MULTILEVEL FIRST-ORDER SYSTEM LEAST SQUARES FOR NONLINEAR ELLIPTIC PARTIAL DIFFERENTIAL EQUATIONS

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

Abstract. A fully variational approach is developed for solving nonlinear elliptic equations that enables accurate discretization and fast solution methods. The equations are converted to a first-order system that is then linearized via Newton's method. First-order system least squares (FOSLS) is used to formulate and discretize the Newton step, and the resulting matrix equation is solved using algebraic multigrid (AMG). The approach is coupled with nested iteration to provide an accurate initial guess for finer levels using coarse-level computation. A general theory is developed that confirms the usual full multigrid efficiency: accuracy comparable to the finest-level discretization is achieved at a cost proportional to the number of finest-level degrees of freedom. In a companion paper, the theory is applied to elliptic grid generation (EGG) and supported by numerical results.

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

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

1. Introduction. We develop a theoretical foundation for a method for solving nonlinear elliptic equations that combines first-order system least squares (FOSLS) with Newton's method, algebraic multigrid (AMG), and nested iteration (NI). The algorithm achieves accuracy comparable to the finest-level discretization at a cost proportional to the number of finest-level degrees of freedom. In a companion paper [17], we apply this theory to the elliptic grid generation equations (EGG) and numerically validate the theory established below.

Our development assumes that the target problem is a first-order system whose associated least-squares functional applied to functions in $H^{1+\delta}$, $\delta \in (0,1)$, has quadratic part that is equivalent to the product H^1 norm. Higher-order differential systems can be recast in the standard way as a first-order system, but care must be taken to ensure such product ellipticity when it is feasible (cf. [15, 14]). Our particular interest is in quasilinear first-order systems where the nonlinearity is a product of variables, at most one of which is a derivative term. For example, if u and v are the variables in a two-dimensional problem, then we admit terms like u, v, $u^2 u_x$, and uvv_y, but not uu_x^2. (Admitting product derivative terms while retaining $H^{1+\delta}$ spaces for the variables would prevent the use of L^2 norms for the equations and inhibit analysis of the linearized equations.)

Our algorithm applies to the first-order system in three separate stages. The outermost stage is nested iteration (NI), which starts on the coarsest level where the discrete nonlinear problem is solved by any appropriate method. The result is then interpolated to the next finer level where it is used as an initial guess for *one* Newton linearization (the middle stage) of the nonlinear problem. A functional is created on that level using a least-squares principle and the resulting matrix equation is solved using *one, two, or three* V-cycles of AMG (the innermost stage). The result is then interpolated up to the next finer level, with the steps repeated until the finest level is

processed. Our theoretical results confirm that this *direct* NI-Newton-FOSLS-AMG scheme converges in *one* overall step to an approximation on the finest level that is accurate to *the level of discretization error*. Numerical experiments for the EGG equations described in a companion paper [17] confirm this result.

One advantage that the FOSLS system has over standard minimization techniques is that the minimum value of the functional is zero at the exact solution of the differential equation. This property has implications for adaptive refinement that are to be explored in future work. Another advantage is that FOSLS used with finite element discretization and Newton linearization of the elliptic equations results in self-adjoint positive-definite matrix problems that themselves correspond to a well-posed elliptic system, so the discrete problems can be efficiently solved by multigrid. We demonstrate this attribute qualitatively in the theory and numerically in the companion paper.

The idea of using Newton iterations coupled with a multilevel scheme is not new. In [20], for example, a multilevel nested iteration Newton scheme was applied to differential eigenproblems. An abstract theory and numerical results confirmed the need for only one Newton step on the finest level. In [5], optimal parameters were calculated for damped approximate Newton to ensure quadratic convergence of a particular finite element approximation of nonlinear elliptic partial differential equations. This was later combined with multilevel techniques [6] to obtain a convergence result that asymptotically required just one Newton linearization per level. This later result is similar to ours, but does not include the derivative terms in the nonlinearity that are present in our target application, elliptic grid generation (EGG). The NI approach is also used in recent work on cascadic multigrid [22], although again their form of the nonlinearity does not include the more complicated case needed here.

The "mesh-independence" theory developed for Newton's method in [21, 4, 3, 2] addresses the same property of NI that we exploit here. Unfortunately, this theory cannot easily be applied to our setting because it requires more smoothness of the infinite-dimensional iterates than ours appear to possess. We are also unable to apply the mesh-independence-based theory developed in [18, 19] because the nonlinearity for the Navier-Stokes equations treated there appears only in the lower-order terms.

This paper is organized as follows. Section 2 introduces the equations and function spaces. In section 3, we describe the NI-Newton-FOSLS method for solving the nonlinear equations. Section 4 contains theory on convergence of the Newton iterates in H^1 and on accuracy estimates for the NI-Newton-FOSLS-AMG scheme. We conclude with some remarks in the final section.

2. Setup. 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 Γ. 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 $|\cdot|_{i,\Omega}$ $(0 \leq i \leq m)$. (We suppress superscript n because dependence of the vector norms on dimension is clear by context.) For $\delta \in (0,1)$, let $H^{m+\delta}(\Omega)$ (c.f. [11]) denote the Sobolev space associated with the norm defined by

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

(This definition allows the use of the "real interpolation" method [1, 11].) Also, let $H^{\frac{1}{2}}(\Gamma)$ denote the trace Sobolev space associated with the norm

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

Finally, let $C^m(\Omega)$ denote the space of functions with continuous derivatives in Ω of up to order $m \geq 0$ and define the $C^0(\Omega)$ norm of $f \in C^0(\Omega)$ by

$$\|f\|_{\infty,\Omega} \equiv \sup_{x\in\Omega} |f(x)|.$$

From Sobolev's lemma [1], there exists a constant, C, depending only on Ω and δ, such that

$$\|f\|_{\infty,\Omega} \leq C\|f\|_{1+\delta,\Omega}, \qquad \forall\, f \in H^{1+\delta}(\Omega).$$

The method we develop applies to elliptic quasilinear partial differential equations, where the highest-order derivative terms appear linearly, with the exception that their coefficients may include lower-order terms. In fact, we assume that the equations have been formulated as a $p \times q$ first-order system with appropriate boundary conditions. It is straightforward to rewrite higher-order equations in first-order form, although care must be taken to ensure that the resulting system is elliptic in the H^1 product norm (assuming that this is even feasible; cf. [15, 14]). This process is exemplified by the reformulation of the EGG equations in the companion paper [17].

