Numerical Approximation of PDE System Fixed Point Maps via Newton's Method

Joseph W. Jerome*

April, 1991

Abstract

Since the fundamental paper of Moser (Ann. Scuola Norm. Pisa XX(1966), 265-315), it has been understood analytically that regularization is necessary as a postconditioning step in the application of approximate Newton methods, based upon the system differential map. A development of these ideas in terms of current numerical methods and complexity estimates was given by the author (Numer. Math. 47(1985), 123-138). It was proposed by the author (Numer. Math. 55(1989), 619-632) to use the fixed point map as a basis for the linearization, and thereby avoid the numerical loss of derivatives phenomenon identified by Moser. Independently, a coherent theory for the approximation of fixed points by numerical fixed points was devised by Krasnosel'skii and his coworkers (Approximate Solution of Operator Equations, Wolters-Noordhoff, 1972). In this paper, the Krasnosel'skii calculus is merged with Newton's method, for the computation of the approximate fixed points, in such a way that the approximation order is preserved with mesh independent constants. Since the application is to a system of partial differential equations, the issue of the implicit nature of the linearized approximation must be addressed as well.

Key words: Nonlinear Systems, Fixed Point Approximation, Krasnosel'skii Calculus, Approximate Newton Methods, Finite Element Methods.
AMS(MOS) Subject Classification: 35J65, 41A35, 65N15, 65N30
Acknowledgements: The author is supported by the National Science Foundation under grant DMS-8922398.

*Department of Mathematics, Northwestern University, Evanston, Il 60208

1 Introduction.

In this paper, we shall study a nonlinear system of partial differential equations from the perspective of its fixed point and approximate fixed point maps. The fixed point map T will be defined by Jacobi system decoupling, and the approximate fixed point map T_n by a corresponding decoupling, based upon piecewise linear finite elements. Superimposed upon the approximate fixed point map, termed the numerical fixed point map, we introduce a systematic procedure, described in terms of Newton's method, for the iterative calculation of the approximate fixed points. The Newton method is defined via the linearization of T_n.

The approach effectively merges two powerful calculi:

- A calculus, developed by the Russian school, for the two-sided estimation of fixed points by approximate fixed points.
- An exact Newton iterative method, based upon the numerical fixed point map, T_n, which functions as an approximate Newton method for T.

By employing the fixed point maps to define Newton's method, we avoid the loss of derivatives' phenomenon, requiring post-conditioning as a residual "catch-up" procedure. A striking aspect of the entire process is that the Newton iteration is global. Convergence is at least linear, and continues until the sufficiently small residual criterion required for quadratic convergence is met. It is possible to correlate the number of Newton iterates with the two-sided estimates given by the fixed point approximation theory.

In Section 2, we shall introduce the model, discuss the existence theory and maximum principles, and introduce the preliminary fixed point map. In Section 3, we shall introduce the fixed point map and the numerical fixed point map, and discuss the convergence properties of the latter. In Section 4, we present the approximation theory, due to [11], but here adapted to a recursive linear theory via Newton's method. In Section 5, we combine the theory of Section 4 with the model studied here, and deduce our major results.

2 The Model and the Preliminary Fixed Point Map

The basic model will be a steady-state system of reaction-diffusion partial differential equations, subject to inhomogeneous Dirichlet boundary conditions. Such models have diverse application. Often these systems are subsystems of larger systems, which benefit from separate analysis. Two which are compatible with the reaction terms selected below include chemically reacting systems (cf. [3]) and the current continuity subsystem associated with the flow of electrons and holes in a semiconductor (cf. [5]) when the Slotboom variables are employed. In the subsections to follow we describe the model specifically, and construct an associated fixed point map, termed the preliminary fixed point map. New maximum principles are established in the process, for we consider the case of sign reversal in reaction terms.

2.1 The System

Let $\mathcal{D}$ be a convex polyhedral domain in R^d, and consider the Dirichlet boundary value problem on $\mathcal{D}$ defined by the system

$$-\nabla \cdot [a(x)\nabla u(x)] + f(u,v) \;=\; 0, \tag{2.1}$$

$$-\nabla \cdot [b(x)\nabla v(x)] + g(u,v) \;=\; 0, \tag{2.2}$$

where u and v satisfy the inhomogeneous Dirichlet boundary conditions,

$$[u - \bar{u}]_{|\partial\mathcal{D}} = 0 = [v - \bar{v}]_{|\partial\mathcal{D}}, \tag{2.3}$$

for $\bar{u}$ and $\bar{v}$ strictly positive functions in $C^2(\bar{\mathcal{D}})$: $\bar{u} \geq u_0 > 0$, $\bar{v} \geq v_0 > 0$. It is assumed here that a and b are bounded measurable functions, also bounded away from zero:

$$a \geq a_0 > 0, \;\; b \geq b_0 > 0. \tag{2.4}$$

The functions f and g are assumed to satisfy the following properties, which generalize those satisfied by the prototypes, $f(u,v) = uv - 1$, $g(u,v) = uv - 1$.

- They are C^2 in their joint arguments, $u \geq 0$ and $v \geq 0$.
- They are monotone in u and v respectively:

$$\frac{\partial f}{\partial u} \geq 0, \;\; \frac{\partial g}{\partial v} \geq 0. \tag{2.5}$$

- The inverse images, $f^{-1}(0)$ and $g^{-1}(0)$, are graphs of continuous, positive, *decreasing* functions f_1 and g_1 for $u > 0$ and $v > 0$:

$$u = f_1(v), \;\; v = g_1(u). \tag{2.6}$$

- For each $u \geq 0$,

$$f(u, \bar{v}_{min}) \leq f(u,v) \leq f(u, \bar{v}_{max}), \;\; \text{for } \; \bar{v}_{min} \leq v \leq \bar{v}_{max}. \tag{2.7}$$

  For each $v \geq 0$,

$$g(\bar{u}_{min}, v) \leq g(u,v) \leq g(\bar{u}_{max}, v), \;\; \text{for } \; \bar{u}_{min} \leq u \leq \bar{u}_{max}. \tag{2.8}$$

Here, we define $\bar{u}_{min}$, $\bar{u}_{max}$, $\bar{v}_{min}$, and $\bar{v}_{max}$ by

$$\bar{u}_{min} = \inf_{\partial\mathcal{D}} \bar{u}, \;\; \bar{u}_{max} = \sup_{\partial\mathcal{D}} \bar{u}, \;\; \bar{v}_{min} = \inf_{\partial\mathcal{D}} \bar{v}, \;\; \bar{v}_{max} = \sup_{\partial\mathcal{D}} \bar{v}. \tag{2.9}$$

