A Trapping Principle and Convergence Result for Finite Element Approximate Solutions of Steady Reaction/Diffusion Systems*

Joseph W. Jerome†

Abstract

We consider nonlinear elliptic systems, with mixed boundary conditions, on a convex polyhedral domain $\Omega \subset \mathbf{R}^N$. These are nonlinear divergence form generalizations of $\Delta \mathbf{u} = \mathbf{f}(\cdot\,, \mathbf{u})$, where $\mathbf{f}$ is outward pointing on the trapping region boundary. The motivation is that of applications to steady-state reaction/diffusion systems. Also included are reaction/diffusion/convection systems which satisfy the Einstein relations, for which the Cole-Hopf transformation is possible. For maximum generality, the theory is not tied to any specific application. We are able to demonstrate a trapping principle for the piecewise linear Galerkin approximation, defined via a lumped integration hypothesis on integrals involving $\mathbf{f}$, by use of variational inequalities. Results of this type have previously been obtained for parabolic systems by Estep, Larson, and Williams, and for nonlinear elliptic equations by Karátson and Korotov. Recent minimum and maximum principles have been obtained by Jüngel and Unterreiter for nonlinear elliptic equations. We make use of special properties of the element stiffness matrices, induced by a geometric constraint upon the simplicial decomposition. This constraint is known as the non-obtuseness condition. It states that the inward normals, associated with an arbitrary pair of an element's faces, determine an angle with nonpositive cosine. Drăgănescu, Dupont, and Scott have constructed an example for which the discrete maximum principle fails if this condition is omitted. We also assume vertex communication in each element in the form of an irreducibility hypothesis on the off-diagonal elements of the stiffness matrix. There is a companion convergence result, which yields an existence theorem for the solution. This entails a consistency hypothesis for interpolation on the boundary, and depends on the Tabata construction of simple function approximation, based on barycentric regions.

Keywords Steady reaction/diffusion systems, variational inequalities, piecewise linear Galerkin approximations, trapping regions, existence, convergence, non-obtuseness condition, irreducible stiffness matrix

AMS classification numbers 35J65, 65N15, 65N30

*This work was supported by the National Science Foundation under grant DMS-0311263.

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

1 Introduction.

Invariant regions, or their generalizations described as trapping regions, have played a significant role in the theory of nonlinear reaction/diffusion systems (see [23] for earlier analytical results; for more recent analytical results, see [13, 3, 4]). Companion results for Galerkin approximations have appeared since the 1970s; however, results for systems, particularly steady systems, are less frequent (see the memoir by D. Estep, M. Larson, and R. Williams [7, Theorem 5.13] for the non-steady case; the non-obtuseness condition is employed by these authors, who also use the maximum principles to improve 'a posteriori' error estimates). For elliptic equations, an important early result was derived by P. Ciarlet and P. Raviart in [5]; see also [16]. There has been substantial recent progress; A. Jüngel and A. Unterreiter extend the scope of [16] in [14], where minimum and maximum principles are derived for non-monotone reaction terms. J. Karátson and S. Korotov [15, Theorems 8, 11] demonstrate the central role of the non-obtuseness condition for quasi-linear elliptic equations. The possible failure of the discrete maximum principle in the obtuse case has been described recently by A. Drăgănescu, T. Dupont, and R. Scott [6] in two dimensions by use of discrete Green's functions. This clearly indicates that the non-obtuseness condition is not simply an artifice of proof. It is not a strict necessary condition, however, as was first observed in [21].

In this paper, for a convex polyhedral spatial domain $\Omega \subset \mathbf{R}^N$, we derive trapping region results for piecewise linear Galerkin approximations defined for nonlinear steady reaction/diffusion systems with mixed boundary conditions. The vector fields are outward pointing on the trapping region boundary, but are not assumed monotone. For the simplicial decompositions, we require the non-obtuseness condition, rather than the acuteness condition. It states that the inward normals, associated with an arbitrary pair of an element's faces, determine an angle with nonpositive cosine (see the following section for elaboration). The interpretation of this condition for the element stiffness matrix is that all off-diagonal elements are nonpositive; strict negativity (acuteness) is not required. This is significant, particularly in three dimensions, since non-obtuseness can be preserved under local and global mesh refinement, as has been demonstrated by S. Korotov and M. Křížek [17] via the '3D yellow refinement'. Since mixed boundary conditions are permitted in the present paper, such mesh refinement is of practical necessity in the neighborhood of portions of the boundary where the character of the boundary condition switches, and gradient singularities may develop. The situation is simpler in two dimensions where the '2D red refinement' may be used to preserve acuteness and/or non-obtuseness. The geometry of individual simplicial elements also enters through the irreducibility hypothesis on the stiffness matrix; see Lemma 2.1 to follow for its use. Irreducibility has the *geometric* interpretation of the existence of a chain of pairs of inward normals, with each pair defining an angle with strictly negative cosine; the chain originates and terminates at a pair of arbitrary distinct faces in the element. When interpreted *algebraically*, the irreducibility hypothesis has been employed by R. Varga in the study of relaxation meth-

ods, particularly for p-cyclic iteration matrices [26]. It is closely connected to M-matrices. The latter have also been employed in the study of maximum principles (see [25], where Delaunay triangulations are used). This strongly suggests the connection between analytical stability and iterative convergence. The irreducibility hypothesis is automatically satisfied for meshes in one dimension. The actual nonlinear algebraic system for the nodal coefficients appears as (10) below. The formulation includes a lumped integration rule motivated by M. Tabata's studies [24].

Analytical proofs are transferred below to the algebraic systems for the nodal coefficients, via the use of variational inequalities. This mirrors the corresponding analytical technique in [13], where the invariant region is enforced through the variational inequality, then the constraints are seen to be non-binding via the outward pointing vector field, which is an abstract statement of the existence of lower and upper solutions. In the finite-dimensional case, there is less flexibility in the choice of test functions; thus, one can derive the nodal equations from the variational inequalities, via the use of a consistent quadrature formula for the evaluation of the integrals involving the vector field. The use of the quadrature formula is required to utilize the outward pointing vector field. The fundamental Theorem 2.1 of this paper is used to deduce Theorem 3.1: the existence of an analytical solution of the PDE system within the trapping region via convergence. Such convergence requires additional regularity of the vector field. Note that the paper contains no uniqueness results. For clarity, we number the assumptions consecutively as they appear in the paper: H1, H2, etc. They are found in the introduction as well as in sections two and three. We discuss at the conclusion of the paper general types of systems to which the theory applies. These include reaction/diffusion/convection systems for which the Einstein relations are valid. The author has studied these in the context of semiconductor modeling [12].

For a convex polyhedral domain $\Omega \subset \mathbf{R}^N$, consider the nonlinear steady-state system of order m, for $\mathbf{u} = (u_1, \ldots, u_m)^t$,

$$\nabla \cdot p_k(\vec{x}, \mathbf{u}) \nabla u_k = f_k(\vec{x}, \mathbf{u}),\ k = 1, \ldots, m. \tag{1}$$

The divergence structure defined by $\mathbf{p} = (p_1, \ldots, p_m)$, and that of the vector field $\mathbf{f} = (f_1, \ldots, f_m)^t$, are significant. We begin here the tabulation of the properties to be satisfied as well as the statement of the system boundary conditions.

System Properties: Hypotheses

H1. Carathéodory Mapping (Property CM) Given a slab,

$$Q = \prod_{j=1}^{m} [a_j, b_j],\ a_j < b_j, j = 1, \ldots, m,$$

in $\mathbf{R}^m$ and the Cartesian product, $\mathcal{D} = \Omega \times Q$, $\mathbf{g}$ is said to be a Carathéodory mapping (CM) on $\mathcal{D}$ if each component $g = g_j$ satisfies:

- $g(\cdot\,, \mathbf{z})$ is measurable in its first argument for each fixed $\mathbf{z} \in Q$.

- $g(\vec{x}, \cdot\,)$ is continuous in its second argument for almost every $\vec{x} \in \Omega$ (a. e.).

The vector functions $\mathbf{p}, \mathbf{f}$ are assumed to satisfy property (CM).

H2. f Outward Pointing on the Boundary of Q (Property OP).

For each $\mathbf{u} \in Q$, if $u_j = a_j,\ j = 1, \ldots, m,$ then $f_j(\cdot\,, u_1, \ldots, u_m) \leq 0,$ a. e. in Ω;

for each $\mathbf{u} \in Q$, if $u_j = b_j,\ j = 1, \ldots, m,$ then $f_j(\cdot\,, u_1, \ldots, u_m) \geq 0,$ a. e. in Ω.

