MULTILEVEL FIRST-ORDER SYSTEM LEAST SQUARES FOR ELLIPTIC GRID GENERATION

A. L. CODD*, T. A. MANTEUFFEL†, S. F. MCCORMICK†, AND J. W. RUGE†

Abstract. A new fully-variational approach is studied for elliptic grid generation (EGG). It is based on a general algorithm developed in a companion paper [10] that involves using Newton's method to linearize an appropriate equivalent first-order system, first-order system least squares (FOSLS) to formulate and discretize the Newton step, and algebraic multigrid (AMG) to solve the resulting matrix equation. The approach is coupled with nested iteration to provide an accurate initial guess for finer levels using coarse-level computation. The present paper verifies the assumptions of [10] and confirms the overall efficiency of the scheme with numerical experiments.

Key words. least-squares discretization, multigrid, nonlinear elliptic boundary value problems

AMS subject classifications. 35J65, 65N15, 65N30, 65N50, 65F10

1. Introduction.

A companion paper [10] develops an algorithm using Newton's method, first-order system least squares (FOSLS), and algebraic multigrid (AMG) for efficient solution of general nonlinear elliptic equations. The equations are first converted to an appropriate first-order system and an approximate solution to the coarsest-grid problem is then computed (by any suitable method such as Newton iteration coupled with perhaps direct solvers, damping, or continuation). The approximation is then interpolated to the next finer level where it is used as an initial guess for one Newton linearization of the nonlinear problem, with a few AMG cycles applied to the resulting matrix equation. This algorithm repeats itself until the finest grid is processed, again by *one* Newton/AMG step. At each Newton step, FOSLS is applied to the linearized system and the resulting matrix equation is solved using just a few V-cycles of AMG.

In the present paper, we apply this algorithm to the elliptic grid generation (EGG) equations. Grid generation is usually based on a map between a relatively simple *computational* region and a possibly complicated *physical* region. It can be used numerically to create a mesh for a discretization method to solve a given system of equations posed on the physical domain. Alternatively, it can be used to transform equations posed on the physical region to ones posed on the computational region, where the transformed equations are then solved. If the Jacobian of the transformation is positive throughout the computational region, the equation type is unchanged [12]. Actually, the relative minimum value of the Jacobian is important in practice because relatively small values signal small angles between the grid lines and large errors in approximating the equations [20].

Our interest is in elliptic grid generation using the Winslow generator [12], which allows us to specify the boundary maps completely. Moreover, by choosing the two-dimensional computational region to be convex, we can ensure that the Jacobian of the map is positive, which in turn ensures that the map is one-to-one and onto and

*Centre for Mathematics and Its Applications, School of Mathematical Sciences, Australian National University, Canberra, ACT 0200, Australia. *email*: andrea.codd@anu.edu.au.,

†Department of Applied Mathematics, Campus Box 526, University of Colorado at Boulder, Boulder, CO 80309–0526. *email*: tmanteuf@boulder.colorado.edu, stevem@boulder.colorado.edu, and jruge@boulder.colorado.edu. This work was sponsored by the National Institute of Health under grant number 1–R01–EY12291–01, the National Science Foundation under grant number DMS–9706866, and the Department of Energy under grant number DE–FG03–93ER25165.,

therefore does not fold [8]. The Winslow generator tends to create smooth grids, with good aspect ratios. The map also tends to control variations in gridline spacing and nonorthogonality of the gridline intersections in the physical space. See Thompson, Warsi, and Mastin [20] and Knupp and Steinberg [12] for background on grid generation in general, and EGG in particular. Several discretization methods of the EGG equations together with their associated errors are discussed in [20]. In [12], the EGG equations are derived and several existing methods are described for solving them.

A brief description of the first-order EGG system is given in section 2. The assumptions needed to apply the theory in [10] are verified in section 3. Section 4 discusses scaling of the functional terms used for the computations as well as numerical results for two representative problems. The last section includes some final remarks.

2. Equations. We use standard notation for the associated spaces. Restricting ourselves to two dimensions, consider a generic open domain, $\Omega \in R^2$, with Lipschitz boundary $\Gamma \in C^{3,1}$. (Superscript 1 indicates Lipschitz continuity of the functions and their derivatives.) Suppose $m \geq 0$ and $n \geq 1$ are given integers. Let $(\cdot,\cdot)_{0,\Omega}$ denote the inner product on $L^2(\Omega)^n$, $\|\cdot\|_{0,\Omega}$ its induced norm, and $H^m(\Omega)^n$ the standard Sobolev space with norm $\|\cdot\|_{m,\Omega}$ and seminorms $|\cdot|_{i,\Omega}$ $(0 \leq i \leq m)$. (Superscript n is omitted when dependence is clear by context.) For $\delta \in (0,1)$, let $H^{m+\delta}(\Omega)$ (c.f. [6]) denote the Sobolev space associated with the norm defined by

$$\|u\|^2_{m+\delta,\Omega} \equiv \|u\|^2_{m,\Omega} + \sum_{|\alpha|=m} \int_\Omega \int_\Omega \frac{|\partial_\alpha u(x) - \partial_\alpha u(y)|^2}{|x-y|^{2(1+\delta)}} dxdy.$$

(This definition allows the use of the "real interpolation" method [1, 6].) Also, let $H^{\frac{1}{2}}(\Gamma)$ denote the trace Sobolev space associated with norm

$$\|u\|_{\frac{1}{2},\Gamma} \equiv \inf\{\|v\|_{1,\Omega} : v \in H^1(\Omega),\ \text{trace } v = u \text{ on } \Gamma\}.$$

We start by mapping a known convex computational region, $\Omega \in R^2$ with boundary $\Gamma \in C^{3,1}$ to a given physical region, $\Omega_{\mathbf{x}} \in R^2$ with boundary $\Gamma_{\mathbf{x}} \in C^{3,1}$. We define map $\boldsymbol{\xi} : \bar{\Omega}_{\mathbf{x}} \to \bar{\Omega}$ and its inverse, $\mathbf{x} : \bar{\Omega} \to \bar{\Omega}_{\mathbf{x}}$. The coordinates in $\Omega_{\mathbf{x}}$ are denoted by the vector of unknowns $\mathbf{x} = (x \quad y)^t$ and those in Ω by $\boldsymbol{\xi} = (\xi \quad \eta)^t$.

For the EGG smoothness or Winslow generator, we choose $\boldsymbol{\xi}$ to be harmonic:

$$\begin{aligned} \Delta\boldsymbol{\xi} &= \mathbf{0}, && \text{in} \quad \Omega_{\mathbf{x}}, \\ \boldsymbol{\xi} &= \mathbf{v}(\mathbf{x}), && \text{on} \quad \Gamma_{\mathbf{x}}, \end{aligned} \tag{2.1}$$

where $\mathbf{v} \in H^{\frac{7}{2}}(\Gamma_{\mathbf{x}})$ is a given homeomorphism (continuous and one-to-one) from the boundary of the physical region onto the boundary of the computational region. ($H^{\frac{7}{2}}(\Gamma_{\mathbf{x}})$ is consistent with our boundary smoothness assumption, $\Gamma \in C^{3,1}$.) With $\Omega_{\mathbf{x}}$ bounded, the weak form of Laplace system (2.1) has one and only one solution, $\boldsymbol{\xi}^*$, in $H^4(\Omega_{\mathbf{x}})^2$ [11] and, by Weyl's Lemma [22], $\boldsymbol{\xi}^* \in C^\infty_{loc}(\Omega_{\mathbf{x}}) \equiv \{\boldsymbol{\xi} \in C^\infty(K)\,,\ \forall\, K \subset \Omega_{\mathbf{x}}\}$.

Map $\boldsymbol{\xi}^*$ is posed on $\Omega_{\mathbf{x}}$, so computing an approximation to it would nominally involve specifying a grid on the physical region. But specifying such a grid is the aim of EGG in the first place, so this formulation is not useful. We therefore choose instead to solve the inverse of problem (2.1), which takes a regular grid in Ω and maps it onto a grid in $\Omega_{\mathbf{x}}$, thus achieving our objective. To this end, we assume $\Omega_{\mathbf{x}}$ and Ω to be simply connected and bounded and $\bar{\Omega}$ to be convex, so Γ and $\Gamma_{\mathbf{x}}$ are simple closed curves. Map $\boldsymbol{\xi}^*$ is continuous and harmonic and $\mathbf{v}$ is a homeomorphism of $\Gamma_{\mathbf{x}}$ onto Γ, so Rado's Theorem (c.f. [16]) implies existence of a unique inverse map, $\mathbf{x}^*$, from

Ω onto $\Omega_{\mathbf{x}}$. An outline of the proof is provided in [14]. It then follows that domain map $\boldsymbol{\xi}^*$ is a diffeomorphism [8, 12] and the associated Jacobian, $J^*_{\mathbf{x}} \equiv \xi^*_x \nu^*_y - \xi^*_y \nu^*_x$, is continuous and uniformly positive and bounded on $\Omega_{\mathbf{x}}$ ($J_0 \leq |J^*_{\mathbf{x}}(x,y)| \leq J_1$ for some constants $J_0, J_1 \in R^+$ and all $(x,y) \in \Omega_{\mathbf{x}}$). The choice for the space for $\mathbf{x}^*$ follows from the assumptions for $\boldsymbol{\xi}^*$, $\Gamma_{\mathbf{x}}$, and $\mathbf{v}$, and is discussed further in section 3.

The inverse map satisfies the following equations (positive Jacobian throughout $\Omega_{\mathbf{x}}$ ensures that the solution of (2.1) is an invertible map):

$$
\begin{aligned}
(x_\eta^2 + y_\eta^2)x_{\xi\xi} - (x_\xi x_\eta + y_\xi y_\eta)(x_{\xi\eta} + x_{\eta\xi}) + (x_\xi^2 + y_\xi^2)x_{\eta\eta} &= 0, && \text{in} \quad \Omega,\\
(x_\eta^2 + y_\eta^2)y_{\xi\xi} - (x_\xi x_\eta + y_\xi y_\eta)(y_{\xi\eta} + y_{\eta\xi}) + (x_\xi^2 + y_\xi^2)y_{\eta\eta} &= 0, && \text{in} \quad \Omega,\\
x &= w_1(\xi,\eta), && \text{on} \quad \Gamma,\\
y &= w_2(\xi,\eta), && \text{on} \quad \Gamma,
\end{aligned}
\tag{2.2}
$$

where function $\mathbf{w} = \begin{pmatrix} w_1(\xi,\eta) \\ w_2(\xi,\eta) \end{pmatrix}$ is the inverse of function $\mathbf{v} = \begin{pmatrix} v_1(x,y) \\ v_2(x,y) \end{pmatrix}$ (i.e., $\mathbf{x} = \mathbf{w}(\mathbf{v}(\mathbf{x}))$). See [12] for more detail. The inverse map, $\mathbf{x}^*$, exists and solves (2.2). We assume that the Fréchet derivative of the operator in (2.2) at $\mathbf{x}^*$ is one-to-one on $H_0^{2+\delta}(\Omega)^2$ (subscript 0 denoting homogeneous Dirichlet conditions on Γ). This is easily verified when $\mathbf{x}^*$ deviates from a constant map by a sufficiently small amount.

To apply our method, we begin by converting equations (2.2) to a first-order system. We could write these equations in a simple way using the standard notation of a 2×2 matrix for the Jacobian matrix, but this is not convenient for the linearized equations treated section 3. Our notation is therefore based primarily on writing the Jacobian matrix as a 4×1 vector:

$$
\mathbf{J} = \begin{pmatrix} x_\xi \\ x_\eta \\ y_\xi \\ y_\eta \end{pmatrix} = \begin{pmatrix} J_{11} \\ J_{21} \\ J_{12} \\ J_{22} \end{pmatrix}.
$$

On the other hand, at times it is useful to refer to the matrix form of the unknowns. We therefore define the block-structured matrix $\underline{\mathbf{J}}$ and its classical adjoint $\hat{\underline{\mathbf{J}}}$ as follows:

$$
\underline{\mathbf{J}} = \begin{pmatrix} J_{11} & J_{21} & 0 & 0 \\ J_{12} & J_{22} & 0 & 0 \\ 0 & 0 & J_{11} & J_{21} \\ 0 & 0 & J_{12} & J_{22} \end{pmatrix} \quad \text{and} \quad \hat{\underline{\mathbf{J}}} = \begin{pmatrix} J_{22} & -J_{21} & 0 & 0 \\ -J_{12} & J_{11} & 0 & 0 \\ 0 & 0 & J_{22} & -J_{21} \\ 0 & 0 & -J_{12} & J_{11} \end{pmatrix}.
$$

Note that the Jacobian of the inverse transformation is given by

$$
J \equiv x_\xi y_\eta - x_\eta y_\xi = J_{11}J_{22} - J_{21}J_{12} = \sqrt{\det \underline{\mathbf{J}}}.
$$

Also, $J = \frac{1}{J_{\mathbf{x}}}$ and $\|J\|_{\infty,\Omega} = \|\frac{1}{J_{\mathbf{x}}}\|_{\infty,\Omega_{\mathbf{x}}} = \frac{1}{\|J_{\mathbf{x}}\|_{\infty,\Omega_{\mathbf{x}}}} > 0$.

