LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-206774

Mesh independent convergence of the modified inexact Newton method for a second order nonlinear problem

T. Kim, J. E. Pasciak , P. S. Vassilevski

September 24, 2004

Numerical Linear Algebra with Applications

MESH INDEPENDENT CONVERGENCE OF THE MODIFIED INEXACT NEWTON METHOD FOR A SECOND ORDER NONLINEAR PROBLEM

TAEJONG KIM, JOSEPH E. PASCIAK, AND PANAYOT VASSILEVSKI

Abstract. In this paper, we consider an inexact Newton method applied to a second order nonlinear problem with higher order nonlinearities. We provide conditions under which the method has a mesh–independent rate of convergence. To do this, we are required to first, set up the problem on a scale of Hilbert spaces and second, to devise a special iterative technique which converges in a higher than first order Sobolev norm. We show that the linear (Jacobian) system solved in Newton's method can be replaced with one iterative step provided that the initial nonlinear iterate is accurate enough. The closeness criteria can be taken independent of the mesh size. Finally, the results of numerical experiments are given to support the theory.

1. Introduction

The purpose of this paper is to provide convergence estimates for inexact Newton methods applied to nonlinear second order problems where the nonlinearity appears in the coefficient of the highest order derivatives. Specifically, we consider the model problem:

$$\begin{aligned} -\mathrm{div}(k(u,x)\boldsymbol{\nabla} u) + \boldsymbol{c}(u,x)\cdot\boldsymbol{\nabla} u + b(u,x)u = f,\ x\in\Omega \\ u(x) = 0,\ x\in\partial\Omega. \end{aligned} \tag{1.1}$$

Here k, b, and $\boldsymbol{c}$ are smooth functions of u and x on $\bar{\Omega}$ which is a bounded polygonal domain in $\mathbb{R}^2$. In addition, k is bounded away from zero. For brevity, in the analysis to follow we will assume $b = 0$ and $\boldsymbol{c}$ to be independent of u. It will be clear that everything we do carries over to the more general form of the problem given in (1.1).

Newton's method is very attractive since its convergence is fast if the initial guess is sufficiently close. However, often the solution of the linear Jacobian system becomes too expensive when the number of unknowns is large. Inexact Newton or Newton Krylov methods instead replace the Jacobian solve by a fixed number of steps in a preconditioned iterative procedure.

Estimates which give rise to a uniform rate of iterative convergence for inexact Newton's method applied to (1.1) in the case when the nonlinearities are restricted to the zeroth order term were provided in [4] (specifically, $k(u,x)\equiv k(x)$ and $\boldsymbol{c}\equiv 0$). This

Date: September 17, 2004.

1991 *Mathematics Subject Classification.* 65N30, 65F10.

Key words and phrases. Inexact Newton Method, mesh independent convergence, preconditioning, second order nonlinear elliptic PDE, finite element approximation.

This material is based upon work supported by the National Science Foundation under grant No. 0311902. This work was also performed under the auspices of the U.S. Department of Energy by the University of California Lawrence Livermore National Laboratory: contract/grant number: W-7405-Eng-48.

restriction enabled them to use the $L^2(\Omega)$-norm for the residuals. Iterative methods with residual convergence in $L^2(\Omega)$ can be constructed in the case of full elliptic regularity by multilevel methods as discussed in [4]. Unfortunately, when the coefficients of the higher derivatives involve the discrete solution, full elliptic regularity no longer holds.

We provide a general theorem for the analysis of inexact Newton methods which is a variant of those given in, e.g., [4, 6, 7]. The application of this theorem requires an iterative scheme which reduces the error in a norm which is related to the stability properties of the partial differential equation. The natural norms which have been used in the stability analysis of nonlinear partial differential equations (PDE's) typically involve the Sobolev space $W^1_p(\Omega)$ for $p > 2$ and its dual (see, e.g., [3, 5]). When this is put into our inexact Newton framework, one requires an iterative scheme which reduces the error in $W^1_p(\Omega)$. Unfortunately, to the best of our knowledge, there do not exist efficient fixed step iterative techniques which guarantee a reduction in this norm. All of the popular techniques, e.g., multigrid and domain decomposition, for analyzing the iterative convergence for the discrete systems resulting from approximations to PDE's are based in Hilbert space and give rise to reductions in the corresponding energy norm.

To deal with this problem, we analyze the PDE and inexact Newton's method in the scale of Sobolev norms $H^{1+\alpha}(\Omega)$, for $0 < \alpha < 1/2$. We restrict to two dimensional problems and so this norm coerces the norm in $L^\infty(\Omega)$. To apply our abstract convergence results for the inexact Newton method, we are required to analyze both the continuous and discrete nonlinear problems in these norms. We then get that a uniform convergence rate for the inexact Newton method will be achieved provided that one uses an iterative procedure which is a reduction operator in a discrete norm equivalent to the norm in $H^{1+\alpha}(\Omega)$. In Section 6, we develop such an iterative method based on the work in [1].

The paper is organized as follows. In Section 2, we define the inexact Newton method and give an abstract theorem for its convergence analysis. In Section 3, we analyze the second order nonlinear model problem in our Hilbert space setting and show existence and uniqueness of the continuous solution in $H^{1+\alpha}(\Omega) \cap H^1_0(\Omega)$. In Section 4, we extend the Hilbert space analysis to the discrete problem and subsequently show the analogous existence and uniqueness properties of the discrete solution. An iterative method for the linearized system which provides a uniform error reduction in a norm equivalent to $H^{1+\alpha}(\Omega)$ is given in Section 6. Finally, numerical results are given in Section 7.

2. The inexact Newton method

In this section, we define the inexact Newton method and give an abstract theorem which can be used to analyze its convergence. Our abstract nonlinear problem is defined in terms of two Banach spaces $\mathcal{V}$ and $\mathcal{W}$ (with norms $\|\cdot\|_{\mathcal{V}}$ and $\|\cdot\|_{\mathcal{W}}$). For a bounded linear operator B from $\mathcal{V}$ into $\mathcal{W}$, let $\|\cdot\|_{[\mathcal{V},\mathcal{W}]}$ denote the operator norm,

$$\|B\|_{[\mathcal{V},\mathcal{W}]} = \sup_{v\in\mathcal{V}} \frac{\|Bv\|_{\mathcal{W}}}{\|v\|_{\mathcal{V}}}.$$

Let $\mathcal{F}$ be a continuous function mapping $\mathcal{V}$ into $\mathcal{W}'$. We consider the problem: Find $u \in \mathcal{V}$ satisfying

$$\mathcal{F}(u) = 0. \tag{2.1}$$

Let $\mathcal{F}'(u)$ denote the Fréchet derivative of $\mathcal{F}$ at u.

The inexact Newton algorithm which we shall consider in this paper is given as follows.

Algorithm 2.1 (Modified Inexact Newton Method)**.** Given an initial iterate $u^0 \in \mathcal{V}$, consider the sequence of iterates $\{u^k\} \subset \mathcal{V}$ generated by

$$u^{k+1} = u^k + \hat{s}^k$$

where $\hat{s}^k$ approximately solves

$$\mathcal{F}'(u^0)s^k = -\mathcal{F}(u^k). \tag{2.2}$$

Specifically, we assume that $\hat{s}^k$ satisfies

$$|||s^k - \hat{s}^k||| \leq \beta \, ||| \, s^k|||$$

for some fixed β in $[0, 1)$. Here $|||\cdot|||$ is a norm on $\mathcal{V}$ which is equivalent to $\|\cdot\|_{\mathcal{V}}$.

Let γ_1 and γ_2 be the constants in the norm equivalence relations between $|||\cdot|||$ and $\|\cdot\|_{\mathcal{V}}$, that is,

$$\gamma_1 \ |||\, v||| \leq \|v\|_{\mathcal{V}} \leq \gamma_2 \ |||\, v \,|||\,. \tag{2.4}$$

Minor variations of this algorithm have been proposed and studied, for example, in [4, 6, 7, 8]. Our analysis is also a slight modification of theirs. We consider the following hypotheses:

(A.1) $\mathcal{F}(u) = 0$ has a solution u^* in $\mathcal{V}$.
(A.2) For given $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ and $M = M(\varepsilon) > 0$ such that the Fréchet derivative $\mathcal{F}'(u)$ exists for $\|u - u^*\|_{\mathcal{V}} < \delta$ and satisfies:
  (A.2.1) $\|\mathcal{F}(u) - \mathcal{F}(u^*) - \mathcal{F}'(u^*)(u - u^*)\|_{\mathcal{W}'} \leq \varepsilon \|u - u^*\|_{\mathcal{V}}$
  (A.2.2) $\|\mathcal{F}'(u) - \mathcal{F}'(u^*)\|_{[\mathcal{V},\mathcal{W}']} \leq \varepsilon$.
  (A.2.3) $\mathcal{F}'(u)^{-1}$ exists and satisfies $\|\mathcal{F}'(u)^{-1}\|_{[\mathcal{W}',\mathcal{V}]} \leq M$.

The above conditions imply the following theorem (compare to Theorem 2.1 in [4] or Theorem 2.3 in [6]). We give its proof for completeness.