H3. L_2 Local Boundedness of f (Property B) There is a constant C such that, if $\mathbf{u}$ is a real measurable function on Ω with range in Q, a. e. , then $\|\mathbf{f}(\cdot\,, \mathbf{u})\|_2 \leq C$.

H4. Divergence Structure (Property DS) . It is assumed further that

$$\mathbf{p} : \mathcal{D} \mapsto Q^*, Q^* = \prod_{j=1}^{m} [a_j^*, b_j^*],\ 0 < a_j^*, j = 1, \ldots, m.$$

We describe now the system boundary conditions.

Mixed System Boundary Conditions.

i. Dirichlet Boundary. There is a (nonempty, relatively open) boundary component Σ_D such that the restriction of $\mathbf{u}$ to Σ_D agrees in the trace sense with a smooth function $\hat{\mathbf{u}} \in C^1(\bar{\Omega})$, with *range* in Q:

$$\Gamma(\mathbf{u} - \hat{\mathbf{u}})|_{\Sigma_D} = 0. \tag{2}$$

Here, Γ denotes the trace operator (see [19] for basic properties).

ii. Neumann Boundary. The normal derivative of $\mathbf{u}$ vanishes in a weak sense on the complement Σ_N of Σ_D with respect to $\partial\Omega$. This is a natural boundary condition subsumed in the weak formulation below. It is expressed:

$$\frac{\partial \mathbf{u}}{\partial \nu} = 0, \text{ on } \Sigma_N. \tag{3}$$

Relation (3) has the interpretation in terms of an abstract Green's theorem (cf. [22, p. 165]), and reduces, when $\mathbf{u}$ is smooth and when the standard Green's formula applies with positive multipliers p_k, to the classical characterization that components have zero directional derivatives along the outward boundary normal vector $\vec{\nu}$ at non-corner points of Σ_N. We recall two useful results.

Lemma 1.1. *If* $\mathbf{g}$ *satisfies hypothesis* H1*, and* $\mathbf{v}$ *is a measurable function defined on* Ω*, with range in* Q*, then the composition* $\mathbf{g}(\cdot\,, \mathbf{v}(\cdot\,))$ *is measurable on* Ω.

(see [11] for underlying concepts)

Lemma 1.2. *For a Carathéodory mapping* $\mathbf{h}$*, if* $\mathbf{h}$ *maps a subset of* $(L_q(\Omega))^m$ *into* $(L_p(\Omega))^m$*, for* $1 \leq p, q < \infty$*, via the composition* $\mathbf{H}(\mathbf{v}) = \mathbf{h}(\cdot\,, \mathbf{v})$*, then* $\mathbf{H}$ *defines a continuous mapping from* $(L_q)^m$ *to* $(L_p)^m$*.*

This was shown by Krasnosel'skii [18, Theorem 2.1, p. 22]. We apply this result to $\mathbf{p}$ and to $\mathbf{f}$ in the following sections. Assumption H3, as well as the mapping *hypothesis* of the preceding Lemma 1.2, are known to hold under the so-called Nemytskii conditions (see [22, p. 48]).

Introduce the notation,

$$Y_0 := \{v \in H^1(\Omega) : \Gamma v|_{\Sigma_D} = 0\},$$

which is assumed to have a dense, zero-trace subspace of infinitely differentiable functions. Then $\mathbf{u}$ is a *weak solution* of (1, 2, 3) if $\mathbf{u}$ satisfies (2) together with the relations,

$$\int_\Omega p_k(\vec{x}, \mathbf{u}(\vec{x}))\nabla u_k \cdot \nabla v \; dx + \int_\Omega f_k(\vec{x}, \mathbf{u}(\vec{x}))v \; dx = 0, \; k = 1, \ldots, m, \; \forall v \in Y_0. \tag{4}$$

The two principal results of this paper appear in the following two sections: a trapping principle for the Galerkin approximation with quadrature (Theorem 2.1) and a convergence/existence theorem for these approximations (Theorem 3.1). Additional hypotheses will be required: in §2 for the N-dimensional simplicial finite element approximations, and in §3 for the convergence results.

2 Piecewise Linear Galerkin Approximations

We will verify a trapping principle for piecewise linear Galerkin approximations of *any* weak solution of (1,2,3). Our approach is non-standard, in that we use a variational inequality as a passage from the function space containing the Galerkin approximations to the appropriate system of equations for the nodal coefficients in Euclidean space. We clarify this as our presentation develops.

Given that Ω is a convex polyhedral domain in $\mathbf{R}^N$, we suppose that a simplicial decomposition $\mathcal{S}$ of elements S is specified, so that each simplex S is the convex hull of its nodes, and $\bar{\Omega}$ is the union of such elements. We refer to each $S \in \mathcal{S}$ as an N-dimensional simplicial finite element. To describe this more fully, we introduce some terminology, which retains the framework of [16]. Let $S \in \mathcal{S}$ be such that

a) $\vec{v}_i$ is an arbitrary vertex of S and $F_i \subset \partial S$ is the opposite $N-1$ dimensional boundary face not containing $\vec{v}_i$;

b) e_{ij} is the edge connecting vertices $\vec{v}_i$ and $\vec{v}_j$;

c) γ_{ij} is the angle between the inward normal vectors to the faces F_i and F_j;

d) ϕ_l is the piecewise linear nodal basis function which is 1 at vertex $\vec{v}_l$;

e) Let $a \geq a_0 > 0$ be a bounded, strictly positive, measurable function on Ω, and define

$$\alpha_{ij} \equiv \int_S a(\vec{x}) \nabla \phi_i(\vec{x}) \cdot \nabla \phi_j(\vec{x})\, dx$$

to be the ijth entry of the *element* stiffness matrix;

f) a_{ij} is the ijth element of the assembled (summed over elements S) stiffness matrix A:

$$a_{ij} = \int_\Omega a(\vec{x}) \nabla \phi_i(\vec{x}) \cdot \nabla \phi_j(\vec{x})\, dx.$$

It is shown in [16, Lemma A.1] that, in any N-dimensional simplicial finite element S, $\cos \gamma_{ij}$ has the same sign as α_{ij}, independently of the function a. Therefore, whenever a is identified with the various components of $\mathbf{p}$ in the PDE divergence operator structure, the statements concerning non-obtuseness and irreducibility in hypothesis H5 below are unchanged. These properties are now defined. They are fundamentally geometric properties, imposed on each S in the simplicial decomposition $\mathcal{S}$.

Geometric Properties of the Decomposition

H5. Non-Obtuseness and Irreducibility (Property N) The simplicial decomposition $\mathcal{S}$ will be said to satisfy property (N) if:

- (non-obtuseness) For each N-dimensional simplicial finite element S, and $\vec{v}_j$ a vertex in S, then

$$\alpha_{ij} \leq 0, \text{ if } \vec{v}_i \text{ is on } e_{ij} \subset S,\ \vec{v}_i \neq \vec{v}_j.$$

- (irreducibility) If $\alpha_{ij} = 0$, there is a chain of distinct indices $i_1 = i, \ldots, i_\ell = j$ such that $\alpha_{i_k i_{k+1}} < 0$ for each successive pair i_k, i_{k+1} in the chain.

It was pointed out in [5] that the relevant off-diagonal elements in the assembled matrix $A = (a_{jk})$ are bounded uniformly smaller than zero if all angles between the outward normal vectors to any two faces of each simplicial element in the mesh are bounded uniformly above by $\frac{\pi}{2} - \delta$. This provides a sufficient (acute) condition. For our purposes here, we require only property (N) itself as expressed in H5. It was shown in [16, Lemma A.1] that the sum of the stiffness matrix entries in any row is zero:

$$\alpha_{ii} + \sum_{j \neq i} \alpha_{ij} = 0. \tag{5}$$

Let d_0 denote the cardinality of the set of nodes defined by the simplicial decomposition, and which occur in $\bar{\Omega}$. We denote by Σ those nodes which intersect Σ_D, and we write d for the cardinality of the set of nodes in $\bar{\Omega} \backslash \Sigma_D$. We assume that $d < d_0$. Since by definition, for $\mathbf{z} \in \mathbf{R}^{d_0}$,

$$\int_\Omega a(\vec{x}) |\sum_{j=1}^{d_0} z_j \nabla \phi_j|^2\, dx = (A\mathbf{z}, \mathbf{z})_{\mathbf{R}^{d_0}}, \tag{6}$$

it follows that the quadratic form,

$$(A\mathbf{z}, \mathbf{z})_{\mathbf{R}^{d_0}},$$

is nonnegative definite. We have the following additional result.