Let the first-order system be represented abstractly by the p-vector equation

$$(2.1) \qquad\qquad \mathbf{p}(\mathbf{J}) = \mathbf{0} \qquad \text{in} \quad \Omega,$$

with boundary conditions

$$(2.2) \qquad\qquad \mathbf{BJ} = \mathbf{g} \qquad \text{on} \quad \Gamma,$$

where $\mathbf{J}$ is the q-vector of unknowns. (p and q are positive integers, generally with $p \geq q$.) To ensure that we can apply least squares to this system, we must have $\mathbf{p}(\mathbf{J}) \in L^2(\Omega)^p$. We want to allow product terms involving a combination of elements of $\mathbf{J}$, one of which may involve a partial derivative. Therefore, we cannot allow $\mathbf{J}$ to roam freely in $H^1(\Omega)^q$ because such products would not necessarily be in $L^2(\Omega)$ and we would thus be prevented from using $L^2(\Omega)$ norms for the functional. However, from Sobolev's Lemma, the $C^0(\Omega)$ norm is bounded by the $H^{1+\delta}(\Omega)$ norm in R^2 when $\delta \in (0, 1)$. Thus, everything in $H^{1+\delta}(\Omega)$ is continuous and our product terms are in $L^2(\Omega)$. We therefore choose the space for $\mathbf{J}$ to be $H^{1+\delta}(\Omega)^q$, ensuring $\mathbf{p}(\mathbf{J}) \in L^2(\Omega)^p$. This obviously places restrictions on the allowable boundary functions, $\mathbf{g}$. In fact, we assume that the solution, $\mathbf{J}^*$, of (2.1)–(2.2) is in $H^{2+\delta}(\Omega)^q$, which places even further restrictions on $\mathbf{g}$. In addition, the coercivity requirement on the first Fréchet derivative of our system influences the allowable spaces for both the boundary and boundary conditions, which in turn influences the solution space for $\mathbf{J}^*$. This issue is addressed implicitly in the abstract theory of section 4 and in detail for the EGG application in the companion paper [17].

In other FOSLS applications [13, 15, 7, 8, 14], both H^{-1} and L^2 norms are used for the domain and $H^{\frac{1}{2}}$ norms for the boundary. With appropriate smoothness [15], FOSLS functionals for general second-order elliptic partial differential equations exhibit H^1 equivalence for the functionals based on L^2 norms for the domain and $H^{\frac{1}{2}}$

norms for the boundary. In practice, while L^2 norms are used for the domain, it is common either to use L^2 norms scaled by $\frac{1}{h}$ for the boundary norms or to impose the boundary conditions. We focus on imposing boundary conditions here for simplicity, although some of our numerical results in [17] take the scaled L^2 norm approach for illustration.

It is more convenient in the analysis to consider homogeneous boundary conditions. To this end, we extend $\mathbf{g}$ smoothly into Ω: assume that we are given a q-vector function, $\mathbf{E}$, defined on Ω that satisfies

$$\mathbf{BE} = \mathbf{g}, \qquad \text{on } \Gamma.$$

Now, writing $\mathbf{J} = \mathbf{D} + \mathbf{E}$ and $\mathbf{P}(\mathbf{D}) = \mathbf{p}(\mathbf{D} + \mathbf{E})$, our target problem becomes

$$\mathbf{P}(\mathbf{D}) = \mathbf{0}, \qquad \text{in } \Omega, \tag{2.3}$$

with homogeneous boundary conditions

$$\mathbf{BD} = \mathbf{0}, \qquad \text{on } \Gamma. \tag{2.4}$$

Generally, $\mathbf{E}$ needs to be as smooth as we require $\mathbf{J}^* = \mathbf{D}^* + \mathbf{E}$ to be, but this requirement is implicit in the following assumptions we make on (2.3)–(2.4).

We start by defining the space on which this system is posed. For any $\nu > 0$, define

$$\mathcal{H}_\nu = \{\mathbf{D} \in H^\nu(\Omega)^q : \mathbf{BD} = \mathbf{0} \text{ on } \Gamma\}. \tag{2.5}$$

We assume that our solution, $\mathbf{D}^*$, resides in $\mathcal{H}_{2+\delta}$ and we look for it in $\mathcal{H}_{1+\delta}$. Note that $\mathbf{D} \in \mathcal{H}_{1+\delta}$ implies $\mathbf{P}(\mathbf{D}) \in L^2(\Omega)^p$.

3. Method. There are several decisions to be made about how (2.3) is solved. Our basic choice is to use Newton's method and FOSLS to obtain a quadratic minimization problem, finite elements for the discretization, and then AMG with NI to solve the resulting matrix equations. Within this basic framework, we need to choose how Newton's method and FOSLS are related. A FOSLS-Newton method would involve forming the least-squares functional, setting its gradient to zero, and then solving this nonlinear problem with Newton's method. A Newton-FOSLS method would involve linearizing the equations first, and then forming the least-squares functional and setting its gradient to zero. The gradient equations that result from these two approaches differ only by a term coming from the second Fréchet derivative of the system operator that, near the solution, is dominated by the other operator terms. Because nested iteration guarantees proximity to the solution on each level, the performance of these two approaches tends to be much the same. We therefore focus on the Newton-FOSLS approach because of its theoretical and numerical simplicity.

Our method involves first applying Newton's method to nonlinear system (2.3) on the coarsest finite element level. We then form an L^2 functional from the linearized equations and minimize a coarsest-level discretization of it by AMG (or perhaps a direct matrix solver). The resulting approximate Newton iterate computed at the coarsest-level scale is then used on the next finer level as an initial guess for an analogous discrete Newton step there: system (2.3) is linearized about this initial guess, an L^2 minimization principle is applied, and then AMG is used to approximate the minimizer on this finer scale. This process continues, with the iterates approximated on successively finer levels, until a desired accuracy is reached.

Applying Newton's method to system (2.3) gives us the following linearized problem: given $\mathbf{D}_n \in \mathcal{H}_{1+\delta}$, find $\mathbf{D}_{n+1} \in \mathcal{H}_{1+\delta}$ such that

$$(3.1) \qquad \mathbf{P}'(\mathbf{D}_n)[\mathbf{D}_{n+1} - \mathbf{D}_n] = -\mathbf{P}(\mathbf{D}_n),$$

where $\mathbf{P}'(\mathbf{D}_n)[\mathbf{D}_{n+1} - \mathbf{D}_n]$ denotes the first Fréchet derivative of $\mathbf{P}(\mathbf{D}_n)$ with respect to $\mathbf{D}_n$ in direction $\mathbf{D}_{n+1} - \mathbf{D}_n$.

