THE APPROXIMATION PROBLEM FOR DRIFT-DIFFUSION SYSTEMS*

Joseph W. Jerome†

June 20, 2011

Abstract

This review surveys a significant set of recent ideas developed in the study of nonlinear Galerkin approximation. A significant role is played by the Krasnosel'skii Calculus, which represents a generalization of the classical inf-sup linear saddle point theory. A description of a proper extension of this calculus, and the relation to the inf-sup theory are part of this review. The general study is motivated by steady-state, self-consistent, drift-diffusion systems. The mixed boundary value problem for nonlinear elliptic systems is studied with respect to defining a sequence of convergent approximations, satisfying requirements of: (1) optimal convergence rate; (2) computability; and, (3) stability. It is shown how the fixed point and numerical fixed point maps of the system, in conjunction with the Newton-Kantorovich method applied to the numerical fixed point map, permit a solution of this approximation problem. A critical aspect of the study is the identification of the breakdown of the Newton-Kantorovich method, when applied to the *differential* system in an approximate way. This is now known as the numerical loss of derivatives. As an antidote, a linearized variant of successive approximation, with locally defined subproblems bounded in number at each iteration, is demonstrated. In (2), a distinction is made between the outer analytical iteration, and the inner iteration, governed by numerical linear algebra. The systems studied are broad enough to include important application areas in engineering and science, for which significant computational experience is available.

Keywords 1 *Nonlinear Systems, Fixed Point Approximation, Krasnosel'skii Calculus, Approximate Newton Methods, Finite Element Methods, Inf-Sup Saddle Point Theory, Complexity*

AMS Classification Numbers 1 *35J65, 41A35, 65N15, 65N30*

---

*This work was supported by the National Science Foundation under grant DMS-9123208.

[2]Department of Mathematics,Northwestern University, Evanston, Il 60208

1 Introduction

The Galerkin and Petrov-Galerkin methods are among the principal procedures for the construction of approximate solutions of elliptic boundary value problems. They are variational methods, well suited to weak formulations. For linear problems, the most comprehensive theory, devoted to the convergence study of these methods, is the inf-sup theory, introduced by Babuška and Aziz [4] (see also [10]). For nonlinear problems the theory of choice, based on fixed point formulations, is that developed by Krasnosel'skii and his collaborators [32]. Although these theories have coexisted for more than two decades, no connection between the two has been made in the literature to the author's knowledge. However, it has recently been shown by the author [25] that the nonlinear calculus is a strict logical extension of the inf-sup theory. In this paper, we shall describe a complex circle of results, initiated some ten years ago, but recently closed, which indicates how the nonlinear calculus serves as both the framework for analysis as well as the indicator for the construction of efficient approximations for nonlinear systems, with the approximations defined by sequential linear systems. In order to provide a nontrivial example of the theory, we present, as the motivating application, the approximation problem for drift-diffusion systems, specifically the semiconductor application. We briefly discuss this now.

Potential driven drift-diffusion systems play a prominent role in pure and applied mathematics. In applications, they have many realizations, such as the Poisson-Nernst-Planck ionic transport model (see [40] for discussion) and semiconductor models [23], as well as models in many other areas of biology, chemistry, engineering, and physics. The study of channels in cell membranes is a recent example (cf. [7]). In analysis, the study of nonlinear elliptic systems is strongly motivated by steady-state systems of this type. A prototypical model, and the one selected for study in this paper, describes two oppositely charged carrier species moving in an electric field, induced by a potential u. This serves as a dependent variable of the system, together with quasi-Fermi levels v and w, for which the carrier concentrations have exponential representations, $\exp(u-v)$, $\exp(w-u)$. Under conditions of constant mobility and diffusion coefficients, connected by the Einstein relations, and zero recombination, the system is given on a Euclidean domain G by

$$\begin{aligned} F_1(u,v,w) &= -\nabla\cdot[\epsilon\nabla u] + e^{u-v} - e^{w-u} - \bar{N} = 0, && (1)\\ F_2(u,v) &= -\nabla\cdot[e^{u-v}\nabla v] = 0, && (2)\\ F_3(u,w) &= -\nabla\cdot[e^{w-u}\nabla w] = 0. && (3) \end{aligned}$$

Here, ϵ and $\bar{N}$ have interpretations as dielectric and ionization functions, respectively. Also included are the boundary conditions, specified by globally defined functions $\bar{u}$, $\bar{v}$, and $\bar{w}$ on the Dirichlet boundary, Σ_D, and homogeneous Neumann conditions on the complement within ∂G. Fortunately, simulation strategies for the solution of (1), (2), (3), are in place (cf. [5]); the latter, in fact,

initiated the analytical studies reported in this review. Relevant background material is taken from the following sources. The model was formulated by Van Roosbroeck [39]; analytical results are contained in [23], numerical results in [26], and comprehensive detail in the monograph [24]. Uniqueness is known to fail (cf. [50]) when $\bar{N}$ defines three or more junctions.

Effective numerical algorithms succeed as a careful blend of discretization and iteration. In this paper, the discretization is effected by piecewise linear finite elements. It is known that, when the proper choice of quadrature rule is made for the evaluation of the integrals, then exponential fitting results (cf. [48]). This has been known for some time to be an appropriate discretization for drift dominated systems (cf. [41]). The emphasis here, then, is upon the proper *iterative* strategy at the analytical level. In current terms, we are studying the convergence properties of the outer iteration, with inner iterations described by numerical linear algebra. It is relevant to emphasize here that the theory we present is a local theory. There is nothing to prevent its being joined to a global theory through continuation, for example, but only in special circumstances, such as gradient systems, does it simultaneously become a global theory. The drift-diffusion model presented here is not an example of a gradient system, however. It is useful to note that there is also an 'a posteriori' theory to complement the 'a priori' theory presented here, so as to ensure that no spurious solutions are computed (see [26]). We begin, then, by describing background results from the linear theory. We present the Galerkin (not Petrov-Galerkin) theory as introduced in [4], but we remark that the theory of [10] was also influential.

1.1 The linear saddle point theory: optimality

The inf-sup theory describes the following situation. It is desired to solve the operator equation, $\mathbf{L}u = f$, approximately. If B denotes the bilinear form of the weak formulation on a Hilbert space E, with inner product, $(\cdot\,,\cdot\,)$, assume:

1. continuity:
$$\mid B(v,w) \mid \leq C_1 \|v\| \, \|w\|. \tag{4}$$

2. sup condition:
$$\text{For } \; w \neq 0, \;\; \sup_{v} \mid B(v,w) \mid > 0. \tag{5}$$

3. inf-sup condition:
$$\inf_{\|v\|=1} \sup_{\|w\| \leq 1} \mid B(v,w) \mid \geq C_2 > 0. \tag{6}$$

Assume also sup and inf-sup conditions on an approximation space, E_n:

$$\text{For } \; \psi \neq 0, \;\; \sup_{\phi} \mid B(\phi,\psi) \mid > 0. \tag{7}$$

$$\inf_{\|\phi\|=1} \sup_{\|\psi\| \leq 1} \mid B(\phi,\psi) \mid \geq c_2 > 0. \tag{8}$$

One concludes the Galerkin approximation, u_n, is well defined by the relation,

$$B(u_n, \psi) = (\mathbf{J}f, \psi), \quad \forall \psi \in E_n, \tag{9}$$

where $\mathbf{J}$ denotes the Riesz map. Moreover, u_n is within a metric distance,

$$\delta_* \{1 + (C_1/c_2)\}, \tag{10}$$

of u, where

$$\delta_* := \|u - u_*\|, \tag{11}$$

and u_* is arbitrary in E_n (cf. [4, pp. 187-188]).

The exceptional nature of this result should not be overlooked. Under the hypotheses just enumerated, the Galerkin approximation from E_n is within a precise constant of the best approximation from E_n, and this is true for all n. But the question arises as to what is an efficient selection principle for the Galerkin subspaces E_n.

In 1936, Kolmogorov [31] introduced the concept of n-dimensional diameter, or n-width, for a set $\mathcal{K}$ in a normed linear space E:

$$d_n(\mathcal{K}) = \inf_{dim \mathcal{M} = n} \sup_{f \in \mathcal{K}} \inf_{g \in \mathcal{M}} \|f - g\|. \tag{12}$$

Even with perfect information about the elements of $\mathcal{K}$, this concept describes the limitation inherent in a finite dimensional approximation procedure which possesses a uniformity property over $\mathcal{K}$. In his pioneering paper, Kolmogorov characterized the diameters of a particularly significant class for differential equations,

$$\mathcal{K} = \{f \in H^k(a,b) : |f|_k = \{\int_a^b |f^{(k)}(x)|^2 \, dx\}^{1/2} \leq 1\}, \tag{13}$$

as viewed in the metric of $L_2(a,b)$. The eigenvalues of a natural Sturm-Liouville problem characterize the diameters; this leads directly to the exact estimate of the asymptotic order. This order reflects n^{-k} dependence of $d_n(\mathcal{K})$ on n. The extension of the Kolmogorov results to classes of functions of several variables was carried out in the author's doctoral thesis [18], and, in Euclidean space of dimension N, the order is $O(n^{-k/N})$. For a more complete description of n-width results, the reader should consult the books [33], [38], and [20], and the detailed papers of Höllig [16], [17].