In keeping with the vector notation, denote grad, div, and curl, respectively, by

$$
\nabla = \begin{pmatrix} \partial_\xi & 0 \\ \partial_\eta & 0 \\ 0 & \partial_\xi \\ 0 & \partial_\eta \end{pmatrix}, \nabla\cdot = \begin{pmatrix} \partial_\xi & \partial_\eta & 0 & 0 \\ 0 & 0 & \partial_\xi & \partial_\eta \end{pmatrix}, \nabla\times = \begin{pmatrix} -\partial_\eta & \partial_\xi & 0 & 0 \\ 0 & 0 & -\partial_\eta & \partial_\xi \end{pmatrix}.
$$

The same calculus notation is used in both $\Omega_{\mathbf{x}}$ and Ω (e.g., ∇, $\nabla\cdot$, and $\nabla\times$). Differentiation in $\Omega_{\mathbf{x}}$ is with respect to x and y and differentiation in Ω is with respect to

ξ and η. Let the boundary unit normal vector be denoted by

$$\mathbf{n} = \begin{pmatrix} n_1 & 0 \\ n_2 & 0 \\ 0 & n_1 \\ 0 & n_2 \end{pmatrix}. \tag{2.3}$$

As in previous applications of the FOSLS methodology (c.f. [7]), the natural first-order system is often augmented with a curl equation to ensure that the system is elliptic in the H^1 product norm. The augmented system also allows for the possibility of solving for the unknowns in two separate stages: we can solve for $\mathbf{J}$ alone in the first stage, then fix $\mathbf{J}$ and solve for $\mathbf{x}$ alone in the second stage, as the following development shows. The curl-augmented system we consider here is

$$\begin{aligned} \mathbf{J} - \nabla \mathbf{x} &= \mathbf{0}, && \text{in} \quad \Omega, \\ (\underline{\hat{\mathbf{J}}\hat{\mathbf{J}}}^t \nabla) \cdot \mathbf{J} &= \mathbf{0}, && \text{in} \quad \Omega, \\ \nabla \times \mathbf{J} &= \mathbf{0}, && \text{in} \quad \Omega, \\ \mathbf{x} &= \mathbf{w}, && \text{on} \quad \Gamma, \\ \mathbf{n} \times \mathbf{J} &= \mathbf{n} \times \nabla \mathbf{w}, && \text{on} \quad \Gamma. \end{aligned} \tag{2.4}$$

To be very clear about our notation, note that derivatives only apply to terms on their right. Thus, for $(\underline{\hat{\mathbf{J}}\hat{\mathbf{J}}}^t \nabla)\cdot$ in the second equation of (2.4), the matrix multiplication is applied first, keeping the order of each entry in the resulting matrix consistent with the multiplication. To perform the dot product, the matrix is transposed without altering the order of the terms in each component. For example, if we write

$$\underline{\hat{\mathbf{J}}\hat{\mathbf{J}}}^t = \begin{pmatrix} \alpha & -\beta & 0 & 0 \\ -\beta & \gamma & 0 & 0 \\ 0 & 0 & \alpha & -\beta \\ 0 & 0 & -\beta & \gamma \end{pmatrix},$$

then

$$(\underline{\hat{\mathbf{J}}\hat{\mathbf{J}}}^t \nabla) \cdot \mathbf{J} = \begin{pmatrix} \alpha \frac{\partial J_{11}}{\partial \xi} - \beta \frac{\partial J_{11}}{\partial \eta} - \beta \frac{\partial J_{21}}{\partial \xi} + \gamma \frac{\partial J_{21}}{\partial \eta} \\ \alpha \frac{\partial J_{12}}{\partial \xi} - \beta \frac{\partial J_{12}}{\partial \eta} - \beta \frac{\partial J_{22}}{\partial \xi} + \gamma \frac{\partial J_{22}}{\partial \eta} \end{pmatrix}.$$

We consider a two-stage algorithm, but focus only on the following first stage:

$$\begin{aligned} (\underline{\hat{\mathbf{J}}\hat{\mathbf{J}}}^t \nabla) \cdot \mathbf{J} &= \mathbf{0}, && \text{in} \quad \Omega, \\ \nabla \times \mathbf{J} &= \mathbf{0}, && \text{in} \quad \Omega, \\ \mathbf{n} \times \mathbf{J} &= \mathbf{n} \times \nabla \mathbf{w}, && \text{on} \quad \Gamma. \end{aligned} \tag{2.5}$$

Note that $\mathbf{x}$ can be recovered from the solution of (2.5) by a second stage that minimizes $\|\nabla \mathbf{x} - \mathbf{J}\|_{0,\Omega}^2 + \|\mathbf{x} - \mathbf{w}\|_{\frac{1}{2},\Gamma}^2$ over $\mathbf{x}$ with the computed $\mathbf{J}$ held fixed. The homogeneous part of the first term in this functional is precisely the $H^1(\Omega)^4$ seminorm of $\mathbf{x}$, so minimizing this functional leads to a simple system of decoupled Poisson equations. The remainder of our analysis therefore focuses on (2.5).

To obtain homogeneous boundary conditions, we rewrite the equations in terms of the perturbation, $\mathbf{D}$, of a smooth extension of $\mathbf{w}$ into Ω. To this end, suppose that some function $\mathbf{w} \in H^4(\Omega)^2$ is given so that its trace agrees with $\mathbf{w}$ on Γ. Defining $\mathbf{E} \equiv \nabla \mathbf{w} \in H^3(\Omega)^4$, we thus have

$$\mathbf{n} \times \mathbf{E} = \mathbf{n} \times \nabla \mathbf{w}, \qquad \text{on } \Gamma. \tag{2.6}$$

(In practice, we do not really need an extension of $\mathbf{w}$, but rather just an extension of its gradient: any $\mathbf{E} \in H^3(\Omega)^4$ that satisfies (2.6) will do. However, if this extension is not necessarily a gradient, then $\mathbf{E}$ must be included in the curl term in (2.7) below.)

In the notation of the companion paper [10], we have

$$\mathbf{P}(\mathbf{D}) \equiv \begin{pmatrix} \left((\underline{\hat{\mathbf{E}}} + \underline{\hat{\mathbf{D}}})(\underline{\hat{\mathbf{E}}} + \underline{\hat{\mathbf{D}}})^t \nabla\right) \cdot (\mathbf{E} + \mathbf{D}) \\ \nabla \times \mathbf{D} \end{pmatrix} = \mathbf{0}, \tag{2.7}$$

with boundary conditions

$$\mathbf{n} \times \mathbf{D} = \mathbf{0}. \tag{2.8}$$

System (2.7)-(2.8) corresponds to the inverse Laplace problem with Dirichlet boundary conditions. Existence of a solution, $\mathbf{D}^*$, that yields a positive Jacobian is guaranteed by Rado's theorem. We show in section 3 that $\mathbf{D}^* \in H^3(\Omega)^4$. One implication of this smoothness property is that $(\underline{\hat{\mathbf{E}}}+\underline{\hat{\mathbf{D}}}^*)(\underline{\hat{\mathbf{E}}}+\underline{\hat{\mathbf{D}}}^*)^t$ is a uniformly positive definite and bounded matrix on Ω.

From the companion paper [10], we define

$$\mathcal{H}_{1+\delta} \equiv \{\mathbf{D} \in H^{1+\delta}(\Omega)^4 : \mathbf{n} \times \mathbf{D} = \mathbf{0} \text{ on } \Gamma\}.$$

Restricting $\mathbf{D}$ to $H^{1+\delta}(\Omega)^4$ ensures that $\left((\underline{\hat{\mathbf{E}}} + \underline{\hat{\mathbf{D}}})(\underline{\hat{\mathbf{E}}} + \underline{\hat{\mathbf{D}}})^t \nabla\right) \cdot (\mathbf{E} + \mathbf{D}) \in L^2(\Omega)^2$, as the results of the next section show.

The first Fréchet derivative of (2.7) in direction $\mathbf{K}$ is

$$\mathbf{P}'(\mathbf{D})[\mathbf{K}] = \begin{pmatrix} \left((\underline{\hat{\mathbf{D}}} + \underline{\hat{\mathbf{E}}})(\underline{\hat{\mathbf{D}}} + \underline{\hat{\mathbf{E}}})^t \nabla\right) \cdot \mathbf{K} + \mathbf{B} \cdot \mathbf{K} \\ \nabla \times \mathbf{K} \end{pmatrix}, \tag{2.9}$$

where

$$\mathbf{B} \cdot \mathbf{K} \equiv \left(\underline{\hat{\mathbf{K}}}(\underline{\hat{\mathbf{D}}} + \underline{\hat{\mathbf{E}}})^t \nabla\right) \cdot (\mathbf{D} + \mathbf{E}) + \left((\underline{\hat{\mathbf{D}}} + \underline{\hat{\mathbf{E}}})\underline{\hat{\mathbf{K}}}^t \nabla\right) \cdot (\mathbf{D} + \mathbf{E}).$$

and the second Fréchet derivative in directions $\mathbf{K}$ and $\mathbf{M}$ is

$$\begin{aligned} \mathbf{P}''(\mathbf{D})[\mathbf{K}, \mathbf{M}] = \;& \begin{pmatrix} \left(\underline{\hat{\mathbf{M}}}(\underline{\hat{\mathbf{D}}} + \underline{\hat{\mathbf{E}}})^t \nabla\right) \cdot \mathbf{K} + \left((\underline{\hat{\mathbf{D}}} + \underline{\hat{\mathbf{E}}})\underline{\hat{\mathbf{M}}}^t \nabla\right) \cdot \mathbf{K} \\ \mathbf{0} \end{pmatrix} \\ & + \begin{pmatrix} \left(\underline{\hat{\mathbf{K}}}(\underline{\hat{\mathbf{D}}} + \underline{\hat{\mathbf{E}}})^t \nabla\right) \cdot \mathbf{M} + \left((\underline{\hat{\mathbf{D}}} + \underline{\hat{\mathbf{E}}})\underline{\hat{\mathbf{K}}}^t \nabla\right) \cdot \mathbf{M} \\ \mathbf{0} \end{pmatrix} \\ & + \begin{pmatrix} \left(\underline{\hat{\mathbf{K}}}\underline{\hat{\mathbf{M}}}^t \nabla\right) \cdot (\mathbf{D} + \mathbf{E}) + \left(\underline{\hat{\mathbf{M}}}\underline{\hat{\mathbf{K}}}^t \nabla\right) \cdot (\mathbf{D} + \mathbf{E}) \\ \mathbf{0} \end{pmatrix}. \end{aligned} \tag{2.10}$$

3. The Assumptions and their Verification. Consider the assumptions made in companion paper [10]. The first is existence of a solution in $\mathcal{H}_{2+\delta}$. From [16], we know that a unique inverse map exists and that it provides a solution, $\mathbf{D}^*$, to (2.7). Recall that $\boldsymbol{\xi}^* \in H^4(\Omega_{\mathbf{x}})^2$. In lemma 3.4 below, we show that $\mathbf{D}^* \in \mathcal{H}_3$. This establishes our first assumption for the EGG equations for any $\delta \in (0, 1)$.

The remaining assumptions we need to establish are, for $\epsilon = 0$ or δ, that $\mathbf{P}[\mathbf{D}] \in H^\epsilon(\Omega)^4$ for every $\mathbf{D} \in B_r$ (lemma 3.5), that $\|\mathbf{P}'(\mathbf{D})[\,\cdot\,]\|_{\epsilon,\Omega}$ is $H^{1+\epsilon}(\Omega)^4$ equivalent (lemma 3.6), and that the second Fréchet derivative of $\mathbf{P}(\mathbf{D})$ is bounded for all $\mathbf{D} \in B_r$

(lemma 3.7). (The discretization assumptions are standard.) But first we need three results that follow directly from a corollary to the Sobolev Embedding theorem [11], which (tailored to our needs) states that the product of a function in $H^{m_1}(\Omega)$ and a function in $H^{m_2}(\Omega)$ is in $H^m(\Omega)$ provided that either $m_1 + m_2 - m \geq 1$, $m_1 > m$, and $m_2 > m$ or $m_1 + m_2 - m > 1$, $m_1 \geq m$, and $m_2 \geq m$.

Assume that $r > 0$ is so small that matrix $(\underline{\hat{\mathbf{E}}} + \underline{\hat{\mathbf{D}}})(\underline{\hat{\mathbf{E}}} + \underline{\hat{\mathbf{D}}})^t$ is positive definite and bounded uniformly on Ω and over $\mathbf{D} \in B_r \equiv \{\mathbf{D} \in \mathcal{H}_{1+\delta} : \|\mathbf{D}^* - \mathbf{D}\|_{1+\delta,\Omega} < r\}$. This assumption is possible because it is true at $\mathbf{D} = \mathbf{D}^*$ and because the matrix is continuous as a function defined on B_r. Assume that $a, b, c \in H^{1+\delta}(\Omega)$. For convenience, we let ∂ denote either ∂_x or ∂_y. In the proof of lemma 3.4, we also use ∂^2 to denote any of the four second partial derivatives and ∂^3 for any combination of third partial derivatives. Note that $(\partial a)^2$ could mean $a_x a_y$, for example.