To solve (3.1) for $\mathbf{D}_{n+1}$, consider the least-squares functional

$$\begin{aligned} \mathbf{G}_0(\mathbf{D}_{n+1}) &\equiv \|\mathbf{P}'(\mathbf{D}_n)[\mathbf{D}_{n+1} - \mathbf{D}_n] + \mathbf{P}(\mathbf{D}_n)\|_{0,\Omega}^2 \\ &= \left(\mathbf{P}'(\mathbf{D}_n)[\mathbf{D}_{n+1} - \mathbf{D}_n] + \mathbf{P}(\mathbf{D}_n), \ \ \mathbf{P}'(\mathbf{D}_n)[\mathbf{D}_{n+1} - \mathbf{D}_n] + \mathbf{P}(\mathbf{D}_n)\right). \end{aligned}$$

Note that $\mathbf{G}_0$ depends on $\mathbf{D}_n$ and $\mathbf{E}$. To minimize $\mathbf{G}_0$, we set to zero its first Fréchet derivative, taken with respect to $\mathbf{D}_{n+1}$ (cancelling the factor 2 for convenience): given $\mathbf{D}_n \in \mathcal{H}_{1+\delta}$, find $\mathbf{D}_{n+1} \in \mathcal{H}_{1+\delta}$ such that

$$(3.2) \qquad \left(\mathbf{P}'(\mathbf{D}_n)[\mathbf{K}], \ \ \mathbf{P}'(\mathbf{D}_n)[\mathbf{D}_{n+1} - \mathbf{D}_n] + \mathbf{P}(\mathbf{D}_n)\right) = \mathbf{0}, \quad \forall \, \mathbf{K} \in \mathcal{H}_{1+\delta}.$$

We illustrate this infinite-dimensional Newton process in figure 3.1, where $\xrightarrow{N}$ indicates one Newton step.

$$\mathbf{D}_0 \xrightarrow{N} \mathbf{D}_1 \xrightarrow{N} \mathbf{D}_2 \xrightarrow{N} \mathbf{D}_3 \xrightarrow{N} \mathbf{D}_4 \ \cdots \ \mathbf{D}_n \xrightarrow{N} \mathbf{D}_{n+1} \xrightarrow{N} \ \cdots \ \mathbf{D}^*$$

FIG. 3.1. *The Newton-FOSLS infinite-dimensional algorithm*

In practice, we need to discretize (3.2) on some given finite element space, H^h. But we then need an approximation for $\mathbf{D}_n$. The main point to keep in mind in this approximation is that early iterates are relatively crude approximations to $\mathbf{D}^*$, so they can be approximated on relatively coarse grids. In general, if final iterate $\mathbf{D}_{n+1}$ is approximated by a best approximation, $\mathbf{U}_{n+1}$, in H^{h_n}, then $\mathbf{D}_n$ need only be approximated on a grid with mesh size $\mathcal{O}(h_n^{\frac{1}{2}})$. This is a natural consequence of quadratic convergence and this premise is served well by a coarse grid of mesh size $2h$. This view gives rise to our nested iteration approach that supplies the initial guess for one Newton iterate on grid h by first iterating on grid $2h$. In particular, consider a nested sequence of $m+1$ finite-dimensional 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 $h_n = 2^{-n}h_0$, $0 \leq n \leq m$. Note that piecewise bilinears on rectangles are in $H^{1+\delta}(\Omega)$ for $\delta \in [0, \frac{1}{2})$ [16]. Let $\mathbf{U}_0$ denote the initial guess in H^{h_0}. For $n = 0, 1, \ldots, m$ in turn, define the grid H^{h_n} problem as follows: given $\mathbf{U}_n \in H^{h_n}$, find $\mathbf{U}_{n+1} \in H^{h_n} \subset H^{h_{n+1}}$ such that

$$(3.3) \qquad \left(\mathbf{P}'(\mathbf{U}_n)[\mathbf{K}^{h_n}], \ \ \mathbf{P}'(\mathbf{U}_n)[\mathbf{U}_{n+1} - \mathbf{U}_n] + \mathbf{P}(\mathbf{U}_n)\right) = \mathbf{0}, \quad \forall \, \mathbf{K}^{h_n} \in H^{h_n}.$$

3.1. NI-Newton-FOSLS. The discretization in (3.3) amounts to approximating the finest-level solution, $\mathbf{U}_m^*$, by a nested iteration on subspaces H^{h_n}, $n = 0, 1, \cdots m$. This NI approach involves first solving problem (3.3) on the coarsest subspace, H^{h_0}. In practice, we can use any sensible solution process here because this space is presumably of very low dimension. We can simply iterate with a (possibly damped) discrete Newton iteration until the error in the approximation is below discretization error. However, because our theory assumes we are sufficiently close to

$\mathbf{D}^*$, we have assumed, for convenience, that this coarsest approximation is computed by only one discrete Newton iteration applied to a sufficiently close approximation, $\mathbf{U}_0 \in H^{h_0}$. Now, the resulting iterate, $\mathbf{U}_1$, is then interpolated to the next finer level where it is used as an initial guess for one discrete Newton step. The resulting approximation, $\mathbf{U}_2$, on subspace H^{h_1} is then used as an initial guess for the next finer level. In general, the initial guess for Newton on level h_n comes from the final Newton step on level h_{n-1}: $\mathbf{U}_n$. The process is repeated until the finest subspace is reached, where one final Newton step is then applied. Note that $\mathbf{U}_{n+1}$ can be interpreted as a discrete approximation to the result, $\widetilde{\mathbf{D}}_{n+1}$, of one infinite-dimensional Newton step applied to $\mathbf{U}_n$. This NI procedure is illustrated in figure 3.2, where $\stackrel{Nh_n}{\rightarrow}$ indicates one Newton step on H^{h_n} and $\stackrel{N}{\searrow}$ indicates one infinite-dimensional Newton step applied to discrete initial guess $\mathbf{U}_n$.

$$\mathbf{U}_0 \stackrel{Nh_0}{\rightarrow} \mathbf{U}_1 \stackrel{Nh_1}{\rightarrow} \mathbf{U}_2 \stackrel{Nh_2}{\rightarrow} \mathbf{U}_3 \stackrel{Nh_3}{\rightarrow} \mathbf{U}_4 \cdots$$
$$\stackrel{N}{\searrow} \quad \stackrel{N}{\searrow} \quad \stackrel{N}{\searrow} \quad \stackrel{N}{\searrow} \quad \ddots$$
$$\widetilde{\mathbf{D}}_1 \quad\quad \widetilde{\mathbf{D}}_2 \quad\quad \widetilde{\mathbf{D}}_3 \quad\quad \widetilde{\mathbf{D}}_4 \quad \cdots$$
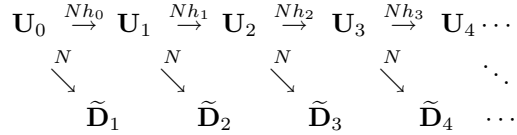
FIG. 3.2. *The NI-Newton-FOSLS algorithm*

One of our main objectives in this paper is to prove that this 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 infinite-dimensional solution.

3.2. AMG. Our theory assumes a standard V-cycle multigrid algorithm because of its superior theoretical basis. However, because of its enhanced robustness, we use AMG in practice as the matrix solver for approximating $\mathbf{U}_{n+1}$. See [12] for basic descriptions of multigrid and AMG.