The impact of this circle of ideas on approximate solutions of linear elliptic boundary value problems was evident within a few years of [18]. Independently, Schultz [43], the author [19], Babuška [3], and Aubin [2] described the optimality approximation conditions required of any Ritz-Galerkin sequence, constructed to converge to the solution u of $\mathbf{L}u = f$, where $\mathbf{L}$ is of order $2m$. The sequence $\{u_n\}$ should satisfy,

$$\|u - u_n\| \leq C d_n(\mathcal{K}), \tag{14}$$

where $\mathcal{K}$ is naturally contained in H^{2m}, and is defined via 'a priori' bounds. In principle, one can require (14) to hold in the metric of H^s, $0 \leq s < 2m$. This particular formulation must be adjusted if geometric or boundary value conditions lead to reduced solution regularity.

1.2 Galerkin approximation theory

By 1970, the finite element method had become firmly established within the Galerkin family. The name refers to decomposition of the physical domain G into cells or elements S of maximal diameter h; these are typically affine translates of a fixed model element. Local basis functions can then be defined with support confined to cells in the vicinity of a node, where the basis function takes on a specified value. These nodal basis functions reduce on each S to members of $\mathcal{P}_k$, the polynomials of total degree less than k. Quasi-uniform partitions, for which inscribed and circumscribed balls are comparable $\forall h$, will be assumed. Given the estimates (10) and (11) of the inf-sup theory, the mathematical foundation of this method rests upon two fundamental results, viz. , the Bramble-Hilbert lemma [8], or the generalization proven in [14], and the Aubin-Nitsche lemma ([1] and [36]). The purpose of the Bramble-Hilbert lemma is to utilize the energy norm best approximation property of the Galerkin approximation, as contained in (10) and (11), for piecewise polynomial trial spaces. It achieves this by examining optimal order interpolation or smoothing linear approximation operators, $\mathbf{Q}$, which reproduce polynomials locally, so that $u_* = \mathbf{Q}u$ in (11). In the application of the inf-sup theory, one typically takes $E = H^m(G)$, so that $\mathbf{Q}u \in H^m(G)$. To guarantee this, the constraint on the piecewise polynomial $\mathbf{Q}u$ is $\mathbf{Q}u \in C^{m-1}(G)$.

We shall describe the approximation theory in more general spaces. Begin by using the Sobolev representation theorem (cf. [44], [14], [20]) to write,

$$u = P_k + R_k, \text{ for each fixed cell } S, \tag{15}$$

where $P_k \in \mathcal{P}_k$, and where R_k is given explicitly by

$$R_k(x) = \sum_{|\alpha|=k} \int_S k_\alpha(x,y) u^{(\alpha)}(y)\, dy, \tag{16}$$

$$k_\alpha(x,y) = (k/\alpha!)(x-y)^\alpha k(x,y), \tag{17}$$

$$k(x,y) = \int_0^1 s^{-N-1} \chi(x + s^{-1}(y-x))\, ds, \tag{18}$$

and where $\chi \in C_0^\infty(S)$, with integral unity. By making use of the inequality,

$$|k(x,y)| \leq \|\chi\|_{L_\infty} diam(S)^N |x-y|^{-N}/N, x \neq y, \tag{19}$$

it is possible to use the Riesz potentials of harmonic analysis [45]. Thus, the authors of [14] proved the critical inequality, for C independent of h and S,

$$|R_k|_{j,p,S} \leq C h^{k-j} |u|_{k,p,S},\ 0 \leq j \leq k, \tag{20}$$

on $W_p^k(S)$, with norm $\|\cdot\|_{k,p,S}$ and seminorms $|\cdot|_{j,p,S}$. An immediate consequence is the inequality,

$$\|R_k\|_{m,p,S} = \{ \sum_{0 \le j \le m} |R_k|_{j,p,S}^p \}^{1/p} \le C h^{k-m} |u|_{k,p,S}. \tag{21}$$

This is used as follows. For any linear operator,

$$\mathbf{Q} : W_p^k(S) \mapsto \mathcal{P}_k,$$

which reproduces members of $\mathcal{P}_k$, for $m < k \le 2m$, we can estimate directly,

$$\|u - \mathbf{Q}u\|_{m,p,S} = \|R_k - \mathbf{Q}R_k\|_{m,p,S} \le \|R_k\|_{m,p,S} + \|\mathbf{Q}R_k\|_{m,p,S}. \tag{22}$$

Inequality (21), together with the assumption that $\mathbf{Q}$ is continuous from $W_p^k/\mathcal{P}_k$ to W_p^m, then yields

$$\|u - \mathbf{Q}u\|_{m,p,S} \le C h^{k-m} |u|_{k,p,S}. \tag{23}$$

The argument in [47, pp. 144-146] generalizes to show that when $\mathbf{Q}$ is the interpolation operator, the continuity condition holds for $N/p < k$. In a separate publication [46], Strang showed how smoothing, followed by interpolation, suffices when $N \ge kp$.

The Aubin-Nitsche lemma improves the order of approximation when measured in a ground space norm, such as the L_2 norm. For example, for second order equations with Neumann boundary conditions, the convergence is of order two in the L_2 metric for Galerkin finite element approximants. For an exposition, the reader is referred to [47]. Since the order just described corresponds, for quasi-uniform triangulations, to the n-dimensional diameter order, both in energy and L_2 norms, it follows that the finite element method achieves a remarkable approximation property for *implicitly* defined members u of $\mathcal{K}$. Moreover, by a well known procedure, the implicit definition of the approximation u_n is no more complicated than solving a sparse matrix equation of order n. The literature on this matrix problem abounds, both for direct (Gaussian elimination) and iterative methods.

1.3 Summation

This is where matters stood in 1972 within the theoretical numerical analysis community. In this country, development in subsequent years dealt, for the most part, with the analysis of linear problems in the case of steady-state systems. Particular emphasis was placed upon complicated linear systems, such as those of linear elasticity, and upon saddle point problems, such as the Stokes problem. A comprehensive account of the finite element theory associated with these general formulations may be found in [9]. For nonlinear parabolic systems, significant work was carried out by Jim Douglas Jr. and his students and colleagues, initiated in the late 1960s. Among the contributors were Todd Dupont,

Mary Wheeler, and Richard Ewing. These ideas did not directly translate into effective methods for steady-state systems, however. Significant issues dealt with in the intervening years, which for the purposes of this paper are second level studies and are not discussed, were the development of numerical methods for strongly convective linear elliptic problems, including the streamline diffusion method and its connection with Petrov-Galerkin methods, domain decomposition methods to facilitate parallelism, the highly successful *h-p* method for singularity resolution, and associated 'a posteriori' estimation, and special finite elements, such as mixed methods. An important trend for the future was the introduction of differential geometry into numerical analysis through the studies of Werner Rheinboldt. An early presentation of some of the ideas discussed here was [26]. We indicate now the independent development of the framework for nonlinear problems. It involves, in a significant way, contributions by Soviet mathematicians.

2 The approximation problem

Write the system (1), (2), (3) as $\mathbf{F}(z) = 0$, and suppose the Sobolev regularity exponent of z is $1+\theta$, $0 < \theta \leq 1$, and 'a priori' bounds locate the solution in a ball of radius ρ in $\prod_1^3 H^{1+\theta}$. A regularity theory for the mixed problem was developed in [35]. As viewed in $\prod_1^3 H^1$, call this set $\mathcal{K}$. We shall require an approximation sequence $\{x_n\}$ for z to satisfy:

1. Optimal Order Approximation;

$$\|z - x_n\| \leq C d_n, \; n \to \infty. \; (d_n \approx n^{-\frac{\theta}{N}}) \tag{24}$$

2. Computability; $\forall n$, x_n is defined by an outer sequence of linear operator equations, bounded in number independently of n, each member of which can be constructed by a (bounded) number of sparse matrix inner sequence calculations.

3. Stability (Discrete Maximum Principle); $\{x_n\}$ is bounded in $\prod L_\infty$.

The determination of such a sequence is what we have termed the approximation problem for drift-diffusion systems. That its resolution is quite subtle is indicated by the following subsections. Because of the limitations on the regularity of solutions, we shall work exclusively in this paper with piecewise linear finite element approximation. This is the case $k = 2$ of the introduction. It may seem curious that we have required a bounded number of outer iterations as part of the computability requirement, although the size of the problem grows with n. The reasoning is as follows. The error tolerance determines the grid size, and ultimately n asymptotically, via (1). For a given n, at most, say, k_0 linear systems of size n must be solved to determine x_n, according to (2). If the computational complexity of the total inner iterations for each outer iteration can be estimated by an invariant function $p(n)$, then an estimate for the

computational complexity of the process is at most $k_0 p(n)$. In this sense, the computational complexity of the numerical linear algebra governs that of the total computation. Thus, one can asymptotically estimate in advance the complexity function, $C(\epsilon)$, for a given error tolerance ϵ. This paper, then, addresses the following two questions:

1. Can the Galerkin approximation be estimated for nonlinear systems?

2. How should the Galerkin approximation be computed so that an efficient 'a priori' estimate for the associated complexity is possible?