Lemma 3.1. *There exists a constant, C, depending only on Ω and δ, such that*

$$\|ab\partial c\|_{\epsilon,\Omega} \leq C\|a\|_{1+\delta,\Omega}\|b\|_{1+\delta,\Omega}\|c\|_{1+\epsilon,\Omega}$$

Proof. Using the corollary to the Sobolev Imbedding theorem [11] with $m_1 = 1 + \delta,\ m_2 = \epsilon$, and $m = \epsilon$ twice yields

$$\begin{aligned}\|ab\partial c\|_{\epsilon,\Omega} &\leq C\|a\|_{1+\delta,\Omega}\|b\partial c\|_{\epsilon,\Omega}\\ &\leq C\|a\|_{1+\delta,\Omega}\|b\|_{1+\delta,\Omega}\|\partial c\|_{\epsilon,\Omega}\\ &\leq C\|a\|_{1+\delta,\Omega}\|b\|_{1+\delta,\Omega}\|c\|_{1+\epsilon,\Omega}.\end{aligned}$$

□

Lemma 3.2. *There exists a constant, C, depending only on Ω and δ, such that*

$$\|ab\partial c\|_{\epsilon,\Omega} \leq C\|a\|_{1+\delta,\Omega}\|b\|_{1+\epsilon,\Omega}\|c\|_{1+\delta,\Omega}$$

Proof. Using the corollary to the Sobolev imbedding theorem [11] first with $m_1 = 1 + \delta,\ m_2 = \epsilon$, and $m = \epsilon$, then with $m_1 = 1 + \epsilon,\ m_2 = \delta$, and $m = \epsilon$ yields

$$\begin{aligned}\|ab\partial c\|_{\epsilon,\Omega} &\leq C\|a\|_{1+\delta,\Omega}\|b\partial c\|_{\epsilon,\Omega}\\ &\leq C\|a\|_{1+\delta,\Omega}\|b\|_{1+\epsilon,\Omega}\|\partial c\|_{\delta,\Omega}\\ &\leq C\|a\|_{1+\delta,\Omega}\|b\|_{1+\epsilon,\Omega}\|c\|_{1+\delta,\Omega}.\end{aligned}$$

□

Lemma 3.3. *Assume that $a, b \in H^{2+\delta}(\Omega)$ and $k \in H^{1+\delta}(\Omega)$. Then there exists a constant, C, depending only on Ω and δ, such that*

$$\|ak\partial b\|_{1,\Omega} \leq C\|a\|_{1+\delta,\Omega}\|b\|_{2+\delta,\Omega}\|k\|_{1,\Omega}.$$

Proof.

$$\begin{aligned}\|ak\partial b\|_{1,\Omega} &\leq C\|a\|_{1+\delta,\Omega}\|k\partial b\|_{1,\Omega}\\ &\leq C\|a\|_{1+\delta,\Omega}\|\partial b\|_{1+\delta,\Omega}\|k\|_{1,\Omega}\\ &\leq C\|a\|_{1+\delta,\Omega}\|b\|_{2+\delta,\Omega}\|k\|_{1,\Omega},\end{aligned}$$

where we used corollary to the Sobolev imbedding theorem from [11] with $m_1 = 1 + \delta,\ m_2 = 1$, and $m = 1$ twice. □

LEMMA 3.4. *The solution, $\mathbf{D}^*$, of (2.7) is in $H^3(\Omega)^4$.*
Proof. We have

$$\mathbf{J}^* \equiv \mathbf{E} + \mathbf{D}^* = \begin{pmatrix} J_{11}^* \\ J_{21}^* \\ J_{12}^* \\ J_{22}^* \end{pmatrix} = \begin{pmatrix} x_\xi^* \\ x_\eta^* \\ y_\xi^* \\ y_\eta^* \end{pmatrix} = \frac{1}{J_{\mathbf{x}}^*} \begin{pmatrix} \eta_y^* \\ -\xi_y^* \\ -\eta_x^* \\ \xi_x^* \end{pmatrix},$$

where $\mathbf{E} \in H^3(\Omega_{\mathbf{x}})^4$ and $\boldsymbol{\xi}^* \in H^4(\Omega_{\mathbf{x}})^2$. We now show that $\mathbf{J}^* \in H^3(\Omega)^4$, from which follows the result that $\mathbf{D}^* \in H^3(\Omega)^4$.

Since $\boldsymbol{\xi}^* \in H^4(\Omega_{\mathbf{x}})^2$, then $\xi_x^*, \xi_y^*, \eta_x^*, \eta_y^* \in H^3(\Omega_{\mathbf{x}})$. From the corollary to the Sobolev imbedding theorem (with $m_1 = 3$, $m_2 = 3$, and $m = 3$), we must have $J_{\mathbf{x}}^* = \xi_x^*\eta_y^* - \xi_y^*\eta_x^* \in H^3(\Omega_{\mathbf{x}})$. Recall from section 2 that $\mathbf{J}^*$ is continuous and uniformly positive and bounded: $J_0 \le |J_{\mathbf{x}}^*(x,y)| \le J_1$ for some constants $J_0, J_1 \in R^+$ and all $(x,y) \in \Omega_{\mathbf{x}}$.

Dropping superscript * for convenience, consider J_{11}. (The other entries are treated similarly.) Using the corollary to the Sobolev imbedding theorem [11] with $m_1 = 3$, $m_2 = 3$, and $m = 3$, we get

$$\|J_{11}\|_{3,\Omega} = \left\| \frac{1}{J_{\mathbf{x}}} \eta_y \right\|_{3,\Omega_{\mathbf{x}}} \le C \left\| \frac{1}{J_{\mathbf{x}}} \right\|_{3,\Omega_{\mathbf{x}}} \|\eta_y\|_{3,\Omega_{\mathbf{x}}}.$$

Therefore, we need only show that $\frac{1}{J_{\mathbf{x}}} \in H^3$. But

$$\left\| \frac{1}{J_{\mathbf{x}}} \right\|_{3,\Omega_{\mathbf{x}}}^2 = \sum_{i \le 3} \left\| \partial^i \frac{1}{J_{\mathbf{x}}} \right\|_{0,\Omega_{\mathbf{x}}}^2.$$

We consider each order separately. By theorem 3.2 in [21], for any $a \in C^0(\Omega_{\mathbf{x}})$ and $b \in L^2(\Omega_{\mathbf{x}})$, we have

$$\|ab\|_{0,\Omega_{\mathbf{x}}} \le \|a\|_{\infty,\Omega_{\mathbf{x}}} \|b\|_{0,\Omega_{\mathbf{x}}}. \tag{3.1}$$

For the zeroth-order term, using (3.1) yields

$$\left\| \frac{1}{J_{\mathbf{x}}} \right\|_{0,\Omega_{\mathbf{x}}} \le \left\| \frac{1}{J_{\mathbf{x}}} \right\|_{\infty,\Omega_{\mathbf{x}}} \|1\|_{0,\Omega_{\mathbf{x}}} \le \frac{1}{J_0} \|1\|_{0,\Omega_{\mathbf{x}}}.$$

For the first-order term, we use (3.1) to get

$$\left\| \partial \frac{1}{J_{\mathbf{x}}} \right\|_{0,\Omega_{\mathbf{x}}} = \left\| \frac{-1}{J_{\mathbf{x}}^2} \partial J_{\mathbf{x}} \right\|_{0,\Omega_{\mathbf{x}}} \le \frac{1}{J_0^2} \|\partial J_{\mathbf{x}}\|_{0,\Omega_{\mathbf{x}}} \le \frac{1}{J_0^2} \|J_{\mathbf{x}}\|_{1,\Omega_{\mathbf{x}}}.$$

For the second-order term, we use the triangle inequality, (3.1), and the Corollary to the Sobolev imbedding theorem with $m_1 = m_2 = 1$ and $m = 0$ to get

$$\begin{aligned} \left\| \partial^2 \frac{1}{J_{\mathbf{x}}} \right\|_{0,\Omega_{\mathbf{x}}} &= \left\| \frac{2}{J_{\mathbf{x}}^3} (\partial J_{\mathbf{x}})^2 + \frac{-1}{J_{\mathbf{x}}^2} \partial^2 J_{\mathbf{x}} \right\|_{0,\Omega_{\mathbf{x}}} \\ &\le \left\| \frac{2}{J_{\mathbf{x}}^3} (\partial J_{\mathbf{x}})^2 \right\|_{0,\Omega_{\mathbf{x}}} + \left\| \frac{1}{J_{\mathbf{x}}^2} \partial^2 J_{\mathbf{x}} \right\|_{0,\Omega_{\mathbf{x}}} \\ &\le \frac{2}{J_0^3} \left\| (\partial J_{\mathbf{x}})^2 \right\|_{0,\Omega_{\mathbf{x}}} + \frac{1}{J_0^2} \left\| \partial^2 J_{\mathbf{x}} \right\|_{0,\Omega_{\mathbf{x}}} \end{aligned}$$

$$\le C\left(\frac{2}{J_0^3}\left\|\partial J_{\mathbf{x}}\right\|_{1,\Omega_{\mathbf{x}}}^2+\frac{1}{J_0^2}\left\|J_{\mathbf{x}}\right\|_{2,\Omega_{\mathbf{x}}}\right)$$
$$\le C\left(\frac{2}{J_0^3}\left\|J_{\mathbf{x}}\right\|_{2,\Omega_{\mathbf{x}}}^2+\frac{1}{J_0^2}\left\|J_{\mathbf{x}}\right\|_{2,\Omega_{\mathbf{x}}}\right).$$

For the third-order term, we use the triangle inequality, (3.1), and the Corollary to the Sobolev imbedding theorem once with $m_1 = 1$, $m_2 = \frac{3}{4}$, and $m = 0$, once with $m_1 = m_2 = 1$ and $m = 0$, and once with $m_1 = m_2 = 1$ and $m = \frac{3}{4}$ to get

$$\begin{aligned}
\left\|\partial^3\frac{1}{J_{\mathbf{x}}}\right\|_{0,\Omega_{\mathbf{x}}} &= \left\|\frac{-6}{J_{\mathbf{x}}^4}(\partial J_{\mathbf{x}})^3+\frac{6}{J_{\mathbf{x}}^3}\partial J_{\mathbf{x}}\partial^2 J_{\mathbf{x}}+\frac{-1}{J_{\mathbf{x}}^2}\partial^3 J_{\mathbf{x}}\right\|_{0,\Omega_{\mathbf{x}}}\\
&\le \left\|\frac{6}{J_{\mathbf{x}}^4}(\partial J_{\mathbf{x}})^3\right\|_{0,\Omega_{\mathbf{x}}}+\left\|\frac{6}{J_{\mathbf{x}}^3}\partial J_{\mathbf{x}}\partial^2 J_{\mathbf{x}}\right\|_{0,\Omega_{\mathbf{x}}}+\left\|\frac{1}{J_{\mathbf{x}}^2}\partial^3 J_{\mathbf{x}}\right\|_{0,\Omega_{\mathbf{x}}}\\
&\le \frac{6}{J_0^4}\left\|(\partial J_{\mathbf{x}})^3\right\|_{0,\Omega_{\mathbf{x}}}+\frac{6}{J_0^3}\left\|\partial J_{\mathbf{x}}\partial^2 J_{\mathbf{x}}\right\|_{0,\Omega_{\mathbf{x}}}+\frac{1}{J_0^2}\left\|\partial^3 J_{\mathbf{x}}\right\|_{0,\Omega_{\mathbf{x}}}\\
&\le C\left(\frac{6}{J_0^4}\left\|\partial J_{\mathbf{x}}\right\|_{1,\Omega_{\mathbf{x}}}^3+\frac{6}{J_0^3}\left\|\partial J_{\mathbf{x}}\right\|_{1,\Omega_{\mathbf{x}}}\left\|\partial^2 J_{\mathbf{x}}\right\|_{1,\Omega_{\mathbf{x}}}+\frac{1}{J_0^2}\left\|J_{\mathbf{x}}\right\|_{3,\Omega_{\mathbf{x}}}\right)\\
&\le C\left(\frac{6}{J_0^4}\left\|J_{\mathbf{x}}\right\|_{2,\Omega_{\mathbf{x}}}^3+\frac{6}{J_0^3}\left\|J_{\mathbf{x}}\right\|_{2,\Omega_{\mathbf{x}}}\left\|J_{\mathbf{x}}\right\|_{3,\Omega_{\mathbf{x}}}+\frac{1}{J_0^2}\left\|J_{\mathbf{x}}\right\|_{3,\Omega_{\mathbf{x}}}\right).
\end{aligned}$$

The result follows from these bounds. ☐

LEMMA 3.5. *$\mathbf{P}[\mathbf{D}] \in H^\epsilon(\Omega)^p$ for every $\mathbf{D} \in B_r$: there exists a constant, C, depending only on $\mathbf{D}^*$, $\mathbf{E}$, r, Ω, and δ, such that*

$$\|\mathbf{P}(\mathbf{D})\|_{\epsilon,\Omega} \le C, \quad \forall\ \mathbf{D} \in B_r. \tag{3.2}$$

Proof. The products in (2.7) are of the form treated in lemma 3.1. In fact, there exists a constant, C, depending only on Ω and δ, such that

$$\|\mathbf{P}(\mathbf{D})\|_{\epsilon,\Omega} \le C(\|\mathbf{D}+\mathbf{E}\|_{1+\delta,\Omega}^2\|\mathbf{D}+\mathbf{E}\|_{1+\epsilon,\Omega}+\|\mathbf{D}\|_{1+\epsilon,\Omega}), \tag{3.3}$$

so (3.2) follows because $\mathbf{D} \in B_r$. ☐

Next we establish uniform coercivity and continuity of $\mathbf{P}'$ in a neighborhood of $\mathbf{D}^*$. This result needs the assumption that $\mathbf{P}'(\mathbf{D}^*)[\,\cdot\,]$ is one-to-one on $\mathcal{H}_{1+\delta}$, which is a consequence of an analogous assumption on the original EGG equations.