If α_u, α_v, β_u, and β_v are defined by

$$0 < \alpha_u = \min(\bar{u}_{min}, f_1(\bar{v}_{max})), \;\; \beta_u = \max(\bar{u}_{max}, f_1(\bar{v}_{min})), \tag{2.10}$$

$$0 < \alpha_v = \min(\bar{v}_{min}, g_1(\bar{u}_{max})), \;\; \beta_v = \max(\bar{v}_{max}, g_1(\bar{u}_{min})), \tag{2.11}$$

then, in particular, f and g satisfy the following inequalities:

$$f(u, \bar{v}_{min}) \;\geq\; 0, \;\forall u \geq \beta_u, \tag{2.12}$$

$$f(u, \bar{v}_{max}) \;\leq\; 0, \;\forall u \leq \alpha_u, \tag{2.13}$$

$$g(\bar{u}_{min}, v) \;\geq\; 0, \;\forall v \geq \beta_v, \tag{2.14}$$

$$g(\bar{u}_{max}, v) \;\leq\; 0, \;\forall v \leq \alpha_v. \tag{2.15}$$

2.2 The Preliminary Fixed Point Map

In anticipation of the maximum principles, we define the domain of the (preliminary) fixed point map T_0 as follows. Set

$$K_0 = \{[u,v] \in \prod_1^2 L_2(\mathcal{D}) : \alpha_u \le u \le \beta_u, \alpha_v \le v \le \beta_v\}, \tag{2.16}$$

where (2.10) and (2.11) have been used. The mapping T_0 proceeds via one Jacobi iteration on elements of K_0. Thus, given $[\tilde{u},\tilde{v}] \in K_0$, define

$$[u,v] = T_0[\tilde{u},\tilde{v}]$$

by *weak* solution of the fully decoupled system,

$$\begin{aligned} -\nabla\cdot(a\nabla u) + f(u,\tilde{v}) &= 0, & (2.17)\\ -\nabla\cdot(b\nabla v) + g(\tilde{u},v) &= 0, & (2.18) \end{aligned}$$

subject to the boundary conditions described in (2.3). These are described in the usual way by the boundary trace operator, Γ.

We shall have need of the convex functions obtained through integration of f and g, where the latter have been evaluated at the fixed function elements, $\tilde{u}$ and $\tilde{v}$, respectively:

$$F(x,s) = \int_0^s f(\sigma,\tilde{v}(x))\,d\sigma, \quad G(x,t) = \int_0^t g(\tilde{u}(x),\tau)\,d\tau. \tag{2.19}$$

Denote by $\mathcal{C}_1$ the convex subset of $L_2(\mathcal{D})$ consisting of functions u for which

$$\int_{\mathcal{D}} F(\cdot,u) < \infty,$$

and by $\mathcal{C}_2$ the corresponding convex set of functions v defined by integration of G. We also require a notation for the affine flats in $H^1(\mathcal{D})$ defining the boundary conditions:

$$L_{\bar{u}} = \{u \in H^1(\mathcal{D}) : \Gamma[u-\bar{u}]\} = 0, \tag{2.20}$$

$$L_{\bar{v}} = \{v \in H^1(\mathcal{D}) : \Gamma[v-\bar{v}]\} = 0. \tag{2.21}$$

Finally, we define

$$\mathcal{C}_u = \mathcal{C}_1 \cap L_{\bar{u}}, \quad \mathcal{C}_v = \mathcal{C}_2 \cap L_{\bar{v}}. \tag{2.22}$$

Theorem 2.1 *The map T_0, defined in (2.17) and (2.18), acts invariantly on K_0. The components u and v may be charactized uniquely as solutions of the convex minimization problems*

$$\begin{aligned} \Phi(u) &= \min_{L_2(\mathcal{D})} \Phi(u^*), & (2.23)\\ \Psi(v) &= \min_{L_2(\mathcal{D})} \Psi(v^*), & (2.24) \end{aligned}$$

where Φ is a proper convex functional defined by

$$\Phi(u^*) = \begin{cases} \int_{\mathcal{D}}[\frac{1}{2}a \mid \nabla u^* \mid^2 + F(\cdot,u^*)], & \text{if } u^* \in \mathcal{C}_u,\\ +\infty, & \text{otherwise,} \end{cases} \tag{2.25}$$

and Ψ is defined similarly.

Proof. We first assume that the system, (2.17), (2.18), and (2.3), possesses a weak solution pair $[u, v]$. We shall demonstrate the maximum principles

$$\alpha_u \le u \le \beta_u, \quad \alpha_v \le v \le \beta_v. \tag{2.26}$$

The argument for v parallels that for u and we shall give only that for the latter. In the weak version of (2.17), i. , e. , the equation obtained by multiplication of (2.17) by a test function $\phi \in H_0^1(\mathcal{D})$, followed by formal integration by parts, we make the substitution $\phi = (u-\beta_u)^+$. It is known that $\phi \in H^1$, and since $\beta_u \ge \bar{u}_{max}$, it follows that $\phi \in H_0^1$ and represents an admissible test function. We obtain, after some simplification, the equation

$$\int_{\mathcal{D}} [a \mid \nabla\phi \mid^2 + f(u, \tilde{v})\phi] = 0. \tag{2.27}$$

The property (2.7) then allows us to conclude that

$$\int_{\mathcal{D}} [a \mid \nabla\phi \mid^2 + f(u, \bar{v}_{min})\phi] \le 0. \tag{2.28}$$

Use of (2.12) allows us to conclude that each term in (2.28) is nonnegative, and hence each integrated product term is, in fact, zero. In particular,

$$\nabla(u - \beta_u)^+ = 0.$$

It follows that $(u-\beta_u)^+$ is constant, and the constant must be zero, since this function vanishes in a generalized sense on the boundary. We have established the upper bound in the first inequality of (2.26). The lower bound in this inequality makes use of the substitution of $\phi = (u - \alpha_u)^-$ into the weak version of (2.17), where $t^- = t - t^+$. The inequalities for v are similar.

We show now that Φ has a minimum by use of standard results in convex analysis (cf. [1]). In the language of [1], Φ is a proper convex functional on $L_2(\mathcal{D})$, and it possesses a minimum $u \in L_{\bar{u}}$ if it is coercive and lower semicontinuous. These properties are defined by

$$\frac{\Phi(w)}{\|w\|} \to \infty \ \text{ as } \ \|w\| \to \infty; \tag{2.29}$$

$$\forall c \in R^1, \quad B = \{w \in L_2(\mathcal{D}) : \Phi(w) \le c\} \ \text{ is closed;} \tag{2.30}$$

respectively. Here, norms are L_2 norms. To prove the lower semicontinuity of Φ, suppose $w_k \to w$ in $L_2(\mathcal{D})$ with $\Phi(w_k) \le c$. By taking a subsequence, if necessary, we may also assume a.e. pointwise convergence. By (2.25) and the inequality,

$$F(x, s) \ge f(0, \tilde{v}(x))s \ge \begin{cases} f(0, \bar{v}_{min})s, & s \ge 0, \\ f(0, \bar{v}_{max})s, & s \le 0, \end{cases} \tag{2.31}$$

we conclude that $\{w_k\}$ is bounded in $H^1(\mathcal{D})$. Here, we may use the norm, equivalent to the standard norm, given by

$$\|w\|_{H^1}^2 = \int_{\mathcal{D}} \mid \nabla w \mid^2 + (\int_{\partial\mathcal{D}} \Gamma w)^2. \tag{2.32}$$

Combining the weak compactness property of bounded H^1 sets with the Rellich theorem characterizing the H^1 injection into L_2 as compact, we conclude that

$$w_k \rightharpoonup w \text{ in } H^1(\mathcal{D}) \text{ (weakly).}$$

By the weak lower semicontinuity of equivalent norm expressions on $H^1(\mathcal{D})$, we conclude that

$$\liminf_{k\to\infty} \int_{\mathcal{D}} a \mid \nabla w_k \mid^2 \geq \int_{\mathcal{D}} a \mid \nabla w \mid^2 . \tag{2.33}$$

Making use of the lower semicontinuity of $F(x,\cdot)$ for fixed x, together with the pointwise convergence of w_k to w, we obtain from Fatou's lemma of integration theory

$$\begin{aligned} \liminf_{k\to\infty} \int_{\mathcal{D}} F(x, w_k(x))\, dx &\geq \int_{\mathcal{D}} \liminf_{k\to\infty} F(x, w_k(x))\, dx \\ &\geq \int_{\mathcal{D}} F(x, \liminf_{k\to\infty} w_k(x))\, dx \\ &= \int_{\mathcal{D}} F(x, w(x))\, dx. \end{aligned} \tag{2.34}$$

Note that the affine bound (2.31) permits the application of Fatou's lemma. By using the fact that the limit infimum of the sum of the two components of Φ numerically dominates the sum of the limit infima of the separate components as $k \to \infty$, together with the inequalities (2.33) and (2.34), we obtain $\Phi(w) \leq c$, which establishes that B is closed and Φ is lower semicontinuous.