To telescope the conclusions of the following sections, we find that the Krasnosel'skii theory allows the estimation of piecewise linear Galerkin approximation, and this process achieves the mandate of (1). At this level, the theory is simply an effective mathematical framework. In principle, one could stop at this point, ignoring (2), in the design of a strategy, and relying on well documented methods for solving the finite dimensional nonlinear systems, obtained via the matrix formulation arising from nodal basis functions. However, the number of outer iterations required to achieve the solution to an accuracy compatible with the error described by (1) will likely depend on h (i.e. , n) in a manner for which 'a priori' estimates are not available, or likely not as efficient as the alternative strategy proposed here. This is particularly true if the differential formulation is used to define Newton's method. A message conveyed in the sequel is that the Krasnosel'skii framework can be used effectively to define the linear mappings of (2), based upon Newton's method applied to the numerical fixed point map. There remains the issue of whether this strategy can actually be achieved in a manner consistent with locally defined mappings, i.e. , via sparse matrix calculations. It turns out that GMRES, based upon residual minimization over Krylov subspaces, is particularly suited. The method is facilitated by the compactness of the derivative of the fixed point map, which forces the spectrum of the identity shifted map to cluster near one. A residual error estimate is available, which suggests that the function $p(n)$ above may be selected to be the work function of a fixed number of GMRES iterations. We return to this in the sequel. Finally, the pointwise stability of (3) may appear to be an attractive, but nonessential, feature of the scheme. Actually, it is a necessary component of the overall stability, compatible with operator differentiability. We emphasize that the pointwise stability is a proven property of the Galerkin approximations themselves, but is *required* of the linearized approximations.

The use of the n-width and the nonlinear calculus masks a limitation of the theory, the use of quasi-uniform grids. The reason why the h-p theory is not easily integrated is that such local refinement is not uniform over balls of functions in smooth spaces, a concept which is implicit in the width estimates, and those of the nonlinear calculus. A natural next step would be the development of a theory simultaneously allowing for complexity estimation and local grid refinement.

2.1 Exact Newton-Kantorovich sequences

A reasonable starting point to construct linear approximations is an operator version of Newton's method. Recall the two fundamental properties required of an exact Newton method, with $\{z^m\}$ the Newton sequence:

$$z^m - z^{m-1} \stackrel{\text{def}}{=} -\mathbf{G}(z^{m-1})\mathbf{F}(z^{m-1}), \tag{25}$$

where z, $\mathbf{F}(z) = 0$, is sought as $z = \lim_{m\to\infty} z^m$, and $\mathbf{G}(z^{m-1}) \equiv (\mathbf{F}')^{-1}(z^{m-1})$. The properties are:

1. Boundedness;
$$\|\mathbf{G}(y)\mathbf{F}(y)\| \leq M\|\mathbf{F}(y)\|. \tag{26}$$
2. Lipschitz Continuous Derivative;
$$\|\mathbf{F}'(x) - \mathbf{F}'(y)\| \leq 2M\|x - y\|. \tag{27}$$

Adjoined is a mechanism ensuring that successive iterates lie within the domain of definition of $\mathbf{F}$. A quadratic convergence result then holds (cf. [27], [28]). The version presented here was stated and proved in [22].

Theorem 2.1 *If* $\|\mathbf{F}(z^0)\| \leq \rho^{-1}$ *and* $\eta = 2M^3\rho^{-1}$, *with* $\eta \leq \frac{1}{2}$, *then*

$$\|z - z^k\| \leq \frac{\theta_k}{\eta\rho}\Big(\prod_{j=0}^{k} \tau_j^{2^{k-j}}\Big)\frac{(1 - \sqrt{1 - 2\eta})^{2^k}}{2^k}. \tag{28}$$

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

It is possible to combine this result with a form of continuation, or homotopy, so that the root tracking procedure always remains within the domain of convergence of Newton's method (cf. [22]; also, the damped Newton iteration strategy of [6]). The hypotheses of Theorem 2.1 have been verified in a slightly different context [13]. At this stage, it appears that we are very close to a solution of the approximation problem as formulated above. This is not the case, however. We elaborate in the next subsection.

2.2 Inexact Newton-Kantorovich sequences

We begin with the innocuous observation that $\mathbf{G}$ is not computable without approximation. Suppose the operator valued function $\mathbf{G}(y)$ is approximated by $\mathbf{G}_h(y)$, via a numerical method of order $O(h^2)$, and (25) is accordingly modified:

$$z^m - z^{m-1} \stackrel{\text{def}}{=} -\mathbf{G}_h(z^{m-1})\mathbf{F}(z^{m-1}). \tag{29}$$

Then each differentiation of

$$[\mathbf{G}(y) - \mathbf{G}_h(y)]\mathbf{F}(y)$$

leads to loss of order one in convergence order:

$$\|D^\alpha[\mathbf{G}(y) - \mathbf{G}_h(y)]\mathbf{F}(y)\|_{L_2} \leq Ch^{2-|\alpha|}\|\mathbf{F}(y)\|_{L_2}$$

if $\theta = 1$. Thus, the approximation of the identity,

$$[\mathbf{F}'(z^{m-1})\mathbf{G}_h(z^{m-1}) - I]\mathbf{F}(z^{m-1}), \tag{30}$$

is of asymptotic order $O(\|\mathbf{F}(z^{m-1})\|)$, since $\mathbf{F}'$ is of order two, and it is not possible to choose h *adaptively* so that the asymptotic order is $O(\|\mathbf{F}(z^{m-1})\|^2)$. This would be required for an extended Kantorovich theory to hold, as outlined and proven in [22]. This breakdown of approximate Newton methods, based upon classical numerical methods applied to the differential formulation, has been termed a numerical loss of derivatives by the author [21]. It was shown that this phenomenon is an exact analogue of that identified by Moser in his fundamental paper [34]. This has led to smoothing via Nash-Moser iteration. It is accordingly possible to modify the iterations (29) to incorporate a smoothing applied to the right hand side at each step. This can be interpreted as high frequency cutoff and yields superlinear convergence, but requires a significant amount of regularity. We shall not pursue this line here, but shall ultimately make use of compact mappings to provide the preconditioning found in computational procedures for solving the approximation problem. In particular, we are able to define an *exact* Newton method for the computation of the numerical fixed point, rendering (30) unnecessary.

3 Mappings and Galerkin approximations

It is natural to examine the Galerkin approximations, though these do not directly satisfy the computability requirement introduced in §2, since the associated algebraic problem is nonlinear. Moreover, through a natural commutativity, attempts at an adaptive choice of h, correlated with the use of Newton's method for these algebraic problems, confronts the same loss of derivatives phenomenon identified in the previous section, so that one does not have an upper bound, independent of n, on the number of linear problems. Nonetheless, we shall find that creating a fixed point mechanism for studying the convergence of the Galerkin approximations will ultimately lead to the resolution of the problem.

We now introduce the system mappings. Define the mapping $\mathbf{U}_f : (v, w) \mapsto u$ by solution of (1) for u if v and w are given. $\mathbf{V}_f : u \mapsto v$ is defined through solution of (2) for given u. $\mathbf{W}_f : u \mapsto w$ is defined similarly. Define a fixed point map $\mathbf{T}_f$ via

$$\mathbf{T}_f = [\mathbf{V}_f, \mathbf{W}_f] \circ \mathbf{U}_f. \tag{31}$$

Note that a fixed point of this map can be identified with the v, w components of a solution triple. These definitions are not mathematical definitions in the strict sense since we have not specified the domains of the mappings. This is facilitated by the maximum principles. These, as well as corresponding discrete principles for the Galerkin approximations, are presented now.

3.1 Maximum and discrete maximum principles

Given the piecewise linear finite element subspace S_h, with members vanishing on the Dirichlet boundary, Σ_D, the finite element equations for the uncoupled potential equation are given, for $i = 1, \cdots, M$, by

$$\langle \epsilon(x)\nabla U_h, \nabla\phi_i\rangle + \langle e^{U_h - v} - e^{w-U_h}, \phi_i\rangle - \langle \bar{N}, \phi_i\rangle = 0, \tag{32}$$

where $U_h = \mathbf{U}_h[v, w]$ is a finite element function, and ϕ_i are test functions comprising a nodal basis of S_h, so that $\phi_i(x_j) = \delta_{ij}$. Select the piecewise linear interpolant $\bar{u}_I$ of $\bar{u}$ so that $U_h \in \bar{u}_I + S_h$. Next, characterize terms $\mathbf{V}_h(U_h) = v_h$, $\mathbf{W}_h(U_h) = w_h$, by

$$\langle e^{U_h - v_h}\nabla v_h, \nabla\phi_i\rangle = 0, \quad \text{for } i = 1, \cdots, M, \tag{33}$$

$$\langle e^{w_h - U_h}\nabla w_h, \nabla\phi_i\rangle = 0, \quad \text{for } i = 1, \cdots, M. \tag{34}$$

Here, $v_h \in \bar{v}_I + S_h$, $w_h \in \bar{w}_I + S_h$, where $\bar{v}_I$ and $\bar{w}_I$ are interpolants of $\bar{v}$ and $\bar{w}$, respectively. The numerical fixed point map, $\mathbf{T}_h$, is defined via,

$$\mathbf{T}_h = [\mathbf{V}_h, \mathbf{W}_h] \circ \mathbf{U}_h. \tag{35}$$

A fixed point is equivalent to a solution of the Galerkin equations. Note that the Galerkin equations are the coupled version of (32), (33), and (34), in analogy with the coupled system, (1), (2), and (3).