Lemma 2.1. *Suppose the simplicial decomposition satisfies hypothesis* H5*.*

$$\textit{If } (A\mathbf{z}, \mathbf{z}) = 0, \textit{ then } \mathbf{z} = \mathbf{c} = \{c, \cdots, c\}, \textit{ for a constant } c.$$

Proof. Define $v = \sum_{j=1}^{d_0} z_j \phi_j$, and consider a fixed simplex S. Define an arbitrary local ordering of the vertices belonging to S, indexed by $i = 1, \ldots, i_0$, and label the corresponding subset of the coordinates of $\mathbf{z}$ as $t_1, \ldots, t_{i_0}$. The restriction of v to S has the representation,

$$v(\vec{x}) = \sum_{i=1}^{i_0} t_i \phi_i(\vec{x}), \ \vec{x} \in S.$$

The following critical 'double sum' identity was proved in [16, Corollary A.1]:

$$\int_S a(\vec{x}) |\nabla v|^2 \, dx = \sum_{i<k} -\alpha_{ik} [t_i - t_k]^2.$$

The hypothesis H5 implies that the vertex evaluations t_i are equal within each S; this is facilitated by the construction of a chain of indices based upon the irreducibility hypothesis. Thus, v is constant on each S. The continuity of v implies the constants have a common value c. In particular, each component of $\mathbf{z}$ is equal to c. □

2.1 Galerkin System

In order to formulate the Galerkin system, we allow for approximate formulations within the vector field integrals. These approximations are defined in terms of a local operator. Local here is interpreted with respect to the union,

$$G_i = \cup \{S \in \mathcal{S} : \vec{v}_i \in S\},$$

of those simplices containing a given vertex $\vec{v}_i$. Note that G_i is simply the support of the nodal basis function ϕ_i contained in $\bar{\Omega}$. We also identify the function $a(\vec{x})$ in the preceding discussion with the individual components of $\mathbf{p}$.

H6. Local Operator Approximation (Property LOA) Fix the simplicial decomposition $\mathcal{S}$, and let $\mathcal{F} = \{\sum_{j=1}^{d_0} z_j \phi_j, \ \mathbf{z} \in \mathbf{R}^{d_0}\}$ represent the finite element space. For each vertex $\vec{v}_i$, there is prescribed an approximation operator J_i such that J_i is continuous from $\mathcal{F}|_{G_i}$ to $L_2(G_i)$, with support in G_i. Furthermore, the following consistency relation is required to hold: If $\mathcal{K}_i$ denotes the support of $J_i \phi_i$, and $\phi \in \mathcal{F}|_{G_i}$ has range in Q_k, then $J_i \phi|_{\mathcal{K}_i}$ has range in Q_k, $\forall k$. These operators are used in (9) below.

For the purposes of the variational inequality derived in the following section, no specific structure is imposed on these operators. Indeed, they may be selected to act as the identity in each G_i, which retrieves the standard definition. However, an important special case, considered in [24], is to define

$$J_i\phi = \phi(\vec{v}_i)\chi_{|E_i},$$

where E_i is the barycentric region, associated with the vertex $\vec{v}_i$. It is a subset of G_i, defined via barycentric coordinates of those simplices containing $\vec{v}_i$. Concisely: $\vec{x} \in E_i$ if the barycentric coordinate λ_i of $\vec{x}$ dominates all coordinates λ_j, for $\vec{v}_j \in G_i$. A critical property of such regions is that $\int_\Omega \phi_i \, dx = \text{meas } (E_i)$. We will not require the precise definition in this paper, and refer to [24] for amplification. The operator J_i is continuous, in the sense defined.

Now, for $k = 1, \ldots, m$, introduce the notation, $(y_{1k}, \ldots, y_{dk})^t$, for the first d components of $\mathbf{y}_k$, and, for the remaining $d_0 - d$ components, specify interpolation boundary values through the requirement that,

$$y_{jk} = \hat{u}_k(\vec{v}_j), \text{ for } \vec{v}_j \in \Sigma.$$

If $\mathbf{u}_h = (u_{1h}, \ldots, u_{mh})^t$ is prescribed via

$$u_{kh}(\vec{x}) = \sum_{j=1}^{d_0} y_{jk}\phi_j(\vec{x}), \; k = 1, \ldots, m, \tag{7}$$

then P_k are the $d \times d_0$ matrices given by, for $k = 1, \ldots, m$,

$$p_{ijk} = \int_\Omega p_k(\vec{x}, \mathbf{u}_h(\vec{x}))\nabla\phi_i(\vec{x}) \cdot \nabla\phi_j(\vec{x}) \, dx, \; i = 1, \ldots, d, \; j = 1, \ldots, d_0, \tag{8}$$

and $\mathbf{g}_k$ are defined by, for $k = 1, \ldots, m$,

$$g_{ik} = \int_\Omega f_k(\vec{x}, J_i\mathbf{u}_h(\vec{x}))J_i\phi_i(\vec{x}) \, dx, \; i = 1, \ldots, d. \tag{9}$$

It is understood that J_i acts on each component of $\mathbf{u}_h$. The system of equations defining the Galerkin method is given by

$$P_k(\mathbf{y}_1, \ldots, \mathbf{y}_m)\mathbf{y}_k + \mathbf{g}_k(\mathbf{y}_1, \ldots, \mathbf{y}_m) = \mathbf{0}, k = 1, \ldots, m. \tag{10}$$

2.2 The Finite Dimensional Variational Inequality

We require some notation. For fixed $k = 1, \ldots, m$, define d copies of $[a_k, b_k]$ by

$$[a_k, b_k]^d := Q_k^d,$$

and, an admissible set in $\mathbf{R}^{d_0}$ by

$$K_k = \{\mathbf{v} \in \mathbf{R}^{d_0} : (v_1, \ldots, v_d)^t \in Q_k^d, \text{ and, for } d < j \leq d_0, \; v_j = \hat{u}_k(\vec{v}_j), \text{ for } \vec{v}_j \in \Sigma\}.$$

We shall employ the coupled variational inequalities: Find $\mathbf{y}_k \in K_k, k = 1, \ldots, m$, satisfying

$$(P_k(\mathbf{y}_1, \ldots, \mathbf{y}_m)\mathbf{y}_k, \mathrm{Tr}(\mathbf{v}_k - \mathbf{y}_k))_{\mathbf{R}^d} + (\mathbf{g}_k(\mathbf{y}_1, \ldots, \mathbf{y}_m), \mathrm{Tr}(\mathbf{v}_k - \mathbf{y}_k))_{\mathbf{R}^d} \geq 0, \ \forall \mathbf{v}_k \in K_k. \tag{11}$$

Here, the inner product in $\mathbf{R}^d$ is the usual one. Also, we have used the notation Tr to denote truncation of the final $d_0 - d$ components from the indicated vector.

Proposition 2.1. *Assume the system hypotheses* H1,H3,H4 *of the introduction as well as hypotheses* H5 *and* H6 *of section two. The variational inequality* (11) *has a solution* $(\mathbf{y}_k, k = 1, \ldots, m)$.

Proof. The proof is divided into three parts. We shall require the $d_0 \times d_0$ matrix extension $\hat{P}_k$ of P_k defined by (8) with the range of i extended to $i = 1, \ldots, d_0$. Also, for fixed $\mathbf{w} = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$, we define the auxiliary bilinear functional, for $\mathbf{u}, \mathbf{v} \in \mathbf{R}^{d_0}$,

$$A_{\mathbf{w},k}(\mathbf{u}_k, \mathbf{v}_k) = (\hat{P}_k(\mathbf{w})\mathbf{u}_k, \mathbf{v}_k)_{\mathbf{R}^{d_0}} + \sum_{j=d+1}^{d_0} u_{jk} v_{jk}.$$

By Lemma 2.1, we conclude that this functional is positive definite in its action on $\mathbf{R}^{d_0}$. Indeed, the final term, identified with summation of product interpolant evaluations on Σ, is included so that the conclusion $c = 0$ may be drawn. Here, c is given by the lemma whenever $A_{\mathbf{w},k}(\mathbf{v}_k, \mathbf{v}_k) = 0$.

The remaining structure of the argument is as follows.