Theorem 2.2. *Assume that (A.1) and (A.2) hold. Let $t \in (\beta, 1)$ be given where β satisfies (2.3). Then there exists $\delta > 0$ such that if $\|u^0 - u^*\|_{\mathcal{V}} < \delta$, then the sequence of iterates $\{u^k\}$ generated by Algorithm 2.1 converges to u^*, in fact,*

$$|||u^{k+1} - u^*||| \leq t \, ||| \, u^k - u^* \, |||\,. \tag{2.5}$$

Remark 2.3. The results of our theorem differ from those of [4] in that we only require that the initial iterate is close in a natural norm ($\|\cdot\|_{\mathcal{V}}$) which, in our subsequent applications, is independent of the mesh size. Indeed, although Theorem 5.2 of [4] gives a convergence rate independent of the mesh size, the initial iterate has to be close in a mesh dependent norm. The problem of finding such a starting iterate appears to require a fairly accurate initial approximation of the nonlinear problem.

Remark 2.4. Note that the algorithm requires an iterative scheme which gives rise to a reduction in a norm $|||\cdot|||$ which is equivalent to $\|\cdot\|_{\mathcal{V}}$. In general, it is not possible to construct fixed step iterative methods which are convergent in arbitrary norms. Indeed, in almost all of the literature on iterative convergence, convergence reductions are achieved in the L^2 or energy norms. We shall further address this issue in Section 6.

Remark 2.5. The above theorem can be used to guarantee convergence rates independent of the mesh size in PDE applications provided that the functions $\delta(\varepsilon)$, $M(\varepsilon)$ and bounds for the constants of norm equivalence between $|||\cdot|||$ and $\|\cdot\|_{\mathcal{V}}$ can all be chosen independently of h.

Proof of Theorem 2.2. It clearly suffices to verify (2.5). We start by observing that

$$\begin{aligned} s^k = &-(\mathcal{F}'(u^0))^{-1}\mathcal{F}'(u^*)(u^k-u^*) \\ &-(\mathcal{F}'(u^0))^{-1}(\mathcal{F}(u^k)-\mathcal{F}(u^*)-\mathcal{F}'(u^*)(u^k-u^*)). \end{aligned} \tag{2.6}$$

Thus,

$$\begin{aligned} |||u^{k+1}-u^*||| = |||&(I-(\mathcal{F}'(u^0))^{-1}\mathcal{F}'(u^*))(u^k-u^*)+\hat{s}^k-s^k \\ &-(\mathcal{F}'(u^0))^{-1}(\mathcal{F}(u^k)-\mathcal{F}(u^*)-\mathcal{F}'(u^*)(u^k-u^*))\,|||\,. \end{aligned}$$

Let $e_k = u^k - u^*$ and δ be such that (A.2) holds for a positive ε to be determined later. Then,

$$\begin{aligned} |||(I-(\mathcal{F}'(u^0))^{-1}\mathcal{F}'(u^*))e_k||| &\le \gamma_1^{-1}\,\|(\mathcal{F}'(u^0))^{-1}\|_{[\mathcal{W}',\mathcal{V}]}\|\mathcal{F}'(u^0)-\mathcal{F}'(u^*)\|_{[\mathcal{V},\mathcal{W}']}\|e_k\|_{\mathcal{V}} \\ &\le \frac{\gamma_2}{\gamma_1}M\,\varepsilon\,|||\,e_k\,|||\,. \end{aligned}$$

Similarly,

$$|||(\mathcal{F}'(u^0))^{-1}(\mathcal{F}(u^k)-\mathcal{F}(u^*)-\mathcal{F}'(u^*)e_k)||| \le \frac{\gamma_2}{\gamma_1}M\varepsilon\;\;|||\,e_k\,|||\,.$$

Using (2.3) and (2.6) gives

$$\begin{aligned} |||\hat{s}^k-s^k||| &\le \beta\,|||\,s^k||| \\ &\le \beta\{|||(\mathcal{F}'(u^0))^{-1}(\mathcal{F}(u^k)-\mathcal{F}(u^*)-\mathcal{F}'(u^*)e_k)||| \\ &\qquad +|||(\mathcal{F}'(u^0))^{-1}\mathcal{F}'(u^*)e_k|||\} \\ &\le \frac{\gamma_2}{\gamma_1}M\beta\varepsilon\,|||\,e_k\,|||+\beta\,|||\,(\mathcal{F}'(u^0))^{-1}\mathcal{F}'(u^*)e_k\,|||\,. \end{aligned}$$

Finally,

$$\begin{aligned} \beta\,|||\,(\mathcal{F}'(u^0))^{-1}\mathcal{F}'(u^*)e_k||| &\le \beta\{|||e_k\,|||+|||\,(I-(\mathcal{F}'(u^0))^{-1}\mathcal{F}'(u^*))e_k|||\} \\ &\le (\beta+\frac{\gamma_2}{\gamma_1}M\beta\varepsilon)\,|||\,e_k\,|||\,. \end{aligned}$$

Combining the above inequalities gives

$$|||u^{k+1}-u^*||| \le \left(\beta+\frac{\gamma_2}{\gamma_1}M(3+\beta)\,\varepsilon\right)|||\,u^k-u^*|||$$

and the theorem follows taking $\varepsilon \le \frac{\gamma_1}{\gamma_2 M(3+\beta)}(t-\beta)$. □

Remark 2.6. It is interesting to note that the continuity constants associated to $\mathcal{F}$ and $\mathcal{F}'$ do not come into the proof. This allows Brown *et. al.* [4] to analyze a discrete problem using the L^2 norm on the discrete space $\mathcal{W}$ even though the discrete Frechét derivatives are not uniformly bounded into this space. The conditions (A.2.1) and (A.2.2) nevertheless hold because the problem considered there only involve linear higher order terms. This fails for our more general application so we are forced to use weaker (negative norm) spaces.

3. Hilbert space setting of (1.1).

We shall ultimately apply Theorem 2.2 to finite element approximations of (1.1). Because of the higher order nonlinearity, (A.2.1) and (A.2.2) cannot hold unless functions in $\mathcal{W}$ have two less Sobolev derivatives than those in $\mathcal{V}$. It is common to use the spaces $\mathcal{V} = W_p^1(\Omega)$ for $p > 2$ (see [3, 5]) for the finite element convergence analysis of (1.1). Under certain hypotheses on the nonlinearities, it is possible to prove (A.1)-(A.2) using these spaces. To the best of our knowledge, there do not exist efficient fixed step iterative methods which are convergent in any norm which is equivalent (independently of the discretization parameter) to the norm in $W_p^1(\Omega)$. To get around this issue, we shall analyze our discrete problem in the scale of Sobolev norms $H^{1+\alpha}(\Omega)$ for $0 < \alpha < 1/2$. To do this, we start with the analysis of the continuous problem. Most of this analysis will involve proving the inequalities of (A.2) on the continuous level.

We first introduce some notation and definitions. Let $\|\cdot\|_s$ denote the norm in $H^s(\Omega)$ and $|\cdot|_s$ denote the corresponding semi-norm. For negative s, the Sobolev space $H^s(\Omega)$ is the set of linear functionals with finite norm,

$$\|u\|_s = \sup_{\phi \in C_0^\infty(\Omega)} \frac{< u, \phi >}{\|\phi\|_{-s}},$$

where $< \cdot, \cdot >$ is the duality pairing. For $s \in (1/2, 1)$, we let $H_0^s(\Omega)$ denote the interpolation space on the scale between $H_0^1(\Omega)$ and $L^2(\Omega)$. We shall also use $(\cdot, \cdot)$ to denote the inner product in both $L^2(\Omega)$ and $(L^2(\Omega))^2$.

To set up the problem, we let φ be in $C_0^\infty(\Omega)$. Then we have

$$(k(u,x)\boldsymbol{\nabla} u, \boldsymbol{\nabla}\varphi) + (\boldsymbol{c}\cdot\boldsymbol{\nabla} u, \varphi) = (f, \varphi)$$

and consider

$$< F(u), \varphi >= (k(u,x)\boldsymbol{\nabla} u, \boldsymbol{\nabla}\varphi) + (\boldsymbol{c}\cdot\boldsymbol{\nabla} u, \varphi) - (f, \varphi). \tag{3.1}$$

To keep the notation from becoming too cumbersome, we have dropped the explicit dependence of u on x above.

We shall use the notation $\boldsymbol{\nabla} k$ to denote the gradient with respect to the x variable considering u independently of x. We shall assume that the quantities

$$k,\ \boldsymbol{\nabla} k,\ \frac{\partial k}{\partial u},\ \frac{\boldsymbol{\nabla}\partial k}{\partial u},\ \frac{\partial^2 k}{\partial^2 u},\ \frac{\boldsymbol{\nabla}\partial^2 k}{\partial^2 u},\ \frac{\boldsymbol{\nabla}\partial k}{\partial x_i} \text{ and } \frac{\partial^3 k}{\partial^3 u}$$

are all uniformly bounded independently of $u \in V$.

Definition 3.1. For a given $\alpha \in (0, \frac{1}{2})$, the spaces of our main interest are: $V = H^{1+\alpha}(\Omega) \cap H_0^1(\Omega)$, $W = H_0^{1-\alpha}(\Omega)$ and its dual $W' = H^{\alpha-1}(\Omega)$.

Note that we have denoted the spaces V and W in contrast to the $\mathcal{V}$ and $\mathcal{W}$ that will be their discrete counterparts, i.e., $\mathcal{V} \subset V$ and $\mathcal{W} \subset W$. The latter pair $(\mathcal{V},\ \mathcal{W})$ takes part in the actual inexact Newton iteration used, in practice, to compute the discrete solution.