The discrete maximum principles are applicable to $\mathbf{U}_h$, $\mathbf{V}_h$ and $\mathbf{W}_h$. We shall illustrate these ideas by reference to a generic gradient equation. For a strictly positive consider:

$$-\nabla\cdot[a(x)\nabla u(x)] + f(x, u(x)) = g(x). \tag{36}$$

Here, $a, g \in L_\infty$, and f is increasing and locally Lipschitz in u for each $x \in G$, with $f^{-1}(x, \cdot)$ the corresponding inverse. We shall state bounds which are satisfied both for the solution of (36) and for its piecewise linear Galerkin approximation. We refer to such bounds as reduced bounds.

Statement of Reduced PDE Bounds:

$$\gamma := \min\{\inf \bar{u}, \gamma'\} \le u \le \delta := \max\{\sup \bar{u}, \delta'\}, \tag{37}$$

$$\gamma' = \inf_{x\in G} f^{-1}(x, \inf_{y\in G} g(y)), \qquad \delta' = \sup_{x\in G} f^{-1}(x, \sup_{y\in G} g(y)). \tag{38}$$

For these reduced bounds to hold for the Galerkin approximation of (36), certain conditions on the mesh must be assumed, as was detailed in [29]. The primary condition requires some terminology and notation. The ideas to follow are closely related to the theory of M-matrices (cf. [49]) and M-functions (cf. [37]), though not directly deducible from these theories.

DEFINITION. Let S be an N-dimensional simplicial finite element such that

1. V is the volume;

2. $\vec{v}_i$ is a vertex;
3. e_{ij} is the edge connecting vertices $\vec{v}_i$ and $\vec{v}_j$;
4. F_k is the face opposite the vertex k, with measure $|F_k|$;
5. h_i is the normal distance of $\vec{v}_i$ to F_i;
6. γ_{ij} is the angle between the inward normal vectors to the faces F_i and F_j;
7. ϕ_l is the piecewise linear nodal basis function which is 1 at vertex $\vec{v}_l$;
8. $$\alpha_{ij} \equiv \int_S a(x)\nabla\phi_i \cdot \nabla\phi_j dx$$
   is the ijth entry of the *element* stiffness matrix;
9. $\langle a(x)\rangle \equiv \int_S a(x)dx/V$, the average of $a(x)$ over the element S;
10. a_{ij} is the ijth element of the assembled stiffness matrix.

Remark 1. It was shown in [29] that

$$\alpha_{ij} \equiv \int_S a(x)\nabla\phi_i \cdot \nabla\phi_j dx = \langle a(x)\rangle \cos(\gamma_{ij})\frac{1}{h_i h_j}V,$$

or

$$\alpha_{ij} = \langle a(x)\rangle \cos(\gamma_{ij})\frac{|F_i||F_j|}{N^2 V}.$$

In [29] it was also shown that L_∞ stability, in terms of satisfying the maximum principle, is a consequence of the following assumption.

Assumption 1.

1. In N dimensions, where $N \geq 2$, we require that for every edge e_{jk} the off-diagonal element a_{jk} in the stiffness matrix satisfies
   $$a_{jk} = \sum_{S \text{ adjacent } e_{jk}} \langle a(x)\rangle_S \cos(\gamma_{jk}^{(S)})\frac{V^{(S)}}{h_j h_k} \leq -\frac{\rho}{h_{\max}^2} \sum_{S \text{ adjacent } e_{jk}} V^{(S)},$$
   with $\rho > 0$.

There is a condition that h must be sufficiently small. This is expressed in terms of a uniform Lipschitz constant D for f in its second argument, in terms of ρ, and in terms of Euclidean dimension N:

$$h^2 \leq \rho\frac{(N+1)(N+2)}{D}. \tag{39}$$

1. A sufficient condition for Assumption 1, in terms of the simplicial decomposition, is the following.

In two dimensions, require for every edge e_{jk} :

$$a_{jk} = \tfrac{1}{2}[\langle a(x)\rangle_{T_1} \cot(\omega_1) + \langle a(x)\rangle_{T_2} \cot(\omega_2)] \geq \rho > 0,$$

where the T_i are the two triangles adjacent to edge e_{jk}, and the ω_i are the two angles opposite to the edge e_{jk}. In higher dimensions, the above condition of Assumption 1 generalizes in a significant way the condition derived by the authors of [12]. These authors require the angle between the vectors normal to any two faces of the same polyhedron to be bounded uniformly from above by $\pi/2 - \eta,\ \eta > 0$.

3.2 The fixed point and numerical fixed point maps

One can now define the domains of $\mathbf{T}_f$ and $\mathbf{T}_h$ in $\prod_1^2 H^1$ by these reduced bounds. Since $f = 0$ in each of the second and third equations, we have the defining bounds, for the domain in both cases,

$$\inf \bar{v} \leq v \leq \sup \bar{v},\ \inf \bar{w} \leq w \leq \sup \bar{w}. \tag{40}$$

It is possible to show that $\mathbf{T}_f$ is compact, continuous, and invariant on this closed and convex set, hence has a fixed point (see [23]) by an application of the Schauder fixed point theorem. Similar remarks apply to $\mathbf{T}_h$. In particular, we are studying systems for which there is a rigorous 'a priori' existence theory. Typically, to conform with the nonlinear calculus, we may restrict the domain of this latter mapping to its intersection with a finite dimensional space.

4 The abstract calculus

Let E be a Banach space, $\mathbf{T}$ a mapping from an open set Ω into E, and x_0 a fixed point,

$$\mathbf{T}x_0 = x_0. \tag{41}$$

In addition, $\{E_n\}$ is a sequence of subspaces of E of dimension $r(n) \geq n$, with $\mathbf{T}_n : \Omega_n \mapsto E_n$, $\Omega_n := \Omega \cap E_n$. Finally, let $\{\mathbf{P}_n\}$ be a family of linear projections onto E_n.

4.1 The approximation estimates

The following theorem is proved in [32, Chapter 19].

Theorem 4.1 *Let the operators* $\mathbf{T}$ *and* $\mathbf{P}_n\mathbf{T}$ *be Fréchet differentiable in* Ω*, and* $\mathbf{T}_n$ *Fréchet differentiable in* Ω_n*. Assume* $\mathbf{I} - \mathbf{T}'(x_0)$ *continuously invertible in* E*. Let*

$$\|\mathbf{P}_n(x_0) - x_0\| \to 0,$$

$$\|\mathbf{P}_n\mathbf{T}\mathbf{P}_n x_0 - \mathbf{T}x_0\| \to 0, \quad \|\mathbf{P}_n\mathbf{T}'(\mathbf{P}_n x_0) - \mathbf{T}'(x_0)\| \to 0,$$

$$\|[\mathbf{T}_n - \mathbf{P}_n\mathbf{T}]\mathbf{P}_n x_0\| \to 0, \quad \|[\mathbf{T}_n' - (\mathbf{P}_n\mathbf{T})'](\mathbf{P}_n x_0)\| \to 0,$$

as $n \to \infty$. Then the numbers α_n, defined by

$$\alpha_n = \|[\mathbf{I} - \mathbf{T}_n'(\mathbf{P}_n x_0)]^{-1}(\mathbf{I} - \mathbf{T}_n)\mathbf{P}_n x_0\|, \tag{42}$$

satisfy $\alpha_n \to 0$. Finally, assume that

$$\|\mathbf{T}_n^{'}(x) - \mathbf{T}_n^{'}(\mathbf{P}_n x_0)\| \leq \epsilon \quad for \quad (n \geq n_\epsilon;\ \|x - \mathbf{P}_n x_0\| \leq \delta_\epsilon,\ x \in \Omega_n). \tag{43}$$

Then there exist n_0 and $\delta_0 > 0$ such that, when $n \geq n_0$, $\mathbf{T}_n x_n = x_n$ has a unique solution x_n in the ball $\|x - x_0\| \leq \delta_0$. Moreover, n_0 can be chosen so that

$$\frac{\alpha_n}{1+q} \leq \|x_n - \mathbf{P}_n x_0\| \leq \frac{\alpha_n}{1-q},\ n \geq n_0, \tag{44}$$

for given fixed $0 < q < 1$. If $\mathbf{T}$ is affine, then $q = 0$. It follows that

$$\|x_n - x_0\| \leq \|[\mathbf{I} - \mathbf{P}_n]x_0\| + \|x_n - \mathbf{P}_n x_0\| \to 0 \quad \text{as} \quad n \to \infty, \tag{45}$$

and

$$c_1\|\mathbf{P}_n\mathbf{T}x_0 - \mathbf{T}_n\mathbf{P}_n x_0\| \leq \|x_n - \mathbf{P}_n x_0\| \leq c_2\|\mathbf{P}_n\mathbf{T}x_0 - \mathbf{T}_n\mathbf{P}_n x_0\|, \tag{46}$$

for positive constants c_1 and c_2.

