not involve second order derivatives. As a result, the stabilizing term can be implemented using a standard L^2 inner product. A stabilized method that uses the least-squares form of the weight function (7.7) was developed in [32]. An interesting feature of this method is the mesh-independence of the stabilizing parameter τ_1 that is set to $1/2$.

Let us examine the structure of parameterized linear systems obtained by residual stabilization. Restriction of (7.2) to a pair $(\mathcal{V}_h, \mathcal{S}_h)$ of discrete spaces gives rise to the parameterized linear system

$$\left\{ \begin{bmatrix} \mathbf{A}_h & \mathbf{B}_h^T \\ \mathbf{B}_h & 0 \end{bmatrix} + \begin{bmatrix} \tilde{\mathbf{A}}_h & \tilde{\mathbf{B}}_h \\ -\mathbf{C}_h & -\mathbf{D}_h \end{bmatrix} \right\} \begin{bmatrix} \mathbf{u}_h \\ \lambda_h \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix} + \begin{bmatrix} \tilde{f} \\ \tilde{g} \end{bmatrix}. \tag{7.9}$$

The first block matrix on the left-hand side is generated by the unmodified bilinear form in (2.5). The terms that were added through the residual contribute the second, stabilizing matrix and a consistency term to the right hand side. The matrices $\mathbf{C}_h$ and $\mathbf{D}_h$ are obtained from $\langle B'\lambda_h, A\mathbf{v}_h\rangle$ and $\langle B'\lambda_h, B'\mu_n\rangle$, respectively. The other two blocks in the stabilizing matrix are given by

$$\tilde{\mathbf{A}}_h = \begin{cases} -\tau_1\mathbf{H}_h + \tau_2\mathbf{G}_h & \text{for (7.5)} \\ -\tau_1\mathbf{H}_h & \text{for (7.6)} \\ +\tau_1\mathbf{H}_h & \text{for RLS} \\ 0 & \text{for SLS} \end{cases} ; \quad \tilde{\mathbf{B}}_h = \begin{cases} -\tau_1\mathbf{C}_h^T & \text{for (7.5) and (7.6)} \\ +\tau_1\mathbf{C}_h^T & \text{for RLS} \\ 0 & \text{for SLS} \end{cases}$$

where $\mathbf{H}_h$ is a matrix generated by $\langle A\mathbf{u}_h, A\mathbf{v}_h\rangle$.

The stabilizing effect is achieved by the (2,2) term of the second matrix in (7.9). The rest of the terms in this matrix are needed to fulfill the consistency of the method and may have a destabilizing effect. The role of the two parameters τ_1 and τ_2 is to prevent destabilization by balancing stabilization and consistency terms.

ETNA
Kent State University
etna@mcs.kent.edu

7.2. Non-residual stabilization. Non-residual stabilization relies upon specific features of the problem solved. As a result, non-residual stabilization does not lend itself so easily to a formal discussion and categorization. Nevertheless, there are some common features that we briefly discuss before moving onto specific examples.

A typical non-residual stabilization method modifies (7.1) to a problem where the right hand side functional is unchanged, while the bilinear form $Q(\cdot,\cdot)$ is replaced by

$$Q_M(\{\mathbf{u},\lambda\},\{\mathbf{v},\mu\}) = Q(\{\mathbf{u},\lambda\},\{\mathbf{v},\mu\}) - \langle D_h\lambda, D_h\mu\rangle_h \tag{7.10}$$

where D_h is a discrete stabilizing operator. As a rule, this operator acts on the discrete Lagrange multiplier space $\mathcal{S}_h$ and has a non-trivial kernel. Compared with (7.9) the linear system associated with (7.10) is much simpler

$$\left\{\begin{bmatrix} \mathbf{A}_h & \mathbf{B}_h^T \\ \mathbf{B}_h & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & -\mathbf{D}_h \end{bmatrix}\right\}\begin{bmatrix} \mathbf{u}_h \\ \lambda_h \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix} \tag{7.11}$$

and retains the symmetry of the original problem. The matrix in (7.11) has the exact same structure as the matrix (6.5) in the penalty method. The key difference is that $\mathbf{D}_h$ is symmetric and semi-definite, which prevents the elimination of the Lagrange multiplier from (7.10) and so (7.11) is not a penalty method. A second, more important difference is that the operator D_h is such that the form (7.10) satisfies (7.3)-(7.4) with pairs $(\mathcal{V}_h, \mathcal{S}_h)$ that are not stable for the penalty method. A general rule of a thumb to achieve stabilization in (7.11) is to design D_h in such a way that $\ker(D_h) \cap \tilde{\mathcal{S}}_h \neq \emptyset$ for some $\tilde{\mathcal{S}}_h \subset \mathcal{S}$ that forms a stable pair with $\mathcal{V}_h$. As a result, the stabilization term in (7.11) "sees" only the unstable component of λ_h. Thus, D_h acts as a filter rather than a penalty term.