The coerciveness is easily established by a comparison of (2.25) and (2.32). When this relation is combined with the affine bound (2.31), we conclude that, for some constant $C > 0$,

$$\liminf_{\|w\|\to\infty} \frac{\Phi(w)}{\|w\|^2} \geq C.$$

Note that this conclusion also makes use of the domination of the L_2 norm by the H^1 norm. The coerciveness relation (2.29) follows immediately. It follows that Φ, and analogously Ψ, has a minimum. These critical points, designated u and v, are weak solutions of (2.17) and (2.18), respectively. The standard technique is to set up the inequality,

$$\Phi(u) \leq \Phi(u \pm \epsilon\phi), \tag{2.35}$$

where ϕ is a test function restricted by the condition, $\phi \in L_\infty(\mathcal{D})$, and $\epsilon > 0$ is arbitrary. Taking limits gives the weak version of (2.17) for ϕ pointwise bounded. The case of a general test function follows by an approximation process. The case of (2.18) is similar.

The uniqueness of solutions of the decoupled system (2.17) and (2.18) follows directly from the monotonicity properties assumed for f and g.

2.3 Existence of Fixed Points

In Theorem 2.1, we presented the arguments which showed that T_0 is well defined, and acts invariantly upon K_0. In this subsection we shall prove that T_0 has a fixed point, which demonstrates that the system (2.1), (2.2), and (2.3) possesses a weak solution. Specifically, we shall prove the following.

Theorem 2.2 *The mapping T_0 is compact and continuous as a mapping on $L_2(\mathcal{D})$, and the closed convex set K_0 is invariant under T_0. In particular, by the Schauder fixed point theorem, T_0 has a fixed point in the closed convex set, K_0. Such a fixed point may be identified with a weak solution of the system (2.1), (2.2), and (2.3). In particular,*

$$\begin{aligned} \langle a\nabla u, \nabla\phi\rangle + \langle f(u,v),\phi\rangle &= 0, \quad (2.36)\\ \langle b\nabla v, \nabla\psi\rangle + \langle g(u,v),\psi\rangle &= 0, \quad (2.37)\end{aligned}$$

where $[u,v]$ is in $\prod_1^2 H^1(\mathcal{D})$ and $[\phi,\psi]$ is an arbitrary pair of test functions in $\prod_1^2 H_0^1(\mathcal{D})$.

Proof. We first establish continuity. Let $[\tilde{u}_1,\tilde{v}_1]$ and $[\tilde{u}_2,\tilde{v}_2]$ be arbitrary members of K_0 and designate their images under T_0 by $[u_1,v_1]$ and $[u_2,v_2]$, respectively. By estimating the difference of these images in $\prod_1^2 H^1$, we shall actually demonstrate the stronger conclusion that

- T_0 is uniformly continuous from K_0 to $\prod_1^2 H^1$.

The estimate proceeds from subtraction of the relevant weak relations defining the respective image points. Use is also made of the identities

$$f(u_1,\tilde{v}_1) - f(u_2,\tilde{v}_2) = [f(u_1,\tilde{v}_1) - f(u_2,\tilde{v}_1)] + [f(u_2,\tilde{v}_1) - f(u_2,\tilde{v}_2)] \quad (2.38)$$

and the corresponding identities for g. The test function identifications are

$$\phi = u_1 - u_2, \quad \psi = v_1 - v_2. \quad (2.39)$$

Altogether, one obtains the inequality

$$\|u_1 - u_2\|_{H^1}^2 \le C\|\tilde{v}_1 - \tilde{v}_2\|^2, \quad (2.40)$$

for some constant C, with a similar inequality for the v-differences. Here, we have used the fact that the first term in (2.38), when multiplied by $u_1 - u_2$, is nonnegative, and the fact that f is Lipschitz continuous on the range described by the maximum principles. Inequality (2.40), and its v-inequality equivalent, imply the uniform continuity part of the theorem. The compactness result follows from the same set of inequalities, or, alternatively, from a bound similar to that given by (3.7) below. The Rellich theorem then implies that this range has compact $\prod_1^2 L_2$ closure, which concludes the proof.

3 The Fixed Point and Numerical Fixed Point Maps

The map T, required to apply the operator calculus of [11], must be defined on an open set in function space. In this context the suitable space is $\prod_1^2 L_2(\mathcal{D})$. However, in the analysis of the fixed point mapping T_0 in Section 2, the assumption was introduced that the preimage, $[\tilde{u},\tilde{v}]$, satisfies the L_∞ bounds specified in (2.16), which the image $[u,v]$ was also shown to satisfy. Because the set K_0 is not open, we modify the definition of T_0 such that this assumption can be removed. To achieve this, we compose a T_0-like map with a truncation operator Tr, which leaves $[\tilde{u},\tilde{v}]$ unaffected within K_0. This necessitates certain pointwise hypotheses, however, which are briefly discussed in Section 5.3 below.

3.1 The Fixed Point Map

We introduce $\zeta_i \in C_0^\infty(\mathbb{R}),\ i = 1, 2,\ \ 0 \le \zeta_i \le 1$, such that *support* $\zeta_i = [0, \beta_i]$, and

$$\begin{aligned} \zeta_1(t) &= 1,\ \inf_{\partial\mathcal{D}} \bar{u} \le t \le \sup_{\partial\mathcal{D}} \bar{u}, \\ \zeta_2(t) &= 1,\ \inf_{\partial\mathcal{D}} \bar{v} \le t \le \sup_{\partial\mathcal{D}} \bar{v}. \end{aligned}$$

Finally, define

$$h_i(t) = t\zeta_i(t),\ \ i = 1, 2. \tag{3.1}$$

We shall define Ω to be an open ball centered at 0 in $\prod_1^2 L_2(\mathcal{D})$, so that it contains both the weak solution pair of (2.17) and (2.18), as well as the solution pair of the corresponding finite element equations. The latter are discussed in the next subsection. We take the radius of Ω to be any number greater than

$$R = \mid \mathcal{D} \mid^{1/2} \sqrt{\beta_1^2 + \beta_2^2},$$

where $\mid \mathcal{D} \mid$ designates the measure of $\mathcal{D}$. We further define,

$$Tr[\tilde{u}, \tilde{v}] := [h_1 \circ \tilde{u}, h_2 \circ \tilde{v}], \quad [\tilde{u}, \tilde{v}] \in \Omega. \tag{3.2}$$

Note that the range of Tr is contained in

$$K \cap \Omega, \tag{3.3}$$

where

$$K = \{[\tilde{u}, \tilde{v}] \in \prod_1^2 L_\infty(\mathcal{D}) : 0 \le \tilde{u} \le \beta_1,\ \ 0 \le \tilde{v} \le \beta_2\}. \tag{3.4}$$

The map T may be defined by

$$T = [U \circ Tr, V \circ Tr]. \tag{3.5}$$

- The component mappings U and V have domain given by (3.3) and are evaluated by the weak solution of (2.17) and (2.18), respectively, subject to the boundary conditions given in (2.3). They are Lipschitz continuous mappings on L_2 (cf. (2.40)). It is also assumed, consistent with regularity and domain considerations, that the range of U and of V is each contained in a bounded subset of H^2. This is essential for the Aubin-Nitsche hypothesis, cited in Section 3.3, to be consistent.