The map T from $\prod_k K_k$ to $\prod_k K_k$. We select a vector $\mathbf{w} = (\mathbf{w}_k) \in \prod_k K_k$, and construct the quadratic functional, for $\mathbf{v}_k \in K_k, k = 1, \ldots, m$,

$$R_{\mathbf{w}}(\mathbf{v}, \mathbf{v}) = \sum_{k=1}^{m} \{A_{\mathbf{w},k}(\mathbf{v}_k, \mathbf{v}_k) + 2(\mathbf{g}_k(\mathbf{w}_1, \ldots, \mathbf{w}_m), \mathrm{Tr}(\mathbf{v}_k))_{\mathbf{R}^d}\}. \tag{12}$$

By standard theory, there is a *unique* minimizer $\mathbf{y}$ in the closed convex set $\prod_k K_k$. This is characterized, via one-sided partial derivatives, by the coupled set of variational inequalities, $\forall \mathbf{v}_k \in K_k, k = 1, \ldots, m$,

$$(P_k(\mathbf{w}_1, \ldots, \mathbf{w}_m)\mathbf{y}_k, \mathrm{Tr}(\mathbf{v}_k - \mathbf{y}_k))_{\mathbf{R}^d} + (\mathbf{g}_k(\mathbf{w}_1, \ldots, \mathbf{w}_m), \mathrm{Tr}(\mathbf{v}_k - \mathbf{y}_k))_{\mathbf{R}^d} \geq 0. \tag{13}$$

The reduction from $\hat{P}_k$ to P_k is possible since $\mathbf{v}_k - \mathbf{y}_k$ has vanishing components with indices $d + 1, \ldots, d_0$. Set $\mathbf{y} = T\mathbf{w}$.

A fixed point for T The map T is continuous, hence has a fixed point $\mathbf{y} = (\mathbf{y}_k)$ in $\prod_k K_k$ by the Brouwer fixed point theorem [9].

The continuity is established as follows. Let $\mathbf{y}^* = T\mathbf{w}^*, \mathbf{y}^{**} = T\mathbf{w}^{**}$. We establish local continuity at $\mathbf{w}^*$, and thus allow $\mathbf{w}^{**}$ to vary. In the process, we will use the fact (see Lemma 2.1) that $\hat{P}_k(\mathbf{w}^*)$ is positive definite on the subspace of $\mathbf{R}^{d_0}$ consisting of vectors $\mathbf{v}_k$ with the last $d_0 - d$ components zero.

Denote by α_k^* the smallest positive eigenvalue on this subspace. Now consider the variational inequalities $*$ and $**$ used to define $\mathbf{y}^*$ and $\mathbf{y}^{**}$ from (13) and make the respective choices in these inequalities for $\mathbf{v}_k \in K_k$,

$$\mathbf{v}_k = \mathbf{y}_k^{**},\ \mathbf{v}_k = \mathbf{y}_k^*.$$

After subsequent multiplication of each inequality by -1, followed by their sum, and addition and subtraction of the term,

$$(P_k(\mathbf{w}^*)\mathbf{y}_k^{**}, \mathrm{Tr}(\mathbf{y}_k^{**} - \mathbf{y}_k^*))_{\mathbf{R}^d},$$

we obtain the inequality:

$$0 < \alpha_k^* \|\mathrm{Tr}(\mathbf{y}_k^{**} - \mathbf{y}_k^*)\|_{\mathbf{R}^d}^2 \leq |((P_k(\mathbf{w}^*) - P_k(\mathbf{w}^{**}))\mathbf{y}_k^{**}, \mathrm{Tr}(\mathbf{y}_k^{**} - \mathbf{y}_k^*))_{\mathbf{R}^d}|$$
$$+|(\mathbf{g}(\mathbf{w}^*) - \mathbf{g}(\mathbf{w}^{**}), \mathrm{Tr}(\mathbf{y}_k^{**} - \mathbf{y}_k^*))_{\mathbf{R}^d}|.$$

The analysis is facilitated by an application of Cauchy's inequality to each r.h.s. term, followed by the use of the Frobenius matrix norm, which is denoted $\|\cdot\|_F$. Moreover, $\mathbf{y}^{**}$ varies within a bounded set. When the inequality,

$$ab \leq Ca^2 + \frac{b^2}{4C},$$

with C sufficiently large, is applied, this leads to

$$\|\mathrm{Tr}(\mathbf{y}_k^* - \mathbf{y}_k^{**})\|_{\mathbf{R}^d}^2 \leq C'\{\|\mathbf{g}(\mathbf{w}^*) - \mathbf{g}(\mathbf{w}^{**})\|_{\mathbf{R}^d}^2 + \|P_k(\mathbf{w}^*) - P_k(\mathbf{w}^{**})\|_F^2\}.$$

Lemma 1.2 now implies the continuous dependence of the map P_k, and this lemma and the continuity of the operators J_i imply the continuous dependence of $\mathbf{g}$. We thus infer the continuity of T. □

2.3 The Equivalence

Hypothesis H6 is not sufficiently robust to utilize the vector field property (OP) of hypothesis H2, so as to guarantee that solutions of the variational inequality (11) necessarily satisfy the finite element Galerkin system equation (10). To guarantee this, we need an explicit nodal approximation, defined by piecewise constants, as introduced rigorously by Tabata [24]. When implemented in integral expressions, it leads to lumped integration formulas.

H7. Nodal Approximation (Property NA) We explicitly define, for $v \in \mathcal{F}|_{E_i}$, and E_i the barycentric region associated with the vertex $\vec{v}_i$:

$$J_i v(\vec{x}) = v(\vec{v}_i)\chi_{E_i}(\vec{x}),\ \vec{x} \in \Omega. \tag{14}$$

Theorem 2.1. *Suppose the hypotheses* H1,H2,H3,H4 *of the introduction are satisfied, together with hypotheses* H5,H7 *of section two. Then the variational inequality* (11) *has a solution in* $\prod_k K_k$*, and any such solution satisfies the Galerkin system* (10)*.*

Proof. We begin with the result of Proposition 2.1. We observe that property (NA) of hypothesis H7 implies property (LOA) of hypothesis H6, so that the proposition may be applied. We plan to show that the dot products ($\mathbf{R}^d$ inner products) of arbitrary vectors ψ with the left hand sides ($k = 1, \ldots, m$) of (10) are zero. The technique is an adaptation of that successfully used at the level of the elliptic system ([13]). The details are given for the first equation ($k = 1$). We shall thus proceed from the first inequality in (11). For notational convenience, we set $\mathbf{y}_1 = \mathbf{y}, p_1 = p, p_{ij1} = p_{ij}$. Now let $\epsilon = \alpha - a_1 = b_1 - \beta > 0$ be such that the interval $[\alpha, \beta]$ contains all components y_j except those for which $y_j = a_1$, or $y_j = b_1$. Given $\psi \in \mathbf{R}^d$, with components normalized not to exceed unity, we select vectors $\mathbf{v} = \mathbf{v}_\pm$ in (11) as follows: For $i = 1, \ldots, d$, set

$$v_i = v_{\pm i} = (y_i \pm \epsilon\psi_i - a_1)^+ + a_1 + (b_1 - y_i \mp \epsilon\psi_i)^-.$$

Here we employ a somewhat nonstandard notation for the negative part: $t^- = t$, if $t < 0, t^- = 0$, otherwise. By considering the three possible cases (positive/positive, positive/negative, negative/positive) for the two terms for which the positive and negative parts are taken, one sees that $(v_1, \ldots, v_d)^t \in Q_1^d$. The verification has two parts (see (17,18) to follow), related to the sign of the dot products. For the first of these, we choose $\mathbf{v} = \mathbf{v}_+$, and introduce index sets $\mathcal{A}$ and $\mathcal{B}$,

$$\mathcal{A} = \{i : 1 \le i \le d : y_i = a_1,\ \psi_i \le 0\},\ \ \mathcal{B} = \{i : 1 \le i \le d : y_i = b_1,\ \psi_i \ge 0\}.$$

The identities:

$$v_{+i} - y_i = 0,\ i \in (\mathcal{A} \cup \mathcal{B}),$$
$$v_{+i} - y_i = \epsilon\psi_i, i \notin (\mathcal{A} \cup \mathcal{B}),$$

permit the rewriting of (11) as

$$\sum_{i \notin \mathcal{A}\cup\mathcal{B}} \sum_{j=1}^{d_0} p_{ij} y_j \psi_i + \sum_{i \notin \mathcal{A}\cup\mathcal{B}} g_i \psi_i \ge 0. \tag{15}$$

For a given element S we use the notation,

$$\pi_{ij}^S = \int_S p(\vec{x}, \mathbf{u}(\vec{x})) \nabla\phi_i(\vec{x}) \nabla\phi_j(\vec{x})\ dx.$$

Recall that, by property (N) of hypothesis H5, $\pi_{ij}^S \le 0$ $i \ne j$, and by (5), $\pi_{ii}^S + \sum_{j \ne i} \pi_{ij}^S = 0$. We wish to show that

$$\sum_{i \in \mathcal{A}\cup\mathcal{B}} \sum_{j=1}^{d_0} p_{ij} y_j \psi_i \ge 0, \sum_{i \in \mathcal{A}\cup\mathcal{B}} g_i \psi_i \ge 0. \tag{16}$$