AMG starts on the coarsest level with initial guess $\mathbf{V}_0 = \mathbf{U}_0$. We apply ν_0 cycles of AMG to the matrix problem arising from (3.3) with $n = 0$. 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 figure 3.3, we illustrate the NI-Newton-FOSLS-AMG algorithm, with $\stackrel{\mathcal{M}h_n}{\rightarrow}$ denoting one approximate Multigrid-Newton step on H^{h_n} and $\stackrel{Nh_n}{\searrow}$ denoting the exact discrete Newton step with initial guess $\mathbf{V}_n$ (with result $\widetilde{\mathbf{U}}_{n+1}$).

$$\mathbf{V}_0 \stackrel{\mathcal{M}h_0}{\rightarrow} \mathbf{V}_1 \stackrel{\mathcal{M}h_1}{\rightarrow} \mathbf{V}_2 \stackrel{\mathcal{M}h_2}{\rightarrow} \mathbf{V}_3 \stackrel{\mathcal{M}h_3}{\rightarrow} \mathbf{V}_4 \cdots$$
$$\stackrel{Nh_0}{\searrow} \quad \stackrel{Nh_1}{\searrow} \quad \stackrel{Nh_2}{\searrow} \quad \stackrel{Nh_3}{\searrow} \quad \ddots$$
$$\widetilde{\mathbf{U}}_1 \quad\quad \widetilde{\mathbf{U}}_2 \quad\quad \widetilde{\mathbf{U}}_3 \quad\quad \widetilde{\mathbf{U}}_4 \quad \cdots$$
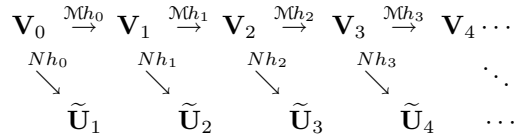
FIG. 3.3. *The NI-Newton-FOSLS-AMG algorithm*

4. Abstract Theory. Bounds on various quantities used in the theory developed here involve many different constants. To avoid proliferation, we use upper case

C to denote a generic constant that, unless otherwise specified, can change meaning with each occurrence. When it is important to track the origin of these constants, we instead use lower case c with unique subscripts. In every occurrence, these constants are independent of h and n, but they may depend on the value of the $H^{1+\delta}(\Omega)^q$ norm of the approximation. (Here, and in what follows, $\delta \in (0,1)$ is a fixed constant.) To make sure that these values are properly bounded, we start with an initial guess in a small $H^{1+\delta}(\Omega)^q$ ball about $\mathbf{D}^*$. We show that the approximations remain in this ball and, in fact, attain order h accuracy in the $H^1(\Omega)^q$ norm. This also controls the values of the $H^{1+\delta}(\Omega)^q$ norm (see Lemma 4.5 below).

Assume that there exists a solution, $\mathbf{D}^*$, of (2.3) in $\mathcal{H}_{2+\delta}$. (Recall that $\mathcal{H}_\nu$ is defined in (2.5) for any $\nu > 0$.) Denote the open $H^{1+\delta}(\Omega)^q$ ball centered at $\mathbf{D}^*$ of radius $r > 0$ by $B_r \equiv \{\mathbf{D} \in \mathcal{H}_{1+\delta} : \|\mathbf{D}^* - \mathbf{D}\|_{1+\delta,\Omega} < r\}$. Several of our norm assumptions and estimates involve both integer and fractional norms. So that our statements apply to both cases, we let $\epsilon = 0$ or δ. Assume now that $\mathbf{P}[\mathbf{D}] \in H^\epsilon(\Omega)^p$ 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

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

Further assume uniform coercivity and continuity of $\mathbf{P}'(\mathbf{D})[\,\cdot\,]$ as a mapping from $\mathcal{H}_{1+\epsilon}(\Omega)$ to $H^\epsilon(\Omega)^p$: for every $\mathbf{D} \in B_r$, there exists constants c_c and c_b, depending only on $\mathbf{D}^*, \mathbf{E}, r, \Omega$, and δ, such that

$$(4.1) \qquad \frac{1}{c_c}\|\mathbf{K}\|_{1+\epsilon,\Omega} \leq \|\mathbf{P}'(\mathbf{D})[\mathbf{K}]\|_{\epsilon,\Omega} \leq c_b\|\mathbf{K}\|_{1+\epsilon,\Omega}, \quad \forall\, \mathbf{K} \in \mathcal{H}_{1+\epsilon}(\Omega).$$

Note that coercivity implies that $\mathbf{P}'(\mathbf{D})[\,\cdot\,]$ is one-to-one on $\mathcal{H}_{1+\epsilon}(\Omega)$ for every $\mathbf{D} \in B_r$, including of course $\mathbf{D} = \mathbf{D}^*$. We also assume boundedness of the second Fréchet derivative of $\mathbf{P}(\mathbf{D})$ 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

$$(4.2) \qquad \|\mathbf{P}''(\mathbf{D})[\mathbf{K},\mathbf{K}]\|_{\epsilon,\Omega} \leq c_2\|\mathbf{K}\|_{1+\delta,\Omega}\|\mathbf{K}\|_{1+\epsilon,\Omega}, \quad \forall\, \mathbf{K} \in \mathcal{H}_{1+\epsilon}(\Omega).$$

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 $\mathbf{K}$.

Let $\mathcal{P}^h$ and $\mathcal{Q}^h$ denote the respective $H^{1+\delta}$ and H^1 projections of $\mathcal{H}_{1+\delta}$ onto H^h. Note that

$$\|\mathcal{P}^h\mathbf{D}\|_{1+\delta,\Omega} \leq \|\mathbf{D}\|_{1+\delta,\Omega}, \quad \forall\, \mathbf{D} \in \mathcal{H}_{1+\delta},$$

and

$$(4.3) \qquad \|\mathcal{Q}^h\mathbf{D}\|_{1,\Omega} \leq \|\mathbf{D}\|_{1,\Omega}, \quad \forall\, \mathbf{D} \in \mathcal{H}_{1+\delta}.$$

Assume that our finite element spaces satisfy the usual *approximation properties* (cf. [9]):

$$(4.4) \qquad \|\mathbf{D}^* - \mathcal{P}^h\mathbf{D}^*\|_{\gamma,\Omega} \leq c_d h^{2+\delta-\gamma}\|\mathbf{D}^*\|_{2+\delta,\Omega}, \quad \forall\, \gamma \in [0, 1+\delta],$$

and

$$(4.5) \qquad \|\mathbf{D}^* - \mathcal{Q}^h\mathbf{D}^*\|_{1,\Omega} \leq c_d h^{1+\delta}\|\mathbf{D}^*\|_{2+\delta,\Omega}.$$

Assume that they also satisfy the *inverse estimate* (cf. [9, 11]):

$$(4.6) \qquad \|\mathbf{U}\|_{\beta,\Omega} \leq \frac{c_i}{h^{\beta-\gamma}} \|\mathbf{U}\|_{\gamma,\Omega}, \quad \forall \, \mathbf{U} \in H^h, \beta \in [0, 1+\delta], \gamma \in [0, \beta].$$

Assume finally that h_0 is so small that $B_r \cap H^{h_0} \neq \phi$ and that initial guess $\mathbf{U}_0$ is in $B_r \cap H^{h_0}$.