3.1. The F mapping.

Lemma 3.2. *Suppose that f is in W', then $F(u)$ given by (3.1) is a well defined map of V into W'.*

To prove the above lemma and others which will be critical to our analysis, we shall use the following lemma which we shall prove at the end of this section.

Lemma 3.3. *For $p > 2$, $g \in W_p^1(\Omega)$, $u \in H^\alpha(\Omega)$ and $v \in W$,*

$$\int_\Omega guv_{x_i}\,dx \le C(\Omega,\alpha,p)\|g\|_{W_p^1(\Omega)}\|u\|_\alpha\|v\|_{1-\alpha},$$

where $v_{x_i} \in H^{-\alpha}(\Omega)$ denotes the partial derivative of v with respect to x_i.

In what follows, we shall use C and c, with or without subscript to denote a generic positive constant. These constants will implicitly depend on the domain, α, p and the assumptions on the coefficients k, $\boldsymbol{c}$ and b. They may depend on additional quantities if explicitly mentioned.

We will also fix p in the interval $(2, 2/(1-\alpha)]$ so that the following two Sobolev inequalities hold:

$$\|w\|_{L^\infty(\Omega)} \le C\|w\|_{W_p^1(\Omega)}, \quad \text{for all } w \in W_p^1(\Omega) \tag{3.2}$$

and

$$\|w\|_{W_p^1(\Omega)} \le C\|w\|_{1+\alpha}, \quad \text{for all } w \in H^{1+\alpha}(\Omega). \tag{3.3}$$

We shall also use the Sobolev inequality

$$\|w\|_{L^q(\Omega)} \le C\|w\|_{1-\alpha}, \quad \text{for all } w \in H^{1-\alpha}(\Omega) \tag{3.4}$$

which holds provided that $q \le 2/\alpha$.

Proof of Lemma 3.2. Let u be in V. It suffices to show that

$$< F(u), \varphi > \le C(u)\|\varphi\|_{1-\alpha} \quad \text{for all } \varphi \in W.$$

Applying Lemma 3.3 gives

$$\begin{aligned} < F(u), \varphi > &= \int_\Omega k(u,x)(\boldsymbol{\nabla} u \cdot \boldsymbol{\nabla}\varphi)\,dx + \int_\Omega (\boldsymbol{c}\cdot\boldsymbol{\nabla} u)\varphi\,dx - \int_\Omega f\varphi\,dx \\ &\le C\{\|k(u,x)\|_{W_p^1(\Omega)}\|u\|_{1+\alpha} + \|u\|_1 + \|f\|_{-1+\alpha}\}\|\varphi\|_{1-\alpha}. \end{aligned}$$

We note that in the distributional sense,

$$\frac{\partial k(u,x)}{\partial x_i} = k_u(u,x)u_{x_i} + k_{x_i}(u,x) \tag{3.5}$$

where the second term obviously denotes differentiation of k with respect to x_i (independent of the u dependence on x). Indeed, (3.5) holds for smooth u and easily follows for general $u \in V$ from the density of smooth functions in V. Thus,

$$\|k(u,x)\|_{W_p^1(\Omega)} \le C(\|u\|_{W_p^1(\Omega)} + 1) \le C(\|u\|_{1+\alpha} + 1). \tag{3.6}$$

This completes the proof of the lemma. □

Remark 3.4. From the above discussion, it is clear that $u \in V$ satisfying $F(u) = 0$ provides a weak solution to (1.1). The existence and uniqueness of solutions to nonlinear problems is always a delicate issue. In [5], existence and uniqueness of a weak solution to (1.1) in $W_p^1(\Omega)$ was verified for $p > 2$ in the case when k only depends on u. In this case, our theory will also give a unique solution $u \in V$ (which coincides with that of [5]).

3.2. **The F' mapping.** To study the local behavior of F, we need to introduce the Fréchet derivative $F'(u)$ (a linear map from V to W'). As we show in the next proposition, its definition is given by

$$
\begin{aligned}
< F'(u)w, \varphi >= & \int_\Omega \frac{\partial k(u,x)}{\partial u} w(\boldsymbol{\nabla} u \cdot \boldsymbol{\nabla}\varphi)\, dx \\
& + \int_\Omega k(u,x)(\boldsymbol{\nabla} w \cdot \boldsymbol{\nabla}\varphi)\, dx + \int_\Omega (\boldsymbol{c}\cdot\boldsymbol{\nabla} w)\varphi\, dx,
\end{aligned} \tag{3.7}
$$

for all $w \in V$ and $\varphi \in W$. The next proposition also proves (A.2.1) in the continuous case.

Proposition 3.5. *For $u \in V$, the Fréchet derivative of F at u is given by (3.7). Moreover, for a given $\delta > 0$, there exists a constant $C = C(\|u\|_{1+\alpha}, \delta)$ such that*

$$
\|F(v) - F(u) - F'(u)(v-u)\|_{-1+\alpha} \le C\|v-u\|^2_{1+\alpha} \tag{3.8}
$$

for all v in the ball $B(u,\delta) \equiv \{v \in V \ : \ \|u-v\|_{1+\alpha} < \delta\}$.

Proof. Using the assumptions on k and a similar argument as used in the proof of Lemma 3.2, it follows that for $u \in V$, $F'(u)$ given by (3.7) is a linear map of V into W'.

To finish the proof, it suffices to verify (3.8). This is equivalent to showing that for all $\phi \in W$,

$$
< F(v) - F(u) - F'(u)(v-u), \phi > \le C\|v-u\|^2_{1+\alpha}\|\phi\|_{1-\alpha}.
$$

A simple computation gives

$$
\begin{aligned}
& < F(v) - F(u) - F'(u)(v-u), \phi > \\
& \qquad = \int_\Omega \frac{\partial k(u,x)}{\partial u}(v-u)((\boldsymbol{\nabla} v - \boldsymbol{\nabla} u)\cdot\boldsymbol{\nabla}\phi)\, dx \\
& \qquad\quad + \int_\Omega \left(k(v,x) - k(u,x) - \frac{\partial k(u,x)}{\partial u}(v-u)\right)(\boldsymbol{\nabla} v \cdot \boldsymbol{\nabla}\phi)\, dx.
\end{aligned} \tag{3.9}
$$

By Lemma 3.3, the first integral of the right hand side of (3.9) is bounded by

$$
C\left\| \frac{\partial k(u,x)}{\partial u}(v-u)\right\|_{W^1_p(\Omega)} \|v-u\|_{1+\alpha}\|\phi\|_{1-\alpha}.
$$

Using techniques similar to those used in the proof of Lemma 3.2 gives

$$
\left\| \frac{\partial k(u,x)}{\partial u}(v-u)\right\|_{W^1_p(\Omega)} \le C\{\|u\|_{W^1_p(\Omega)} + 1\}\|v-u\|_{W^1_p(\Omega)} \le C\|v-u\|_{1+\alpha}.
$$

We next bound the second integral of the right hand side of (3.9). By Lemma 3.3, it suffices to show that

$$
\begin{aligned}
& \left\| k(v(x),x) - k(u(x),x) - \frac{\partial k(u,x)}{\partial u}(v(x)-u(x))\right\|_{W^1_p(\Omega)} \\
& \qquad = \left\| \int_u^v \frac{\partial^2 k(s,x)}{\partial s^2}(v(x)-s)\, ds\right\|_{W^1_p(\Omega)} \le C\|v-u\|^2_{1+\alpha}.
\end{aligned} \tag{3.10}
$$

We obviously have

$$\left\| \int_u^v \frac{\partial^2 k(s,x)}{\partial s^2}(v(x)-s)\,ds \right\|_{L^p(\Omega)} \le C\|v-u\|^2_{L^\infty(\Omega)} \le C\|v-u\|^2_{1+\alpha}. \tag{3.11}$$

Finally, we will bound the semi-norm in (3.10). We have

$$\begin{aligned} \left\| \boldsymbol{\nabla} \int_u^v \frac{\partial^2 k(s,x)}{\partial s^2}(v(x)-s)\,ds \right\|_{L^p(\Omega)} \le & \left\| \int_u^v \frac{\boldsymbol{\nabla}\partial^2 k(s,x)}{\partial^2 s}(v(x)-s))\,ds \right\|_{L^p(\Omega)} \\ & + \left\| \boldsymbol{\nabla} v \left(\frac{\partial k(v,x)}{\partial v} - \frac{\partial k(u,x)}{\partial u} \right) - \boldsymbol{\nabla} u \frac{\partial^2 k(u,x)}{\partial u^2}(v-u) \right\|_{L^p(\Omega)}. \end{aligned} \tag{3.12}$$

The first term on the right hand side is bounded analogously to (3.11). For the second, we note that

$$\begin{aligned} \boldsymbol{\nabla} v(x) \left(\frac{\partial k(v,x)}{\partial v} - \frac{\partial k(u,x)}{\partial u} \right) - \boldsymbol{\nabla} u(x) \frac{\partial^2 k(v,x)}{\partial v^2}(v(x)-u(x)) & \\ = \boldsymbol{\nabla}(v-u) \int_{u(x)}^{v(x)} \frac{\partial k(s,x)}{\partial s}\,ds & \\ + \boldsymbol{\nabla} u \int_{u(x)}^{v(x)} \frac{\partial^3 k(s,x)}{\partial s^3}(v(x)-s)\,ds. & \end{aligned}$$

As above we have

$$\left\| \boldsymbol{\nabla}(v-u) \int_{u(x)}^{v(x)} \frac{\partial k(s,x)}{\partial s}\,ds \right\|_{L^p(\Omega)} \le C\|u-v\|^2_{1+\alpha}$$

and

$$\left\| \boldsymbol{\nabla} u \int_{u(x)}^{v(x)} \frac{\partial^3 k(s,x)}{\partial s^3}(v(x)-s)\,ds \right\|_{L^p(\Omega)} \le C\|u-v\|^2_{1+\alpha}.$$

This completes the proof of the proposition. □

We can also show that (A.2.2) holds on the continuous level using similar techniques.