We estimate:

$$\sum_{i \in \mathcal{A}\cup\mathcal{B}} \sum_{j=1}^{d_0} p_{ij} y_j \psi_i = \sum_{i \in \mathcal{A}} \sum_{S : \vec{v}_i \in S} \left(\pi_{ii}^S + \sum_{j : i \ne j} \pi_{ij}^S \right) a_1 \psi_i - \sum_{i \in \mathcal{A}} \sum_{S : \vec{v}_i \in S} \sum_{j : i \ne j} \pi_{ij}^S (a_1 - y_j) \psi_i$$

$$+\sum_{i\in\mathcal{B}}\sum_{S:\vec{v}_i\in S}\left(\pi_{ii}^S+\sum_{j:i\neq j}\pi_{ij}^S\right)b_1\psi_i-\sum_{i\in\mathcal{B}}\sum_{S:\vec{v}_i\in S}\sum_{j:i\neq j}\pi_{ij}^S(b_1-y_j)\psi_i\geq 0.$$

Also, by the hypotheses H2 and H7, and the definition of g_i via (9), we have $g_i\psi_i\geq 0$ whenever $i\in\mathcal{A},\ i\in\mathcal{B}$. Hence, $\sum_{i\in\mathcal{A}\cup\mathcal{B}}g_i\psi_i\geq 0$. It follows from (15) and (16) that

$$\sum_{j=1}^{d_0}\sum_{i=1}^{d}p_{ij}y_j\psi_i+\sum_{i=1}^{d}g_i\psi_i\geq 0. \tag{17}$$

The companion inequality,

$$\sum_{j=1}^{d_0}\sum_{i=1}^{d}p_{ij}y_j\psi_i+\sum_{i=1}^{d}g_i\psi_i\leq 0, \tag{18}$$

follows from the selection of $\mathbf{v}_-$ and the redefinition of $\mathcal{A},\mathcal{B}$ to reflect a reversal of the sign of ψ_i. Indeed, the new identities become

$$v_{-i}-y_i=0,\ i\in(\mathcal{A}\cup\mathcal{B}),$$

$$v_{-i}-y_i=-\epsilon\psi_i, i\notin(\mathcal{A}\cup\mathcal{B}),$$

leading to

$$\sum_{i\notin\mathcal{A}\cup\mathcal{B}}\sum_{j=1}^{d_0}p_{ij}y_j\psi_i+\sum_{i\notin\mathcal{A}\cup\mathcal{B}}g_i\psi_i\leq 0.$$

The cases $i\in\mathcal{A}\cup\mathcal{B}$ lead to a similar nonpositive sum, and then, finally, to (18). It follows that (10) holds. □

2.4 Graded Mass Approximations

Theorem 2.1 was proven under the explicit assumption H7, which is implemented in the definition (9). This is a general interpretation of a lumped mass approximation, but nonetheless allows for variation in $\vec{x}\in E_i$. Notice that, at this level, one has not introduced considerations of computability or convergence. For the former, one might wish to restrict the mass approximation further by defining:

$$g'_{ik}=\int_\Omega J_i(f_k(\vec{x},\mathbf{u}_h(\vec{x})))J_i\phi_i(\vec{x})\,dx,\ i=1,\ldots,d. \tag{19}$$

Notice that the primed quantities are readily computable, since, locally,

$$J_i(f_k(\vec{x},\mathbf{u}_h(\vec{x})))=f_k(\vec{v}_i,\mathbf{u}_h(\vec{v}_i))\chi_{E_i}.$$

However, increased assumptions on $\mathbf{f}$ are required to analyze this more specialized approximation. In the following section dealing with convergence, we will be required to make further assumptions on $\mathbf{f}$. The result will be that the graded definitions (9) and (19) are in some sense indistinguishable when these assumptions are made.

3 Convergence and Existence

The framework introduced in [5] and [24] identifies the interpolant as a key component of the convergence analysis, requiring certain regularity of the approximation class. Specifically, $W^{1,p}(\Omega)$ regularity is required for $p > N$. Very general theorems for this case are to be found in [2, Section 4.4]. For a simplicial decomposition $\mathcal{S}$, such that h is the maximal diameter of simplices in $\mathcal{S}$, and for $v \in C(\bar{\Omega})$, we employ the notation $v \mapsto I_h v$ for the interpolant:

$$I_h v = \sum_{\vec{v}_i \in \bar{\Omega}} v(\vec{v}_i)\phi_i.$$

Closely associated with the interpolant is a simple function approximation, generated by assembling linear combinations of the characteristic functions χ_{E_i} of the barycentric regions E_i introduced earlier. Specifically, for $v \in C(\bar{\Omega})$, we define the mapping $v \mapsto \bar{v}$ via

$$\bar{v} = \sum_{\vec{v}_i \in \bar{\Omega}} v(\vec{v}_i)\chi_{E_i}.$$

The usefulness of these simple function approximations was developed in [24]. We have retained the notation $\bar{v}$ of that reference. The following lemma requires the notion of a quasi-uniform family $\{\mathcal{S}\}$ of simplicial decompositions, which we concisely define [2]: there exists $\rho > 0$ such that, for each $\mathcal{S}$ and each $S \in \mathcal{S}$, the diameter of the largest ball contained in S has ρh as a lower bound, where h varies with $\mathcal{S}$. Standard seminorm notation is employed in the lemma: for $1 \leq p < \infty$,

$$|v|_{1,p} = \left(\sum_{|\alpha|=1} \|D^\alpha v\|_p^p \right)^{1/p}.$$

Lemma 3.1. *For a quasi-uniform family $\{\mathcal{S}\}$ of simplicial decompositions, the interpolation sequence and the simple function sequence converge to v for $v \in W^{1,p}(\Omega), p > N$, as $h \to 0$. More precisely, the following estimates hold:*

$$\|v - I_h v\|_p \leq Ch|v|_{1,p}, \tag{20}$$

$$\|v - \overline{I_h v}\|_p \leq Ch|v|_{1,p}, \tag{21}$$

where C in each inequality is a positive constant not depending on h or v.

A proof of inequality (20) may be found in [2, Theorem 4.4.20] and in [5]. For (21), cf. [24, Lemma 5.1]. We make no 'a priori' assumptions regarding solution regularity; the weak solution derived in this section has components in $H^1(\Omega)$. It will be necessary to mollify, however, in order to avoid regularity assumptions. Although mollification (convolution) requires functions to be defined on a superset of $\bar{\Omega}$, the Calderón extension theorem (see [1, Theorem 4.32]) permits this for Sobolev class functions, via a linear continuous extension

operator to the Sobolev space defined on $\mathbf{R}^N$. For Lebesgue class functions, the trivial extension by zero is adequate. In the remaining part of this section, we will tacitly make use of these extensions. When explicit notation is required, we designate $\tilde{v}$ as the extension to $\mathbf{R}^N$ of v defined on Ω. As a preliminary, we define the formal convolution:

$$g_1 * g_2(\vec{x}) = \int_{\mathbf{R}^N} g_1(\vec{x} - \vec{y}) g_2(\vec{y})\, dy.$$

For rigorous mollification results, we appeal to [20, Theorem 2.16]. We state a condensed version of it here.

Lemma 3.2. *Suppose $j \in L_1(\mathbf{R}^N)$ with $\int j = 1$. For $\varepsilon > 0$, set $j_\varepsilon = \varepsilon^{-N}(\cdot\,/\varepsilon)$ and define the mollification, for $g \in L_p(\mathbf{R}^N), 1 \le p < \infty$:*

$$g_\varepsilon(\vec{x}) = j_\varepsilon * g(\vec{x}).$$

Then g_ε converges strongly to g in L_p as $\varepsilon \to 0$. Moreover, if $j \in C_0^\infty(\mathbf{R}^N)$, then g_ε is in $C^\infty(\mathbf{R}^N)$. If g is in the Sobolev space $H^1(\mathbf{R}^N)$, then $D^\alpha g_\varepsilon$, for $|\alpha| = 1$, is expressed by applying D^α to either member of the convolution. It follows from this property that g_ε converges strongly to g in $H^1(\mathbf{R}^N)$ as $\varepsilon \to 0$.

We will assume the use of standard exponential mollifiers in $C_0^\infty(\mathbf{R}^N)$. In the application of Lemma 3.1, it is necessary to evaluate the Sobolev p-seminorm of v for $p > N$. In the proof of Theorem 3.1 to follow, we will be required to consider the case $\tilde{v} = j_\varepsilon * \tilde{u}$, where $u \in H^1(\Omega)$, and where derivatives in the seminorm are applied to j_ε. The cases $N = 1, 2$ are elementary. For $N > 2$, we need a version of Young's inequality.