Let us examine some of the non-residual stabilization methods for the Stokes equations (3.2) and the Darcy problem (3.4). The *pressure gradient projection* method [6] relaxes the incompressibility constraint by the difference between the pressure gradient and its L^2 projection onto the velocity space $\mathcal{V}_h$. For this method, the last term in (7.10) is given by

$$\alpha\langle (P-I)\nabla\lambda_h, (P-I)\nabla\mu_h\rangle_0$$

where P is the L^2 projection $\mathbf{L}^2(\Omega) \mapsto \mathcal{V}_h$. This method is motivated by fractional step solution techniques [6] for transient incompressible flows. Because the gradient is projected onto the C^0 velocity space, computation of the stabilization term is a global problem. In practice the method is implemented by introducing the projected gradient as a new dependent variable in the mixed form.

The *polynomial pressure projection method* of [15] also uses an L^2 projection. However, the projected variable is the pressure, and the projection operator P is onto a local polynomial space $\mathcal{P}(K)$. This operator maps the pressure space $\mathcal{S}_h$ into a discontinuous space $\cup_K \mathcal{P}(K)$. The stabilizing term for this method is given by

$$\alpha\langle (P-I)\lambda_h, (P-I)\mu_h\rangle_0$$

and can be computed locally in an element by element fashion. Another attractive property of this stabilization is the possibility to extend it to discontinuous pressure spaces and to the Darcy problem.

Two non-residual stabilized formulations designed specifically for low order finite elements with discontinuous pressure spaces $\mathcal{S}_h$ are the *global pressure jump* and the *local pressure jump* [37, 36] stabilization methods. Let Γ_h denote the set of all internal edges in a

ETNA
Kent State University
etna@mcs.kent.edu

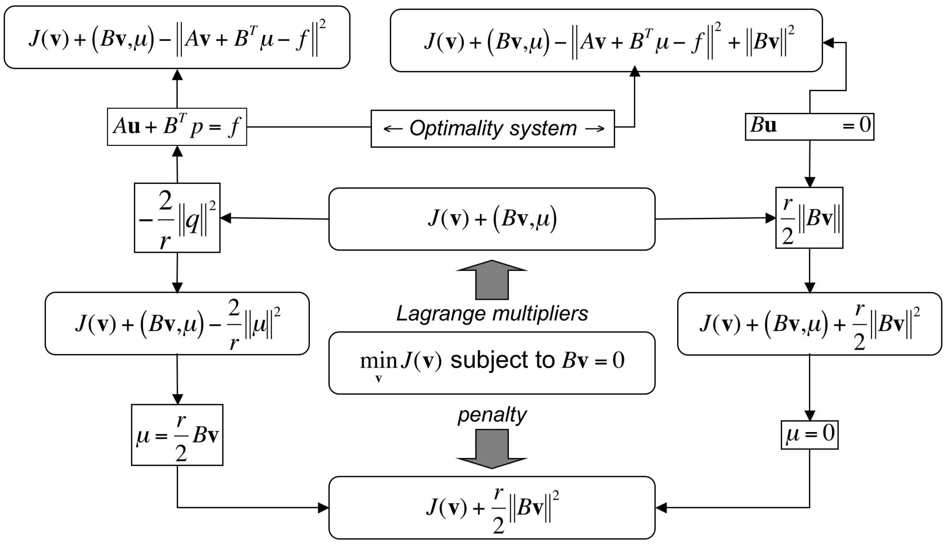

FIG. 8.1. *A schematic connection between penalty, augmented Lagrangian and residual modifications of saddle-point problems.*

finite element partition $\mathcal{T}_h$ of Ω and let $[\lambda_h]$ denote the jump of a function $\lambda_h \in \mathcal{S}_h$ across an edge e. In the first method, the stabilizing term is given by

$$\alpha h \sum_{e \in \Gamma_h} \int_e [\lambda_h][\mu_h] d\Gamma \, .$$

To define the second method, suppose that the elements in $\mathcal{T}_h$ can be arranged in non-overlapping patches $\mathcal{T}_h^i$ with approximately the same number of elements. The stabilizing term for the local pressure jump stabilization method is given by

$$\alpha h \sum_{\mathcal{T}_h^i} \sum_{e \in \Gamma_h^i} \int_e [\lambda_h][\mu_h] d\Gamma$$

where Γ_h^i is the set of all internal edges in patch $\mathcal{T}_h^i$. In this method jumps are integrated only on those edges from Γ_h that are internal to a patch $\mathcal{T}_h^i$. As a result, the stabilized formulation enforces local conservation on each patch.