The argument makes use of the contractive mapping,

$$\mathbf{B}x = x_* - [\mathbf{A}_n'(x_*)]^{-1}\{\mathbf{A}_n x_* + [\mathbf{A}_n x - \mathbf{A}_n x_* - \mathbf{A}_n'(x_*)(x - x_*)]\}, \tag{47}$$

where $x_* = \mathbf{P}_n x_0$ and $\mathbf{A} = \mathbf{I} - \mathbf{T}_n$, to deduce a fixed point, x_n.

The following questions present themselves.

1. Is this theorem related to the inf-sup theory?
2. Does it provide a mechanism for solving the approximation problem?

We shall take up these questions in turn. The answer to the first question is affirmative, as we shall outline in the remainder of this section. The second question requires a qualified response, involving an extension of Theorem 4.1. This extension requires the contraction of (47) to hold on a ball defined by a stronger norm (cf. (66)) than that used to describe the convergence inequalities (46), as well as the construction of an asymptotic sequence of such balls. Its discussion will take us through to the end of this paper.

4.2 Reformulation of the inf-sup theory

In this section, we reformulate the inf-sup theory in terms of an affine fixed point mapping. We show that the numbers α_n below have several characterizations. These are the numbers estimated by the inf-sup theory, and the multiple realizations permit the application of the Krasnosel'skii framework. We note briefly the role of the hypotheses of the inf-sup theory.

1. (4) $\Rightarrow$ $\mathbf{L}$ may be identified with a continuous linear map $\mathbf{R}$ on E.

2. (6) $\Rightarrow$ $\mathbf{R}^{-1}$ exists on a closed domain of E.

3. (5) $\Rightarrow$ Domain and range of $\mathbf{R}$ are all of E.

There is a fixed point formulation:

$$\mathbf{T}u = u, \;\; \mathbf{T}v := (\mathbf{I} - \mathbf{R})^{-1}(v - \mathbf{J}f), \;\; \mathbf{R} = \mathbf{JL}. \tag{48}$$

$\mathbf{T}$ is affine. Its domain independent operator derivative is defined by

$$\mathbf{T}'v = (\mathbf{I} - \mathbf{JL})^{-1}v, \;\; \forall v \in E. \tag{49}$$

The results to be described now were first obtained in [25]. There is a substantial intersection with this paper, but it was our intent to provide for the reader an integrated presentation of the abstract calculus, its application to a significant problem, as well as its natural connection to the saddle point theory. We begin by itemizing key properties of the construction of [4]. It is now understood that E is a Hilbert space. In applications, E is typically the Sobolev space, H^1.

1. The fixed point map, $\mathbf{T}$, is defined by (48). The mapping $\mathbf{R}$ in the definition of $\mathbf{T}$ is determined by the relation,

$$B(u, v) = (\mathbf{R}u, v), \;\; \forall u, v \in E. \tag{50}$$

2. If the numerical fixed point map on E_n is defined by

$$\mathbf{T}_n v = (\mathbf{I} - \mathbf{P}_n\mathbf{R})^{-1}(v - \mathbf{P}_n\mathbf{J}f), \; \forall v \in E_n, \tag{51}$$

it follows that the Galerkin approximation, written as u_n, is characterized as a fixed point of $\mathbf{T}_n$. Indeed, it is shown in ([4, pp. 187-188]) that $\mathbf{P}_n\mathbf{R}$ maps E_n into itself according to

$$B(\phi, \psi) = (\mathbf{P}_n\mathbf{R}\phi, \psi), \;\; \forall \phi, \psi \in E_n. \tag{52}$$

The fixed point property then follows immediately from (9) and (52).

3. The domain independent derivative of $\mathbf{T}_n$ is given by

$$\mathbf{T}'_n v = (\mathbf{I} - \mathbf{P}_n\mathbf{R})^{-1}v, \;\; \forall v \in E_n. \tag{53}$$

4. If v is an eigenvector of $\mathbf{T}'$, given by (49), corresponding to eigenvalue, 1, then v is also an eigenvector of $\mathbf{R}$, corresponding to eigenvalue, 0. This follows from the relation,

$$\mathbf{T}'v = (\mathbf{I} - \mathbf{R})^{-1}v = v.$$

Since the latter contradicts both the sup condition (5) and the inf-sup condition (6), 1 is not an eigenvalue. This represents the principal non-singularity hypothesis of Theorem 4.1.

Theorem 4.2 *Let u denote the unique solution of the operator equation, $\mathbf{L}u = f$. Set*

$$v_n = [\mathbf{I} - \mathbf{T}_n']^{-1}(\mathbf{I} - \mathbf{T}_n)\mathbf{P}_n u. \tag{54}$$

Define

$$\alpha_n = \|v_n\|. \tag{55}$$

Then the numbers α_n are characterized by the following equivalent statements.

1. *From* (44)*, with $q = 0$,*

$$\alpha_n = \|u_n - \mathbf{P}_n u\|. \tag{56}$$

2. *If $(\mathbf{P}_n\mathbf{R})^{-1}$ denotes inversion on E_n,*

$$\alpha_n = \|(\mathbf{P}_n\mathbf{R})^{-1}[\mathbf{P}_n\mathbf{R}(\mathbf{P}_n u - u)]\|. \tag{57}$$

Independently, the relations,

$$\mathbf{P}_n\mathbf{R}v_n = \mathbf{P}_n\mathbf{R}\mathbf{P}_n u - \mathbf{P}_n\mathbf{J}f, \tag{58}$$

$$\mathbf{P}_n\mathbf{R}u = \mathbf{P}_n\mathbf{R}u_n = \mathbf{P}_n\mathbf{J}f, \tag{59}$$

hold. We may infer, therefore, the inequalities,

$$\|u_n - \mathbf{P}_n u\| \leq \frac{C_1}{c_2}\|\mathbf{P}_n u - u\|, \tag{60}$$

$$\|u_n - \mathbf{P}_n u\| \geq \frac{C_2}{C_1}\|\mathbf{P}_n u - u\|. \tag{61}$$

The inf-sup estimate follows directly from (60)*.*

Proof We begin with (54), apply the mapping, $[\mathbf{I}-\mathbf{T}_n']$, insert the representations given by (51) and (53), and simplify to obtain (58). In order to obtain the second equality in (59), begin with the relation, $\mathbf{T}_n u_n = u_n$, substitute (51), and simplify. In a similar manner, the relation, $\mathbf{T}u = u$, implies

$$\mathbf{R}u = \mathbf{J}f,$$

so that both equalities in (59) are seen to hold. The identity (57) is now a simple consequence of (56) and (59). Note that (56) is the consequence of the nonlinear calculus, and it is in this sense that the inf-sup theory is subsumed. Alternatively, a direct proof of (57) uses (58) and (59), and the definition (55) of α_n. Inequalities (60) and (61) follow from standard mapping properties of the inf-sup theory, as applied to $\mathbf{R}$ and $\mathbf{P}_n\mathbf{R}$. *Box*

5 Calculus for the extended system maps

Recall that the fixed point and numerical fixed point mappings associated with the system, (1), (2), (3), are originally defined only on a closed convex set,

$$K = \{[v, w] : \alpha_v \leq v \leq \beta_v,\ \alpha_w \leq w \leq \beta_w\}, \tag{62}$$

where

$$\alpha_v = \inf_{\Sigma_D} \bar{v}, \qquad \alpha_w = \inf_{\Sigma_D} \bar{w}, \tag{63}$$

$$\beta_v = \sup_{\Sigma_D} \bar{v}, \qquad \beta_w = \sup_{\Sigma_D} \bar{w}. \tag{64}$$

We now describe the framework for the explicit calculus. We set $E = \prod_1^2 H^1(G)$ and E_n = linear span $\{\bar{v}_I, S_h\}\otimes$ linear span $\{\bar{w}_I, S_h\}$, with $\mathbf{P}_n$ the orthogonal projection onto E_n. The map $\mathbf{T}$, required to apply the operator calculus of [32], must be defined on an open set in function space. In this context the suitable space is $\prod_1^2 H^1(G)$. However, a number of the results require 'a priori' bounds on the extrema of the functions u, v, and w, similar to the maximum principles presented earlier. Thus, $\mathbf{T}$ will be an extension map of $\mathbf{T}_f$. Because the set K is not open, we modify the definition of $\mathbf{T}$ such that the assumption that the preimage $[v, w]$ lies in K can be removed. To achieve this, we compose a $\mathbf{T}$-like map with a truncation operator $\mathbf{Tr}$, which leaves $[v, w]$ unaffected within K (where the solution lies), but which restricts the range to a set K_1 (cf. (65)) which is only slightly larger than K. Thus, we introduce $h_i \in C_0^\infty(R)$, $i = 1, 2$, such that *support* $h_i = [\alpha_i, \beta_i]$, and

$$\begin{aligned} h_1(t) &= t, \qquad \inf_{\Sigma_D} \bar{v} \leq t \leq \sup_{\Sigma_D} \bar{v}, \\ h_2(t) &= t, \qquad \inf_{\Sigma_D} \bar{w} \leq t \leq \sup_{\Sigma_D} \bar{w}. \end{aligned}$$