T is Lipschitz continuous. Moreover, since the range of T is contained in

$$K_0 \cap \Omega, \tag{3.6}$$

it follows that T is a proper extension of $T_0 \mid_{range\, T_0}$. The mapping T is also compact as the following argument shows. By simple substitution of the test functions $u - \bar{u}$ in the weak form of (2.17), and $v - \bar{v}$ in the weak form of (2.18), followed by routine estimation, we obtain the upper bound $R_{[u,v]}$ for the norm derived from (2.32), where its square is given by

$$R_{[u,v]}^2 = (\frac{\sup a}{\inf a})^2 \|\bar{u}\|_{H^1}^2 + (\frac{\sup b}{\inf b})^2 \|\bar{v}\|_{H^1}^2 + \frac{C_f}{\inf a} + \frac{C_g}{\inf b}, \tag{3.7}$$

and where

$$C_f = 2 \mid \mathcal{D} \mid \max\{\mid f(s,t) \mid : 0 \le s \le \beta_1,\ \ 0 \le t \le \beta_2\}[\beta_1 + \sup_{\mathcal{D}} \bar{u}], \tag{3.8}$$

$$C_g = 2 \mid \mathcal{D} \mid \max\{\mid g(s,t) \mid : 0 \le s \le \beta_1,\ \ 0 \le t \le \beta_2\}[\beta_2 + \sup_{\mathcal{D}} \bar{v}]. \tag{3.9}$$

The compactness follows from these estimates, together with the Rellich theorem.

3.2 The Discretized Model and the Finite Element Maps

In this section we introduce the piecewise linear finite element method which permits the construction of the components of the numerical fixed point map. These components are designated U_h and V_h and are approximations of U and V, as defined in Section 3.1. We also describe the associated approximation properties.

Let $\{\phi_i\}_1^N$ comprise a nodal basis of a given piecewise linear finite element space M_h. The functions of M_h are continuous, and linear in each simplex, S. As usual, $h = \max_S\{\text{diam } S\}$. It is required that the members of M_h vanish on the boundary of the polyhedral domain $\mathcal{D}$. Analogs of the affine flats $L_{\bar{u}}$ and $L_{\bar{v}}$, introduced in the definitions (2.20) and (2.21), are obtained through interpolation of boundary data. Accordingly, we define

$$L_{\bar{u}_I} = \{u_h \in H^1(\mathcal{D}) : [u_h - \bar{u}_I] \in M_h\}, \tag{3.10}$$

$$L_{\bar{v}_I} = \{v_h \in H^1(\mathcal{D}) : [v_h - \bar{v}_I] \in M_h\}, \tag{3.11}$$

where we have selected the piecewise linear interpolant $\bar{u}_I$ of $\bar{u}$, so that the finite element approximation of u is taken from $\bar{u}_I + M_h$, with a similar statement for v.

- The domain of each of the mappings U_h and V_h is taken to be the convex set specified in (3.3). Given $[\tilde{u}, \tilde{v}]$ in this set, we characterize the components

$$u_h = U_h([\tilde{u}, \tilde{v}]), \quad v_h = V_h([\tilde{u}, \tilde{v}]),$$

via solution of the equations

$$\langle a\nabla u_h, \nabla\phi_i\rangle + \langle f(u_h, \tilde{v}), \phi_i\rangle = 0, \quad \text{for} \quad i = 1, \cdots N, \tag{3.12}$$

and

$$\langle b\nabla v_h, \nabla\phi_i\rangle + \langle g(\tilde{u}, v_h), \phi_i\rangle = 0, \quad \text{for} \quad i = 1, \cdots N. \tag{3.13}$$

Here, $u_h \in L_{\bar{u}_I}$, $v_h \in L_{\bar{v}_I}$. It follows from the results of [10] that the component mappings just defined satisfy the same pointwise bounds as defined in (2.16). These discrete maximum principles require certain mesh hypotheses on the simplicial decomposition. These are discussed at length in [10] and are too detailed to repeat here. The explicit hypotheses follow.

- We assume that $\frac{1}{h}\text{diam } (S) \geq h_0 > 0$ as well as the discrete maximum principle:

$$[u_h, v_h] \in K_0.$$

By use of this principle, we may obtain the existence of fixed points in complete analogy with the analysis of Section 2. This section may now be closed by the formal definition of T_n:

$$T_n = [U_h \circ Tr, V_h \circ Tr]. \tag{3.14}$$

As with U and V, the mappings U_h and V_h are Lipschitz continuous on L_2; it follows that T_n is Lipschitz continuous.

3.3 Approximation Theory for the Finite Element Maps and Convergence Properties of T_n

Prior to describing the approximation properties of T_n, it is essential to discuss the linear approximation properties of the H_0^1 projection S_h onto M_h. For $H^2(\mathcal{D}) \cap H_0^1(\mathcal{D})$ functions, with uniform norm bound in this space, an estimate of this projection procedure is described adequately in [12]. The piecewise linear interpolant of an extension/smoothing process gives the requisite energy upper bound of order $O(h)$, but the smoothing should be done only in the tangential variables on $\partial\mathcal{D}$, so that the smoothed function also vanishes on $\partial\mathcal{D}$.

The next result is a generic result for gradient equations which will be used to deduce the approximation properties of U_h and V_h. Those of T_n follow. A proof may be found in [8].

Proposition 3.1 *Suppose $B(\cdot,\cdot)$ is a continuous symmetic bilinear form on H^1, L_2 coercive on H_0^1. For $u \in H^1$, let $\mathcal{F}(u)$ denote the continuous linear functional on H_0^1 defined by*

$$\mathcal{F}(u)(v) = \int_{\mathcal{D}} F(\cdot, u)v, \tag{3.15}$$

for F increasing in its second argument and $\frac{\partial F(\cdot,s)}{\partial s} \leq C$. Suppose that u and u_h satisfy the gradient relations

$$\begin{aligned} B(u,v) + \mathcal{F}(u)(v) &= \langle q, v\rangle, \quad \forall v \in H_0^1, &(3.16)\\ B(u_h, v_h) + \mathcal{F}(u_h)(v_h) &= \langle q, v_h\rangle, \quad \forall v_h \in M_h, &(3.17)\end{aligned}$$

where $u \in \bar{u} + H_0^1, u_h \in \bar{u}_I + M_h, \bar{u} \in C^2(\bar{\mathcal{D}})$. Here, $q \in L_2$ is prescribed. Then there exist constants C_1 and C_2, independent of h, such that

$$B(u-u_h, u-u_h) \leq C_1 \inf_{v_h \in M_h} B(u - \bar{u}_I - v_h, u - \bar{u}_I - v_h) + C_2 \|\bar{u} - \bar{u}_I\|_{H^1}^2. \tag{3.18}$$

In order to obtain an upper bound for (3.18), as well the order of approximation of P_n to be introduced later, we use the following inequality: (cf. [9, p. 85] and [13]):

$$\|D^\alpha(\bar{u} - \bar{u}_I)\|_{L_\infty} \leq \frac{Ch^{2-|\alpha|+1}}{\min_{i=1,\cdots,N} h_i} |\bar{u}|_{W^{2,\infty}}, \quad |\ \alpha\ | \leq 1. \tag{3.19}$$

Note that the second term on the r.h.s. of (3.18) is of order h^2, by use of this inequality. Since the first term is also of order h^2, the finite element approximation in (3.17) converges in the energy norm (2.32) with order h. We use this in conjunction with the following *hypothesis*:

- An adaptation of the Aubin-Nitsche duality argument, making use of the weak form (3.16), gives an L_2 approximation order of h^2, as in the linear theory, for the convergence of u_h to u. One requires the result for homogeneous boundary conditions, in terms of which the standard auxiliary problem with such data would be defined.