The following theory shows that $\mathbf{U}_n$ is in an $H^1(\Omega)$ ball about $\mathbf{D}^*$ of radius $(1+\eta)c_d h_n$, where η is any predetermined positive constant and c_d is the constant in (4.4) and (4.5). For simplicity, we choose $\eta = 1$ and thus define

$$(4.7) \qquad \mathcal{S}_n = \{\mathbf{U} \in H^{h_n} : \|\mathbf{D}^* - \mathbf{U}\|_{1,\Omega} \leq 2c_d h_n\}$$

and

$$\mathcal{S} = \cup_{n=0}^m \mathcal{S}_n.$$

Lemma 4.5 shows that $\mathcal{S}$ is bounded in $H^{1+\delta}(\Omega)^q$ and, hence, compact in $H^1(\Omega)^q$.

We first state our three central theorems. Their proofs follow from a series of results developed in the next subsection.

For all three theorems, we assume that $r > 0$ is sufficiently small. For theorem 2, with r fixed, we assume further that $h_0 > 0$ is sufficiently small, especially so that $\mathcal{S}_0 \subset B_r$. For theorem 3, with r fixed, we assume that $h_0 > 0$ is possibly smaller still. We do this so that, in addition to $\mathcal{S}_0 \subset B_r$, we are sure that the exact discrete iterate, $\mathbf{U}_{n+1}$, is even closer to $\mathbf{D}^*$ than $2c_d h_n$, which in turn allows us to deduce that the multigrid approximation, $\mathbf{V}_{n+1}$, is within $2c_d h_n$ of $\mathbf{D}^*$. Finally, theorem 3 also assumes that, on each level, the discrete Newton problem is approximately solved with a sufficient but fixed number, ν_0, of multigrid V-cycles.

THEOREM 4.1 (Newton). *With $\mathbf{D}_n \in B_r$ given, let $\mathbf{D}_{n+1}$ be the exact infinite-dimensional Newton step defined by (3.2). Then $\mathbf{D}_{n+1} \in B_r$ and there exists a constant, c_q, depending only on $\mathbf{D}^*$, $\mathbf{E}$, Ω, and δ, such that*

$$(4.8) \qquad \|\mathbf{D}^* - \mathbf{D}_{n+1}\|_{1+\epsilon,\Omega} \leq c_q \|\mathbf{D}^* - \mathbf{D}_n\|_{1+\epsilon,\Omega} \|\mathbf{D}^* - \mathbf{D}_n\|_{1+\delta,\Omega}, \quad \epsilon = 0, \delta.$$

THEOREM 4.2 (discrete Newton). *Assume that $\mathbf{U}_0 \in S_0$. Then $\mathbf{U}_{n+1} \in \mathcal{S}_{n+1}$: the Newton approximation on level h_n based on initial guess $\mathbf{U}_n$ satisfies the error bound*

$$(4.9) \qquad \|\mathbf{D}^* - \mathbf{U}_{n+1}\|_{1,\Omega} \leq 2c_d h_{n+1} = c_d h_n.$$

THEOREM 4.3 (inexact discrete Newton). *Assume that $\mathbf{V}_0 \in S_0$. Then $\mathbf{V}_{n+1} \in \mathcal{S}_{n+1}$: the multigrid approximation on level h_n based on initial guess $\mathbf{V}_n$ satisfies the error bound*

$$(4.10) \qquad \|\mathbf{D}^* - \mathbf{V}_{n+1}\|_{1,\Omega} \leq 2c_d h_n.$$

4.1. Preliminaries. Although pose problem (2.3) on $\mathcal{H}_{1+\delta} \subset H^{1+\delta}(\Omega)^q$, we prove convergence in the weaker $H^1(\Omega)^q$ norm. Since $\mathcal{H}_{1+\delta}$ is not complete in the $H^1(\Omega)^q$ norm, we cannot appeal to standard Newton convergence theory. Fortunately, the result we need (theorem 4.1) is weaker.

LEMMA 4.4. *Let $\mathbf{D} \in B_r$ and $\widetilde{\mathbf{D}} = \theta\mathbf{D} + (1-\theta)\mathbf{D}^*$, with $\theta \in [0,1]$. Then*

$$\|\widetilde{\mathbf{D}}\|_{1+\delta,\Omega} \leq r + \|\mathbf{D}^*\|_{1+\delta,\Omega}.$$

Proof. The result follows directly from the triangle inequality and is thus omitted.
□

LEMMA 4.5. *Suppose that $\mathbf{U}_n \in \mathcal{S}_n$. Then*

(4.11) $$\|\mathbf{D}^* - \mathbf{U}_n\|_{1+\delta,\Omega} \leq c_\delta h_n^{1-\delta},$$

where $c_\delta = 2c_i c_d + (1+c_i)c_d h_0^\delta \|\mathbf{D}^*\|_{2+\delta,\Omega}$. Thus, $\mathcal{S} = \cup_{n=0}^m \mathcal{S}_n \subset B_r$ provided h_0 is so small that $c_\delta h_0^{1-\delta} \leq r$.

Proof. The bound follows the triangle inequality, (4.6), the triangle inequality again, (4.4), the definition of $\mathcal{S}_n$ in (4.7), and (4.4) again:

$$\begin{aligned}
\|\mathbf{D}^* - \mathbf{U}_n\|_{1+\delta,\Omega} &\leq \|\mathbf{D}^* - \mathcal{P}^{h_n}\mathbf{D}^*\|_{1+\delta,\Omega} + \|\mathcal{P}^{h_n}\mathbf{D}^* - \mathbf{U}_n\|_{1+\delta,\Omega} \\
&\leq c_d h_n \|\mathbf{D}^*\|_{2+\delta,\Omega} + \frac{c_i}{h_n^\delta}\|\mathcal{P}^{h_n}\mathbf{D}^* - \mathbf{U}_n\|_{1,\Omega} \\
&\leq c_d h_n \|\mathbf{D}^*\|_{2+\delta,\Omega} + \frac{c_i}{h_n^\delta}\left[\|\mathbf{D}^* - \mathcal{P}^{h_n}\mathbf{D}^*\|_{1,\Omega} + \|\mathbf{D}^* - \mathbf{U}_n\|_{1,\Omega}\right] \\
&\leq c_d h_n \|\mathbf{D}^*\|_{2+\delta,\Omega} + \frac{c_i}{h_n^\delta}\left[c_d h_n^{1+\delta}\|\mathbf{D}^*\|_{2+\delta,\Omega} + 2c_d h_n\right] \\
&\leq c_\delta h_n^{1-\delta}.
\end{aligned}$$

This lemma confirms max norm $O(h^{1-\delta})$ convergence. □

LEMMA 4.6. *Suppose that $\mathbf{V}_n \in \mathcal{S}_n$ for sufficiently small r. Let $\widetilde{\mathbf{U}}_{n+1}$ denote one exact discrete Newton step with initial guess $\mathbf{V}_n$ and let $\mathbf{V}_{n+1}$ denote its multigrid approximation. Then*

(4.12) $$\|\widetilde{\mathbf{U}}_{n+1} - \mathbf{V}_{n+1}\|_{1,\Omega} \leq \rho^{\nu_0}\|\widetilde{\mathbf{U}}_{n+1} - \mathbf{V}_n\|_{1,\Omega}.$$

Here, $\rho \in [0,1)$ is a bound on the multigrid convergence factor for any level n and any initial guess $\mathbf{V} \in \mathcal{S}_n$; it depends only on $\mathbf{D}^*, \mathbf{E}, r, \Omega, \delta, c_d$, and c_i.