Below we shall define an open ball Ω, centered at zero in $\prod_1^2 H^1$, on which

$$\mathbf{Tr}[v, w] := [\mathbf{h}_1(v), \mathbf{h}_2(w)], \qquad \forall [v, w] \in \Omega.$$

Note that the range of $\mathbf{Tr}$ is contained in $K_1 \subset \prod_1^2 L_\infty$, where

$$K_1 = \{[v, w] \in \prod_1^2 L_\infty : \alpha_1 \leq v \leq \beta_1, \alpha_2 \leq w \leq \beta_2\}. \tag{65}$$

$\mathbf{Tr}$ is Lipschitz continuously differentiable when restricted to $\Omega \cap \prod_1^2 L_\infty$, but is simply Fréchet differentiable on Ω. This entails the introduction of a stronger component norm,

$$\|u\| = \max(\|u\|_{H^1}, \|u\|_{L_\infty}). \tag{66}$$

One wishes to measure convergence (the bounds of (46)) in the weaker norm, while maintaining certain mapping invariance and assumptions in the stronger

norm. This has been carried out fully in [24]. Here, we simply outline the general features of the theory. We consider the extension maps $\mathbf{U}$ of $\mathbf{U}_f$, $\mathbf{V}$ of $\mathbf{V}_f$, and $\mathbf{W}$ of $\mathbf{W}_f$ defined as previously, with elements in the domain of $\mathbf{U}$ now taken from $K_1 \supset K$. In terms of these quantities, $\mathbf{T}$ may be defined by

$$\mathbf{T} = [\mathbf{V} \circ \mathbf{U} \circ \mathbf{Tr}, \mathbf{W} \circ \mathbf{U} \circ \mathbf{Tr}]. \tag{67}$$

The domain Ω of the map $\mathbf{T}$ is defined in tandem with the composition maps defining $\mathbf{T}$ in such a way as to ensure that Ω is invariant under $\mathbf{T}$ and contains a fixed point. $\mathbf{T}_n$ may be defined analogously; for consistency with the previous subsection, we may wish to consider $\mathbf{T}_n$ as restricted to Ω_n, but this is unimportant. Note that $\mathbf{T}_n$ may be viewed as an extension of the mapping $\mathbf{T}_h$, introduced in an earlier section (§3.1).

An important approximation property of $\mathbf{P}_n$ on the union of the convex hull, *co* $R_{\mathbf{T}}$, of the range of $\mathbf{T}$, with $\prod_1^2 H^{1+\theta}(G) \cap H^1_{0,\Sigma_D}(G)$, is

$$\|\mathbf{P}_n\tau - \tau\|_{\prod H^1} \leq ch^\theta, \qquad \|\tau\|_{\prod H^{1+\theta}} \leq 1. \tag{68}$$

This is a consequence of standard approximation theory. Here, $1 < \theta \leq 1$ depends upon Euclidean dimension N, and reflects the transition point boundary condition singularities. Note that, in (68), τ is a member of the set, *co* $R_{\mathbf{T}} \cup \prod_1^2(H^{1+\theta} \cap H^1_{0,\Sigma_D})$. Similar approximation results hold for the approximation of $\mathbf{T}$ by $\mathbf{T}_n$.

5.1 Domains and ranges

It is essential to identify carefully the domains and ranges of the composition maps used to define $\mathbf{T}$. Since

$$|\nabla[\mathbf{h}_1(v)]|^2 = |\mathbf{h}_1^{'}(v)\nabla v|^2 \leq c|\nabla v|^2,$$

with a similar inequality for $|\nabla[\mathbf{h}_2(w)]|^2$, it follows that the mapping $\mathbf{Tr}$ has range contained in the set $K_1 \cap \{C\mu : \mu \in \Omega \subset \prod_1^2 H^1\}$, for some positive constant C. Thus, we select the domain of $\mathbf{U}$ to be $K_1 \cap (C\Omega)$. By employing the H^1 norm defined below in (71), we see that the range of $\mathbf{U}$ is contained in a bounded set Γ in $H^1(G) \cap L_\infty(G)$; indeed, the pointwise bounds, involved in defining Γ, have already been quoted, while the H^1 bounds for $\mathbf{U}$ and its finite element approximation, also involved in defining Γ, require separate analysis. In particular, the following pointwise bounds (maximum principles) hold for $U = \mathbf{U}[v, w]$:

$$\gamma \leq U \leq \delta,$$

$$\begin{aligned} \gamma &= \min(\gamma^{'}, \inf_{\Sigma_D} \bar{u}), \qquad \delta = \max(\delta^{'}, \sup_{\Sigma_D} \bar{u}), \\ \gamma^{'} &= \sinh^{-1}[(1/2)\inf_G k_1\, e^{(\alpha_1-\alpha_2)/2}] + (\alpha_1+\alpha_2)/2, \\ \delta^{'} &= \sinh^{-1}[(1/2)\sup_G k_1\, e^{(\beta_1-\beta_2)/2}] + (\beta_1+\beta_2)/2. \end{aligned} \tag{69}$$

Finally, the joint domain of $\mathbf{V}$ and $\mathbf{W}$ is Γ, while the range of these maps is contained in the intersection of K (not K_1), with the domain Ω, which is now defined. Since, for $U \in \Gamma$ (the pointwise bounds suffice),

$$\begin{aligned} \int_G |\nabla v|^2 dx &\leq e^{\beta_v - \gamma} \int_G e^{U-v} |\nabla v|^2 dx \\ &= e^{\beta_v - \gamma} \int_G e^{U-v} \nabla v \cdot \nabla \bar{v} dx \\ &\leq (1/2)[\int_G |\nabla v|^2 dx + e^{2(\delta + \beta_v - \gamma - \alpha_v)} \int_G |\nabla \bar{v}|^2 dx], \end{aligned} \tag{70}$$

it follows, for $\|\cdot\|_{H^1}^2$ given by

$$\|v\|_{H^1}^2 = \|\nabla v\|_{L_2}^2 + (\int_{\Sigma_D} v dx)^2, \tag{71}$$

that

$$\|v\|_{H^1}^2 < e^{2(\delta - \gamma + (\beta_v + \beta_w) - (\alpha_v + \alpha_w))} \|\bar{v}\|_{H^1}^2 \ \ (\text{Note:} v \mid_{\Sigma_D} = \bar{v} \mid_{\Sigma_D}),$$

with a similar estimate for $\|w\|_{H^1}^2$. Thus, we initially choose Ω to contain the open ball centered at 0 of radius,

$$\rho = e^{(\delta - \gamma + (\beta_v + \beta_w) - (\alpha_v + \alpha_w))} \|[\bar{v}, \bar{w}]\|_{\prod H^1}.$$

It is evident that Ω contains the range of $\mathbf{T}$. However, it is essential for our purposes that Ω also contain the range of $\mathbf{T}_n$, which is defined analogously in terms of composite mappings, $\mathbf{U}_h$, $\mathbf{V}_h$, and $\mathbf{W}_h$. Energy estimation of $v_h = \mathbf{V}_h(U_h)$ and $w_h = \mathbf{W}_h(U_h)$ is required. Such estimates yield yield the result that the number ρ just defined need only be perturbed by a term of order $O(h)$. This gives us, finally, an admissible radius of Ω. We continue with a study of the differentiability.

5.2 Differentiability of composition mappings

Several assumptions are required for a derivation of the theory. We refer the reader to [24, Chapter 5] for a full accounting. However, we mention here perhaps the most significant such assumption. It may be interpreted as the assumption that singularities are not the most general possible.

Assumption 2. The inequality,

$$\theta > N(\frac{1}{2} - \frac{1}{N}), \tag{72}$$

holds.

We present without proof the basic differentiability results.

Lemma 5.1 *Let* $\mathbf{U} : (v, w) \mapsto u$ *be the mapping defined implicitly through the solution of the boundary value problem,*

$$\langle \nabla u, \nabla \phi \rangle + \langle e^{u-v} - e^{w-u} - \bar{N}, \phi \rangle = 0, \tag{73}$$

where $\phi \in H^1_{0,\Sigma_D}$*, subject to suitable mixed boundary conditions in* N *dimensions. Then the derivative,* $D_{(v,w)}\mathbf{U}(v, w) : (\sigma, \tau) \mapsto \mu$*, is defined through the solution of*

$$\langle \nabla \mu, \nabla \phi \rangle + \langle e^{u-v}[\mu - \sigma] + e^{w-u}[\mu - \tau], \phi \rangle = 0, \tag{74}$$

where $\mu|_{\Sigma_D} \equiv 0$*, and for each* $[v, w]$ *is a uniformly bounded linear operator from* $\prod_1^2 L_2$ *to* H^1_{0,Σ_D}*, and, if* $N \leq 3$*, from* $\prod_1^2 H^1$ *to* L_∞*. Moreover, the mapping is Lipschitz continuous from* $\prod_1^2 H^1$ *to the mappings from* $\prod_1^2 L_2$ *to* H^1_{0,Σ_D} *if* $N \leq 4$*.*

We continue with the mappings $\mathbf{V}$ and $\mathbf{W}$.