On the basis of this hypothesis, we may assume that there exists an approximation order for U_h and V_h:

$$\|(U - U_h)[\tilde{u}, \tilde{v}]\| \leq Ch^2, \quad \|(V - V_h)[\tilde{u}, \tilde{v}]\| \leq Ch^2, \tag{3.20}$$

for some constant C and $[\tilde{u}, \tilde{v}] \in \Omega$. We may now close this section with a description of the approximation properties of T_n.

Theorem 3.1 *The estimate,*

$$\|(T - T_n)[\tilde{u}, \tilde{v}]\| \leq Ch^2, \tag{3.21}$$

holds for some constant C, uniformly over the domain Ω on which T and T_n are defined. The approximation estimates are assumed as described in (3.20).

Proof. Immediate from the definitions and from the approximation estimates (3.20).

4 The Approximation Calculus and Newton's Method

As we have seen in the preceding sections, the model is formed by a system of two coupled partial differential equations (PDEs) for which maximum principles exist. A fixed point mapping T can be defined by solving each of these PDEs for its corresponding component and substituting these components in successive PDEs in a Jacobi fashion. By use of the maps U and V, it is possible to achieve complete decoupling, via gradient equations. Fixed points of such a mapping then coincide with solutions to the model.

In order to analyze piecewise linear finite element discretizations, a companion approximation map is induced if the variational procedure, inherent in defining the successive gradient equations, is taken over piecewise linear, finite dimensional affine subspaces. The fixed points of the companion map are clearly candidates for approximation of the fixed points of the solution map for the original system of PDEs. In this section, we deduce an approximation theory, described by two-sided estimates, for this discretization procedure. Our theory is based upon an operator calculus developed by Krasnosel'skii and his collaborators (cf. [11]), which we now develop.

4.1 The Krasnosel'skii Calculus

Given a fixed point x_0 of a smooth mapping T, a numerical approximation map T_n, and a linear projection map P_n, a theory is constructed to estimate $\|x_n - P_n x_0\|$ where $T_n x_n = x_n$. The manner in which the estimates are derived is to deduce a zero of the map $I - T_n$, in a ball centered at $P_n x_0$, by constructing an equivalent contraction map: The methodology involves derivative inversion and a mean value calculus. The result is stated as Theorem 4.1 below, and follows from the general Lemma 4.1. In our application of this theory we shall work with L_2 norms. Also, we shall provide a sketch of how to prove the result because some of these details will be required in Section 4.2.

Thus, let E be a Banach space, and suppose T is a mapping from an open set Ω in E into E. We assume the existence of a fixed point x_0 for T:

$$Tx_0 = x_0. \tag{4.1}$$

If $\{E_n\}$ denotes a sequence of linear subspaces of E of dimension $r(n) \geq n$, suppose that $T_n : \Omega_n \mapsto E_n, \Omega_n \subset E_n$, has a fixed point:

$$T_n x_n = x_n. \tag{4.2}$$

Finally, let $\{P_n\}$ be a family of linear projections of E onto E_n.

We examine the degree to which (4.2) approximates (4.1) by examining the size of the operators

$$R_n = T_n P_n - P_n T, \tag{4.3}$$

defined in E.

Our first convergence result is adapted from Theorem **19.1** in [11]:

Theorem 4.1 *Let the operators T and P_nT be Fréchet-differentiable in Ω, and T_n Fréchet-differentiable in Ω_n. Assume that (4.1) has a solution $x_0 \in \Omega$ and the linear operator $I - T'(x_0)$ is continuously invertible in E. Suppose T' is continuous at x_0 in the uniform operator topology; and,*

$$\|P_n(x_0) - x_0\| \to 0,$$
$$P_n x_0 \in \{x \in \Omega_n : \|x - x_0\| \le \delta_*\}, \quad n \ge n_*,$$
$$\|R_n x_0\| \to 0,$$
$$\|R_n'(P_n x_0)\| \to 0,$$

as $n \to \infty$. Finally, assume that for any $\epsilon > 0$ there exist n_ϵ and $\delta_\epsilon > 0$ such that

$$\|T_n'(x) - T_n'(P_n x_0)\| \le \epsilon \quad \text{for} \quad (n \ge n_\epsilon;\ \|x - P_n x_0\| \le \delta_\epsilon,\ x \in \Omega_n). \tag{4.4}$$

Then there exist n_0 and $\delta_0 > 0$ such that when $n \ge n_0$ equation (4.2) has a unique solution x_n in the set $\{x \in \Omega_n : \|x - x_0\| \le \delta_0\}$. Moreover,

$$\|x_n - x_0\| \le \|[I - P_n]x_0\| + \|x_n - P_n x_0\| \to 0 \quad \text{as} \quad n \to \infty, \tag{4.5}$$

and $\|x_n - P_n x_0\|$ satisfies the following two-sided estimate ($c_1, c_2 > 0$):

$$c_1 \|R_n x_0\| \le \|x_n - P_n x_0\| \le c_2 \|R_n x_0\|. \tag{4.6}$$

Note that in this theorem the actual rate of convergence depends only on the terms in the two sided estimate (4.6). The additional convergence assumptions need not hold with this same rate.

Proof. We present the major components. The proof proceeds in three steps, summarized here.

- There exist constants κ and κ' such that

  $$\|[I - T_n'(P_n x_0)]^{-1}\| \le \kappa, \quad \|I - T_n'(P_n x_0)\| \le \kappa', \tag{4.7}$$

  for $n \ge n_*$; here n_* is described in the statement of the theorem.

- The numbers α_n, for α_n defined by

  $$\alpha_n = \|[I - T_n'(P_n x_0)]^{-1}(I - T_n)P_n x_0\|, \tag{4.8}$$

  satisfy

  $$\frac{1}{\kappa'}\|R_n(x_0)\| \le \alpha_n \le \kappa \|R_n(x_0)\| \tag{4.9}$$

  for sufficiently large n.

- The statement of the theorem concerning x_n, n_0, and δ_0 holds, and we have the bounds

$$\frac{\alpha_n}{1+q} \leq \|x_n - P_n x_0\| \leq \frac{\alpha_n}{1-q}, \tag{4.10}$$

  for some $0 < q < 1$.

The inequalities (4.9) in the second item follow from routine calculations, while the first item, and the associated inequalities contained in (4.7), follow from systematic use of successive perturbation, beginning with $I - T'(x_0)$ (cf. Lemma 4.2), and measurement of the perturbation via the assumption on R'_n. The third item above is a consequence of Lemma 4.1, stated at the conclusion of the proof (cf. [11, Lemma 19.1]). The identifications

$$A = I - T_n, \quad x_* = P_n x_0, \quad X = E_n,$$

are made.

If the hypothesis (4.4) of the theorem is employed with $\epsilon_0 = q/\kappa$, $0 < q < 1$ arbitrary, then the first hypothesis of the lemma is satisfied for some $\delta_0 \leq \delta_*$; the second hypothesis is satisfied for n_0 sufficiently large by the second relation in (4.9). Since the bounds of the theorem follow from the conjunction of (4.10) and (4.9), the proof is completed. Note that, by selecting δ_0 sufficiently small, we may assume that x_n is the unique element in Ω_n within a distance of δ_0 from x_0.