Lemma 3.3. *Suppose $g_1 \in L_q(\mathbf{R}^N), g_2 \in L_r(\mathbf{R}^N)$, with*

$$2 > \frac{1}{q} + \frac{1}{r} > 1.$$

Then $g_1 * g_2 \in L_{p'}(\mathbf{R}^N)$, where

$$\frac{1}{p'} = \frac{1}{q} + \frac{1}{r} - 1.$$

Furthermore, the $L_{p'}$ norm is bounded by

$$\|g_1 * g_2\|_{p'} \le \|g_1\|_q \|g_2\|_r. \tag{22}$$

It follows that, if $g_1 \in C_0^\infty(\mathbf{R}^N)$ and $g_2 \in H^1(\mathbf{R}^N)$, for $N > 2$, then the choices

$$\frac{1}{q} = \frac{N+2}{2N} + \frac{1}{N+\delta},\ \frac{1}{r} = \frac{N-2}{2N},$$

are admissible, and lead to $p' = N + \delta$ in Young's inequality (22)*. Here, $\delta > 0$ is arbitrary.*

Proof. Young's inequality (22) is a duality restatement of [20, Theorem 4.2, p. 98]. If $g_2 \in H^1(\mathbf{R}^N)$, then the Sobolev embedding theorem [1, Theorem 5.4, Case A, p. 97] yields $g_2 \in L_r(\mathbf{R}^N)$, and permits the application of Young's inequality as stated. □

We continue with the consistency *hypothesis*, required for boundary-value trace consistency.

Consistency Hypothesis

H8. Linear Boundary-Value Interpolation (Property LBI) There are two parts.

- The function $\hat{\mathbf{u}}$ is gradient stable on the family $\mathcal{S}$ under interpolation:

$$\|\nabla \hat{\mathbf{u}}_{I_h}\|_2 \leq C\|\nabla \hat{\mathbf{u}}\|_2,$$

  where C does not depend on h.

- If $\bar{h} = \sup_{S\in\mathcal{S}} \operatorname{diam}(S \cap \Sigma_D)$, and $I_{\bar{h}}\hat{\mathbf{u}}$ denotes the interpolant of $\hat{\mathbf{u}}$ on the nodes of $\bar{\Omega}$, then:

$$\Gamma I_{\bar{h}}\hat{\mathbf{u}} \to \Gamma\hat{\mathbf{u}},\ \bar{h} \to 0,\ \text{in } L_2(\Sigma_D).$$

It is understood that the diameters are computed w.r.t. $N-1$ dimensional measure in the definition of $\bar{h}$ in property (LBI). For a thorough study of the underlying ideas involved in these hypotheses, the reader may consult [2].

The convergence arguments to follow also depend on regularity properties for $\mathbf{f}$ as well as a local intersection property associated with the simplicial decompositions. We state the specific properties for $\mathbf{f}$.

Vector Field Regularity

H9. Mean Value and Composition Property (MVC) There are two parts. It is assumed that there exists a positive constant C, such that, for each $k = 1, \ldots, m$, we have:

- The pointwise bound,

$$|f_k(\vec{x},\mathbf{v}) - f_k(\vec{y},\mathbf{w})| \leq C(\|\vec{x}-\vec{y}\|_{\mathbf{R}^N} + \|\mathbf{v}-\mathbf{w}\|_{\mathbf{R}^m}),\ \forall (\vec{x},\mathbf{v}), (\vec{y},\mathbf{w}) \in \mathcal{D}.$$

- For any $1 \leq p < \infty$, if $\mathbf{v} \in (W^{1,p}(\Omega))^m$, with range in Q, the composition $f_k(\cdot\,,\mathbf{v})$ is in $W^{1,p}(\Omega)$, and the following semi-norm bound holds:

$$|f_k(\cdot\,,\mathbf{v})|_{1,p} \leq C(1 + |\mathbf{v}|_{1,p}).$$

A sufficient condition for both parts of this property is the continuity and bounded differentiability of f_k in the joint variables $(\vec{x},\mathbf{u})$ (see [8, p. 86] for a statement of the multidimensional mean value theorem; the second part is implied by standard chain rule theorems for the differential map).

The final property assumed in this paper relates to the 'economy' of the barycentric cover.

H10. Simplex Intersection Property (SI) There is an integer κ_0 such that, for any simplicial decomposition $\mathcal{S}$, and any $S \in \mathcal{S}$,

$$S \cap E \neq \emptyset,$$

can occur at most κ_0 times for E a barycentric region defined by $\mathcal{S}$.

3.1 Strong Convergence to a Trapping Region Solution of a Galerkin Subsequence

The convergence result depends upon an auxiliary lemma, which we postpone to the conclusion of the proof of Theorem 3.1.

Theorem 3.1. *Suppose the hypotheses of Theorem* 2.1 *hold for a quasi-uniform family* $\{\mathcal{S}\}$ *of simplicial decompositions, and, additionally, assume the hypotheses* H8,H9,H10. *Then a weak solution* $\mathbf{u}$ *of* (1) *exists, satisfying the Dirichlet boundary condition* (2) *on* Σ_D *and the weak form of the Neumann boundary conditions* (3). *It may be characterized as the limit of a subsequence of* $\{\mathbf{u}_{h_n}\}$, *associated with* $\{\mathcal{S}_{h_n}\}$, *for which* $h_n \to 0$. *This subsequence is weakly convergent in* $(H^1(\Omega))^m$ *and strongly convergent in* $(L_2(\Omega))^m$.

Proof. Given $\mathbf{v}$ in a smooth dense subset of $(Y_0)^m$, and a sequence $\mathcal{S}_{h_n}$ such that $h_n \to 0$, let $\mathbf{u}_{h_n}$ be corresponding Galerkin solutions, assembled from (10) and (7) via Theorem 2.1. Set $\mathbf{v}_n = I_{h_n}\mathbf{v}$. For notational simplicity, we write the components of $\mathbf{v}_n$ as v_{kn} and the components of $\mathbf{u}_{h_n}$ as u_{kn}. The idea is to take a rigorous limit in the relations, for $k = 1, \ldots, m$,

$$\int_\Omega p_k(\vec{x}, \mathbf{u}_{h_n}(\vec{x}))\nabla u_{kn}(\vec{x}) \cdot \nabla v_{kn}(\vec{x})\, dx + \int_\Omega \sum_i f_k(\cdot\,, \mathbf{u}_{h_n}(\vec{v}_i))v_{kn}(\vec{v}_i)\chi_{E_i}\, dx = 0. \tag{23}$$

Note that (23) is an analytical restatement of (10) when (14) is employed. The limiting relation will be identified with (4). The argument proceeds via weak compactness, as applied to the Galerkin approximations $\mathbf{u}_{h_n}$. To obtain the preliminary estimate, we note that (23) holds more generally under the replacement $\mathbf{v}_n \mapsto \mathbf{w}_n$, where the latter is any piecewise linear trial function vanishing on the nodes of Σ_D; we make the choice

$$\mathbf{w}_n = \mathbf{u}_{h_n} - I_{h_n}\hat{\mathbf{u}}.$$

We thus obtain, for $k = 1, \ldots, m$:

$$\begin{aligned}&\int_\Omega p_k(\vec{x}, \mathbf{u}_{h_n}(\vec{x}))\nabla u_{kn}(\vec{x}) \cdot \nabla(u_{kn}(\vec{x}) - I_{h_n}\hat{u}_k(\vec{x}))\, dx \\ &+ \int_\Omega \sum_i f_k(\cdot\,, \mathbf{u}_{h_n}(\vec{v}_i))(u_{kn} - I_{h_n}\hat{u}_k)(\vec{v}_i)\chi_{E_i}\, dx = 0.\end{aligned} \tag{24}$$

$\{I_{h_n}\hat{\mathbf{u}}\}$ is assumed gradient stable, hence is a bounded sequence in $(H^1)^m$. An L_2 gradient estimate for $\mathbf{u}_{h_n}$ is then obtained from (24) by making use of

the L_2 boundedness of the functions appearing in the second integrals and the L_∞, strictly positive, boundedness of the p_k multipliers. This is deduced from hypotheses H3, H4, and H10, as well as the uniform pointwise boundedness properties of $\mathbf{u}_{h_n}$ and $I_{h_n}\hat{u}_k$. We conclude that the Galerkin approximations lie in a fixed ball in $(H^1)^m$. By weak compactness, we infer the existence of an H^1-weakly convergent subsequence, which by the Rellich theorem, is strongly L_2 convergent to a limit $\mathbf{u}$. For simplicity, we relabel the convergent subsequence as $\mathbf{u}_{h_n}$. The first term in (23) involves a triple product in L_2: two strongly convergent sequences (one L_∞-bounded), and one weakly convergent sequence. By Lemma 3.4 to follow, we draw the conclusion that the first term in (23) converges to the corresponding term in (4). To conclude that the boundary condition (2) holds, we note that the trace mapping Γ is continuous, and even compact, from $H^1(\Omega)$ into $L_2(\Sigma_D)$ ([19]). We now invoke hypothesis H8 to conclude that the limit $\mathbf{u}$ coincides with $\Gamma\hat{\mathbf{u}}$ on Σ_D. This yields (2).