Proof. Convergence estimate (4.12) follows from standard multigrid theory (cf. [10]) using the $H^1(\Omega)^q$ equivalence result in (4.1) with $\mathbf{D} = \mathbf{V}_n$. □

We have now established the tools that allow us to prove our central theorems.

4.2. Proofs of theorems 4.1, 4.2, and 4.3 . *Proof of theorem 4.1.* Consider a Taylor expansion for $\mathbf{P}(\mathbf{D}^*)$ about $\mathbf{D}_n$:

$$\mathbf{0} = \mathbf{P}(\mathbf{D}^*) = \mathbf{P}(\mathbf{D}_n) + \mathbf{P}'(\mathbf{D}_n)[\mathbf{D}^* - \mathbf{D}_n] + \frac{1}{2}\mathbf{P}''(\widetilde{\mathbf{D}})[\mathbf{D}^* - \mathbf{D}_n, \mathbf{D}^* - \mathbf{D}_n],$$

where $\widetilde{\mathbf{D}} = \theta\mathbf{D}_n + (1-\theta)\mathbf{D}^*$ for some $\theta \in [0,1]$ is bounded in the $H^{1+\delta}(\Omega)^q$ norm (lemma 4.4). Then the lower bound in (4.1), combining the above expansion with (3.1), and using (4.2) proves (4.8):

$$\begin{aligned}
\|\mathbf{D}^* - \mathbf{D}_{n+1}\|_{1+\epsilon,\Omega} &\leq c_c\|\mathbf{P}'(\mathbf{D}_n)[\mathbf{D}^* - \mathbf{D}_{n+1}]\|_{\epsilon,\Omega} \\
&= \frac{c_c}{2}\|\mathbf{P}''(\widetilde{\mathbf{D}})[\mathbf{D}^* - \mathbf{D}_n, \mathbf{D}^* - \mathbf{D}_n]\|_{\epsilon,\Omega} \\
(4.13)\qquad &\leq \frac{c_c c_2}{2}\|\mathbf{D}^* - \mathbf{D}_n\|_{1+\delta,\Omega}\|\mathbf{D}^* - \mathbf{D}_n\|_{1+\epsilon,\Omega}.
\end{aligned}$$

To show that $\mathbf{D}_{n+1} \in B_r$, consider (4.13) with $\epsilon = \delta$. For $\mathbf{D}_n \in B_r$, this reduces to

$$\|\mathbf{D}^* - \mathbf{D}_{n+1}\|_{1+\delta,\Omega} \leq \frac{c_c c_2}{2} r^2,$$

and we just require

(4.14)
$$r \leq \frac{2}{c_c c_2}.$$

☐

Proof of theorem 4.2. First assume that r is so small that (4.14) is satisfied. Assume also that h_0 is so small that

$$c_\delta h_0^{1-\delta} = 2c_i c_d h_0^{1-\delta} + (1 + c_i)c_d \|\mathbf{D}^*\|_{2+\delta,\Omega} h_0 \leq r.$$

Hence, by lemma 4.5 and because we assume that $\mathbf{U}_0 \in \mathcal{S}_0$, we must have $\mathbf{U}_0 \in B_r$. Suppose now that we could show that $\mathbf{U}_n \in \mathcal{S}_n$ implies that $\mathbf{U}_{n+1} \in \mathcal{S}_{n+1}$ for all $n \geq 0$. Then, since $\mathbf{U}_0 \in \mathcal{S}_0$, we would know that $\mathbf{U}_1 \in \mathcal{S}_1$, which in turn would imply that $\mathbf{U}_2 \in \mathcal{S}_2$. Continuing in this way would show that (4.9) holds for all $n \geq 0$.

To this end, assume that (4.9) holds for n replaced by $n-1$, with $\mathbf{U}_n \in \mathcal{S}_n$:

(4.15)
$$\|\mathbf{D}^* - \mathbf{U}_n\|_{1,\Omega} \leq 2c_d h_n,$$

From lemma 4.5, we see that

(4.16)
$$\|\mathbf{D}^* - \mathbf{U}_n\|_{1+\delta,\Omega} \leq c_\delta h_n^{1-\delta}.$$

We bound the left side of (4.9) by first using the triangle inequality:

(4.17)
$$\|\mathbf{D}^* - \mathbf{U}_{n+1}\|_{1,\Omega} \leq \|\mathbf{D}^* - \widetilde{\mathbf{D}}_{n+1}\|_{1,\Omega} + \|\widetilde{\mathbf{D}}_{n+1} - \mathbf{U}_{n+1}\|_{1,\Omega},$$

where $\widetilde{\mathbf{D}}_{n+1}$ is the result of infinite-dimensional Newton step (3.2) based on initial guess $\mathbf{U}_n$.

Consider the first term on the right of (4.17). By theorem 4.1 with $\epsilon = 0$, (4.11), and (4.15), we have that

(4.18)
$$\begin{aligned}\|\mathbf{D}^* - \widetilde{\mathbf{D}}_{n+1}\|_{1,\Omega} &\leq c_q \|\mathbf{D}^* - \mathbf{U}_n\|_{1,\Omega} \|\mathbf{D}^* - \mathbf{U}_n\|_{1+\delta,\Omega} \\ &\leq 2c_d c_q c_\delta h_n^{2-\delta}.\end{aligned}$$

For the second term on the right of (4.17), we now show that

(4.19)
$$\|\widetilde{\mathbf{D}}_{n+1} - \mathbf{U}_{n+1}\|_{1,\Omega} \leq c_b c_c (4c_d c_q c_\delta h_0^{1-\delta} + c_d h_0^\delta \|\mathbf{D}^*\|_{2+\delta,\Omega}) h_n.$$

To this end, let $\mathbf{f} \equiv \mathbf{P}'(\mathbf{U}_n)[\mathbf{U}_n] - \mathbf{P}(\mathbf{U}_n)$. Then discrete Newton step (3.3) becomes: given $\mathbf{U}_n \in H^{h_n}$, find $\mathbf{U}_{n+1} \in H^{h_n}$ such that

(4.20)
$$\left(\mathbf{P}'(\mathbf{U}_n)[\mathbf{K}^{h_n}], \ \mathbf{P}'(\mathbf{U}_n)[\mathbf{U}_{n+1}]\right) = \left(\mathbf{P}'(\mathbf{U}_n)[\mathbf{K}^{h_n}], \ \mathbf{f}\right), \quad \forall \, \mathbf{K}^{h_n} \in H^{h_n}.$$