8. Conclusions. Finite element solution of PDEs associated with constrained minimization problems leads to parameterized saddle point systems of equations. The well-posedeness of the associated sequence of the saddle point matrices is subject to the discrete compatibility conditions (4.7)-(4.8). These conditions guarantee that discrete solutions converge to the exact solution as the mesh size h approaches zero. Penalty and augmented Lagrangian modifications are regularization methods that lead to saddle point systems of equations that are more amenable to preconditioned iterative solvers but do not eliminate the stringent compatibility conditions on the discrete spaces. Residual and non-residual stabilization are modifications that circumvent the stability conditions and allow definition of a well-posed parameterized family of algebraic equations by a wider range of discrete spaces. Different types of modifications can be obtained using the same terms and may be related to equivalent variational problems at the continuum level. Several important cases are illustrated in Figure 8.1. However, modification and discretization steps *do not commute* and the resulting method will depend upon the order of these steps. In other words, modifications to the discrete saddle-point system lead to problems that are different from those obtained by the discretization of the modified variational equations.

Acknowledgment. We are indebted to the anonymous referees for their thorough and careful reading of the paper and the detailed remarks that helped to improve significantly the manuscript. We also thank one of the referees for pointing out the reference [10].

REFERENCES

[1] D. N. ARNOLD, R. S. FALK, AND R. WINTHER, *Multigrid in H(div) and H(curl)*, Numer. Math., 85 (2000), pp. 197–218.
[2] D. N. ARNOLD AND R. WINTHER, *Mixed finite elements for elasticity*, Numer. Math., 42 (2002), pp. 401–419.
[3] I. BABUŠKA, *The finite element method with Lagrange multipliers*, Numer. Math., 20 (1973), pp. 179–192.
[4] T. BARTH, P. BOCHEV, M. GUNZBURGER, AND J.N.SHADID, *A taxonomy of consistently stabilized finite element methods for the Stokes problem*, SIAM J. Sci. Comput., 25 (2004), pp. 1585–1607.
[5] M. BENZI, G. GOLUB, AND J. LIESEN, *Numerical solution of saddle point problems*, Acta Numer., 14 (2005), pp. 1–137.
[6] J. BLASCO AND R. CODINA, *Stabilized finite element method for the transient Navier-Stokes equations based on a pressure gradient projection*, Comput. Methods Appl. Math., 182 (2000), pp. 277–300.
[7] P. BOCHEV AND M. GUNZBURGER, *An absolutely stable pressure-Poisson stabilized method for the Stokes equations*, SIAM J. Numer. Anal., 42 (2005), pp. 1189–1207.
[8] J. BOLAND AND R. NICOLAIDES, *Stable and semistable low order finite elements for viscous flows*, SIAM J. Numer. Anal., 22 (1985), pp. 474–492.
[9] F. BREZZI, *On existence, uniqueness and approximation of saddle-point problems arising from Lagrange multipliers*, Mathematical Modeling and Numerical Analysis, 21 (1974), pp. 129–151.
[10] ———, *Stability of saddle-points in finite dimensions*, in Frontiers in Numerical Analysis, J. Blowey, A. Craig, and T. Shardlow, eds., Springer Verlag, 2002, pp. 17–61.
[11] F. BREZZI AND K.-J. BATHE, *A discourse on the stability conditions for mixed finite element formulations*, Comput. Methods Appl. Math., 82 (1990), pp. 27–57.
[12] F. BREZZI AND M. FORTIN, *Mixed and Hybrid Finite Element Methods*, Springer, Berlin, 1991.
[13] E. BURMAN AND P. HANSBO, *A unified stabilized method for Stokes and Darcy's equations*, Tech. Report 2002-15, Chalmers University of Technology, Chalmers Finite Element Center, 2002.
[14] G. F. CAREY, G. BICKEN, V. CAREY, C. BERGER, AND J. SANCHEZ, *Locally constrained projections*, Internat. J. Numer. Methods Engrg., 50 (2001), pp. 549–577.
[15] C. DOHRMANN AND P. BOCHEV, *A stabilized finite element method for the Stokes problem based on polynomial pressure projections*, International Journal on Numerical Methods in Engineering, 46 (2004), pp. 183–201.
[16] J. DOUGLAS AND J. WANG, *An absolutely stabilized finite element method for the Stokes problem*, Math. Comp., 52 (1989), pp. 495–508.
[17] H. C. ELMAN, D. J. SILVESTER, AND A. J. WATHEN, *Finite Elements and Fast Iterative Solvers with applications in incompressible fluid dynamics*, Numerical Mathematics and Scientific Computation, Oxford University Press, 2005.
[18] R. FALK, *An analysis of the penalty method and extrapolation for the stationary Stokes equations*, in Advances in Computer Methods for Partial Differential Equations, R. Vichnevetsky, ed., New Brunswick,NJ, 1975, AICA, pp. 66–69.
[19] M. FORTIN AND R. GLOWINSKI, *Augmented Lagrangian methods: applications to numerical solution of boundary value problems*, in Studies in Mathematics and its applications, J. L. Lions, G. Papanicolaou, R. T. Rockafellar, and H. Fujita, eds., vol. 15, North-Holland, 1983.