LEMMA 3.6 (Ellipticity Property). *$\|\mathbf{P}'(\mathbf{D})[\,\cdot\,]\|_{\epsilon,\Omega}$ is $H^{1+\epsilon}(\Omega)^4$ equivalent: there exists constants c_c and c_b, depending only on $\mathbf{D}^*, \mathbf{E}, r, \Omega$, and δ, such that*

$$\frac{1}{c_c}\|\mathbf{K}\|_{1+\epsilon,\Omega} \le \|\mathbf{P}'(\mathbf{D})[\mathbf{K}]\|_{\epsilon,\Omega} \le c_b\|\mathbf{K}\|_{1+\epsilon,\Omega}, \quad \forall\ \mathbf{K} \in \mathcal{H}_{1+\epsilon}. \tag{3.4}$$

Proof. The products in (2.9) are of the form treated in lemmas 3.1 and 3.2. In fact, there exists a constant, C, depending only on Ω and δ, such that

$$\|\mathbf{P}'(\mathbf{D})[\mathbf{K}]\|_{\epsilon,\Omega} \le C(\|\mathbf{D}+\mathbf{E}\|_{1+\delta,\Omega}^2\|\mathbf{K}\|_{1+\epsilon,\Omega}+\|\mathbf{K}\|_{1+\epsilon,\Omega}). \tag{3.5}$$

Proof of the lower bound follows from theorem 10.5 of ADN2 [2] as we now show. We first need to prove H^{m+1} boundedness and coercivity for $\mathbf{P}'(\mathbf{D}^*)[\mathbf{K}]$: there exists constants, c_1 and c_3, depending only on $\mathbf{D}^*, \mathbf{E}, m$, and Ω, such that

$$\frac{1}{c_1}\|\mathbf{K}\|_{m+1,\Omega} \le \|\mathbf{P}'(\mathbf{D}^*)[\mathbf{K}]\|_{m,\Omega} \le c_3\|\mathbf{K}\|_{m+1,\Omega}, \quad \forall\ \mathbf{K} \in \mathcal{H}_{m+1}, \tag{3.6}$$

for any $m \in [0,1]$. The upper bound is simply an application of the corollary to the Sobolev imbedding theorem similar to lemmas 3.1 and 3.2. Consider the lower bound. It would be a simple matter to just assume $\mathbf{D}^* \in \mathcal{H}_{3+\delta}$ and then, because the coefficients would be sufficiently smooth, apply ADN2 theory for both $m=0$ and $m=1$. Instead, we just have $\mathbf{D}^* \in \mathcal{H}_{2+\delta}$, so while the higher-order coefficients are in C^1, the lower order coefficients are only in C^0. This means that we need more care.

First consider $m = 0$. What follows for this case is a straightforward application of ADN2 theory to the entire system because all of the coefficients are sufficiently smooth. Recall that Ω is a bounded open subset of R^2 with $C^{3,1}$ boundary Γ. We write the system as

$$\begin{array}{lcc} \mathcal{L}\mathbf{K} = \mathbf{f}, & \text{in} & \Omega, \\ \mathcal{B}\mathbf{K} = \mathbf{g}, & \text{on} & \Gamma, \end{array} \tag{3.7}$$

where $\mathcal{L} \equiv \mathbf{P}'(\mathbf{D}^*)$ and $\mathcal{B} = \mathbf{n}\times$. (Recall that $\mathbf{n}$ is the outward unit normal on Γ (2.3).) For convenience, we write the coefficients using $\mathbf{J}^* = \mathbf{D}^* + \mathbf{E}$ and drop the * from the components. Note that $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$ and $l_{ij} = l'_{ij} + l''_{ij}$, where

$$\mathcal{L}_1 = (l'_{ij}(\boldsymbol{\xi}, \partial)) = \begin{pmatrix} \alpha\partial_\xi - \beta\partial_\eta & -\beta\partial_\xi + \gamma\partial_\eta & 0 & 0 \\ 0 & 0 & \alpha\partial_\xi - \beta\partial_\eta & -\beta\partial_\xi + \gamma\partial_\eta \\ -\partial_\eta & \partial_\xi & 0 & 0 \\ 0 & 0 & -\partial_\eta & \partial_\xi \end{pmatrix},$$

$$\begin{aligned} \mathcal{L}_2 &= (l''_{ij}(\boldsymbol{\xi}, \partial)) \\ &= \begin{pmatrix} 2J_{11}J_{21,\eta} - J_{21}(J_{11,\eta} + J_{21,\xi}) & 2J_{11}J_{22,\eta} - J_{21}(J_{12,\eta} + J_{22,\xi}) & 0 & 0 \\ 2J_{21}J_{11,\xi} - J_{11}(J_{11,\eta} + J_{21,\xi}) & 2J_{21}J_{12,\xi} - J_{11}(J_{12,\eta} + J_{22,\xi}) & 0 & 0 \\ 2J_{12}J_{21,\eta} - J_{22}(J_{11,\eta} + J_{21,\xi}) & 2J_{12}J_{22,\eta} - J_{22}(J_{12,\eta} + J_{22,\xi}) & 0 & 0 \\ 2J_{22}J_{11,\xi} - J_{12}(J_{11,\eta} + J_{21,\xi}) & 2J_{22}J_{12,\xi} - J_{12}(J_{12,\eta} + J_{22,\xi}) & 0 & 0 \end{pmatrix}^t, \end{aligned}$$

$$\alpha = J_{21}^2 + J_{22}^2, \qquad \beta = J_{11}J_{21} + J_{12}J_{22}, \qquad \gamma = J_{11}^2 + J_{12}^2.$$

Note that

$$\mathcal{B} = (b_{ij}(\boldsymbol{\xi}, \partial)) = \begin{pmatrix} -n_2 & n_1 & 0 & 0 \\ 0 & 0 & -n_2 & n_1 \end{pmatrix}. \tag{3.8}$$

In ADN2 theory, three types of integer weights are used to determine the leading-order terms for boundary value problem (3.7). Weight $s_i \le 0$ refers to the ith equation, weight $t_j \ge 0$ to the jth dependent variable, and weight r_k to the kth boundary condition. These weights are chosen as small as possible but so that

$$\deg l_{ij}(\boldsymbol{\xi}, \partial) \le s_i + t_j, \quad \deg b_{kj}(\boldsymbol{\xi}, \partial) \le r_k + t_j, \quad i, j = 1,2,3,4, \quad k = 1,2,$$

where deg refers to the order of the derivatives. Our weights are

$$s_i = 0, \qquad t_j = 1, \qquad r_k = -1, \qquad i, j = 1,2,3,4, \quad k = 1,2.$$

The leading-order part of $\mathcal{L}$ consists of the elements l_{ij} for which $\deg l_{ij}(\boldsymbol{\xi}, \partial) = s_i + t_j = 1$. Therefore, $\mathcal{L}_1$ is the leading-order (in this case, first-order) part. The leading-order part of $\mathcal{B}$ consists of elements b_{kj} for which $\deg b_{kj}(\boldsymbol{\xi}, \partial) = r_k + t_j = 0$. Therefore, the leading-order (in this case, zeroth-order) part of $\mathcal{B}$ is $\mathcal{B}$ itself.

We must show that $\mathcal{L}_1$ satisfies two ADN2 conditions: the *Supplementary Condition on its determinant* and *uniform ellipticity.* ($\mathcal{L}_1$ will then automatically be elliptic.) ADN2 also requires that the system of equations and boundary conditions be well-posed. This means that $\mathcal{L}_1$ and $\mathcal{B}$, when combined, must satisfy the *Complementing Boundary Condition.* Let L denote the determinant of $\mathcal{L}_1$:

$$L(\boldsymbol{\xi}, \partial) = \det(l'_{ij}) = -(\alpha\partial_\xi^2 - 2\beta\partial_\xi\partial_\eta + \gamma\partial_\eta^2)^2.$$

Since $J > 0$ (see section 2), then

$$\begin{aligned} \beta^2 - \alpha\gamma &= (x_\xi x_\eta + y_\xi y_\eta)^2 - (x_\eta^2 + y_\eta^2)(x_\xi^2 + y_\xi^2) \\ &= -(x_\xi y_\eta - y_\xi x_\eta)^2 = -J^2 < 0. \end{aligned} \tag{3.9}$$

Let $\mathbf{d} = (d \quad e)^t$ and $\mathbf{p} = (p \quad q)^t$ be any two linearly independent vectors. To aid clarity of the following discussion, we first define some quantities:

$$\begin{aligned} \mathcal{A} &= \alpha dp - \beta(pe + dq) + \gamma eq, \\ \mathcal{B} &= \sqrt{\mathcal{DC} - \mathcal{A}^2} = J|pe - dq|, \\ \mathcal{C} &= \alpha d^2 - 2\beta de + \gamma e^2, \\ \mathcal{D} &= \alpha p^2 - 2\beta pq + \gamma q^2. \end{aligned}$$

Note that $\mathcal{B} > 0$ since $J > 0$ and linear independence of $\mathbf{p}$ and $\mathbf{d}$ implies $|pe - dq| > 0$.

The *Supplementary Condition on* L requires the equation

$$L(\boldsymbol{\xi}, \mathbf{d} + \tau\mathbf{p}) = -\{\mathcal{C} + 2\tau\mathcal{A} + \tau^2\mathcal{D}\}^2 = 0 \tag{3.10}$$

to have exactly two roots in τ with positive imaginary part. Polynomial (3.10) has two double roots ($\iota = \sqrt{-1}$),

$$\tau = \frac{-\mathcal{A} \pm \iota\mathcal{B}}{\mathcal{D}},$$

which form two complex conjugate pairs. Thus, (3.10) does indeed have exactly two roots with positive imaginary part (one such double root).

To satisfy *uniform ellipticity*, we need to show that

$$\frac{1}{C}\|\mathbf{d}\|^4 \le |L(\boldsymbol{\xi}, \mathbf{d})| \le C\|\mathbf{d}\|^4, \tag{3.11}$$

with $\|\mathbf{d}\| = \sqrt{d^2 + e^2}$, for all vectors $\mathbf{d} \neq \mathbf{0}$ and points $\boldsymbol{\xi}$ in Ω.