Proposition 3.6. *Let u be in V. For a given $\delta > 0$, there exists a constant $C = C(\|u\|_{1+\alpha}, \delta)$ such that, for all v in the ball $B(u,\delta)$ in V,*

$$\|F'(v) - F'(u)\|_{[V,W']} \le C\|v-u\|_{1+\alpha}.$$

Proof. We need to show that for all w in V and φ in W,

$$< (F'(v) - F'(u))w, \varphi > \le C\|v-u\|_{1+\alpha}\|w\|_{1+\alpha}\|\varphi\|_{1-\alpha}. \tag{3.13}$$

Now

$$\begin{aligned} < (F'(v) - F'(u))w, \varphi > = & \int_\Omega \left(\frac{\partial k(v,x)}{\partial v} - \frac{\partial k(u,x)}{\partial u} \right) (\boldsymbol{\nabla} v \cdot \boldsymbol{\nabla}\varphi) w\,dx \\ & + \int_\Omega \frac{\partial k(u,x)}{\partial u} (\boldsymbol{\nabla}(v-u) \cdot \boldsymbol{\nabla}\varphi) w\,dx \\ & + \int_\Omega (k(v,x) - k(u,x))(\boldsymbol{\nabla} w \cdot \boldsymbol{\nabla}\varphi)\,dx. \end{aligned}$$

The inequality (3.13) can be derived by applying similar techniques as in the proof of Proposition 3.5 to the above identity. This completes the proof of the proposition. □

3.3. **Existence and uniqueness of solutions.** There are no results available to guarantee the existence of solutions to Problem (1.1) in the generality which we have posed it. To proceed with the analysis, we shall need to make the following additional assumptions.

(B.1) (1.1) has a solution u^* in V,
(B.2) $F'(u^*)$ satisfies the uniqueness property:

$$< F'(u^*)w, \varphi >= 0 \quad \text{for all } \varphi \in W \text{ implies } w = 0. \tag{3.14}$$

Remark 3.7. If we take $k(u(x), x) = k(u)$, $b = 0$, and $\boldsymbol{c}$ to be divergence free, then it is possible to verify the above assumptions. In this case, (3.14) follows from the proof of uniqueness in Theorem 5.1 of [5], and (B.1) follows from the analysis there.

Using (B.2), we will show that $F'(u^*)$ is an isomorphism (see Proposition 3.10 below). This fact will be used to verify the existence of a discrete solution in Section 4. To prove this isomorphism, we need the following two lemmas.

Lemma 3.8. *Let* $D(u, \varphi) = \displaystyle\int_\Omega \boldsymbol{\nabla} u \cdot \boldsymbol{\nabla} \varphi \, dx$. *If* u *is in* $H_0^1(\Omega)$ *and satisfies*

$$\sup_{\varphi \in W} \frac{D(u, \varphi)}{\|\varphi\|_{1-\alpha}} < \infty, \tag{3.15}$$

then u *is also in* V. *Furthermore (3.15) provides an equivalent norm on* V.

Proof. Let u satisfy the above conditions and define the functional f by

$$< f, \varphi >= D(u, \varphi) \quad \text{for all } \varphi \in W.$$

Clearly, u is the solution to the Dirichlet problem, $u \in H_0^1(\Omega)$ satisfying

$$D(u, \varphi) =< f, \varphi > \quad \text{for all } \varphi \in H_0^1(\Omega). \tag{3.16}$$

Elliptic regularity for (3.16) implies that $u \in H^{1+\alpha}(\Omega)$ and satisfies

$$\|u\|_{1+\alpha} \le C \|f\|_{-1+\alpha}.$$

This shows that $\|u\|_{1+\alpha}$ is bounded by a multiple of the supremum in (3.15). The bound in other direction follows from Lemma 3.3. □

Lemma 3.9. *For a fixed* $u \in V$, *the map* $w \to k(u, x)w$ *is an isomorphism from* V *onto itself.*

Proof. We first show that for $w \in V$, $k(u, x)w$ is also in V. Since both w and u are in V, estimates similar to those in (3.6) give

$$\|k(u, x)w\|_1 \le C(1 + \|u\|_{1+\alpha})\|w\|_{1+\alpha}$$

from which it immediately follows that $k(u, x)w \in H_0^1(\Omega)$. Finally by Lemma 3.3,

$$\begin{aligned} D(k(u, x)w, \varphi) &= \int_\Omega \boldsymbol{\nabla}(k(u, x)w) \cdot \boldsymbol{\nabla} \varphi \, dx \\ &= \int_\Omega \left[\frac{\partial k(u, x)}{\partial u} w(\boldsymbol{\nabla} u \cdot \boldsymbol{\nabla} \varphi) + w(\boldsymbol{\nabla} k(u, x) \cdot \boldsymbol{\nabla} \varphi) + k(u, x)(\boldsymbol{\nabla} w \cdot \boldsymbol{\nabla} \varphi) \right] dx \\ &\le C \|u\|_{1+\alpha} \|w\|_{1+\alpha} \|\varphi\|_{1-\alpha}. \end{aligned}$$

Applying Lemma 3.8 shows that $k(u,x)w$ is in V and satisfies

$$\|k(u,x)w\|_{1+\alpha} \leq C\|w\|_{1+\alpha}.$$

We note that $k(u,x)^{-1}$ satisfies the same assumptions as $k(u,x)$ so that boundedness of the inverse map follows by the same reasoning. This completes the proof of the lemma. □

The next result shows that $F'(u^*)$ is an isomorphism.

Proposition 3.10. *$F'(u^*) : V \to W'$ is an isomorphism, i.e., $F'(u^*)^{-1}$ exists and there exists a positive constant M such that $\|F'(u^*)^{-1}\|_{[W',V]} \leq M$.*

Proof. In this proof, we adapt the idea in the proof of Theorem 5.2 in [5]. Let $T : W' \to V$ be the solution operator for

$$\int_\Omega \boldsymbol{\nabla}(k(u^*,x)w) \cdot \boldsymbol{\nabla}\varphi \, dx =< f, \varphi > \quad \text{for all } \varphi \in W,$$

that is, $Tf = w$. Let $u \in V$ solve (3.16). Since the map $f \to u$ is an isomorphism from W' onto V, Lemma 3.8 implies that T is also.

We define an operator $A_1 : V \to W'$ by

$$< A_1 w, \varphi >= -\int_\Omega (\boldsymbol{\nabla} k(u^*,x) \cdot \boldsymbol{\nabla}\varphi) w \, dx + \int_\Omega (\boldsymbol{c} \cdot \boldsymbol{\nabla} w)\varphi \, dx \quad \text{for all } \varphi \in W.$$

Then, for all $\varphi \in W$,

$$\begin{aligned}\int_\Omega \boldsymbol{\nabla}(k(u^*,x)TF'(u^*)w) \cdot \boldsymbol{\nabla}\varphi \, dx &=< F'(u^*)w, \varphi > \\ &= \int_\Omega \boldsymbol{\nabla}(k(u^*,x)w) \cdot \boldsymbol{\nabla}\varphi \, dx + < A_1 w, \varphi > \\ &= \int_\Omega \boldsymbol{\nabla}(k(u^*,x)w) \cdot \boldsymbol{\nabla}\varphi \, dx + \int_\Omega \boldsymbol{\nabla}(k(u^*,x)TA_1 w) \cdot \boldsymbol{\nabla}\varphi \, dx.\end{aligned}$$

Therefore, for $w \in V$,

$$TF'(u^*)w = w + TA_1 w. \tag{3.17}$$

It suffices to show that $TF'(u^*)$ is an isomorphism of V onto V. By Lemma 3.3,

$$< A_1 w, \varphi > \leq C\{\|w\|_\alpha \|\varphi\|_{1-\alpha} + \|w\|_1 \|\varphi\|_0\} \leq C\|w\|_1 \|\varphi\|_{1-\alpha}.$$

Since $H^{1+\alpha}(\Omega)$ is compactly imbedded in $H^1(\Omega)$, A_1 is a compact operator from V into W'. Thus, TA_1 is also compact from V into V. Hence the mapping $TF'(u^*)$ is a linear Fredholm operator with index zero. Since (B.2) implies $F'(u^*)$ is injective, $TF'(u^*)$ is injective and bijective also. Since it is also continuous, it is an isomorphism. This completes the proof of the proposition. □

To be able to apply finite element duality, we shall need regularity for the adjoint problem. We consider the adjoint operator $(F'(u^*))^*$ defined by

$$< (F'(u^*))^* v, \varphi > \equiv < F'(u^*)\varphi, v > .$$

Clearly, this is well defined for $v \in W$ and $\varphi \in V$. The next proposition shows that it is also well defined for $v \in V$ and $\varphi \in W$ and gives rise to an isomorphism.