The convergence analysis of the second term in (23) is simplified by the replacement of the integrand by a function with the same limit:

$$\overline{f_k(\cdot\,,\mathbf{u}_{h_n})v_{kn}}. \tag{25}$$

The simplex by simplex estimation of the difference of the two integrands employs hypotheses H9 and H10. Indeed, for $S \in \mathcal{S}$, one has

$$\int_S \left| \sum_i f_k(\cdot\,,\mathbf{u}_{h_n}(\vec{v}_i))v_{kn}(\vec{v}_i)\chi_{E_i} - \overline{f_k(\cdot\,,\mathbf{u}_{h_n})v_{kn}} \right| \, dx \leq ch_n \text{meas }(S)\|v\|_{L_\infty},$$

where $c = C\kappa_0$. The constants C and κ_0 are defined in H9 and H10, respectively. Subsequent integration over Ω shows that the difference tends to zero as $h_n \to 0$.

Once this reduction has been achieved, the limit for the replacement function (25) uses a telescoping representation in terms of differences. The estimates are facilitated by notational simplification: $v_k \mapsto v$, $v_{kn} \mapsto v_{I_{h_n}}$. We have:

$$f_k(\cdot\,,\mathbf{u})v - \overline{f_k(\cdot\,,\mathbf{u}_{h_n})v_{I_{h_n}}} = \sum_{i=1}^{4} T_i,$$

where

$$\begin{aligned}
T_1 &:= f_k(\cdot\,,\mathbf{u})v - f_k(\cdot\,,\mathbf{u}_\varepsilon)v \\
T_2 &:= f_k(\cdot\,,\mathbf{u}_\varepsilon)v - f_k(\cdot\,,\mathbf{u}_\varepsilon)v_{I_{h_n}} \\
T_3 &:= f_k(\cdot\,,\mathbf{u}_\varepsilon)v_{I_{h_n}} - \overline{f_k(\cdot\,,(\mathbf{u}_\varepsilon)_{I_{h_n}})v_{I_{h_n}}} \\
T_4 &:= \overline{f_k(\cdot\,,(\mathbf{u}_\varepsilon)_{I_{h_n}})v_{I_{h_n}}} - \overline{f_k(\cdot\,,\mathbf{u}_{h_n})v_{I_{h_n}}}.
\end{aligned}$$

The order of analysis is: selection of ε, followed by selection of h. Suppose that $\eta > 0$ is specified. Choose $\varepsilon_1 > 0$ such that $\|T_1\|_{L_1} < \eta/4$ for $\varepsilon \leq \varepsilon_1$. This is possible via Lemma 3.2, the pointwise boundedness of v, and the general strong convergence of composition. We next estimate $\|T_4\|_{L_1}$. We will obtain

an estimate via the representation of $\bar{\Omega}$ as the union of elements S of $\mathcal{S}$. On each simplex S of a given $\mathcal{S}$, by hypothesis H10, there are at most κ_0 barycentric regions E intersecting S, so that all other characteristic functions vanish in the representation of T_4 on S. This yields, via the first inequality of hypothesis H9 and the triangle inequality:

$$\int_S |T_4|\, dx \leq C\kappa_0 \int_S \{\|\mathbf{u}_\varepsilon(\vec{v}_i) - \mathbf{u}_\varepsilon(\vec{x})\|_{\mathbf{R}^m} + \|\mathbf{u}_\varepsilon(\vec{x}) - \mathbf{u}(\vec{x})\|_{\mathbf{R}^m}$$
$$+\|\mathbf{u}(\vec{x}) - \mathbf{u}_{h_n}(\vec{x})\|_{\mathbf{R}^m} + \|\mathbf{u}_{h_n}(\vec{x}) - \mathbf{u}_{h_n}(\vec{v}_i)\|_{\mathbf{R}^m}\}|v_{I_{h_n}}|\, dx$$
$$= \int_S (T_4^1 + T_4^2 + T_4^3 + T_4^4)\, dx.$$

Here, $\vec{v}_i$ is understood as a typical vertex in S; the cardinality has already been incorporated in the multiplier κ_0. The L_1 estimation of T_4^2 on Ω parallels the estimation of T_1. Thus, there exists $\varepsilon_2 > 0$ such that $\|T_4^2\|_{L_1(\Omega)} < \eta/16$ for $\varepsilon \leq \varepsilon_2$. The restrictions on ε are now completed by the requirement that $\varepsilon = \min(\varepsilon_1, \varepsilon_2)$. The estimations, $\|T_4^1\|_{L_1(\Omega)} < \eta/16$ and $\|T_4^3\|_{L_1(\Omega)} < \eta/16$ for $h_n < h_1'$, follow routinely from the uniform continuity of $\mathbf{u}_\varepsilon$ on $\bar{\Omega}$ and the strong convergence of $\mathbf{u}_{h_n}$, respectively. The estimation of $\|T_4^4\|_{L_1}$ is more subtle. It makes use of a Poincaré-type inequality as applied to the difference represented by T_4^4 on S (see the remark at the conclusion of the proof). Such an inequality expresses the L_2 norm of T_4^4 on S as bounded by h_n times the L_2 gradient norm of T_4^4 on S. This allows for an estimate over Ω proportional to h_n, and permits the choice of h_2': $\|T_4^4\|_{L_1(\Omega)} < \eta/16$ if $h_n < h_2'$.

Similarly,

$$\|T_2\|_{L_1} \leq \int_\Omega |f_k(\cdot\,, \mathbf{u}_\varepsilon)||v - v_{I_{h_n}}|\, dx.$$

For $p > N$, and $1/q = 1 - 1/p$, Hölder's inequality can be applied with the first inequality in Lemma 3.1. Note that v has been selected in a smooth dense subspace of Y_0, permitting the application of Lemma 3.1 as applied to v. The multiplier term involving f_k is in L_2, hence L_q (the case $N = 1$ is immediate). Thus, h_3' exists such that $h_n < h_3'$ implies that this term is bounded above by $\eta/4$.

To obtain the bound for $\|T_3\|_{L_1}$, we apply the second inequality of Lemma 3.1 to the indicated product. Hölder's inequality permits the L_p estimation of T_3, and by the indicated lemma this estimate is proportional to the product of h_n and the Sobolev p-seminorm of $f_k(\cdot\,, \mathbf{u}_\varepsilon)v_{I_{h_n}}$. Two types of terms arise via the product rule of differentiation. They may be analyzed separately via the triangle inequality. Terms involving D^α applied to $v_{I_{h_n}}$ are resolved by the gradient p-stability for interpolants of smooth (say, $W^{2,p}$) functions ([5]). Thus, if the Schwarz inequality is applied to the resulting p-th power integrand, one is utilizing the gradient $2p$-stability for $v_{I_{h_n}}$ and ordinary $2p$ stability for $f_k(\cdot\,, \mathbf{u}_\varepsilon)$. Terms involving D^α applied to $f_k(\cdot\,, \mathbf{u}_\varepsilon)$ are resolved by application of the second part of hypothesis H9. Here, we use the 'a priori' uniform boundedness of $v_{I_{h_n}}$, not the Schwarz inequality in the estimation. The Sobolev p-seminorm

of $\mathbf{u}_\varepsilon$ is estimated, via an application of derivatives to the mollifier used in the convolution. This is direct for $N = 1, 2$. For $N > 2$, lemma 3.3 is applied to the differentiated convolution, with p' and p identified. Thus, by choosing h_n sufficiently small, say, $h_n < h_4'$, one may ensure that $\|T_3\|_{L_1} < \eta/4$. Since η is arbitrary, we have established (4) for a smooth dense subset of Y_0, hence for all $v \in Y_0$. □