To prove the left bound in (3.11), let $\rho = \frac{|\beta|}{\sqrt{\alpha\gamma}} < 1$ (see (3.9)). Then

$$\begin{aligned} |L(\boldsymbol{\xi}, \mathbf{d})| &= (\alpha d^2 - 2\beta de + \gamma e^2)^2 \\ &\ge (\alpha d^2 - 2|\beta||d||e| + \gamma e^2)^2 \\ &= (\alpha d^2 - 2\rho\sqrt{\alpha\gamma}|d||e| + \gamma e^2)^2 \\ &= ((1-\rho)(\alpha d^2 + \gamma e^2) + \rho(\sqrt{\alpha}d - \sqrt{\gamma}e)^2)^2 \\ &\ge ((1-\rho)(\alpha d^2 + \gamma e^2))^2 \\ &\ge \min((1-\rho)^2\alpha^2, (1-\rho)^2\gamma^2)(d^2 + e^2)^2 \\ &= \min((1-\rho)^2\alpha^2, (1-\rho)^2\gamma^2)\|\mathbf{d}\|^4. \end{aligned}$$

To prove the right bound in (3.11), note that Hölder's inequality implies that

$$\begin{aligned} |L(\boldsymbol{\xi}, \mathbf{d})| &\le (\alpha d^2 + 2|\beta||d||e| + \gamma e^2)^2 \\ &\le (\alpha d^2 + |\beta|(d^2 + e^2) + \gamma e^2)^2 \\ &\le \max((\alpha + |\beta|)^2, (\gamma + |\beta|)^2)(d^2 + e^2)^2 \\ &= \max((\alpha + |\beta|)^2, (\gamma + |\beta|)^2)\|\mathbf{d}\|^4. \end{aligned}$$

We then establish (3.11) by choosing

$$C = \max\left\{\frac{1}{(1-\rho)^2\alpha^2}, \frac{1}{(1-\rho)^2\gamma^2}, (\alpha+|\beta|)^2), (\gamma+|\beta|)^2)\right\}.$$

This shows that operator L satisfies the two conditions of ADN2. We now prove that the problem is well-posed by showing that $\mathcal{L}_1$ and $\mathcal{B}$ satisfy the *Complementing Boundary Condition*. This condition involves comparing of two polynomials. We consider a point on the boundary with normal $\mathbf{d} = (d \;\; e)^t$ and tangent $\mathbf{p} = (p \;\; q)^t$ vectors. The first polynomial is formed from the roots of (3.10) with positive imaginary parts:

$$M^+(\boldsymbol{\xi}, \mathbf{d}, \tau) = \left[\tau + \frac{\mathcal{A} - \iota\mathcal{B}}{\mathcal{D}}\right]^2. \tag{3.12}$$

The second polynomial is formed from the leading order elements of $\mathcal{L}$ and $\mathcal{B}$:

$$\sum_{k=1}^{2} a_k (BL)_{km}, \tag{3.13}$$

where

$$(BL)_{km} = \sum_{j=1}^{4} b_{kj}(\boldsymbol{\xi}, \mathbf{d}+\tau\mathbf{p}) l^{jm}(\boldsymbol{\xi}, \mathbf{d}+\tau\mathbf{p}),$$

and $l^{jm}(\boldsymbol{\xi}, \mathbf{d}+\tau\mathbf{p})$ are the elements of the (classical) adjoint ($l'_{ij} l^{jm} = \delta_i^m L,\ i, j, m = 1, 2, 3, 4$) of $l'_{ij}(\boldsymbol{\xi}, \mathbf{d}+\tau\mathbf{p})$ and b_{kj} is defined in (3.8).

The polynomials for (3.13) are

$$\left(\sum_{k=1}^{2} a_k (BL)_{km}\right) = \mathcal{D}\left[\tau + \frac{\mathcal{A}+\iota\mathcal{B}}{\mathcal{D}}\right]\left[\tau + \frac{\mathcal{A}-\iota\mathcal{B}}{\mathcal{D}}\right]\begin{pmatrix} a_1(pe-qd) \\ a_2(pe-qd) \\ a_1(\mathcal{A}+\tau\mathcal{D}) \\ a_2(\mathcal{A}+\tau\mathcal{D}) \end{pmatrix}^t. \tag{3.14}$$

Comparing polynomials (3.12) and (3.14) and noting that $\mathcal{B} > 0$, we have that (3.12) is not a factor of (3.14). Thus, the *Complementing Boundary Condition* is satisfied.

Theorem 10.5 from ADN2 [2] implies that, for $l_{ij} \in C^m(\overline{\Omega}), b_{kj} \in C^{m+1}(\Gamma)$, there exists a constant, c_1, that depends only on $\mathbf{D}^*, \mathbf{E}, r, \Omega$ and δ, such that, if $K_j \in H^1(\Omega)$, $1 \le j \le 4$, solves (3.7) and is unique, then $K_j \in H^{m+1}(\Omega)$ and

$$\|K_j\|_{m+1,\Omega} \le \frac{c_1}{4}\left[\sum_{i=1}^{4}\|f_i\|_{m,\Omega} + \sum_{i=1}^{2}\|g_i\|_{m+\frac{1}{2},\Gamma}\right],$$

where f_i and g_i are the components of $\mathbf{f}$ and $\mathbf{g}$, respectively, in (3.7). The coefficients of $\mathcal{L}$ are at least in $H^{1+\delta}(\Omega)$ and $C^0(\Omega)$. For the boundary conditions, we have $\mathcal{B} = \mathbf{n}\times$ and $\Gamma \in C^{3,1}$, so we get $b_{kj} \in C^2(\Gamma)$. The boundary conditions are homogeneous, so we can drop the boundary term in the inequality. We therefore have

$$\|\mathbf{K}\|_{1,\Omega} \le c_1\|\mathcal{L}(\mathbf{D}^*)[\mathbf{K}]\|_{0,\Omega}, \quad \forall\ \mathbf{K} \in \mathcal{H}_1(\Omega)^4. \tag{3.15}$$

Now consider $m = 1$. We cannot simply apply ADN2 to the whole system because the coefficients are not sufficiently smooth. Instead, we split the operator according to $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$ and restrict our ADN2 result to reduced system

$$\begin{aligned} \mathcal{L}_1\mathbf{K} &= \mathbf{f}, \quad \text{in} \quad \Omega, \\ \mathcal{B}\mathbf{K} &= \mathbf{g}, \quad \text{on} \quad \Gamma. \end{aligned}$$

Operator $\mathcal{L}_1$ satisfies the ADN2 conditions (as illustrated for case $m=0$), so

$$\|\mathbf{K}\|_{2,\Omega} \le c_1 \|\mathcal{L}_1(\mathbf{D}^*)[\mathbf{K}]\|_{1,\Omega}, \quad \forall\, \mathbf{K} \in \mathcal{H}_2. \tag{3.16}$$

The coefficients of $\mathcal{L}_1$ are in $H^{2+\delta}(\Omega)$ and $C^1(\Omega)$. For the boundary conditions, we have $\mathcal{B} = \mathbf{n}\times$ and $\Gamma \in C^{3,1}$, so we get $b_{kj} \in C^2(\Gamma)$. The boundary conditions are homogeneous, so we can drop the boundary term in the inequality.

We use lemma 3.3 to obtain

$$\|\mathcal{L}_2(\mathbf{D}^*)[\mathbf{K}]\|_{1,\Omega} \le c_2 \|\mathbf{K}\|_{1,\Omega}, \quad \forall\, \mathbf{K} \in \mathcal{H}_2. \tag{3.17}$$

Note that c_2 depends continuously on $\sup_{\mathbf{D}\in B_r} \|\mathbf{D}^* + \mathbf{E}\|_{2+\delta,\Omega}$, so it depends on $\mathbf{D}^*, \mathbf{E}, r$, Ω, and δ.

Combining (3.16) and (3.17) yields

$$\|\mathbf{K}\|_{2,\Omega} \le C(\|\mathbf{P}'(\mathbf{D}^*)[\mathbf{K}]\|_{1,\Omega} + \|\mathbf{K}\|_{1,\Omega}), \quad \forall\, \mathbf{K} \in \mathcal{H}_2. \tag{3.18}$$

This is a Gårdings inequality (cf. [13, 19]), which allows us now to prove that

$$\frac{1}{c_c}\|\mathbf{K}\|_{2,\Omega} \le \|\mathbf{P}'(\mathbf{D}^*)[\mathbf{K}]\|_{1,\Omega}, \quad \forall\, \mathbf{K} \in \mathcal{H}_2. \tag{3.19}$$

To this end, assume that (3.19) is not true. Then there exists a sequence, $\mathbf{K}_j \in \mathcal{H}_2$, such that

$$\|\mathbf{K}_j\|_{2,\Omega} = 1 \tag{3.20}$$

and

$$\|\mathbf{P}'(\mathbf{D}^*)[\mathbf{K}_j]\|_{1,\Omega} = \frac{1}{j} \quad j = 1, 2 \dots . \tag{3.21}$$

Now, because $H^2(\Omega)$ is compactly imbedded in $H^1(\Omega)$ (cf. the Rellich selection theorem [5]), then (3.20) implies that there exists a limit, $\hat{\mathbf{K}} \in \mathcal{H}_2$, of a subsequence, $\mathbf{K}_{j_k} \to \hat{\mathbf{K}}$, in the $H^1(\Omega)$ norm. Combining this with (3.18) and (3.21), we know that $\mathbf{K}_{j_k}$ must also be a Cauchy sequence in the $H^2(\Omega)$ norm with some limit $\bar{\mathbf{K}}$. But, from the upper bound in (3.4), we have

$$\begin{aligned} \lim_{j_k\to\infty} \|\mathbf{P}'(\mathbf{D}^*)[\mathbf{K}_{j_k}] - \mathbf{P}'(\mathbf{D}^*)[\bar{\mathbf{K}}]\|_{1,\Omega} &= \lim_{j_k\to\infty} \|\mathbf{P}'(\mathbf{D}^*)[\mathbf{K}_{j_k} - \bar{\mathbf{K}}]\|_{1,\Omega} \\ &\le \lim_{j_k\to\infty} \|\mathbf{K}_{j_k} - \bar{\mathbf{K}}\|_{2,\Omega} = 0. \end{aligned}$$

From (3.21), we thus obtain

$$\|\mathbf{P}'(\mathbf{D}^*)[\bar{\mathbf{K}}]\|_{1,\Omega} \le \lim_{j_k\to\infty} \{\|\mathbf{P}'(\mathbf{D}^*)[\mathbf{K}_{j_k}] - \mathbf{P}'(\mathbf{D}^*)[\bar{\mathbf{K}}]\|_{1,\Omega} + \|\mathbf{P}'(\mathbf{D}^*)[\mathbf{K}_{j_k}]\|_{1,\Omega}\} = 0.$$

But $\mathbf{P}'(\mathbf{D}^*)[\,\cdot\,]$ is one-to-one. Hence, $\mathbf{P}'(\mathbf{D}^*)[\bar{\mathbf{K}}] = \mathbf{0}$ implies that $\bar{\mathbf{K}} = \mathbf{0}$, which in turn implies that $\|\bar{\mathbf{K}}\|_{1,\Omega} = 0$, contradicting (3.20). Thus, (3.4) and the lemma are established for $m = 1$. We have thus established (3.6) for both $m = 0$ and $m = 1$.

For the general case of $m \in [0, 1]$, bound (3.6) follows from the results in [17], [3], and [18] and the following proof of elliptic regularity of the formal adjoint problem.

Consider boundary value problem (3.7). For $\mathbf{K} \in \mathcal{H}_{m+1}$ and $\mathcal{L}\mathbf{K} \in \mathcal{V}_m = \{\mathbf{D} \in H^m(\Omega)^4\}$. From [2], we know that $\mathcal{L}\mathbf{K} = \mathbf{f}$ is onto. This system has normal boundary conditions and, hence, the formal adjoint problem has normal boundary conditions of the same type [17]. We thus consider

$$\begin{aligned} \mathcal{L}^*\mathbf{M} &= \mathbf{f}_1, \quad \text{in} \quad \Omega, \\ \mathcal{B}^*\mathbf{M} &= \mathbf{0}, \quad \text{on} \quad \Gamma, \end{aligned}$$

for $\mathbf{M} \in (\mathcal{V}_m)^* = \{\mathbf{M} \in H^{-m}(\Omega)^4 : \mathcal{B}^*\mathbf{M} = \mathbf{0} \text{ on } \Gamma\}$ and $\mathcal{L}^*\mathbf{M} \in (\mathcal{H}_{m+1})^*$. The system is both Petrovskii elliptic, because $s_1 = s_2 = s_3 = s_4 = 0$, and homogeneous elliptic, because $t_1 = t_2 = t_3 = t_4$; cf. [17]. Thus, the adjoint system is elliptic [17] and has a similar ellipticity result in the dual space: for all $\mathbf{M} \in (\mathcal{V}_m)^*$, we have

$$\begin{aligned} \|\mathbf{M}\|_{-m,\Omega} &= \sup_{\mathbf{V}\neq 0 \in \mathcal{V}_m} \tfrac{(\mathbf{M},\mathbf{V})}{\|\mathbf{V}\|_{m,\Omega}} \\ &= \sup_{\mathbf{K}\neq 0 \in \mathcal{H}_{m+1}} \tfrac{(\mathbf{M},\mathcal{L}\mathbf{K})}{\|\mathcal{L}\mathbf{K}\|_{m,\Omega}} \\ &\leq c_1 \sup_{\mathbf{K}\neq 0 \in \mathcal{H}_{m+1}} \tfrac{(\mathcal{L}^*\mathbf{M},\mathbf{K})}{\|\mathbf{K}\|_{m+1,\Omega}} \\ &= c_1 \|\mathcal{L}^*\mathbf{M}\|_{-(m+1),\Omega} \end{aligned}$$

and

$$\begin{aligned} \|\mathbf{M}\|_{-m,\Omega} &= \sup_{\mathbf{V}\neq 0 \in \mathcal{V}_m} \tfrac{(\mathbf{M},\mathbf{V})}{\|\mathbf{V}\|_{m,\Omega}} \\ &= \sup_{\mathbf{K}\neq 0 \in \mathcal{H}_{m+1}} \tfrac{(\mathbf{M},\mathcal{L}\mathbf{K})}{\|\mathcal{L}\mathbf{K}\|_{m,\Omega}} \\ &\geq \tfrac{1}{c_3} \sup_{\mathbf{K}\neq 0 \in \mathcal{H}_{m+1}} \tfrac{(\mathcal{L}^*\mathbf{M},\mathbf{K})}{\|\mathbf{K}\|_{m+1,\Omega}} \\ &= \tfrac{1}{c_3} \|\mathcal{L}^*\mathbf{M}\|_{-(m+1),\Omega}. \end{aligned}$$

The result for all $m \in [0, 1]$ now follows from interpolation [15] and use of local maps and a partition of unity. (If we had assumed $\mathbf{D}^* \in C^\infty(\Omega)^4$ and $\Gamma \in C^\infty$, then the ellipticity result would hold for all real m; we only need this result for $m \in [0, 1]$, so we are able to reduce the continuity requirements of [15] as we have.)