Proposition 3.11. *$(F'(u^*))^* : V \to W'$ is an isomorphism.*

Proof. By definition,

$$< (F'(u^*))^*\varphi, w >= \int_\Omega \frac{\partial k(u^*,x)}{\partial u^*} w(\boldsymbol{\nabla} u^* \cdot \boldsymbol{\nabla}\varphi)\, dx + \int_\Omega k(u^*,x)(\boldsymbol{\nabla} w \cdot \boldsymbol{\nabla}\varphi)\, dx + \int_\Omega (\boldsymbol{c}\cdot\boldsymbol{\nabla} w)\varphi\, dx,$$

Using arguments similar to those above and Lemma 3.3, it is easy to show that $(F'(u^*))^*$ is a well defined linear map of V into W'.

Since $F'(u^*)$ is an isomorphism from V onto W', $(F'(u^*))^*$ is an isomorphism from W onto V' and

$$\|((F'(u^*))^*)^{-1}\|_{[V',W]} = \|(F'(u^*))^{-1}\|_{[W',V]}.$$

Thus, by Proposition 3.10,

$$\|\varphi\|_{1-\alpha} \le M \sup_{u\in V} \frac{< F'(u^*)u, \varphi >}{\|u\|_{1+\alpha}}. \tag{3.18}$$

The above inequality implies that $(F'(u^*))^*$ is injective on V.

Define $A_2 : V \to W'$ by

$$\begin{aligned} < A_2 w, \varphi >= & \int_\Omega \frac{\partial k(u^*,x)}{\partial u^*}(\boldsymbol{\nabla} u^*\cdot\boldsymbol{\nabla} w)\varphi\, dx + \int_\Omega (\boldsymbol{c}\cdot\boldsymbol{\nabla}\varphi) w\, dx \\ & - \int_\Omega \frac{\partial k(u^*,x)}{\partial u^*}(\boldsymbol{\nabla} u^*\cdot\boldsymbol{\nabla}\varphi) w\, dx - \int_\Omega (\boldsymbol{\nabla} k(u^*,x)\cdot\boldsymbol{\nabla}\varphi) w\, dx, \end{aligned} \tag{3.19}$$

then

$$< (F'(u^*))^* w, \varphi >= \int_\Omega \boldsymbol{\nabla}(k(u^*,x)w)\cdot\boldsymbol{\nabla}\varphi\, dx + < A_2 w, \varphi > .$$

Thus, as in the proof of Proposition 3.10,

$$T(F'(u^*))^* w = w + TA_2 w$$

and it suffices to show that TA_2 is compact on V. This, in turn, will follow if we show that

$$< A_2 w, \varphi > \le C\|w\|_{1+\beta}\|\varphi\|_{1-\alpha}. \tag{3.20}$$

for some β with $0 < \beta < \alpha$.

By Lemma 3.3, the last three terms of (3.19) can be bounded by the right hand side of (3.20). For example, the third term is bounded by

$$C\left\| w\frac{\partial k(u^*,x)}{\partial u^*}\right\|_{W^1_q(\Omega)} \|u^*\|_{1+\alpha}\|\varphi\|_{1-\alpha} \le C\|w\|_{1+\beta}\|\varphi\|_{1-\alpha}$$

provided that q is taken so that

$$H^{1+\beta}(\Omega) \subset W^1_q(\Omega).$$

For the first term in (3.19), we consider $p = 1/(1-\alpha)$ and $q = 1/\alpha$. Then

$$\begin{aligned} \int_\Omega \frac{\partial k(u^*,x)}{\partial u^*}(\boldsymbol{\nabla} u^*\cdot\boldsymbol{\nabla} w)\varphi\, dx &\le \left\| \frac{\partial k(u^*,x)}{\partial u^*}\varphi\boldsymbol{\nabla} u^*\right\|_0 \|\boldsymbol{\nabla} w\|_0 \\ &\le C\|w\|_1\|\boldsymbol{\nabla} u^*\|_{L^{2p}(\Omega)}\|\varphi\|_{L^{2q}(\Omega)} \\ &\le C\|w\|_{1+\beta}\|\varphi\|_{1-\alpha}. \end{aligned}$$

□

3.4. **Proof of Lemma 3.3.**

Proof of Lemma 3.3. We have

$$\int_\Omega guv_{x_i}\,dx = \int_\Omega u(gv_{x_i})\,dx \le \|u\|_\alpha \|gv_{x_i}\|_{-\alpha}.$$

To get the desired bound for $\|gv_{x_i}\|_{-\alpha}$, we use interpolation between $H^{-1}(\Omega)$ and $L^2(\Omega)$. By the Sobolev inequality (3.2),

$$\|gv_{x_i}\|_0 \le \|g\|_{L^\infty(\Omega)}\|v_{x_i}\|_0 \le C\|g\|_{W^1_p(\Omega)}\|v\|_1. \tag{3.21}$$

We shall show that

$$\|gv_{x_i}\|_{-1} \le C\|g\|_{W^1_p(\Omega)}\|v\|_0 \tag{3.22}$$

and the lemma will follow by interpolation.

Let v be in $C_0^\infty(\Omega)$. Then

$$\|gv_{x_i}\|_{-1} = \sup_{w\in C_0^\infty(\Omega)} \frac{< v, (gw)_{x_i} >}{\|w\|_1} \le C\|v\|_0 \sup_{w\in C_0^\infty(\Omega)} \frac{\|(gw)_{x_i}\|}{\|w\|_1}.$$

Using Hölder's inequality, we have

$$\begin{aligned}\|g_{x_i}w + gw_{x_i}\|_0 &\le \|g_{x_i}w\|_0 + \|gw_{x_i}\|_0\\ &\le \|g_{x_i}\|_{L^p(\Omega)}\|w\|_{L^{2q}(\Omega)} + \|g\|_{L^\infty(\Omega)}\|w_{x_i}\|_0,\end{aligned}$$

where $\dfrac{2}{p} + \dfrac{1}{q} = 1$. Applying the Sobolev inequality, $\|w\|_{L^{2q}(\Omega)} \le C\|w\|_1$, gives

$$\|(gw)_{x_i}\| \le C\|g\|_{W^1_p(\Omega)}\|w\|_1.$$

This completes the proof of the lemma. □

4. Existence of a discrete solution

In this section, we define the finite element approximation of (1.1). By applying the results of [5], we will conclude the existence of a finite element solution which is close to the solution u^*.

Let $\mathcal{T}_h$ be a triangulation of Ω with mesh size h. Let $V_h \subset H^1_0(\Omega)$ be a finite element space of continuous piecewise polynomials with respect to the triangulation. We also assume that the corresponding mesh is quasi-uniform which implies inverse inequalities for the finite element functions. The discrete counterpart of (1.1) reads: Find $u_h^* \in V_h$ such that

$$\int_\Omega k(u_h^*, x)(\boldsymbol{\nabla} u_h^* \cdot \boldsymbol{\nabla}\varphi)\,dx + \int_\Omega (\boldsymbol{c}\cdot\boldsymbol{\nabla} u_h^*)\varphi\,dx = \int_\Omega f\varphi\,dx \quad \text{for all } \varphi \in V_h. \tag{4.1}$$

For $u, \varphi \in V$, let $A(u,\varphi) =< F'(u^*)u, \varphi >= \hat{A}(u,\varphi) + \hat{D}(u,\varphi)$, where

$$\begin{cases} \hat{A}(u,\varphi) = \displaystyle\int_\Omega k(u^*,x)(\boldsymbol{\nabla} u\cdot\boldsymbol{\nabla}\varphi)\,dx,\\ \hat{D}(u,\varphi) = \displaystyle\int_\Omega \frac{\partial k(u^*,x)}{\partial u^*}(\boldsymbol{\nabla} u^*\cdot\boldsymbol{\nabla}\varphi)u\,dx + \int_\Omega (\boldsymbol{c}\cdot\boldsymbol{\nabla} u)\varphi\,dx.\end{cases}$$

We will show that the form $A(\cdot,\cdot)$ satisfies a discrete inf-sup condition. To do this, we need the following lemmas:

Lemma 4.1. *There exist two constants $C_1 > 0$ and C_2 such that*

$$C_1\|u\|_1^2 - C_2\|u\|_{1-\alpha}^2 \le A(u,u) \quad \textit{for all } u \in V. \tag{4.2}$$

Proof. By the assumption on k and Poincaré's inequality, there exists $C > 0$ satisfying

$$\hat{A}(u,u) \ge C\|u\|_1^2 \quad \text{for all } u \in H_0^1(\Omega). \tag{4.3}$$

By the Schwarz inequality,

$$\left| \int_\Omega (\boldsymbol{c}\cdot\boldsymbol{\nabla} u)u\,dx \right| \le C\|\boldsymbol{\nabla} u\|_0\|u\|_0 \le C\|u\|_1\|u\|_{1-\alpha}. \tag{4.4}$$