Note that $\widetilde{\mathbf{D}}_{n+1} \in \mathcal{H}_{1+\delta}$, which is generally not in H^{h_n}, is defined by

(4.21)
$$\left(\mathbf{P}'(\mathbf{U}_n)[\mathbf{K}], \ \mathbf{P}'(\mathbf{U}_n)[\widetilde{\mathbf{D}}_{n+1}]\right) = (\mathbf{P}'(\mathbf{U}_n)[\mathbf{K}] \, , \ \mathbf{f}), \quad \forall \, \mathbf{K} \in \mathcal{H}_{1+\delta}.$$

Combining (4.20) and (4.21) for $\mathbf{K} = \mathbf{K}^{h_n} \in H^{h_n}$, we have that

$$\left(\mathbf{P}'(\mathbf{U}_n)[\mathbf{K}^{h_n}], \ \mathbf{P}'(\mathbf{U}_n)[\widetilde{\mathbf{D}}_{n+1} - \mathbf{U}_{n+1}] \right) = 0$$

from which follows

$$\left(\mathbf{P}'(\mathbf{U}_n)[\mathbf{K}], \ \mathbf{P}'(\mathbf{U}_n)[\widetilde{\mathbf{D}}_{n+1} - \mathbf{U}_{n+1}] \right)$$
$$= \left(\mathbf{P}'(\mathbf{U}_n)[\mathbf{K} - \mathbf{K}^{h_n}], \ \mathbf{P}'(\mathbf{U}_n)[\widetilde{\mathbf{D}}_{n+1} - \mathbf{U}_{n+1}] \right).$$

Letting $\mathbf{K} = \widetilde{\mathbf{D}}_{n+1} - \mathbf{U}_{n+1}$ and $\mathbf{K}^{h_n} = \mathcal{Q}^{h_n}(\widetilde{\mathbf{D}}_{n+1} - \mathbf{U}_{n+1}) = \mathcal{Q}^{h_n}\widetilde{\mathbf{D}}_{n+1} - \mathbf{U}_{n+1}$, so that $\mathbf{K} - \mathbf{K}^{h_n} = \widetilde{\mathbf{D}}_{n+1} - \mathcal{Q}^{h_n}\widetilde{\mathbf{D}}_{n+1}$, then yields

$$\left(\mathbf{P}'(\mathbf{U}_n)[\widetilde{\mathbf{D}}_{n+1} - \mathbf{U}_{n+1}], \ \mathbf{P}'(\mathbf{U}_n)[\widetilde{\mathbf{D}}_{n+1} - \mathbf{U}_{n+1}] \right)$$

(4.22)
$$= \left(\mathbf{P}'(\mathbf{U}_n)[\widetilde{\mathbf{D}}_{n+1} - \mathcal{Q}^{h_n}\widetilde{\mathbf{D}}_{n+1}], \ \mathbf{P}'(\mathbf{U}_n)[\widetilde{\mathbf{D}}_{n+1} - \mathbf{U}_{n+1}] \right).$$

Applying the Cauchy-Schwarz inequality to the right side of (4.22), then cancelling the term $\|\mathbf{P}'(\mathbf{U}_n)[\widetilde{\mathbf{D}}_{n+1} - \mathbf{U}_{n+1}]\|_{0,\Omega}$ that results on both sides, yields

$$\|\mathbf{P}'(\mathbf{U}_n)[\widetilde{\mathbf{D}}_{n+1} - \mathbf{U}_{n+1}]\|_{0,\Omega} \leq \|\mathbf{P}'(\mathbf{U}_n)[\widetilde{\mathbf{D}}_{n+1} - \mathcal{Q}^{h_n}\widetilde{\mathbf{D}}_{n+1}]\|_{0,\Omega}.$$

But (4.1) confirms that $\mathbf{P}'(\mathbf{U}_n)[\mathbf{M}]$ is coercive and bounded in the $H^1(\Omega)^q$ norm for $\mathbf{M} \in \mathcal{H}_{1+\delta}$, so the above bound becomes

(4.23)
$$\|\widetilde{\mathbf{D}}_{n+1} - \mathbf{U}_{n+1}\|_{1,\Omega} \leq c_b c_c \|\widetilde{\mathbf{D}}_{n+1} - \mathcal{Q}^{h_n}\widetilde{\mathbf{D}}_{n+1}\|_{1,\Omega}.$$

We now bound the right side of (4.23) using the triangle inequality, (4.3), (4.5), (4.18), and relation $h_n \leq h_0$:

$$\|\widetilde{\mathbf{D}}_{n+1} - \mathcal{Q}^{h_n}\widetilde{\mathbf{D}}_{n+1}\|_{1,\Omega} \leq \|\widetilde{\mathbf{D}}_{n+1} - \mathbf{D}^*\|_{1,\Omega} + \|\mathbf{D}^* - \mathcal{Q}^{h_n}\mathbf{D}^*\|_{1,\Omega}$$
$$+ \|\mathcal{Q}^{h_n}(\mathbf{D}^* - \widetilde{\mathbf{D}}_{n+1})\|_{1,\Omega}$$
$$\leq 2\|\widetilde{\mathbf{D}}_{n+1} - \mathbf{D}^*\|_{1,\Omega} + c_d h_n^{1+\delta}\|\mathbf{D}^*\|_{2+\delta,\Omega}$$
(4.24)
$$\leq (4c_d c_q c_\delta h_0^{1-\delta} + c_d h_0^\delta \|\mathbf{D}^*\|_{2+\delta,\Omega})h_n.$$

Bound (4.19) now follows from bounds (4.23) and (4.24).

Combining (4.17), (4.18), (4.19), and relation $h_n \leq h_0$ yields

$$\|\mathbf{D}^* - \mathbf{U}_{n+1}\|_{1,\Omega} \leq (2(1 + 2c_b c_c)c_q c_\delta h_0^{1-\delta} + c_b c_c \|\mathbf{D}^*\|_{2+\delta,\Omega} h_0^\delta)(c_d h_n).$$

Theorem 4.2 now follows by choosing h_0 perhaps smaller still so that

$$2(1 + 2c_b c_c)c_q c_\delta h_0^{1-\delta} + c_b c_c \|\mathbf{D}^*\|_{2+\delta,\Omega} h_0^\delta \leq 1.$$

☐

Proof of theorem 4.3 As in theorem 4.2, we need h_0 sufficiently small, but even smaller yet to account for the fact that we do not solve the Newton steps exactly: we need h_0 so small that the error in approximating $\widetilde{\mathbf{U}}_{n+1}$ (the exact discrete Newton step) by $\mathbf{V}_{n+1}$ keeps these iterates in S_{n+1}.