We now generalize the result for $\mathbf{D} \in B_r$. Using a Taylor expansion, the triangle inequality, lemmas 3.5 and 3.7, and (3.6), we have (for $\epsilon = 0$ or δ)

$$\begin{aligned} \|\mathbf{P}'(\mathbf{D})[\mathbf{K}]\|_{\epsilon,\Omega} &= \|\mathbf{P}'(\mathbf{D}^*)[\mathbf{K}] + \mathbf{P}''(\widetilde{\mathbf{D}})[\mathbf{K}, \mathbf{D}^* - \mathbf{D}]\|_{\epsilon,\Omega} \\ &\geq \|\mathbf{P}'(\mathbf{D}^*)[\mathbf{K}]\|_{\epsilon,\Omega} - \|\mathbf{P}''(\widetilde{\mathbf{D}})[\mathbf{K}, \mathbf{D}^* - \mathbf{D}]\|_{\epsilon,\Omega} \\ &\geq c_1\|\mathbf{K}\|_{1+\epsilon,\Omega} - c_4\|\widetilde{\mathbf{D}} + \mathbf{E}\|_{1+\delta,\Omega}\|\mathbf{K}\|_{1+\epsilon,\Omega}\|\mathbf{D}^* - \mathbf{D}\|_{1+\delta,\Omega} \\ &\geq \|\mathbf{K}\|_{1+\epsilon,\Omega}(c_1 - c_4 r(\|\mathbf{D}^*\|_{1+\delta,\Omega} + \|\mathbf{E}\|_{1+\delta,\Omega} + r)) \\ &\geq \tfrac{1}{c_c}\|\mathbf{K}\|_{1+\epsilon,\Omega}, \end{aligned} \tag{3.22}$$

for sufficiently small r. The lemma now follows. ☐

LEMMA 3.7. *The second Fréchet derivative of* $\mathbf{P}(\mathbf{D})$ *is bounded for all* $\mathbf{D} \in B_r$*: for every* $\mathbf{D} \in B_r$*, there exists a constant,* c_2*, depending only on* $\mathbf{D}^*$*,* $\mathbf{E}$*,* r*,* Ω*, and* δ*, such that*

$$\|\mathbf{P}''(\mathbf{D})[\mathbf{K}, \mathbf{K}]\|_{\epsilon,\Omega} \leq c_2\|\mathbf{K}\|_{1+\delta,\Omega}\|\mathbf{K}\|_{1+\epsilon,\Omega}, \quad \forall\ \mathbf{K} \in \mathcal{H}_{1+\epsilon}(\Omega). \tag{3.23}$$

Here, $\mathbf{P}''(\mathbf{D})[\mathbf{K}, \mathbf{K}]$ *denotes the second Fréchet derivative of* $\mathbf{P}(\mathbf{D}_n)$ *with respect to* $\mathbf{D}_n$ *in directions* $\mathbf{K}$ *and* $\mathbf{K}$*.*

Proof. The products in (2.10) are of the form treated in lemmas 3.1 and 3.2. In fact, there exists a constant, C, depending only on Ω and δ, such that

$$\|\mathbf{P}''(\mathbf{D})[\mathbf{K}, \mathbf{K}]\|_{\epsilon,\Omega} \leq C\|\mathbf{D} + \mathbf{E}\|_{1+\delta,\Omega}\|\mathbf{K}\|_{1+\delta,\Omega}\|\mathbf{K}\|_{1+\epsilon,\Omega}\ \forall\ \mathbf{K}, \in H^{1+\delta}(\Omega).$$

The lemma now follows. ☐

4. Numerical Results. Here we validate our algorithm with numerical tests. Define H^h as the space of continuous piecewise bilinear functions corresponding to a uniform grid. Note that $H^h \subset H^{1+\delta}(\Omega)$ for any $\delta < \frac{1}{2}$. The functional to be minimized is

$$\begin{aligned}
&\mathbf{G}(\mathbf{x}_{n+1}, \mathbf{J}_{n+1}; \mathbf{x}_n, \mathbf{J}_n, \mathbf{w}) \\
&\quad = \varepsilon \|\mathbf{J}_{n+1} - \nabla \mathbf{x}_{n+1}\|_{0,\Omega}^2 \\
&\quad + (1-\varepsilon) \left\| \frac{1}{J_n} \left[(\underline{\hat{\mathbf{J}}_n \hat{\mathbf{J}}_n^t} \nabla) \cdot \mathbf{J}_{n+1} + (\underline{\hat{\mathbf{J}}_{n+1} \hat{\mathbf{J}}_n^t} \nabla) \cdot \mathbf{J}_n + (\underline{\hat{\mathbf{J}}_n \hat{\mathbf{J}}_{n+1}^t} \nabla) \cdot \mathbf{J}_n - 2(\underline{\hat{\mathbf{J}}_n \hat{\mathbf{J}}_n^t} \nabla) \cdot \mathbf{J}_n \right] \right\|_{0,\Omega}^2 \\
&\quad + (1-\varepsilon) \left\| \nabla \times \mathbf{J}_{n+1} \right\|_{0,\Omega}^2 + \varepsilon \|\mathbf{x}_{n+1} - \mathbf{w}\|_{\frac{1}{2},\Gamma}^2 + (1-\varepsilon) \|\mathbf{n} \times \mathbf{J}_{n+1} - \mathbf{n} \times \nabla \mathbf{w}\|_{\frac{1}{2},\Gamma}^2 .
\end{aligned} \tag{4.1}$$

There are three aspects of (4.1) worth noting. The first is that we are solving for $\mathbf{x}_{n+1}, \mathbf{J}_{n+1}$ and not $\mathbf{D}_{n+1}$ as we did for the theory. While it was more convenient in the theory to incorporate the boundary conditions into the equations, here we enforce them, so that the last two terms in (4.1) vanish. A second aspect is the inter-stage scale factor, ε. In [10], we discussed the two-stage algorithm, where, in the first stage, we set $\varepsilon = 0$ and solve for $\mathbf{J}_{n+1}$ and, in the second, we set $\varepsilon = 1$ and solve for $\mathbf{x}_{n+1}$. The second stage amounts to a simple system of decoupled Poisson equations. For $\varepsilon \in (0,1)$, minimizing (4.1) amounts to a single-stage algorithm. In section 4.1, we compare performance of the first stage of the two-stage algorithm with the single-stage algorithm for the pinched square (figure 4.1). For the single stage, we set $\varepsilon = \frac{1}{2}$ and multiply the entire functional by 2 for fair comparison. Results for both algorithms are similar, so the remaining tests are for the single-stage algorithm. The third aspect of (4.1) to notice is the presence of the equation scale factor, $\frac{1}{J_n}$, in the second functional term. The EGG equations are derived from the well-understood Laplace equations, so we exploit this correspondence now to guide the choice of scales. First note that the augmented first-stage of the first-order system [7] associated with the Laplace equations (2.1) that define $\boldsymbol{\xi}$ is (ignoring boundary conditions and with $\boldsymbol{\Psi} = \nabla \boldsymbol{\xi}$)

$$\begin{pmatrix} \nabla \cdot \boldsymbol{\Psi} \\ \nabla \times \boldsymbol{\Psi} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}.$$

Transforming this system to that for the gradient, $\mathbf{J}$, of the inverse map (without cancelling terms) yields

$$\frac{1}{J^2} \underline{\hat{\mathbf{J}}} \begin{pmatrix} \frac{1}{J} \left[(\underline{\hat{\mathbf{J}} \hat{\mathbf{J}}^t} \nabla) \cdot \mathbf{J} \right] \\ \nabla \times \mathbf{J} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}. \tag{4.2}$$

The key in scaling this new system is to understand the relative balance between its two equations. Thus, $\frac{1}{J^2} \underline{\hat{\mathbf{J}}}$ can be dropped in deference to the relative scale reflected in the $\frac{1}{J}$ term in the first equation of (4.2). To mimic the Laplace scaling for the EGG system, we thus choose to scale the second term in (4.1) by $\frac{1}{J}$. This expresses the scaling we use in the numerical experiments. To improve performance in practice, we use $\frac{1}{J}$ just to scale the norm: it is *not* involved in the linearization process. (Presence of scale factor $\frac{1}{J}$ in this way does not affect the theoretical results, so it was omitted in the analysis to simplify the calculations.) The scaling effect is demonstrated in section 4.3, where we study the convergence factors for increasingly distorted maps for the pinched square using both unscaled and scaled functionals. In both sections 4.1 and 4.2, we measure actual errors as well as functional values and validate the equivalence of the square root of the functional and the H^1 errors as proved theoretically in section 3.

In section 4.2, we first test the performance of AMG on the one-sided pinched square with grid size $h = \frac{1}{64}$. We compare the performance of V(q,s)-cycles with $q+s \leq 3$, where q is the number of relaxation steps before coarse grid correction and s is the number after. We use V(1,1)-cycles for the rest of our tests because these initial results suggest that it is one of the most efficient of these choices. We then test dependence of the linear solver on grid size. We study how the convergence factor for linear solves suffers with increasingly large perturbations from the identity map for several different grid sizes.

The method we use to obtain an approximation to $\mathbf{D}^*$ (or $\mathbf{J}^*$) is discussed in some detail in [10]. Here we give a brief overview. We use a nested sequence of $m+1$ rectangular grids with continuous piecewise bilinear function subspaces of $\mathcal{H}_{1+\delta}$ denoted by $H^{h_0} \subset H^{h_1} \subset \ldots \subset H^{h_m} \subset \mathcal{H}_{1+\delta}$, where $h_n = 2^{-n}h_0,\ \ 0 \leq n \leq m$. Let $\mathbf{V}_0$ denote the initial guess in H^{h_0} obtained by solving the problem on the coarsest subspace, H^{h_0}. In practice, we simply iterate with a discrete Newton iteration until the error in the approximation is below discretization error. The result, $\mathbf{V}_1$, becomes the initial guess for level h_1, where the process continues. In general, the initial guess for AMG on level h_n comes from the final AMG approximation on level h_{n-1}: $\mathbf{V}_n$.

In sections 4.2 and 4.4, we study performance of the NI algorithm. Here we use transfinite interpolation (TFI) to form the initial guess, which is analogous to linear interpolation. The basic principle is to add the linear interpolant between the north and south boundary maps to the linear interpolant between the east and west boundary maps, then subtract the interpolant between the four corners. The computational domain is the unit square. For boundary conditions $\mathbf{x} = \mathbf{w}(\boldsymbol{\xi})$, we get

$$\begin{aligned}\mathbf{x} = \eta\mathbf{w}_n(\xi)\ &+(1-\eta)\mathbf{w}_s(\xi) + \xi\mathbf{w}_e(\eta) + (1-\xi)\mathbf{w}_w(\eta)\\ &-\left[\xi\eta\mathbf{w}_a + \xi(1-\eta)\mathbf{w}_b + (1-\xi)(1-\eta)\mathbf{w}_c + (1-\xi)\eta\mathbf{w}_d\right],\end{aligned}$$

where we define $\mathbf{w}_n$, $\mathbf{w}_s$, $\mathbf{w}_e$, and $\mathbf{w}_w$ as the boundary maps on the north, south, east, and west boundaries, respectively, and $\mathbf{w}_a$, $\mathbf{w}_b$, $\mathbf{w}_c$, and $\mathbf{w}_d$ as the values on the northeast, southeast, southwest, and northwest corners, respectively. The initial condition we use for $\mathbf{J}$ is the Jacobian of this map.

On the north and south boundaries, boundary conditions are needed for x, y, J_{11}, and J_{12}. On the east and west boundaries, boundary conditions are needed for x, y, J_{21}, and J_{22}. Boundary conditions are imposed on the finite element space.

We first establish the similarity between the first stage of the two-stage algorithm ($\varepsilon = 0$) and the single-stage algorithm ($\varepsilon = \frac{1}{2}$ and the functional in (4.1) multiplied by 2). Second, we test the effect of different numbers of relaxation sweeps for multigrid V-cycles to suggest a good choice for the remainder of the tests. Third, we study performance of the AMG solver for increasingly distorted grids for the pinched square. Finally, we study the algorithm on the arch. Further results can be found in [9].