For $1/p + 1/q = 1$, applying Hölder's inequality gives

$$\left| \int_\Omega \frac{\partial k(u^*,x)}{\partial u^*}(\boldsymbol{\nabla} u^*\cdot\boldsymbol{\nabla} u)u\,dx \right| \le C\|\boldsymbol{\nabla} u^*\|_{L^p(\Omega)}\|\boldsymbol{\nabla} u\ u\|_{L^q(\Omega)}. \tag{4.5}$$

Using Hölder's inequality again for $1/r + 1/s = 1$,

$$\|\boldsymbol{\nabla} u\ u\|_{L^q(\Omega)} \le \|u\|_{L^{qr}(\Omega)}\|\boldsymbol{\nabla} u\|_{L^{qs}(\Omega)}. \tag{4.6}$$

We take $p = 2/(1-\alpha)$, $q = 2/(1+\alpha)$ and $r = (1+\alpha)/\alpha$ and apply the Sobolev inequalities (3.3) and (3.4) to get

$$\left| \int_\Omega \frac{\partial k(u^*,x)}{\partial u}(\boldsymbol{\nabla} u^*\cdot\boldsymbol{\nabla} u)u\,dx \right| \le C\|u^*\|_{1+\alpha}\|u\|_1\|u\|_{1-\alpha}. \tag{4.7}$$

Combining $ab \le (\epsilon a^2)/2 + b^2/(2\epsilon)$, (4.3), (4.4) and (4.7), we get the result. □

Remark 4.2. The proof of the above lemma implies that there exists a constant C such that

$$|A(u,v)| \le C\|u\|_1\|v\|_1 \quad \text{for all } u,\ v \in V.$$

Lemma 4.3. *For each $u \in V$, there exists $h_0 > 0$ such that for $h \le h_0$, the problem : Find $u_h \in V_h$ satisfying*

$$A(u_h,\varphi) = A(u,\varphi) \quad \textit{for all } \varphi \in V_h \tag{4.8}$$

has a unique solution which we denote by $\Pi_h u$. Furthermore,

$$\|u - \Pi_h u\|_1 \le Ch^\alpha\|u\|_{1+\alpha}. \tag{4.9}$$

Proof. The proof immediately follows Lemma 4.1 and the finite element duality argument [9] with Proposition 3.11. □

Lemma 4.4. *Let h be less than h_0 defined in Lemma 4.3 then $\|\Pi_h u\|_{1+\alpha} \le C\|u\|_{1+\alpha}$.*

Proof. Let $e = u - \Pi_h u$ and $Q_h : V \to V_h$ be the $L^2(\Omega)$ projection onto V_h. Then,

$$\begin{cases} \|u - Q_h u\|_1 \le Ch^\alpha\|u\|_{1+\alpha}, \\ \|Q_h u\|_{1+\alpha} \le C\|u\|_{1+\alpha}. \end{cases} \tag{4.10}$$

By an inverse inequality, (4.10) and (4.9)

$$\begin{aligned} \|e\|_{1+\alpha}^2 &\le \|\Pi_h u - Q_h u\|_{1+\alpha}^2 + \|Q_h u - u\|_{1+\alpha}^2 \\ &\le Ch^{-2\alpha}\|\Pi_h u - Q_h u\|_1^2 + C\|u\|_{1+\alpha}^2 \le C\|u\|_{1+\alpha}^2. \end{aligned}$$

Hence,

$$\|e\|_{1+\alpha} \leq C\|u\|_{1+\alpha}.$$

The triangle inequality completes the proof of this lemma. □

Proposition 4.5 (Discrete Inf-Sup Condition)**.** *Let h be less than h_0 defined in Lemma 4.3. Then there exists a constant C such that*

$$\|\varphi\|_{1-\alpha} \leq C \sup_{u_h \in V_h} \frac{A(u_h, \varphi)}{\|u_h\|_{1+\alpha}} \quad \text{for all } \varphi \in V_h. \tag{4.11}$$

Proof. By (3.18) and Lemma 4.4, for all $\varphi \in V_h$

$$\begin{aligned}\|\varphi\|_{1-\alpha} &\leq C \sup_{u \in V} \frac{A(u, \varphi)}{\|u\|_{1+\alpha}} = C \sup_{u \in V} \frac{A(\Pi_h u, \varphi)}{\|u\|_{1+\alpha}} \\ &\leq C \sup_{u \in V} \frac{A(\Pi_h u, \varphi)}{\|\Pi_h u\|_{1+\alpha}} \leq C \sup_{u_h \in V_h} \frac{A(u_h, \varphi)}{\|u_h\|_{1+\alpha}}.\end{aligned}$$

□

Remark 4.6. Similar to the derivation of (3.18), (4.11) implies

$$\|u_h\|_{1+\alpha} \leq C \sup_{\varphi \in V_h} \frac{A(u_h, \varphi)}{\|\varphi\|_{1-\alpha}} \quad \text{for all } u_h \in V_h. \tag{4.12}$$

Finally, we are ready to prove the existence of the discrete solution for the model problem by applying a result in [5].

Theorem 4.7 (Existence of the discrete solution)**.** *With assumptions (B.1) and (B.2), there exist two constants $\delta > 0$ and $h_0 > 0$ such that for $h \leq h_0$ there exists a unique solution u_h^* for problem (4.1) in the ball $B(u^*, \delta)$. Moreover there exists a constant C independent of h such that*

$$\|u^* - u_h^*\|_{1+\alpha} \leq C \inf_{\xi \in V_h} \|u^* - \xi\|_{1+\alpha}. \tag{4.13}$$

Proof. Proposition 3.5 and 3.6 show $F'(u)$ exists for all $u \in V$ and is Lipschitz continuous in a neighborhood of u^*. Moreover, $F'(u^*)$ is an isomorphism from V to W'(see Proposition 3.10). The theorem follows from the discrete inf-sup condition (4.12) and Theorem 7.1 in [5]. □

Remark 4.8. The results of the present section show that the solution has regularity, $u^* \in H^{1+s}$ for $s < 1/2$. Taking $s > \alpha$ and applying the above theorem gives

$$\|u^* - u_h^*\|_{1+\alpha} \leq Ch^{s-\alpha}\|u^*\|_{1+s},$$

i.e., u_h^* converges to u^* in $H^{1+\alpha}(\Omega)$.

5. The discrete problem in the framework of Section 2

In this section, we set up the discrete problem in the framework of Section 2. We start by defining $\mathcal{V} = \mathcal{W} = V_h$ with norms $\|\cdot\|_{1+\alpha}$ on $\mathcal{V}$ and $\|\cdot\|_{1-\alpha}$ on $\mathcal{W}$. We identify $\mathcal{W}'$ with V_h and define for $G \in \mathcal{W}'$,

$$< G, \varphi > \equiv (G, \varphi) \quad \text{for all } \varphi \in \mathcal{W}.$$

We then define $F_h : \mathcal{V} \mapsto \mathcal{W}'$ by

$$(F_h(v), \varphi) = \int_\Omega k(v,x)(\boldsymbol{\nabla} v \cdot \boldsymbol{\nabla}\varphi)\, dx + \int_\Omega (\boldsymbol{c}\cdot\boldsymbol{\nabla} v)\varphi\, dx - \int_\Omega f\varphi\, dx, \tag{5.1}$$

for all $v \in \mathcal{V}$ and $\varphi \in \mathcal{W}$. Clearly, the problem of finding $u_h^* \in \mathcal{V}$ satisfying $F_h(u_h^*) = 0$ coincides with the discrete problem (4.1). Thus, (A.1) is contained in Theorem 4.7.

For $u \in V$, we define the linear map $F_h' : \mathcal{V} \to \mathcal{W}'$ by

$$\begin{aligned}(F_h'(u)v, \varphi) = &\int_\Omega k(u,x)(\boldsymbol{\nabla} v \cdot \boldsymbol{\nabla}\varphi)\, dx \\ &+ \int_\Omega \frac{\partial k(u,x)}{\partial u}(\boldsymbol{\nabla} u \cdot \boldsymbol{\nabla}\varphi) v\, dx + \int_\Omega (\boldsymbol{c}\cdot\boldsymbol{\nabla} v)\varphi\, dx,\end{aligned} \tag{5.2}$$

for all $v \in \mathcal{V}$ and $\varphi \in \mathcal{W}$. Note that

$$(F_h'(u)v, \varphi) =< F'(u)v, \varphi >, \quad \text{for all } v \in \mathcal{V},\ \varphi \in \mathcal{W},$$

i.e., $F_h'(u)$ is the restriction of $F'(u)$ to $\mathcal{V}\times\mathcal{W}$. Because of this, Proposition 3.5 immediately implies that for $u \in \mathcal{V}$, $\delta > 0$ and $v \in B(u,\delta) \subset \mathcal{V}$,

$$\|F_h(v) - F_h(u) - F_h'(u)(v-u)\|_{\mathcal{W}'} \le C(\|u\|_{1+\alpha}, \delta)\|v-u\|_{\mathcal{V}}^2.$$

When restricted to $\mathcal{V}$, this is (A.2.1). Similarly, by Proposition 3.6, for u and v as above,

$$\|F_h'(v) - F_h'(u)\|_{[\mathcal{V},\mathcal{W}']} \le C(\|u\|_{1+\alpha}, \delta)\|v-u\|_{\mathcal{V}} \tag{5.3}$$

and (A.2.2) follows directly.