Remark 2. The derivative $D\mathbf{V}(u) : \mu \mapsto \sigma$, is defined through solution of the boundary value problem,

$$\langle e^{u-v}[(\mu - \sigma)\nabla v + \nabla \sigma], \nabla \phi \rangle = 0, \tag{75}$$

where $\mu \in H^1$, $\sigma \in H^1_{0,\Sigma_D}$, and ϕ is a test function. Similarly for $D\mathbf{W}(u)$.

Lemma 5.2 *The derivative* $D\mathbf{V}$ *of the mapping* $\mathbf{V}$ *from* u *to* v*, defined through the equation* (75)*, is uniformly bounded over its domain as a family of linear operators from* H^1_{0,Σ_D} *to itself. The derivative* $D\mathbf{V}$ *is a locally Lipschitz continuous mapping from* H^1 *to the mappings from* $H^1 \cap L_\infty$ *to* H^1_{0,Σ_D}*, for Euclidean dimension* $N \leq 3$*. A similar statement holds for* $\mathbf{W}$*.*

The role of Assumption 2 is critical to this result, particularly in the derivation of the L_∞ statements. The restriction on N corresponds precisely to the requirements of the Moser regularity theory as presented in [15, Chapter 8].

We consider now $\mathbf{U}_h$ and $\mathbf{V}_h$; for simplicity, the same symbols are used for the extensions.

Lemma 5.3 *The derivative* $D_{(v,w)}\mathbf{U}_h(v, w) : (\sigma, \tau) \mapsto \mu_h$ *is defined through the solution of the projection relation,*

$$\langle \nabla \mu_h, \nabla \phi \rangle + \langle e^{U_h - v}[\mu_h - \sigma] + e^{w-U_h}[\mu_h - \tau], \phi \rangle = 0, \tag{76}$$

where μ_h *and* ϕ *are in* S_h*. The derivative* $D\mathbf{V}_h(u) : \mu \mapsto \sigma_h$ *may be defined by*

$$\langle e^{u-v_h}[(\mu - \sigma_h)\nabla v_h + \nabla \sigma_h], \nabla \phi \rangle = 0, \tag{77}$$

where $\sigma_h, \phi \in S_h$*.*

We shall not give a detailed listing of the properties of these mappings. It is interesting that Assumption 2 plays the same role in the discrete L_∞ theory as in the continuous case. In particular, the L_∞ convergence of the finite element approximations is of order $O(h^\rho)$, where

$$\rho = \theta - N(\frac{1}{2} - \frac{1}{N}). \tag{78}$$

5.3 Convergence of the Galerkin approximations

The preceding subsection describes results which may be used to infer the conclusion of a strengthened Theorem 4.1, and hence the inequalities represented by (46). In turn, the order of (46) may be deduced. This strengthening of Theorem 4.1 is required since the Lipschitz continuity of $\mathbf{T}'$ does not hold unless the domain is normed by (66). Altogether, we have the following, where the norm is the energy norm.

Theorem 5.1 *The Galerkin approximations introduced in* §3 *converge with order* $O(h^\theta)$ *to a solution of the drift-diffusion system. This order is optimal.*

Although we have labored considerably to obtain this result, it does not yet constitute a resolution of the approximation problem defined earlier. Until this point, we have used the nonlinear calculus as a *framework*. The remainder of the paper will exploit it as a *carrier* of new ideas.

6 Linearization: numerical fixed point map

6.1 Asymptotic linearity

The result expressed in Theorem 4.1, (46), depends upon the following inequality:

$$\frac{\alpha_n}{1+q} \leq \|x_n - \mathbf{P}_n x_0\| \leq \frac{\alpha_n}{1-q}. \tag{79}$$

The number q is fixed here. It serves as the contraction constant of the mapping (47); once a fixed point has been determined, inequalities (79) are routine. These inequalities, in turn, directly imply inequalities (46), via inverse bounding hypotheses. We note that the estimates described in §5.2 allow one to determine q with respect to the stronger norm (66), yet, once the fixed point is determined, to estimate with respect to the weaker energy norm, where convergence is of higher order. In particular, the concrete interpretation of (79) is in terms of energy norms.

It is an interesting question whether a refined version of (79) holds, with a sequence q_n replacing q, $q_n \to 0$. In this case, we can call the approximations defined by x_n asymptotically linear, since

$$\alpha_n \sim \|x_n - \mathbf{P}_n x_0\|, \tag{80}$$

with the usual meaning that the quotients tend to one. This constitutes a natural extension of the relation of equality, which holds in the affine case, and accounts for the use of the description of asymptotic linearity. What may not be immediately apparent, however, is that an affirmative answer is linked to the replacement of the numerical fixed point by a sequence of Newton iterates, bounded in number independent of n. Recall that, in the definition of the approximation problem, item two required this computability property. By

our previous remarks, the existence of such a sequence $\{q_n\}$ is tied to the estimate of the contraction constants of the mapping (47) on a sequence of balls of shrinking diameters. We recall that $\alpha_n \to 0$. We have the following. For ease of understanding, the result is stated for a generic numerical fixed point map, then specialized.

Theorem 6.1 *Suppose that $\mathbf{T}'_n$ is Lipschitz continuous with constant $2C > 0$, independent of n, so that* (43) *holds with $\delta_\epsilon = \epsilon/(2C)$. Then the numbers q_n, defined for n such that $\alpha_n < \frac{1}{8\kappa C}$ by,*

$$q_n = \frac{1}{2}(1 - \sqrt{1 - 8\alpha_n \kappa C}), \tag{81}$$

satisfy $q_n \to 0$. Here, $\|[\mathbf{I} - \mathbf{T}'_n(\mathbf{P}_n x_0)]^{-1}\| \leq \kappa$. Moreover,

$$\frac{\alpha_n}{1 + q_n} \leq \|x_n - \mathbf{P}_n x_0\| \leq \frac{\alpha_n}{1 - q_n} = \delta_n. \tag{82}$$

In particular, the asymptotic relation (80) *holds. If the Lipschitz derivative continuity requires the introduction of a stronger norm such as* (66)*, then α_n and κ are replaced by maxima over both norms in* (81)*, and* (82) *continues to hold for α_n defined only with respect to the energy norm.*

Proof Set $\epsilon_n = \frac{q_n}{\kappa}$. With the choice of $\delta_n = \frac{\epsilon_n}{2C}$, we have the equation,

$$\alpha_n = \frac{q_n(1 - q_n)}{2\kappa C} = \delta_n(1 - q_n). \tag{83}$$

This implies that the mapping $\mathbf{B}$ of (47) maps $\{x : \|x - x_*\| \leq \delta_n\}$ into itself, with contraction constant q_n. The inequalities (82) are then immediate. The extension to the stronger norm case is similar. *Box*

6.2 A bounded number of Newton iterates

The next result of this paper deals with maintaining the essence of the upper bound in (82), while transferring the determination of x_n to the solution of a sequence of approximate linear problems, via Newton iteration; it will be critical to use only k_0 of these iterates, where k_0 does not depend on n. Because the application requires the use of the stronger norm (66), the convergence, for technical reasons, must be measured by the linear, rather than quadratic convergence, of Newton iteration. We consider Newton's method for the approximation of x_n, defined via

$$u_n^{k+1} - u_n^k = -[\mathbf{A}'_n(u_n^k)]^{-1}\mathbf{A}_n(u_n^k). \tag{84}$$

Here we have written $\mathbf{A}_n$ for the mapping $\mathbf{I} - \mathbf{T}_n$, and intend to employ (84) inductively for $k \leq k_0$ only, with starting value, $\tilde{x}_{n-1}$. The determination of k_0 is now discussed.

Theorem 6.2 *Assume Lipschitz continuity and uniform inverse boundedness properties for* $\mathbf{A}_n'$*, the latter as in Theorem* 6.1*. Assume also the uniform boundedness property,*

$$\|[\mathbf{A}_n'(y)]\| \leq \kappa'. \tag{85}$$

Without loss of generality, suppose that the sequence $\{q_n\}$ *is monotone decreasing. Then Newton's method is q-linearly convergent:*

$$\|x_n - u_n^k\| \leq \frac{q_{*n}^k}{1 - q_{*n}} \|u_n^1 - u^0\|. \tag{86}$$

Here, we have set $q_{*n} = \frac{2q_n}{1-q_n}$*. In particular, we may select an integer* k_0*, not depending on* n*, such that*

$$\|x_n - u_n^k\| \leq \delta_n := \frac{q_n}{2\kappa C},\ k \geq k_0, \tag{87}$$

and we may define, $\tilde{x}_n = u_n^{k_0}$*. The error estimate,* $\| \tilde{x}_n - \mathbf{P}_n x_0\| \leq 2\delta_n$*, holds. We may choose* k_0 *to be the smallest integer* k *satisfying,*

$$2\delta_0 \frac{q_{*n}^k}{1 - q_{*n}} \leq \delta_n. \tag{88}$$

Proof By use of the definition of the Newton increment, (84), we immediately obtain the estimate,

$$\|u_n^{k+1} - u_n^k\| \leq \frac{\kappa}{1 - q_n} \|\mathbf{A}_n(u_n^k)\|,$$

where we have used a global inverse bound, derived from the bound assumed in Theorem 6.1. In order to estimate the residual term, we employ the integral representation,

$$\mathbf{A}_n(u_n^k) = \tag{89}$$

$$\int_0^1 [\mathbf{A}_n'(u_n^{k-1} + t(u_n^k - u_n^{k-1})) - \mathbf{A}_n'(u_n^{k-1})](u_n^k - u_n^{k-1})\, dt,$$

and estimate (89) by use of the inequality,

$$\|\mathbf{A}_n'(x) - \mathbf{A}_n'(y)\| \leq \frac{2q_n}{\kappa},$$

which follows from the hypotheses, via the triangle inequality. We obtain, finally,

$$\|u_n^{k+1} - u_n^k\| \leq \frac{2q_n}{1 - q_n} \|u_n^k - u_n^{k-1}\|. \tag{90}$$