4.1. First-stage and single-stage algorithms. The one-sided pinched square map has the following exact solution:

$$\begin{array}{ll} x = \xi, & y = \frac{\eta}{a\xi+1}, \\ J_{11} = 1, & J_{21} = 0, \\ J_{12} = \frac{a\eta}{(a\xi+1)^2}, & J_{22} = \frac{1}{a\xi+1}, \end{array} \tag{4.3}$$

where $a \in [0,1]$. The physical domain is a square for $a = 0$, with the pinch increasing

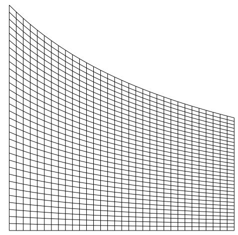

FIG. 4.1. *Pinched square with $a = 1$.*

Newton	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{128}$
1	0.96	0.93	0.90	0.95
2	0.49	0.85	0.56	0.50
3	0.24	0.47	0.41	0.40
4	0.25	0.33	0.40	0.43
5	0.25	0.32	0.40	0.41
6	0.25	0.33	0.40	0.44

TABLE 4.1
Asymptotic convergence factors for V(1,1)-cycles, with varying grid size and Newton iterations.

as a increases. See figure 4.1.

To test performance of the first-stage and single-stage algorithms for standard Newton iterations and NI on the pinched square with $a = 1.0$, we add a varying amount of small error at each grid point (except for those on the boundary) to TFI (the exact solution in this case) to form the initial guess:

$$\begin{array}{ll} x = \xi + \mathbf{g}\sin(b\xi + c\eta), & y = \frac{\eta}{\xi+1} + \mathbf{g}\sin(d\xi + e\eta), \\ J_{11} = 1 + \mathbf{g}\sin(b\xi + c\eta), & J_{21} = \mathbf{g}\sin(b\xi + c\eta), \\ J_{12} = \frac{\eta}{(\xi+1)^2} + \mathbf{g}\sin(d\xi + e\eta), & J_{22} = \frac{1}{\xi+1} + \mathbf{g}\sin(d\xi + e\eta), \end{array}$$

where

$$\begin{array}{llll} \mathbf{g} & = f\xi\eta(1-\xi)(1-\eta) & & \\ b & = 12967493.946193764, & c & = 491843027.481264509, \\ d & = 184625498.4710938, & e & = 174365204.5761938, \end{array}$$

with $f = 2$ for grid $h = \frac{1}{128}$, $f = 4$ for grids $h = \frac{1}{64}$ and $h = \frac{1}{32}$, $f = 7$ for grid $h = \frac{1}{16}$, and $f = 24$ for NI (with coarsest grid $h = \frac{1}{4}$). Note that the exact solutions for x, J_{11}, and J_{21} are in the finite-dimensional subspaces.

Consider the first-stage one-sided pinched square. (Recall that there are no x or y terms.) Table 4.1 depicts asymptotic convergence factors for the AMG solver. Note the poor performance shown in the early Newton steps. This degradation is probably because the functional is suffering from loss of elliptic character due to the crude initial guess inheriting poor values for the Jacobian map. Nested iteration tends to ameliorate this potential difficulty, so we may focus on later Newton iterations, where these results suggest that 2 $V(1,1)$-cycles yield overall convergence factors of about 0.2. This is what we use in the tests that follow.

Figure 4.2 depicts Newton convergence results for grids $h = \{\frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{128}\}$. We study performance in terms of both the functional error measure (i. e., square

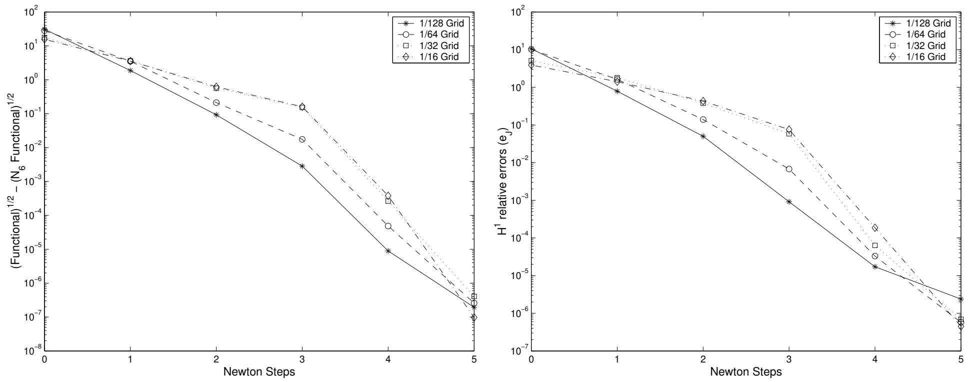

FIG. 4.2. *First-stage functional. Newton convergence, using standard Newton iterations with imposed boundary conditions, in both functional and H^1 error measures. Differences between the values at the current and sixth Newton steps are plotted.*

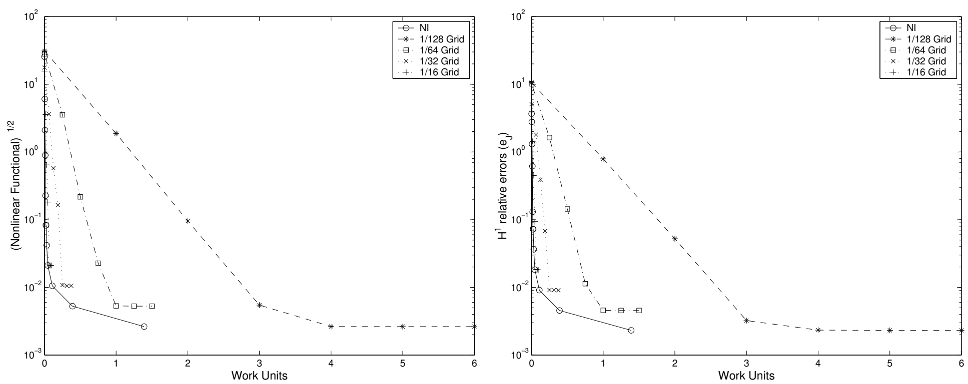

FIG. 4.3. *First-stage functional. Functional and H^1 error measures for standard Newton iterations and NI with imposed boundary conditions. One work unit is the equivalent of one step on the $h = \frac{1}{128}$ grid using 2 V(1,1)-cycles.*

root of the functional) and the relative H^1 errors in $\mathbf{J}$, $e_{\mathbf{J}} \equiv \frac{\|\mathbf{J}^* - \mathbf{J}_n\|_{1,\Omega}}{\sqrt{\|\mathbf{J}^*\|_{1,\Omega}\|\mathbf{J}_n\|_{1,\Omega}}}$. The graphs show the differences between the values at the current and sixth Newton steps. We are interested in the functional measure because it is equivalent to the H^1 norm of the errors, as we established theoretically in [10] and as these graphs suggest. The left graph contains the functional values and the right graph contains the errors. Convergence appears to be approximately linear, which is consistent with the theoretical result. The factors also appear to be bounded independent of grid size.

Figure 4.3 compares Standard Newton and NI. We again report on the functional and relative H^1 error measures in $\mathbf{J}$. For proper comparison of cost, we now base the data on a *work unit*, defined to be the equivalent of one Newton step on the $h = \frac{1}{128}$ grid. (One Newton step has 2 V(1,1)-cycles.) We thus count one Newton step on the $h = \frac{1}{64}$ grid as $\frac{1}{4}$ of a work unit, $\frac{1}{16}$ on the next coarser grid, and so on. After about the sixth standard Newton step for each of the grid sizes, the change in the functional value (and the H^1 error) at each iteration is very small relative to the functional value itself. The exact solution is only approximated by the finite-dimensional subspace.

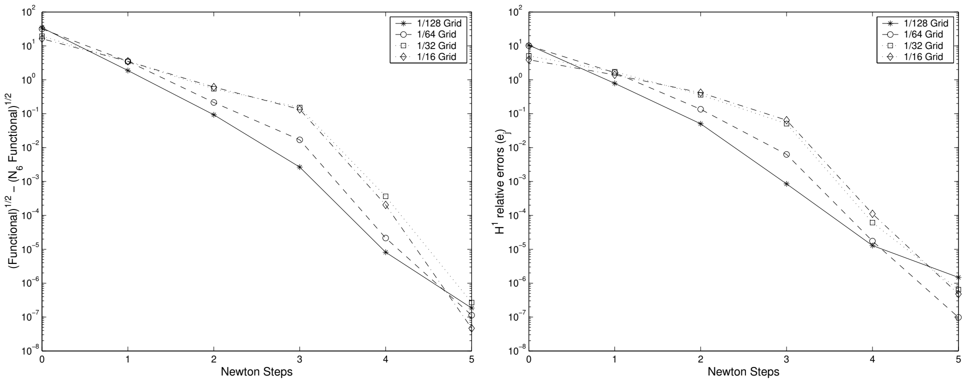

FIG. 4.4. *Newton convergence, using standard Newton iterations with imposed boundary conditions, in both functional and H^1 error measures. Difference between the value at the current and sixth Newton steps are plotted.*

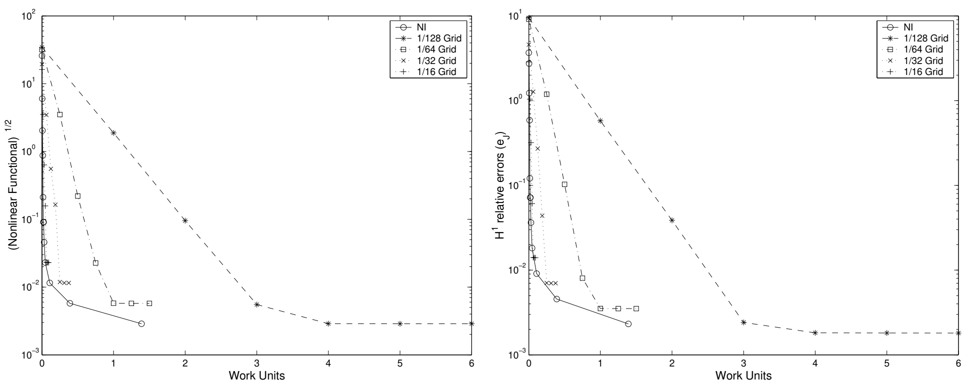

FIG. 4.5. *Newton vs. NI, with imposed boundary conditions, in both functional and H^1 error measures.*

Thus, while the functional value for the exact solution is zero, the minimum on the finite-dimensional subspace is not. With more Newton steps, we can thus get as close as we choose to the finite-dimensional approximation of the exact solution, but the decrease in the functional and, hence, the error stalls because discretization error is reached. The ratios of the functional and the H^1 error measures in $\mathbf{J}$ are about 1.16 near the solution for grids $h = \{\frac{1}{16}, \frac{1}{32}, \frac{1}{64}\}$ and 1.14 for grid $h = \frac{1}{128}$. After the third Newton step, this ratio is a constant for all grid sizes, which affirms H^1 equivalence.

Next we study performance of the single-stage algorithm, with the same map and initial guess. Again, we report on functional and relative H^1 error measures in $\mathbf{J}$. Figure 4.4 contains graphs of differences between these values at the current and sixth Newton steps. Consistent with the theory, convergence using this measure appears to be approximately linear, with factors bounded independent of grid size.

Figure 4.5 compares standard Newton iterations with NI based on works units as defined above. Behavior of the errors for the single stage is essentially the same as for the first stage. The ratios of functional measures of the single stage to the first stage varies between 0.87 and 1.12, fixing at 1.09 after Newton step 4 for each grid. For NI, the ratio is 1.09 for all the finer grids.

The relative error in the computed solution does not appear to vary with grid size because the variation is small compared to the error. Again, on any finite-dimensional subspace, we cannot expect to reduce the error to zero because of discretization error. The ratios of the functional and H^1 error measures in $\mathbf{J}$ are about 1.3 near the solution for grids $h = \{\frac{1}{16}, \frac{1}{32}, \frac{1}{64}\}$ and 1.2 for grid $h = \frac{1}{128}$. After the third Newton step, this ratio is a constant for all grid sizes, which affirms H^1 equivalence. We need at least 4 standard Newton steps to reduce the functional to about the same level as for NI that needed an equivalent of only about 1.5 steps. We expect this difference to widen for larger problems, where the required steps for standard Newton would tend to grow but NI would probably remain below an equivalent of two.