Lemma 4.1 *Let A be an operator in a Banach space X which is Fréchet differentiable in a closed ball centered at x_*. Suppose $[A'(x_*)]^{-1}$ exists as a bounded linear operator, and that*

$$\sup_{\|x - x_*\| \leq \delta_0} \|[A'(x_*)]^{-1}[A'(x) - A'(x_*)\| \leq q, \tag{4.11}$$

$$\alpha := \|[A'(x_*)]^{-1} A(x_*)\| \leq \delta_0(1-q) \tag{4.12}$$

for some δ_0 and $0 < q < 1$. Then the equation $Ax = 0$ has a unique solution x_0 in the ball satisfying the estimate

$$\frac{\alpha}{1+q} \leq \|x_0 - x_*\| \leq \frac{\alpha}{1-q}. \tag{4.13}$$

4.2 Approximate Fixed Points via Newton's Method

We consider the extent to which the theory of the preceding subsection persists when the fixed points are computed by a systematic approximation procedure, viz. , when Newton's method is applied to the numerical map, T_n. What is interesting is that the hypotheses, which account for the success of the Krasnosel'skii calculus, also guarantee a corresponding replacement theory in terms of Newton's method.

In order to set the stage for a detailed study, we briefly summarize the essential properties allowing for an R-quadratically convergent Newton iteration. For conciseness, we set $H_n = I - T_n$. Then we require:

1. H'_n is Lipschitz continuous on its domain, with Lipschitz constant $L^{H'_n} \leq L$.

2. The family of inverses of H'_n is uniformly bounded in norm, say, by κ_*.

3. The initial residual, $H_n(u^0)$, does not exceed in norm the quantity, $[2L\,\kappa_*^2]^{-1}$.

The derivation of the convergence result under these hypotheses is described in [4, Section 2], with slight changes. Note that an exact Newton method for T_n is analyzed, which indirectly, via the Krasnosel'skii framework, translates into an approximate Newton method for T. In the items listed above, the reader will find a striking overlap with the previous subsection; what may appear to be missing there is a condition guaranteeing the sufficiently small residual required for R-quadratic convergence. What is remarkable, however, is that the Newton iterates converge q-linearly, under the hypotheses of the preceding subsection. It follows that such linear convergence, stated as R-linear convergence in Theorem 4.2, will eventually guarantee the residual condition required for R-quadratic convergence.

We state now the perturbation lemma used in the proof of Theorem 4.1, as well as the linear convergence theorem to follow.

Lemma 4.2 *Suppose that A and B are bounded linear operators on a Banach space X such that A^{-1} exists with $\|A^{-1}\|\|B\| < 1$. Then $A+B$ is invertible and an inverse bound is given by*

$$\|[A+B]^{-1}\| \leq \frac{\|A^{-1}\|}{1-\|A^{-1}\|\|B\|}. \tag{4.14}$$

Theorem 4.2 *Suppose that the bounds (4.7) hold and that the choices, $\epsilon = q/\kappa$, δ_0, and n_0 are made in (4.4), as in the proof of Theorem 4.1, for fixed $0 < q < 1$. Thus, for $n \geq n_0$, let x_n be an appropriate uniquely determined fixed point of T_n within the set, $\mathcal{U}_n = \{x \in \Omega_n : \|x - x_0\| \leq \delta_0\}$. Then Newton's method, beginning with any $u^0 \in \mathcal{U}_n$ within distance $\delta_0/2$ of x_0, such that the first iterate satisfies*

$$\|u_n^1 - u^0\| \leq \frac{1}{2}(1-q_*)\delta_0,$$

where

$$q_* := 2\frac{q}{1-q} < 1,$$

is contractive, and the estimate,

$$\|x_n - u_n^k\| \leq \frac{q_*^k}{1-q_*}\|u_n^1 - u^0\|, \tag{4.15}$$

holds for the Newton sequence,

$$u_n^{k+1} - u_n^k = -[H_n'(u_n^k)]^{-1}H_n(u_n^k). \tag{4.16}$$

In this case, the entire Newton sequence is in $\mathcal{U}_n$.

Proof. By use of Lemma 4.2, we may conclude the existence of a uniform bound on the inverse derivative mappings given by

$$|[H_n'(x)]^{-1}\| \leq \frac{\kappa}{1-q}, \quad x \in \mathcal{U}_n. \tag{4.17}$$

By use of the definition of the Newton increment, (4.16), we immediately obtain the estimate,

$$\|u_n^{k+1} - u_n^k\| \leq \frac{\kappa}{1-q}\|H_n(u_n^k)\|,$$

where we have used (4.17). In order to estimate the residual term, we employ the integral representation,

$$H_n(u_n^k) = \int_0^1 [H_n'(u_n^{k-1} + t(u_n^k - u_n^{k-1})) - H_n'(u_n^{k-1})](u_n^k - u_n^{k-1})\, dt, \tag{4.18}$$

and estimate (4.18) by use of the inequality,

$$\|H_n'(x) - H_n'(y)\| \leq \frac{2q}{\kappa},$$

which follows from the hypotheses, via the triangle inequality, for x and y in $\mathcal{U}_n$. We obtain, finally,

$$\|u_n^{k+1} - u_n^k\| \leq \frac{2q}{1-q}\|u_n^k - u_n^{k-1}\|. \tag{4.19}$$

By the definition of q_*, and the repeated use of (4.19), we obtain a standard Cauchy sequence estimate for $\|u_n^l - u_n^k\|$. Passage to the limit then yields (4.15). Note that here we have used the uniqueness of x_n in $\mathcal{U}_n$ and the behavior of the residuals as estimated in the course of the proof.

The R-quadratic convergence estimate is based upon the following result, which is quoted from [4].

Lemma 4.3 *Suppose that the initial residual satisfies*

$$\|H_n(u^0)\| \leq \rho^{-1}. \tag{4.20}$$

For the Newton sequence defined by (4.16), suppose that the inequalities,

$$\|u_n^k - u_n^{k-1}\| \leq \kappa_*\|H_n(u_n^{k-1})\|, \quad k \geq 1, \tag{4.21}$$

$$\|H_n(u_n^k)\| \leq \frac{\eta\rho}{2}\|H_n(u_n^{k-1})\|^2, \quad k \geq 1, \tag{4.22}$$

hold for some $\eta \leq \frac{1}{2}$. Then the convergence is described by the error estimate,

$$\|x_n - u_n^k\| \leq \frac{\theta_k\kappa_*}{\eta\rho}\Big(\prod_{j=0}^{k}\tau_j^{2^{k-j}}\Big)\frac{(1-\sqrt{1-2\eta})^{2^k}}{2^k}. \tag{4.23}$$

Here, $\{\theta_k\}$ and $\{\tau_k\}$ are decreasing sequences bounded by 1.

The manner in which this result is used is similar to the structure of the proof of Theorem 4.2. There, the Newton increment was estimated by an inverse bound, and the residual was estimated by the integral representation (4.18). Here, the inverse estimate is already built into (4.21), and a sharpened version of (4.19) will be employed, which uses the Lipschitz continuity of the differentiated map. In fact, we have the following result.

Theorem 4.3 *Define*

$$\eta := \frac{L\kappa_*^2}{\rho} \tag{4.24}$$

and suppose that $\eta \leq \frac{1}{2}$. Under the hypothesis (4.20), it follows that the conditions (4.21) and (4.22) hold. In particular, the R-quadratic convergence estimate (4.23) holds with $\eta\rho$ as defined in (4.24) above.