Let g be in $\mathcal{W}'$ and extend g to a functional on W by

$$< g, \varphi > \equiv (g, \varphi) \quad \text{for all } \varphi \in W.$$

Then

$$\begin{aligned}\|g\|_{W'} &= \sup_{\varphi\in W} \frac{(g,\varphi)}{\|\varphi\|_{1-\alpha}} \\ &\le C \sup_{\varphi\in W} \frac{(g, Q_h\varphi)}{\|Q_h\varphi\|_{1-\alpha}} = C\|g\|_{\mathcal{W}'}.\end{aligned}$$

We used the fact that Q_h is a bounded operator on $\|\cdot\|_{1-\alpha}$. Let $u = (F'(u^*))^{-1}g$. Using the fact that $F'(u^*)$ is an isomorphism and (4.12) gives that the solution $u_h \in V_h$ of

$$A(u_h, \theta) = A(u, \theta) = (g, \theta) \quad \text{for all } \theta \in V_h$$

satisfies

$$\|u_h\|_{\mathcal{V}} \le M\|g\|_{\mathcal{W}'}, \tag{5.4}$$

i.e., $\|(F_h'(u^*))^{-1}\|_{[\mathcal{W}',\mathcal{V}]} \le M$. Here M can be chosen independent of h if h_0 is small enough.

The final condition (A.2.3) required for the application of the results of Section 2 is contained in the following proposition.

Proposition 5.1. *There exist $h_0 > 0$ and $\delta > 0$ such that if h is less than h_0 and u_h is in the ball $B(u_h^*, \delta)$ in $\mathcal{V}$, $F_h'(u_h)^{-1} : \mathcal{W}' \to \mathcal{V}$ exists and satisfies*

$$\|F_h'(u_h)^{-1}\|_{[\mathcal{W}',\mathcal{V}]} \le 2M. \tag{5.5}$$

Proof. By (5.3) and (5.4), there exists δ_0 such that

$$\begin{aligned}\|I-(F_h'(u^*))^{-1}F_h'(u)\|_{[\mathcal{V},\mathcal{V}]} &\le \|(F_h'(u^*))^{-1}\|_{[\mathcal{W}',\mathcal{V}]}\|F_h'(u^*)-F_h'(u)\|_{[\mathcal{V},\mathcal{W}']}\\ &\le CM\|u^*-u\|_{1+\alpha},\end{aligned}$$

for all u in $B(u^*,\delta_0)\subset V$. If we choose $2\delta<\min\{\frac{1}{2CM},\delta_0\}$, then

$$\|I-(F_h'(u^*))^{-1}F_h'(u)\|_{[\mathcal{V},\mathcal{V}]}<\frac{1}{2},\quad \text{for all } u\in B(u^*,2\delta)\subset V.$$

By the Neumann series argument, $F_h'(u)$ is nonsingular and $\|F_h'(u)^{-1}\|_{[\mathcal{W}',\mathcal{V}]}\le 2M$.

By Remark 4.8, we can choose h_0 sufficiently small so that

$$\|u^*-u_h^*\|_{1+\alpha}\le\delta$$

when h is less than h_0. Then for $u_h\in B(u_h^*,\delta)\subset\mathcal{V}$,

$$\|u_h-u^*\|_{1+\alpha}\le 2\delta$$

and the conclusion of the proposition follows. □

The above results show that (A.1) and (A.2) hold for our discrete framework. Moreover, the functions $\delta(\varepsilon)$ and $M(\varepsilon)$ can be chosen independent of h if h_0 is small enough. Thus, a modified inexact Newton algorithm of the form given in Section 2 will converge at a uniform rate (independently of mesh size h) provided that an iteration satisfying (2.3) is used.

Remark 5.2. The above proposition shows that $F_h'(u_h)$ is an isomorphism if $u_h\in V_h$ is close enough to u_h^*, i.e., there are two constants C_1 and C_2 independent of h such that

$$C_1\|\varphi\|_{1+\alpha}\le\|F_h'(u_h)\varphi\|_{-1+\alpha}\le C_2\|\varphi\|_{1+\alpha}\quad\text{for all }\varphi\in V_h. \tag{5.6}$$

6. An iteration satisfying (2.3)

In this section, we define an iteration which satisfies (2.3) when $\mathcal{V}$ is defined as in the previous section. We start by defining computable Sobolev norms by using a variation of the approach from [1, 2]. An iteration satisfying (2.3) is then constructed in terms of these norms.

We assume that the space V_h results from a multilevel sequence of meshes. Specifically, we assume that we have a sequence of nested triangulations, e.g., the triangles in $\mathcal{T}_{j+1}$ are formed by subdividing those in $\mathcal{T}_j$ into four by connecting the midpoints of the edges. We require that $\mathcal{T}_1$ is of unit size and set V_j to be the finite element space corresponding to $\mathcal{T}_j$. We assume that $V_h\equiv V_J$.

We next define a sequence of approximation operators $\widehat{Q}_j: L^2(\Omega)\to V_j$. Let $\{\phi_i\}_{i=1}^m$ be the nodal basis for V_j. For $j>0$, set

$$\widehat{Q}_j u=\sum_{i=1}^m\frac{(u,\phi_i)}{(1,\phi_i)}\phi_i. \tag{6.1}$$

Define

$$T_s u=\sum_{j=1}^J h_j^{-2s}\widehat{Q}_j u\quad\text{for all } u\in V_h. \tag{6.2}$$

Then, there are two constants $C_0 = C_0(s)$ and $C_1 = C_1(s)$ not depending on h such that for $-3/2 < s < 0$,

$$C_0\|u\|_s \le (T_s u, u)^{1/2} \le C_1\|u\|_s \quad \text{for all } u \in V_h. \tag{6.3}$$

We note that if we set $\widehat{Q}_0 = 0$ and define T_s by

$$T_s u = \sum_{j=1}^{J} h_j^{-2s}(\widehat{Q}_j - \widehat{Q}_{j-1})^2 u \quad \text{for all } u \in V_h, \tag{6.4}$$

then (6.3) still holds (cf. [1]).

Let $\widehat{F} = F_h'(u^0)$ where $u^0 \in V_h$ is the starting iterate of Algorithm 2.1 and satisfies $\|u^0 - u_h^*\|_{1+\alpha} \le \delta$ so that (A.1) and (A.2) hold. In addition, define $a(\cdot,\cdot)$ on $V_h \times V_h$ and $A : V_h \to V_h$ by

$$\begin{cases} a(u,v) = (T_{-1+\alpha}\widehat{F}u, \widehat{F}v), \\ (Au, v) = a(u,v) \quad \text{for all } u, v \in V_h. \end{cases} \tag{6.5}$$

Then $a(\cdot,\cdot)$ is clearly symmetric and positive definite by Proposition 5.1. We define the norm $|||\cdot|||$ on V_h by

$$|||u_h||| \equiv a(u_h, u_h)^{1/2} \quad \text{for all } u_h \in V_h.$$

It follows from (5.6) and (6.3) that $|||\cdot|||$ is uniformly (independently of h) equivalent to $\|\cdot\|_{1+\alpha}$ on V_h.

We then have the following lemma.

Lemma 6.1. *There exist two positive constants $\tilde{C}_0$ and $\tilde{C}_1$ independent of h such that*

$$\tilde{C}_0 a(u,u) \le a(T_{-1-\alpha}Au, u) \le \tilde{C}_1 a(u,u) \quad \textit{for all } u \in V_h. \tag{6.6}$$

Proof. For all u in V_h, using Remark 5.2 gives

$$\begin{aligned} a(T_{-1-\alpha}Au, u) &= (T_{-1-\alpha}Au, Au) \ge C\|Au\|_{-1-\alpha}^2 \\ &= C\sup_{v\in V}\frac{(Au,v)^2}{\|v\|_{1+\alpha}^2} \ge C\frac{a(u,u)^2}{\|u\|_{1+\alpha}^2} \ge \tilde{C}_0 a(u,u). \end{aligned}$$

On the other hand, by (6.3)

$$a(T_{-1-\alpha}Au, u) \le C\sup_{v\in V}\frac{(Au,v)^2}{\|v\|_{1+\alpha}^2} = C\sup_{v\in V}\frac{(Au, Q_h v)^2}{\|v\|_{1+\alpha}^2}.$$

Using the boundedness of Q_h on V gives

$$a(T_{-1-\alpha}Au, u) \le C\sup_{v_h\in V_h}\frac{a(u,v_h)^2}{\|v_h\|_{1+\alpha}^2} \le \tilde{C}_1 a(u,u).$$

This completes the proof of the lemma. □

We consider the following problem: Find s^k in V_h satisfying

$$a(s^k, \varphi) = (-T_{-1+\alpha}F_h(u^k), \widehat{F}\varphi) \quad \text{for all } \varphi \in V_h. \tag{6.7}$$

The solution of (6.7) and (2.2) in Algorithm 2.1 coincide. To define $\hat{s}^k$, we apply the m step PCG method to (6.7) with the zero initial iterate. The preconditioner used is $T_{-1-\alpha}$. Then

$$|||s^k - \hat{s}^k||| \leq \frac{2q^m}{1+q^{2m}} \; |||\, s^k|||, \tag{6.8}$$

where $q = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} < 1$ and $\kappa = cond(T_{-1-\alpha}A) \leq \tilde{\kappa} \equiv \tilde{C}_1/\tilde{C}_0$. Thus, (2.3) holds for

$$\beta = \frac{2\tilde{q}^m}{1+\tilde{q}^{2m}} < 1$$

where $\tilde{q} = \frac{\sqrt{\tilde{\kappa}}-1}{\sqrt{\tilde{\kappa}}+1} < 1$ and independent of h.