ETNA
Kent State University
etna@mcs.kent.edu

[20] M. FORTIN AND R. PIERRE, *Stability analysis of discrete generalized Stokes problems*, Numer. Methods for Partial Differential Equations, 8 (1992), pp. 303–323.
[21] L. FRANCA, T. HUGHES, AND R. STENBERG, *Stabilized finite element methods*, in Incompressible Computational Fluid Dynamics, M. Gunzburger and R. Nicolaides, eds., Cambridge, University Press, 1993, pp. 87–108.
[22] L. FRANCA AND R. STENBERG, *Error analysis of some Galerkin least-squares methods for the elasticity equations*, SIAM J. Numer. Anal., 28 (1991), pp. 1680–1697.
[23] V. GIRAULT AND P. RAVIART, *Finite Element Methods for Navier-Stokes Equations*, Springer, Berlin, 1986.
[24] M. GUNZBURGER, *Finite Element Methods for Viscous Incompressible Flows*, Academic, Boston, 1989.
[25] ———, *Iterative penalty methods for the Stokes and Navier-Stokes equations*, in Finite Element Analysis in Fluids, University of Alabama, Huntsville, 1989, pp. 1040–1045.
[26] ———, *The inf-sup condition in mixed finite element methods with application to the Stokes system*, in Collected lectures on the preservation of stability under discretization, D. Estep and S. Tavener, eds., SIAM Proceedings series list, SIAM, 2002, pp. 93–122.
[27] T. HUGHES AND L. FRANCA, *A new finite element formulation for computational fluid dynamics: VII. the Stokes problem with various well-posed boundary conditions: symmetric formulations that converge for all velocity pressure spaces*, Comput. Methods Appl. Math., 65 (1987), pp. 85–96.
[28] T. HUGHES, L. FRANCA, AND M. BALESTRA, *A new finite element formulation for computational fluid dynamics: Circumventing the Babuska-Brezzi condition: A stable Petrov-Galerkin formulation of the Stokes problem accommodating equal-order interpolations*, Computer Methods in Applied Mechanics and Engineering, 59 (1986), pp. 85–99.
[29] T. HUGHES, W. LIU, AND A. BROOKS, *Finite element analysis of incompressible viscous flows by the penalty function formulation*, J. Comput. Phys., 30 (1979), pp. 1–60.
[30] T. HUGHES AND D. MALKUS, *Mixed finite element methods—reduced and selective integration techniques—a unification of concepts*, Computer Methods in Applied Mechanics and Engineering, 15 (1978).
[31] K. MARDAL, X. TAI, AND R. WINTHER, *A robust finite element method for Darcy-Stokes flow*, SIAM J. Numer. Anal., 40 (2002), pp. 1605–1631.
[32] A. MASUD AND T. J. R. HUGHES, *A stabilized finite element method for Darcy flow*, Comput. Methods Appl. Math., 191 (2002), pp. 4341–4370.
[33] J. NOCEDAL AND S. WRIGHT, *Numerical optimization*, Springer Verlag, 1999.
[34] J. ODEN AND N. KIKUCHI, *Finite element methods for constrained problems in elasticity*, Internat. J. Numer. Methods Engrg., 18 (1982), pp. 701–725.
[35] T. J. ODEN, *Rip-methods for Stokesian flow*, in Finite elements in fluids, H. Gallagher, D. Norrie, T. Oden, and O. Zienkiewicz, eds., vol. 4, John Wiley, Chichester, 1982, pp. 305–318.
[36] D. SILVESTER, *Optimal low order finite element methods for incompressible flow*, Comput. Methods Appl. Math., 111 (1994), pp. 357–368.
[37] D. J. SILVESTER AND N. KECHKAR, *Stabilized bilinear-constant velocity-pressure finite elements for the conjugate gradient solution of the Stokes problem*, Comput. Methods Appl. Math., 79 (1990), pp. 71–86.
[38] A. TOSELLI AND O. WIDLUND, *Domain Decomposition Methods—Algorithms and Theory*, Springer Series in Computational Mathematics, Springer, 2005.