Remark 3.1. *The Poincaré inequality for a simplex S of diameter h uses the so-called star-shaped property of S with respect to any vertex. Thus, if one fixes a vertex $\vec{v}$ and selects a member ψ of $\mathcal{F}$ which vanishes at $\vec{v}$, one writes the standard proof as follows. Fix $\vec{x}$ on the face opposite $\vec{v}$ and let $\vec{y} = (1-s)\vec{v}+s\vec{x}$ be an arbitrary point on the line segment connecting $\vec{v}$ and $\vec{x}$. Here, $0 \le s \le 1$. We have, by the fundamental theorem of calculus for absolutely continuous functions and the chain rule:*

$$|\psi(\vec{y})|^2 = \left| \int_0^s \nabla\psi((1-\sigma)\vec{v} + \sigma\vec{x}) \cdot (\vec{x} - \vec{v})\, d\sigma \right|^2$$

$$\le h^2 \int_0^1 |\nabla\psi((1-s)\vec{v} + s\vec{x}))|^2\, ds.$$

Integration over S, via the iterated Fubini theorem as applied to rays from $\vec{v}$ and the opposite face, in that order, gives the result:

$$\int_S |\psi(\vec{y}|^2\ dy \le h^2 \int_S |\nabla\psi(\vec{y})|^2\ dy.$$

Lemma 3.4. *Suppose sequences f_n, g_n, q_n in $L_2(\Omega)$ satisfy the following conditions:*

$$f_n \rightharpoonup f, \text{ weakly; } g_n \to g, \text{ strongly; } q_n \to q, \text{ strongly, } \|q_n\|_\infty \le c.$$

Then

$$\int_\Omega f_n g_n q_n\, dx \to \int_\Omega fgq\, dx, \text{ as } n \to \infty.$$

Proof. It suffices to show that

$$f_n q_n \rightharpoonup fq, \text{ weakly in } L_2(\Omega), \text{ as } n \to \infty. \tag{26}$$

This, coupled with the strong convergence of $\{g_n\}$, implies the conclusion of the lemma. To establish (26), one notes that the hypotheses imply:

$$\lim_{n\to\infty} (v, f_n q_n)_2 = (v, fq)_2,\ \forall v \in C_0^\infty(\Omega).$$

This actually holds more generally for bounded measurable functions v. Now suppose $v \in L_2(\Omega)$ and $\varepsilon > 0$. By the denseness of $C_0^\infty(\Omega)$ in $L_2(\Omega)$, there exists a function $v_\varepsilon \in C_0^\infty(\Omega)$ such that $\|v - v_\varepsilon\|_2 < \varepsilon$. For some C a positive constant independent of n, we have

$$|(v, f_n q_n)_2 - (v_\varepsilon, f_n q_n)_2| \le \|v - v_\varepsilon\|_2 \|f_n\|_2 \|q_n\|_\infty \le C\varepsilon,$$

$$|(v, fq)_2 - (v_\varepsilon, fq)_2| \leq \|v - v_\varepsilon\|_2 \|f\|_2 \|q\|_\infty \leq C\varepsilon.$$

Finally, for v_ε as selected, there exists N_0 such that $n \geq N_0$ implies

$$|(v_\varepsilon, f_n q_n)_2 - (v_\varepsilon, fq)_2| \leq C\varepsilon.$$

These inequalities imply

$$\lim_{n\to\infty} |(v, f_n q_n)_2 - (v, fq)_2| \leq 3C\varepsilon,$$

so that

$$\lim_{n\to\infty} |(v, f_n q_n)_2 - (v, fq)_2| = 0.$$

This establishes (26) and concludes the proof. □

3.2 Further Applicability

Although the theory is directly applicable to mixed boundary problems for steady reaction/diffusion systems, the theory also applies to certain steady reaction/diffusion/convection systems in transport theory, in which: (i) the Einstein relations hold, relating diffusion and mobility, and, (ii) the drift is potential driven. In this case, a change to logarithmic variables via the Hopf-Cole transformation reduces the system to the self-adjoint form considered in this paper. The reader may consult [12] for elaboration.

We now discuss the role of variational inequalities. We confine the discussion to a brief comment, since the development is outside the scope of the paper. If one eliminates the nodal approximation as introduced in H7, and selects for the operators J_i the local identity operators, then the piecewise linear approximation has coefficients which satisfy the (system) variational inequality (10). If 'a priori' bounds can be derived in H^1, as in Theorem 3.1, then a subsequence of these approximations converges strongly to a solution of an analytical variational inequality by a straightforward convergence analysis utilizing Lemma 3.4 and the strong convergence of composition with $\mathbf{f}$. One requires only hypotheses H1–H5 and H8 in this case. However, the analytical variational inequality is seen to be equivalent to the system (4), via arguments utilizing hypothesis H2 (see [13] for illustration).

Finally, no statements regarding uniqueness are made in this article.

Acknowledgment

The author is grateful to the referees for suggestions which greatly improved the accuracy and exposition of this paper.

References

[1] R.A. Adams. *Sobolev Spaces*. Academic Press, 1975.

[2] S.C. Brenner and L.R. Scott. *The Mathematical Theory of Finite Elements*. Springer-Verlag, Texts in Applied Mathematics 15, 1994.

[3] S. Carl and J.W. Jerome. Trapping region for discontinuous quasilinear elliptic systems of mixed monotone type. Nonlinear Analysis 51 (2002), pp. 839–859.

[4] S. Carl, S. Heikkilä, and J.W. Jerome. Trapping regions for discontinuously coupled systems of evolution variational inequalities and application. J. Math. Anal. Appl. 282 (2003), pp. 424–438.

[5] P.G. Ciarlet and P.-A. Raviart. Maximum principle and uniform convergence for the finite element method. Comp. Methods Appl. Mech. Eng. 2 (1973), pp. 17–31.

[6] A. Drăgănescu, T.F. Dupont, and L.R. Scott. Failure of the discrete maximum principle for an elliptic finite element problem. Math. Comp. 74 (2005), pp. 1–23.

[7] D. Estep, M. Larson, and R.D. Williams. *Estimating the Error of Numerical Solutions of Systems of Reaction-Diffusion Equations*. Memoirs of the American Mathematical Society, vol. 696, AMS, Providence, 2000.

[8] W. Fleming. *Functions of Several Variables*, 2nd. ed. Springer-Verlag, 1977.

[9] J. Franklin. *Methods of Mathematical Economics*. Springer-Verlag, 1980.

[10] D. Gilbarg and N.S. Trudinger *Elliptic Partial Differential Equations of Second Order*. Grundlehren der mathematischen Wissenschaften 224, Springer-Verlag, 1977.

[11] E. Hewitt and K. Stromberg. *Real and Abstract Analysis*. Springer-Verlag, 1965.

[12] J.W. Jerome. The mathematical study and approximation of semiconductor models, in *Advances in Numerical Analysis: Large Scale Matrix Problems and the Numerical Solution of Partial Differential Equations*. (J. Gilbert and D. Kershaw, editors), Oxford University Press, 1994, pp. 157–204.

[13] J.W. Jerome. A trapping principle for discontinuous elliptic systems of mixed monotone type. J. Math. Anal. Appl. 262 (2001), pp. 700–721.

[14] A. Jüngel and A. Unterreiter. Discrete minimum and maximum principles for finite element approximations of non-monotone elliptic equations. Numerische Math. 99 (2005), pp. 485–508.

[15] A. Karátson and S. Korotov. Discrete maximum principles for finite element solutions of nonlinear elliptic problems with mixed boundary conditions. Numerische Math. 99 (2005), pp. 669–698.

[16] T. Kerkhoven and J.W. Jerome. L_∞ stability of finite element approximations to elliptic gradient equations. Numerische Math. 57 (1990), pp. 561–575.

[17] S. Korotov and M. Křížek. Acute type refinements of tetrahedral partitions of polyhedral domains. SIAM J. Numer. Anal. 39 (2001), pp. 724–733.

[18] M. Krasnosel'skii. *Topological Methods in the Theory of Non-Linear Integral Equations*. Pergamon Press, 1964.

[19] A. Kufner, O. John, and S. Fucik. *Function Spaces*. Nordhoff, 1977.

[20] E. Lieb and M. Loss. *Analysis*, second ed. Graduate Studies in Math. 14, Amer. Math. Soc., 2001.

[21] V. Ruas Santos. On the strong maximum principle for some piecewise linear finite element approximate problems of non-positive type. J. Fac. Sci. Univ. Tokyo Sect. IA Math. 29 (1982), pp. 473–491.

[22] R. Showalter. *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*. Mathematical Surveys and Monographs 41, American Mathematical Society, Providence, 1997.

[23] J. Smoller. *Shock Waves and Reaction-Diffusion Equations*. Grundlehren der Mathematischen Wissenschaften 258, Springer, 1983.

[24] M. Tabata. Uniform convergence of the upwind finite element approximation for semilinear parabolic problems. J. Math. Kyoto Univ. 18 (1978), no. 2, pp. 327–351.

[25] R. Vanselow. Relations between FEM and FVM applied to the Poisson equation. Computing 57 (1996), pp. 93–104.

[26] R.S. Varga. *Matrix Iterative Analysis*, expanded edition. Springer series in Computational Math. vol. 27, Springer-Verlag, 2000.