Proof. Since (4.21) is immediate, it remains to verify (4.22) with the choice of $\eta\rho$ as defined in (4.24). For this end, we use the representation (4.18) as in the proof of Theorem 4.2. By use of the full Lipschitz continuity of H_n', we are able to deduce the stronger result (4.22).

We have now developed a tight linearization theory. We summarize the essential features as follows.

1. Newton's method, based upon the approximate map H_n, is globally convergent at the level of the calculus developed in the previous subsection. In particular, there is a systematic procedure for determining x_n approximately.

2. The convergence is at least linear, as described in Theorem 4.2. The switch to quadratic convergence takes place no later than when the residual condition (4.20) is met. The latter convergence is described by Theorem 4.3.

3. Only as many Newton iterates are required to approximate x_n as matches the approximation estimate for the dispersion between $P_n x_0$ and x_n as given in (4.6).

5 Representation and Properties of the PDE System Maps

The theory of the previous section requires certain properties to be satisfied for the fixed point and numerical fixed point maps. Some of these were presented in section 3. The remaining will be discussed in this section together with the issue of computability for the approximation scheme. Ultimately, both of these issues reduce to representations derived and examined below.

5.1 Representation for T' and the Eigenvalue Hypothesis

The invertibility of the operator, $I - T'(x_0)$, the major hypothesis of Section 4, is equivalent, for T defined in (3.5), to the following. At a fixed point $[u, v]$ of T,

$$1 \notin sp(T'[u, v]). \tag{5.1}$$

This is a standard result of the resolvent calculus of compact operators (cf. [14]). In this subsection, we shall describe the analytical condition which guarantees that this eigenvalue condition holds.

Suppose for convenience that we represent T by the composition mapping (cf. (3.5)):

$$T = S \circ Tr, \tag{5.2}$$

where $S = [U, V]$. By the chain rule, we have

$$T'[\tilde{u}, \tilde{v}] = S'(Tr[\tilde{u}, \tilde{v}]) \circ T'r[\tilde{u}, \tilde{v}] = S'[h_1 \circ \tilde{u}, h_2 \circ \tilde{v}](h_1' \circ \tilde{u}, h_2' \circ \tilde{v}). \tag{5.3}$$

$T'r$ acts via the multipliers h_i', while S' may be displayed by the following matrix tableau:

$$\left[\begin{array}{c|c} 0 & U_v \\ \hline V_u & 0 \end{array}\right]. \tag{5.4}$$

Here, we have anticipated the fact that $U_u \equiv 0$ and $V_v \equiv 0$. The expressions for the remaining operator partials are readily computed, after application to the test pair (ϕ, ψ), to be

$$\omega = U_v[u^*, v^*](\psi) = [-\nabla\cdot(a\nabla) + f_u(U(u^*, v^*), v^*)]^{-1}(-f_v(U(u^*, v^*), v^*)\psi), \quad (5.5)$$

$$\chi = V_u[u^*, v^*](\phi) = [-\nabla\cdot(b\nabla) + g_v(u^*, V(u^*, v^*))]^{-1}(-g_u(u^*, V(u^*, v^*))\phi). \quad (5.6)$$

For use in the computations to follow in this subsection, we note that the map Tr is the identity on fixed points of T. We now proceed to the eigenvalue hypothesis. Let us suppose that

$$T'[u, v](\phi, \psi) = (\phi, \psi), \quad (5.7)$$

where $[u, v]$ is a fixed point of T and $(\phi, \psi) \in H_0^1(\mathcal{D})$. Making use of (5.3), (5.4), (5.5), and (5.6), we obtain the relations

$$\phi = [-\nabla\cdot(a\nabla) + f_u(u, v)]^{-1}(-f_v(u, v)\psi), \quad (5.8)$$

$$\psi = [-\nabla\cdot(b\nabla) + g_v(u, v)]^{-1}(-g_u(u, v)\phi), \quad (5.9)$$

as equivalent to the equation (5.7).

- The eigenvalue hypothesis, and the consequent invertibility property, are implied by the hypothesis that the solutions of (5.8), (5.9) are exhausted by $\phi = 0$, $\psi = 0$, and hence (ϕ, ψ) cannot be an eigenvector.

5.2 Implementation of Newton's Method

In this subsection, we consider the explicit inversion of the mappings $H_n' = I - T_n'$, required for the application of Newton's method as a computational procedure to determine the fixed points of T_n. In the next subsection, we shall discuss the pointwise hypotheses which underlie the application of the theory. In the final subsection, we shall complete our theoretical analysis of the hypotheses of the Krasnosel'skii/Newton theory developed in the previous section. The general theory relies upon the existence and uniform boundedness of $[H_n']^{-1}$ and it is to these we now turn. We begin with the representation for H_n'. First, we set

$$A = (U_h)_v \circ T'r, \quad B = (V_h)_u \circ T'r.$$

Then

$$H_n' = \left[\begin{array}{c|c} I & -A \\ \hline -B & I \end{array}\right]. \quad (5.10)$$

The inverse of (5.10) is easily computed to be

$$[H_n']^{-1} = \left[\begin{array}{c|c} (I - AB)^{-1} & A(I - BA)^{-1} \\ \hline B(I - AB)^{-1} & (I - BA)^{-1} \end{array}\right]. \quad (5.11)$$

Motivated by Neumann series considerations, we consider the approximation of the inverse operator matrix (5.11) by

$$\left[\begin{array}{c|c} I & A \\ \hline B & I \end{array}\right]. \quad (5.12)$$

The remaining remarks of this subsection are summary in nature, and are not intended to substitute for a careful analysis, which is outside the scope of this paper. By direct computation, one sees that the product of (5.10) and (5.12) is given by

$$\left[\begin{array}{c|c} I - AB & 0 \\ \hline 0 & I - BA \end{array}\right]. \tag{5.13}$$

Theories which deal with approximate Newton methods of this type have been considered in [6]. Specifically, the difference between the identity and (5.13) must be of the order of the residual in order to maintain R-quadratic convergence.

We shall close the subsection by noting that the action of $(U_h)_v$ and $(V_h)_u$ is readily determined by replacing U and V in (5.5), (5.6) by U_h and V_h, and then computing the finite element solution of the homogeneous Dirichlet boundary value problems.

5.3 Pointwise Hypotheses

In this subsection, we shall summarize the assumptions made in this paper at the level of pointwise approximation and stability (both in the sense of L_∞). A very careful investigation of these questions will appear in the monograph [7], for models more complicated and inclusive than the one considered here. Therefore, for reasons of economy, we simply state the properties assumed.

The reason for the necessity of pointwise properties is that the truncation map fails to be Lipschitz continuously differentiable on L_2, or its open subsets. Therefore, the Krasnosel'skii theory does not directly apply, and a modification is required, in which closed subsets of the estimation balls replace the balls themselves. The required modifications of this theory are elaborated in [7]. When the modified theory is applied, it is found that L_∞ subsets must be considered, to permit the truncation map to be appropriately smooth. In particular, this allows the regularization map, employed in the proof of Lemma 4.1, to be invariant on such a set, and thus to possess a (numerical) fixed point near the fixed point of T. The additional hypotheses are as follows.

1. T_n and P_n are pointwise convergent (in the sense of L_∞) at the fixed point.

2. T_n and P_n are stable (pointwise norm) operator sequences. This is implied by underlying discrete maximum principles.