Remark 6.2. We can apply the one step PCG(Steepest Descent) method also. In this case, (2.3) holds for $\beta = \left(\frac{\tilde{\kappa}-1}{\tilde{\kappa}+1}\right) < 1$.

We can now conclude the following theorem.

Theorem 6.3. *Suppose that we use the iterative method described above for computing the approximation $\hat{s}^k$. There is a positive number h_0 and a $\delta > 0$ such that if $h \leq h_0$ and $\|u^0 - u_h^*\|_{1+\alpha} \leq \delta$, then the inexact Newton algorithm converges with a rate of linear convergence which is independent of h.*

7. Numerical Results

In this section, numerical results supporting Theorem 6.3 for a model problem are given. We present the results for (2.3) defined using the algorithm of the previous section for the two cases when T_s is given by (6.2) and (6.4) applied to the following problem:

$$\begin{aligned} -\mathrm{div}(k(u,x)\boldsymbol{\nabla} u) + \boldsymbol{c}\cdot\boldsymbol{\nabla} u = f, \;\; x \in \Omega \\ u(x) = 0, \;\; x \in \partial\Omega. \end{aligned} \tag{7.1}$$

Here, $\Omega = (0,1)\times(0,1)$, $k(u,x) = 1/(u^2+1) + e^{-x_1^2-x_2^2}$, $\boldsymbol{c} = (1,1)$, and the exact solution is $u^* = u(x_1,x_2) = (x_1 - x_1^2)(x_2 - x_2^2)$. The right hand side f is defined by applying the left hand side to the exact solution.

The discrete problem is obtained by using linear basis functions on triangles of mesh size $h = \frac{1}{2^n}$, $n = 5, 6, 7, 8, 9$. To define $\hat{s}^k$ in the Modified Inexact Newton Algorithm, we used the five step PCG method for (6.7). We used $\alpha = 0.05$. We stopped the algorithm when the normalized discrete l_2 norm of of the nodal values of $(u_h^{k+1} - u_h^k)$ was less than 10^{-6}. The initial nonlinear iterate u_h^0 was 0.1 on the interior nodes of the mesh and zero at the boundary nodes.

Table 1 and 2 illustrate the number of nonlinear iterations required to reach the above-mentioned convergence criteria. We also report the L^2 and H^1 norm error between the discrete solution u_h^n and the interpolant of the exact solution $I_h u^*$. The number of nonlinear iterations increase slightly as h decreases. The rate of increase decreases for smaller h. This is not surprising since similar increases are observed when T_1 is applied as a preconditioner for the Laplacian, an iterative procedure which can also be bounded independently of the number of unknowns. The operator $T_{-1-\alpha}A$ is better conditioned

when T_s is given by (6.4) and so it is not surprising to see that the results of Table 1 are better than Table 2.

h^{-1}	nonlinear iterations	$\|u_h^n - I_h u^*\|_0$	$\|u_h^n - I_h u^*\|_1$
32	18	5.50e-05	2.50e-04
64	23	1.38e-05	6.25e-05
128	26	3.44e-06	1.56e-05
256	32	8.59e-07	3.89e-06
512	33	2.14e-07	9.77e-07

TABLE 7.1. $T_s = \sum h_j^{-2s}(\widehat{Q}_j - \widehat{Q}_{j-1})^2$

h^{-1}	nonlinear iterations	$\|u_h^n - I_h u^*\|_0$	$\|u_h^n - I_h u^*\|_1$
32	32	5.50e-05	2.50e-04
64	43	1.38e-05	6.24e-05
128	56	3.43e-06	1.56e-05
256	66	8.51e-07	3.89e-06
512	79	2.12e-07	1.01e-06

TABLE 7.2. $T_s = \sum h_j^{-2s}\widehat{Q}_j$

To illustrate the linear convergence in Theorem 6.3, we set up a problem where the exact discrete solution was known. To do this, we applied the discrete nonlinear operator to $I_h u^*$ so that $I_h u^*$ was the exact discrete solution. Figure 1 shows the linear convergence with respect to the norm $|||\cdot|||$ when $n = 6$ and $T_s = \sum h_j^{-2s}(\widehat{Q}_j - \widehat{Q}_{j-1})^2$. For these results, we took five steps of PCG to define $\hat{s}^k$.

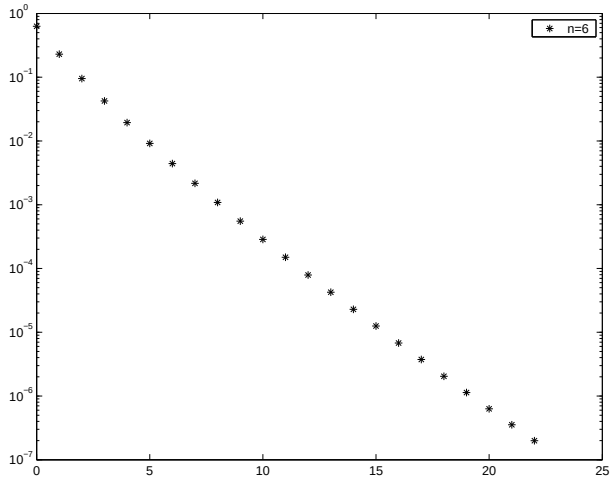

FIGURE 1. $|||u_h^n - u_h^*|||$ vs. iteration numbers

The next figure illustrates that the algorithm converges even if only one step of the PCG is taken in the definition of $\hat{s}^k$. (Remark 6.2). In this case we used $n = 6$

and $T_s = \sum h_j^{-2s}(\widehat{Q}_j - \widehat{Q}_{j-1})^2$. One step of PCG results in $\hat{s}^k$ being a fairly crude approximation to s^k and it is not surprising to see that the resulting rate of convergence for the nonlinear problem is slow. Nevertheless, the figure illustrates the monotone convergence behavior guaranteed in Theorem 2.2.

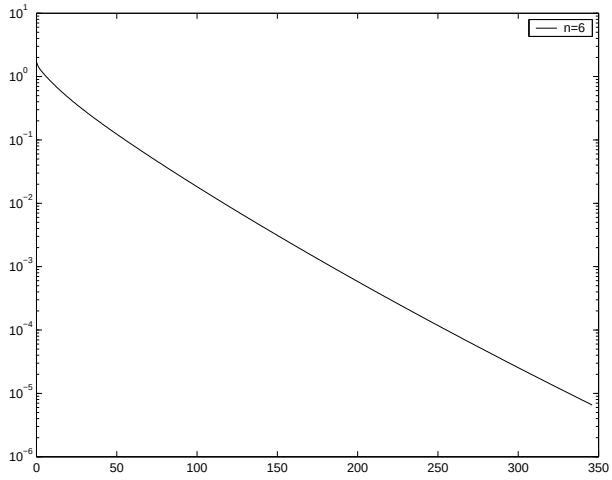

FIGURE 2. $|||u_h^n - u_h^*|||$ vs. iteration numbers

REFERENCES

[1] J. H. BRAMBLE, J. E. PASCIAK, AND P. S. VASSILEVSKI, *Computational scales of Sobolev norms with application to preconditioning*, Mathematics of Computation, 69 (2000), pp. 463–480.

[2] J. H. BRAMBLE, J. E. PASCIAK, AND J. XU, *Parallel multilevel preconditioners*, Mathematics of Computation, 55 (1990), pp. 1–22.

[3] S. C. BRENNER AND L. R. SCOTT, *The mathematical theory of finite element methods*, Springer-Verlag, New York, 1994.

[4] P. N. BROWN, P. VASSILEVSKI, AND C. S. WOODWARD, *On mesh independent convergence of an inexact Newton-Multigrid algorithm*, SIAM J. Sci. Comput., 25 (2003), pp. 570–590.

[5] G. CALOZ AND J. RAPPAZ, *Numerical analysis for nonlinear and bifurcation problems*. In P. G. Ciarlet and J. L. Lions, editors, Handbook of numerical analysis, Volume V of Techniques of Scientific Computing(Part 2), 1998.

[6] R. S. DEMBO, S. C. EISENSTAT, AND T. STEIHAUG, *Inexact Newton methods*, SIAM J. Numer. Anal., 19 (1982), pp. 400–408.

[7] P. DEUFLHARD, *Newton methods for nonlinear problems. Affine invariance and adaptive algorithms*, Springer-Verlag, Berlin, 2004.

[8] B. MORINI, *Convergence behavior of inexact Newton methods*, Mathematics of Computation, 66 (1999), pp. 1605–1613.

[9] A. H. SCHATZ, *An observation concerning Ritz-Galerkin methods with indefinite bilinear forms*, Mathematics of Computation, 28 (1974), pp. 959–962.

Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, U.S.A.
E-mail address: tjkim@math.tamu.edu.

Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, U.S.A.
E-mail address: pasciak@math.tamu.edu.

Center for Applied Scientific Computing, UC Lawrence Livermore National Laboratory, P. O. Box 808, L-560, Livermore, CA 94550, U.S.A.
E-mail address: panayot@llnl.gov.