From theorem 4.2, we have that if $\mathbf{U}_n \in \mathcal{S}_n$, then $\mathbf{U}_{n+1} \in \mathcal{S}_{n+1}$. This result can be tightened by choosing a smaller value for h_0: choosing h_0 such that, say,

$$2(1 + 2c_b c_c) c_q c_\delta h_0^{1-\delta} + c_b c_c \|\mathbf{D}^*\|_{2+\delta,\Omega} h_0^\delta \le \frac{2}{3},$$

means that $\mathbf{U}_n \in \mathcal{S}_n$ implies that

$$(4.25) \qquad \|\mathbf{D}^* - \mathbf{U}_{n+1}\|_{1,\Omega} \le \frac{4}{3} c_d h_{n+1} = \frac{2}{3} c_d h_n.$$

As for theorem 4.2, to prove that (4.10) holds for n, we may assume that it holds for n replaced by $n-1$:

$$(4.26) \qquad \|\mathbf{D}^* - \mathbf{V}_n\|_{1,\Omega} \le 2 c_d h_n.$$

Letting $\widetilde{\mathbf{U}}_{n+1}$ as before denote one exact discrete Newton step with initial guess $\mathbf{V}_n$, then

$$(4.27) \qquad \|\mathbf{D}^* - \mathbf{V}_{n+1}\|_{1,\Omega} \le \|\mathbf{D}^* - \widetilde{\mathbf{U}}_{n+1}\|_{1,\Omega} + \|\widetilde{\mathbf{U}}_{n+1} - \mathbf{V}_{n+1}\|_{1,\Omega}.$$

For sufficiently small h_0, we know that $\mathbf{V}_n$ is in $\mathcal{S}_n \subset B_r$ and, with the reduced value of h_0, the first term on the right is bounded according to (4.25):

$$(4.28) \qquad \|\mathbf{D}^* - \widetilde{\mathbf{U}}_{n+1}\|_{1,\Omega} \le \frac{2}{3} c_d h_n.$$

For the second term, we use (4.12), the triangle inequality, (4.28), and (4.26):

$$
\begin{aligned}
\|\mathbf{U}_{n+1} - \mathbf{V}_{n+1}\|_{1,\Omega} &\le \rho^{\nu_0} \|\widetilde{\mathbf{U}}_{n+1} - \mathbf{V}_n\|_{1,\Omega} \\
&\le \rho^{\nu_0} \left[\|\widetilde{\mathbf{U}}_{n+1} - \mathbf{D}^*\|_{1,\Omega} + \|\mathbf{D}^* - \mathbf{V}_n\|_{1,\Omega} \right] \\
&\le \rho^{\nu_0} \left[\frac{2}{3} c_d h_n + 2 c_d h_n \right] \\
&= \frac{8}{3} c_d \rho^{\nu_0} h_n \\
(4.29) \qquad &\le \frac{1}{3} c_d h_n,
\end{aligned}
$$

where ν_0 is chosen so large that $\rho^{\nu_0} \le \frac{1}{8}$. Combining bounds (4.27), (4.28), and (4.29) then yields

$$\|\mathbf{D}^* - \mathbf{V}_{n+1}\|_{1,\Omega} \le c_d h_n,$$

which proves Theorem 4.3.
□

5. Conclusion. The general theory developed here applies to virtually any set of quasilinear partial differential equations that can be reformulated as a first-order system, provided it is amenable to an H^1-elliptic least-squares principle. The approach uses nested iteration based on one Newton step per level implemented using a fixed number of multigrid V-cycles. The theory shows that, for a sufficiently fine coarsest grid, the method produces a final approximation to the solution of the first-order system that is H^1 accurate to the level of discretization error. Use of this general theory is illustrated in the companion paper [17] by applying it to a first-order system for the elliptic grid generation equations. The companion paper also reports on numerical experiments that support the theory.

REFERENCES

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

[2] E.L. Allgower and Böhmer. Application of the mesh independence principle to mesh refinement strategies. *SIAM J. Numer. Anal.*, 24(6):1335–1351, 1987.

[3] E.L. Allgower, K. Böhmer, F.A. Portra, and W.C. Rheinboldt. A mesh-independence principle for operator equations and their discretizations. *SIAM J. of Numer. Anal.*, 23(1):160–169, 1986.

[4] E.L. Allgower, S. McCormick, and D. Pryor. A general mesh independence principle for Newton's method applied to second order boundary value problems. *Computing*, 23:223–246, 1979.

[5] R.E. Bank and D.J. Rose. Global approximate Newton methods. *Numer. Math.*, 37:279–295, 1981.

[6] R.E. Bank and D.J. Rose. Analysis of a multilevel iterative method for nonlinear finite element equations. *Math. Comp., 39 (1982), 435-465.*, 39:435–465, 1982.

[7] P. Bochev, Z. Cai, T.A. Manteuffel, and S.F. McCormick. Analysis of velocity-flux least-squares principles for the Navier-Stokes equations: Part I. *SIAM J. of Numer. Anal.*, 34:425–454, 1997.

[8] P. Bochev, T.A. Manteuffel, and S.F. McCormick. Analysis of velocity-flux least-squares principles for the Navier-Stokes equations: Part II. *SIAM J. of Numer. Anal.*, 34:425–454, 1997.

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

[10] J.H. Bramble. *Multigrid Methods*, volume 294 of *Pitman Research Notes in Mathematics Series*. Longman Scientific and Technical, Harlow, Essex, England, 1993.

[11] 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.

[12] W.L. Briggs, V.E. Henson, and S.F. McCormick. *A Multigrid Tutorial, 2nd ed.* SIAM, Philadelphia, 2000.

[13] Z. Cai, R. Lazarov, T.A. Manteuffel, and S.F. McCormick. First-order system least squares for second-order partial differential equations: Part I. *SIAM J. of Numer. Anal.*, 30:921–935, 1993.

[14] Z. Cai, T.A. Manteuffel, and S.F McCormick. First-order system least squares for the Stokes equations, with application to linear elasticity. *ETNA*, 3:150–159, 1995.

[15] 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.

[16] 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.

[17] A.L. Codd, T. Manteuffel, S. McCormick, and W. Ruge. Multilevel first-order system least squares for elliptic grid generation: Ellipticity and computational results.

[18] W. Layton. A two-level discretization method for the Navier-Stokes equations. *Comput. Math. Appl.* 26:33–38, 1993.

[19] W. Layton and H. W. J. Lenferink. A multilevel mesh independence principle for the Navier-Stokes equations. *SIAM J. Numer. Anal.* 33:17–30, 1996.

[20] S. McCormick. A mesh refinement method for $Ax = \lambda Bx$. *Math. Comp.*, 36(154):485–498, 1981.

[21] S. F. McCormick. A revised mesh refinement strategy for Newton's method applied to nonlinear two-point boundary value problems. In *Numerical Treatment of Differential Equations in Applications (Proc. Meeting, Math. Res. Center, Oberwolfach, 1977)*, pages 15–23. Springer, Berlin, 1978.

[22] G Timmermann. A cascadic multigrid algorithm for semilinear elliptic problems. *Numer. Math.*, 86:717–731, 2000.