4.2. V-cycle tests. To determine which V(q,s)-cycle is most efficient, we study asymptotic convergence factors with $q + s \leq 3$ and $h = \frac{1}{64}$. Here we linearize the equations about the exact solution, set the right side to zero, start with a random initial guess, and then observe residual reduction factors after many V-cycles. Table 4.2 shows the observed V-cycle convergence factors for different values of a. The cycles

a	V(0,1)	V(1,0)	V(0,2)	V(1,1)	V(2,0)	V(0,3)	V(1,2)	V(2,1)	V(3,0)
0.0	0.33	0.23	0.20	0.13	0.13	0.16	0.11	0.11	0.11
0.1	0.33	0.25	0.21	0.14	0.14	0.16	0.12	0.12	0.12
0.4	0.38	0.30	0.26	0.18	0.19	0.21	0.15	0.15	0.16
0.7	0.47	0.39	0.34	0.26	0.25	0.28	0.22	0.22	0.22
1.0	0.60	0.53	0.45	0.40	0.39	0.37	0.32	0.32	0.28

TABLE 4.2

Asymptotic convergence factors for different V-cycles and values of a, grid size $h = \frac{1}{64}$.

with more relaxation sweeps naturally have better convergence factors, but involve more computation. We thus consider a measure of the time required to reduce the initial residual by a factor of 10. Since we are only interested in comparisons, we choose the *relative* measure $t \equiv (q + s + c)\ln(0.1)/\ln(r)$, where r is the observed asymptotic convergence factor for the V(q,s)-cycle and c estimates the fixed cost of a cycle. We choose $c = 2$ because of residual calculations and intergrid transfers. Observed values for t for the $h = \frac{1}{64}$ grid and different values of a are given in table 4.3. While performance of the V(1,1)- or V(2,0)-cycles were similar, we chose the

a	V(0,1)	V(1,0)	V(0,2)	V(1,1)	V(2,0)	V(0,3)	V(1,2)	V(2,1)	V(3,0)
0.0	6.2	4.7	5.7	4.5	4.5	6.3	5.2	5.2	5.2
0.1	6.2	5.0	5.9	4.7	4.7	6.3	5.4	5.4	5.4
0.4	7.1	5.7	6.8	5.4	5.5	7.4	6.1	6.1	6.3
0.7	9.1	7.3	8.5	6.8	6.6	9.0	7.6	7.6	7.6
1.0	13.5	10.9	11.5	10.1	9.8	11.6	10.1	10.1	9.0

TABLE 4.3

Relative time to reduce residual by a factor of 10 with various a for different V-cycles and $h = \frac{1}{64}$.

V(1,1)-cycle for the remainder of our tests.

4.3. AMG tests. We next test the performance of the linear solver with varying h. Again, the equations are linearized about the solution and the right side is set to zero. We study the deterioration in asymptotic convergence factors as a increases from zero to one. The results are plotted in figure 4.6. In all but one test, we

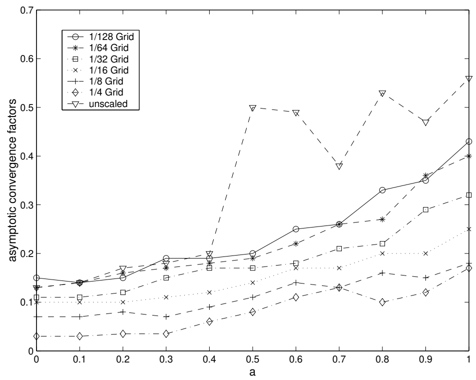

FIG. 4.6. *Asymptotic convergence factors for various grid sizes and values of a. The scaled functional was used in all but the test for $h = \frac{1}{64}$.*

used the scaled functional in (4.1) with the boundary conditions enforced so that the boundary terms vanish. For the test marked "unscaled" and for which $h = \frac{1}{64}$, factor $\frac{1}{J_n}$ in the second term of the functional in (4.1) was omitted. For the scaled functional, asymptotic convergence factors increase as a increases, as expected. At $a = 0$, which corresponds to the identity map, convergence factors are similar to those for the Laplace problem. There is some variation with respect to the grid size, although for smaller h $\left(\frac{1}{64} \text{ and } \frac{1}{128}\right)$ the factors are similar. The results for the unscaled and scaled convergence factors for the $h = \frac{1}{64}$ grid confirm that scaling the second term of the functional in (4.1) by $\frac{1}{J_n}$ significantly improves convergence factors: the unscaled factors are significantly larger than the scaled factors for $a \geq .5$.

4.4. Nested iteration for the Arch. We compare standard Newton iterations for $h = \frac{1}{128}, \frac{1}{64}, \frac{1}{32}$, and $\frac{1}{16}$ to NI with $h = \frac{1}{4}$ for the coarsest grid and $h = \frac{1}{128}$ for the finest grid. The initial guess for the arch is

$$\begin{aligned}
x &= 1.5 + (1.5 - \xi)\cos(\pi(1-\eta)), & y &= (1.5 - \xi)\sin(\pi(1-\eta)), \\
J_{11} &= -\cos(\pi(1-\eta)), & J_{21} &= \pi(1.5 - \xi)\sin(\pi(1-\eta)), \\
J_{12} &= -\sin(\pi(1-\eta)), & J_{22} &= -\pi(1.5 - \xi)\cos(\pi(1-\eta)).
\end{aligned}$$

Choices of the numbers of V(1,1)-cycles per iteration and Newton steps on each grid are currently made by observation. In the theoretical section of [10], we suggested $\rho^{\nu_0} \leq \frac{1}{8}$ as a criterion, where ρ is the convergence factor and ν_0 is the number of V-cycles. A significantly larger value would allow the iterates to wander too far from the true solution as the grid is refined. We could chose a smaller value for ρ^{ν_0} so that the multigrid solutions shadow the exact finite-dimensional solutions more closely. But too small of value would likely be less efficient than simply proceeding to finer meshes. NI required significantly less work to obtain the same discretization error as standard Newton. Standard Newton needed just a few steps to reach discretization error for our tests anyway, but the savings afforded by NI for smaller h should be much larger still. More results can be found in [9].

Table 4.4 depicts asymptotic convergence factors for standard Newton iterations. These factors are not small enough to allow just one V cycle per Newton step. We thus used 3 V(1,1) cycles to solve each Newton step. So one work unit is 3 V(1,1) cycles on the $\frac{1}{128}$ grid. Here we performed two coarsest-grid Newton iterations, with only one on all finer grids. Three standard Newton steps were required to reach

$Newton$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{128}$
1	0.59	0.50	0.43	0.59
2	0.63	0.65	0.67	0.61
3	0.66	0.67	0.68	0.67
4	0.63	0.67	0.66	0.67
5	0.66	0.68	0.66	0.68

Table 4.4

Asymptotic convergence factors for the V(1,1) cycle, with varying grid size and Newton iterations, for the arch.

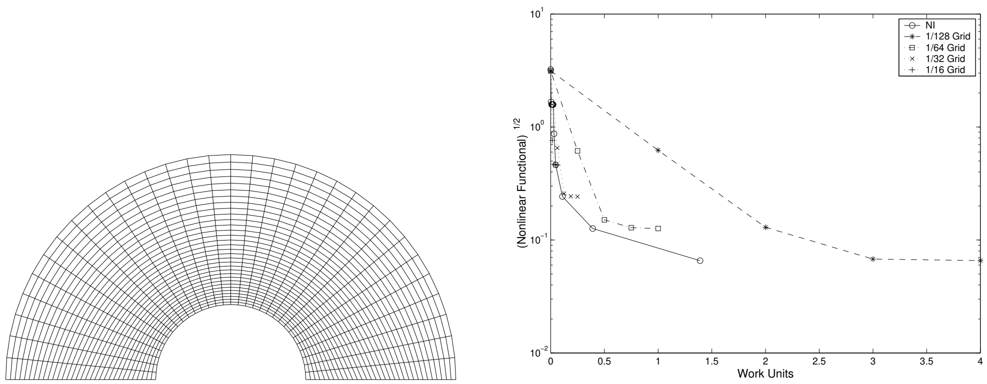

Fig. 4.7. *Nested iteration and standard Newton for the arch.*

discretization error, while NI required less than one and a half equivalents. The final functional value decreases by about a factor of four as the grid size is halved, which confirm $\mathcal{O}(h)$ approximation in the $H^1(\Omega)$ norm.

5. Conclusion. We showed theoretically that the nested iteration process involving only one discrete Newton step on each level produces a result on the finest level that is within discretization error of the exact solution. We also showed this result numerically using an $H^{1+\delta}(\Omega)$ discrete space for each of the unknowns. Future directions involve automating the numerical tests to include the following choices: number of relaxations before and after coarsening, number of V-cycles, number of Newton steps on each grid, size and choice of solvers for the coarsest grid, parameterization of the boundary maps, and adaptive mesh refinement.

The first three choices dictate the overall efficiency of the algorithm and should be considered carefully for maximum effectiveness. Automation would require heuristics to sense performance of smoothing and coarse-grid correction, as well as linearization trade-offs. We used one Newton step on all but the coarsest grid in our examples and theory, but severely distorted regions may dictate more such steps to improve effectiveness and possibly other continuation methods to address Newton's local convergence characteristics. In any case, the special ability of the FOSLS functional to signal errors could be exploited to make these choices in an effective and automatic way. The fourth coarsest-grid choice rests heavily on the geometry of the particular map. Complex regions may require a fairly small coarsest grid and a significant amount of effort to solve the nonlinear problem there. Damped Newton methods and various forms of continuation techniques may come into play. Of course, complicated regions generally require very fine meshes to supply meaningful simulations, so the relative cost of such coarsest-grid effort may again be fairly minimal. Moreover, the special properties of the FOSLS functional may also be exploited for these choices. The fifth choice would be to use a parameterization of the boundary in the associated terms of the functional that would allow concentration of gird points near special boundary features. The final choice of adaptive mesh refinement can be served by noting that the functional value on each element is a sharp measure of the error on that element, which makes it suitable as a measure to determine which elements need to be further subdivided (cf. [4]).

REFERENCES

[1] R.A Adams. *Sobolev Spaces*, volume 65 of *Pure and Applied Mathematics*. Academic Press, New York, 1975.

[2] S. Agmon, A. Douglis, and L. Nirenberg. Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions II. *Comm. Pure Appl. Math.*, 17:35–92, 1963.

[3] A.K. Aziz. *The Mathematical Foundations of the Finite Element Method with Application to Partial Differential Equations*. Academic Press, New York, 1972.

[4] M. Berndt, T. A. Manteuffel, and S. F. McCormick. Local error estimates and adaptive refinement for first-order system least-squares (FOSLS). *E.T.N.A.*, 6:35–43, 1997.

[5] D. Braess. *Finite Elements Theory, fast solvers, and applications in solid mechanics.* Cambridge University Press, Cambridge, 1997.

[6] S.C. Brenner and L.R. Scott. *The Mathematical Theory of Finite Elememt Methods*, volume 15 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1994.

[7] Z. Cai, T.A. Manteuffel, and S.F. McCormick. First-order system least squares for second-order partial differential equations: Part II. *SIAM J. of Numer. Anal.*, 34:425–454, 1997.

[8] J.E. Castillo, editor. *Mathematical Aspects of Numerical Grid Generation*, volume 8 of *Frontiers in Applied Mathematics*. SIAM, Philadelphia, 1991.

[9] A.L. Codd. *Elasticity-Fluid Coupled Systems and Elliptic Grid Generation (EGG) based on First-Order System Least Squares (FOSLS).* PhD thesis, University of Colorado at Boulder, 2001.

[10] A.L. Codd, T. Manteuffel, and S. McCormick. Multilevel first-order system least squares for nonlinear elliptic partial differential equations.

[11] V. Girault and P.-A. Raviart. *Finite Element Methods for Navier-Stokes Equations*. Springer, Berlin, 1986.

[12] P. Knupp and S. Steinberg. *Fundamentals of Grid Generation*. CRC Press, Boca Raton, 1993.

[13] B. Lee, T. A. Manteuffel, S. F. McCormick, and J. Ruge. First-order system least-squares (FOSLS) for the Helmholtz equation. *SIAM J of Numer. Anal.*, 21(5):1927–1949, 2000.

[14] G. Liao. On harmonic maps. In J.E. Castillo, editor, *Mathematical Aspects of Numerical Grid Generation*, volume 8 of *Frontiers in Applied Mathematics*, chapter 9, pages 123–130. SIAM, Philadelphia, 1991.

[15] J.L. Lions and E. Magenes. *Non-Homogeneous Boundary Value Problems and Applications v 1*, volume 181 of *Die Grundlehren der mathematischen Wissenschlaften in Einzeldarstellungen*. Springer-Verlag, Berlin, english edition, 1972.

[16] T. Rado. Aufgabe 41. *Jahresbericht der Deutschen Mathematiker Vereinigung*, 35:49, 1926.

[17] Ja. A. Roitberg and Z.G. Seftel. A theorem on homeomorphisms for elliptic systems and its applications. *Mathematics of the USSR-Sbornik*, 7(3):439–465, 1969.

[18] Ya. A. Roitberg. A theorem about the complete set of isomorphisms for systems elliptic in the sense of Douglis and Nirenberg. *Ukrainian Mathematical Journal*, 25(4):396–405, 1973.

[19] M. Schechter. Solution of dirichlet problem for systems not necessarily strongly elliptic. *Comm. Pure Appl. Math.*, 12:241–247, 1959.

[20] J.F. Thompson, Z.U.Z. Warsi, and C.W. Mastin. *Numerical Grid Generation*. North Holland, New-York, 1985.

[21] J. Wloka. *Partial Differential Equations*. Cambridge University Press,, Cambridge, 1987.

[22] K Yosida. *Functional Analysis*. Springer-Verlag, Berlin, 6 edition, 1965.