3. Items (1) and (2) of the next subsection hold for the pointwise norm.

We remark, in closing, that the L_∞ version of hypothesis (1) of the next subsection is a direct consequence of elliptic regularity theory, in particular, the technique of Moser iteration (cf. [2, Section 8.5]). We now proceed to the final subsection.

5.4 Main Result

In this final subsection of the paper, we state the main result and then provide the verification of the remaining hypotheses.

Theorem 5.1 *Assume the eigenvalue hypothesis, (5.1), and the L_∞ hypotheses of the previous subsection. For the fixed point and numerical fixed point mappings defined by (3.5) and (3.14), the hypotheses of the Krasnosel'skii Calculus (cf. Theorem 4.1) and of the Newton calculus (cf. (1) and (2) of Section 4.2) hold. The truncation error specified by the two-sided estimate (4.6) is of order h^2 in L_2, and hence the exact numerical scheme converges with order h^2. The linear/quadratic Newton iteration of (4.16) is continued, as dictated by (4.15) and (4.23), until this resolution is attained.*

Proof. The following identifications remain to be made:

$$E \to \prod_1^2 L_2(\mathcal{D}), \quad E_n \to span\{\bar{u}_I \otimes \bar{v}_I, \prod_1^2 M_h\},$$

with P_n the L_2 projection onto E_n. By (3.19) and the duality lemma applied to S_h, one concludes that the order of convergence of P_n to I, on bounded subsets of $\prod_1^2 H^2$, is $O(h^2)$.

By use of the triangle inequality, one sees that $R_n(x_0)$ is of order $O(h^2)$:

$$\|R_n(x_0)\| \leq \|T_n P_n x_0 - T_n x_0\| + \|T_n x_0 - T x_0\| + \|T x_0 - P_n T x_0\|. \tag{5.14}$$

Indeed, the Lipschitz property of T_n is used for the first term of (5.14), while the approximation of T by T_n is used for the second term. The remaining hypotheses are implied by the following two properties:

1. T' is continuous in the operator topology on Ω_n.
2. T' is uniformly approximated by T_n' on Ω_n. In particular, $\|T_n'(P_n x_0)\|$ is bounded.

Note that the condition (4.4) is then implied and the estimation of $\|R_n'(x_0)\|$ proceeds via the identity:

$$R_n'[P_n x_0](\phi) = \{T_n'[P_n x_0] - P_n T_n'[P_n x_0]\}(P_n \phi) + \{P_n T_n'[P_n x_0] - P_n T'[P_n x_0]\}(P_n \phi). \tag{5.15}$$

The proof is complete, subject to the proof of (1) and (2) above. This is isolated in the following lemma.

Lemma 5.1 *The continuity and uniform approximation items (1) and (2) are valid.*

Proof. The uniform approximation property (2) may be established by use of the lifting operators, which were employed systematically in [3], applied to the matrix operator (5.4). We shall use these to describe the operator approximation of U_v by $(U_h)_v$; the approximation of V_u is similar. To describe the procedure, let J denote the Riesz operator, which functions as the inverse of

$$-\nabla \cdot (a \nabla) : H_0^1(\mathcal{D}) \mapsto H^{-1}(\mathcal{D}). \tag{5.16}$$

If E_h denotes the orthogonal projection when H_0^1 employs the inner product defined by the operator of (5.16), then set $J_h = E_h J$. It is shown in [3] that J_h functions as the inverse finite element map. More precisely, if we form the difference, $JU_v - J_h(U_h)_v$, then we obtain the following representation, where ω and ψ are specified in (5.5), where we have made obvious abbreviations, and where ω_h is the finite element analog of ω:

$$(\omega - \omega_h) + J\{f_u(\omega - \omega_h)\} =$$

$$J(f_u^h\omega_h - f_u\omega_h) + (J_h - J)(f_u^h\omega_h) + J(f_v^h\psi - f_v\psi) + (J_h - J)(f_v^h\psi). \tag{5.17}$$

Two important properties of J, developed in [3, Ch. 1], are its pointwise nonnegativity on L_2, and the fact that the bilinear form (Jy, z) on L_2 may be estimated via H^{-1} norms of y and z.

The representation (5.17) is now multiplied by $\omega - \omega_h$ and integrated; the second term on the left hand side of the resultant is nonnegative by the pointwise nonnegativity of the operator J. If one uses the H^{-1} estimation specified above, the Lipschitz properties of f_u and of f_v, the continuous and discrete maximum principles, the L_2 approximation order of $O(h^2)$ for the approximation of J by J_h, and the order of approximation of U by U_h, one obtains the second item. There is one subtle point: the estimation of the third term on the r.h.s. of (5.17) requires knowledge that the functions ψ may be assumed bounded by β_1, because of the truncation operator. Item (1) is verified by a technique familiar in the resolvent calculus. We simply state the identity, since the estimation is routine. In order to use an abbreviated notation, write (5.5) (with an equivalent identity for (5.6)) as

$$\omega_1 = R_1(-f_v^1\psi_1), \;\; \omega_2 = R_2(-f_v^2\psi_2),$$

when two distinct evaluation points for T', $[\tilde{u}_1, \tilde{v}_1]$ and $[\tilde{u}_2, \tilde{v}_2]$, are compared. Item (1) follows, via operator boundedness and Lipschitz properties, from the fundamental identity,

$$\omega_1 - \omega_2 = R_1[f_u^2 - f_u^1]R_2(-f_v^1\psi_1) + R_2(f_v^2\psi_2 - f_v^1\psi_1). \tag{5.18}$$

References

[1] Ivar Ekeland and Roger Temam. *Convex Analysis and Variational Problems*. North Holland and American Elsevier, 1976.

[2] D. Gilbarg and N. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Springer-Verlag, New York, 1977.

[3] Joseph W. Jerome. *Approximation of Nonlinear Evolution Systems*. Academic Press, 1983.

[4] Joseph W. Jerome. Approximate Newton methods and homotopy for stationary operator equations. *Constructive Approximation*, 1:271–285, 1985.

[5] Joseph W. Jerome. Consistency of semiconductor modeling: An existence/stability analysis for the stationary Van Roosbroeck system. *SIAM J. Appl. Math.*, 45:565–590, 1985.

[6] Joseph W. Jerome. Newton's Method for Gradient Equations Based Upon The Fixed Point Map. *Numerische Mathematik*, 55:619–632, 1989.

[7] Joseph W. Jerome. *Analysis of Charge Transport*. Springer, Berlin, 1996.

[8] Joseph W. Jerome and Thomas Kerkhoven. A Finite Element Approximation Theory for the Drift-Diffusion Semiconductor Model. *SIAM J. Numer. Anal.*, 28:403–422, 1991.

[9] Claes Johnson. *Numerical Solutions of Partial Differential Equations by the Finite Element Method*. Cambridge University Press, 1987.

[10] Thomas Kerkhoven and Joseph W. Jerome. L_∞ Stability of Finite Element Approximations to Elliptic Gradient Equations. *Numerische Mathematik*, 57:561–575, 1990.

[11] M. A. Krasnosel'skii, G. M. Vainikko, P. P. Zabreiko, Ya. B. Rutitskii, and V. Ya. Stetsenko. *Approximate Solution of Operator Equations*. Wolters-Noordhoff, 1972.

[12] G.W. Strang. Approximation in the finite element method. *Numerische Mathematik*, 19:81–98, 1972.

[13] G.W. Strang and G. Fix. *An Analysis of the Finite Element Method*. Prentice-Hall, Englewood Cliffs, N.J., 1973.

[14] Angus E. Taylor. *Introduction to Functional Analysis*. Wiley, 1961.