By the definition of q_{*n}, and the repeated use of (90), with n fixed, we obtain a standard Cauchy sequence estimate for $\|u_n^l - u_n^k\|$. Passage to the limit with respect to k then yields (86). Note that here we have used the local uniqueness of x_n. *Box*

6.3 Computability of the Newton iterates

One requires the iterative solution of the linear system,

$$-\mathbf{A}'_n(u_n^k)[u_n^{k+1} - u_n^k] = \mathbf{A}_n(u_n^k). \tag{91}$$

The left hand side of the sequence (91) can be computed from (76), with the components of the Newton increment, σ_n^{k+1} and τ_n^{k+1}, undetermined linear combinations of basis functions, followed by use of (77), and the corresponding system for the second component. This makes direct use of the chain rule for the composition mapping, and suppresses the truncation map in the situation where maximum principles hold. From this discussion, we see that the first component of the left hand side of (91) is given by the expression,

$$-[\mathbf{I} - \mathbf{V}'_h(\mathbf{U}_h(u_n^k)) \circ \mathbf{U}'_h(u_n^k)](\sigma_n^{k+1}, \tau_n^{k+1}), \tag{92}$$

where

$$[\sigma_n^{k+1}, \tau_n^{k+1}] = u_n^{k+1} - u_n^k. \tag{93}$$

The right hand side of (91) involves the use of the operator nonlinear Gauss-Seidel procedure, used to define the numerical fixed point map. Thus, the system can be finally solved by simple linear inversion. The latter fixes the values, then, of the basis coefficients for σ_n^{k+1} and τ_n^{k+1}. Thus, the successive solutions induced by the compositions, when viewed as a symbolic calculation, are defined by sparse matrix calculations. However, this does not represent an easily verifiable complexity estimate. Toward this end, this system has been investigated in great detail by Kerkhoven and Saad in [30] by use of GMRES (and other methods as well). GMRES (generalized minimum residual) is based upon Arnoldi's process, and was introduced in [42]. The useage here may be thought of as a way of approximately implementing the solution of (91), which defines the exact Newton sequence for the numerical fixed point. In fact, the Krylov subspace of dimension m employed at the n-th stage is

$$K_m = \{v_1, \mathbf{J}_n v_1, ..., \mathbf{J}_n^{m-1} v_1\},$$

where $\mathbf{J}_n$ is the derivative (Jacobian) of the identity shifted $\mathbf{T}_n$, and v_1 is a normalized residual. Kerkhoven and Saad are able to implement this algorithm in a Jacobian free manner, i.e. , using only values generated by $\mathbf{T}$, and derive an essential *superlinear* convergence estimate for the residual; this suggests, though not with the conclusiveness of proof, that the work function of a single GMRES iteration, depending on n, when multiplied by a fixed number of inner iterations, yields the function $p(n)$ discussed in §2. Some features distinguish the present setting from that of [30]:

1. $\mathbf{T}$ is defined differently;
2. [30] is concerned with the acceleration of the fixed point iteration;
3. outer and inner iterations are consolidated in [30].

Nonetheless, this approach complements the present paper, and adheres to the philosophical principle that the iterative strategy contains its own preconditioning (see also [11] for this idea in a different context).

Remark 3. The previous subsection has shown that solution of (91) for $k \leq k_0$ suffices to guarantee the rate of convergence of Theorem 5.1. Since this already is known to be optimal, and since computability is clearly satisfied, we have a solution of the approximation problem. Note that the theory requires uniform pointwise boundedness of the approximation sequence, though this sequence does not necessarily satisfy the discrete maximum principles derived for the Galerkin sequence. Altogether, we have obtained a solution for the approximation problem introduced in §2.

Acknowledgment The author expresses appreciation to the referees for their significantly helpful insights and comments. Also, the author expresses appreciation to Ivo Babuška for raising the problem, in 1987, of synchronization of linear iteration with nonlinear error analysis. This occurred during the author's visit to John Osborn and Ivo at the University of Maryland, and ultimately led to the solution proposed in this review.

References

[1] J.-P. Aubin, *Approximation des espaces de distributions et des opérateurs differéntials*, Bull. Soc. Math. France Suppl. Mém., 12 (1967).

[2] ——, *Approximation of Elliptic Boundary-Value Problems*, Wiley, New York, 1972.

[3] I. Babuška, *The rate of convergence for the finite element method*, SIAM J. Numer. Anal., 8 (1971), pp. 304–315.

[4] I. Babuška and A. Aziz, *Survey lectures on the mathematical foundations of the finite element method*, in The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations, A.K. Aziz, ed., Academic Press, 1972, pp. 5–359.

[5] R. Bank, D. Rose, and W. Fichtner, *Numerical methods for semiconductor device simulation*, IEEE Trans. Electron Devices, 30 (1983), pp. 1031–1041.

[6] R. E. Bank and D. J. Rose, *Global approximate Newton methods*, Numerische Mathematik, 37 (1981), pp. 279–295.

[7] V. Barcilon, D.-P. Chen, and R. Eisenberg, *Ion flow through narrow membrane channels: part II*, SIAM J. Appl. Math., 52 (1992), pp. 1405–1425.

[8] J. Bramble and S. Hilbert, *Estimation of linear functionals on Sobolev spaces with application to Fourier transforms and spline interpolation*, SIAM J. Numer. Anal., 7 (1970), pp. 112–124.

[9] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, Springer-Verlag, New York, 1994.

[10] F. Brezzi, *On the existence, uniqueness, and approximation of saddle point problems arising from Lagrangian multipliers*, R.A.I.R.O., Anal. Numér., 12 (1974), pp. 129–151.

[11] T. Chan and K. Jackson, *Nonlinearly preconditioned Krylov subspace methods for discrete Newton algorithms*, SIAM J. Sci. Stat. Comput., 5 (1984), pp. 533–542.

[12] P. Ciarlet and P.-A. Raviart, *Maximum principle and uniform convergence for the finite element method*, Computer Methods in Applied Mechanics and Engineering, 2 (1973), pp. 17–31.

[13] W. Coughran and J. Jerome, *Modular algorithms for transient semiconductor device simulation, Part I: Analysis of the outer iteration*, in Computational Aspects of VLSI Design with an Emphasis on Semiconductor Device Simulation, R.E. Bank, ed., American Mathematical Society, Lectures in Applied Mathematics 25, 1990, pp. 107–149.

[14] T. Dupont and R. Scott, *Polynomial approximations of functions in Sobolev spaces*, Math. Comp., 34 (1980), pp. 441–463.

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

[16] K. Höllig, *Approximationszahlen von Sobolev-Einbettungen*, Math. Ann., 242 (1979), pp. 273–281.

[17] ——, *Diameters of classes of smooth functions*, in Quantitative Approximation, Academic Press, Orlando, 1980.

[18] J. W. Jerome, *On the L_2 n-Width of Certain Classes of Functions of Several Variables*, PhD thesis, Purdue University, Lafayette, Indiana, 1966.

[19] ——, *On n-widths in Sobolev spaces with applications to elliptic boundary value problems*, Journal of Mathematical Analysis and Applications, 29 (1970), pp. 201–215.

[20] ——, *Approximation of Nonlinear Evolution Systems*, Academic Press, 1983.

[21] ——, *An adaptive Newton algorithm based on numerical inversion: Regularization as postconditioner*, Numerische Mathematik, 47 (1985), pp. 123–138.

[22] ——, *Approximate Newton methods and homotopy for stationary operator equations*, Constructive Approximation, 1 (1985), pp. 271–285.

[23] ——, *Consistency of semiconductor modeling: An existence/stability analysis for the stationary van Roosbroeck system*, SIAM J. Appl. Math., 45 (1985), pp. 565–590.

[24] ——, *Mathematical Theory and Approximation of Semiconductor Models*, Springer-Verlag, 1995.

[25] ——, *An asymptotically linear fixed point extension of the inf-sup theory of Galerkin approximation*, Numer. Functional Anal. Optim., 45 (1995), pp. 345–361.

[26] J. W. Jerome and T. Kerkhoven, *A finite element approximation theory for the drift-diffusion semiconductor model*, SIAM J. Numer. Anal., 28 (1991), pp. 403–422.

[27] L. Kantorovich, *Functional analysis and applied mathematics*, Tech. Report 1509, National Bureau of Standards, 1952. C. Benster, transl.

[28] L. Kantorovich and G. Akilov, *Functional Analysis in Normed Spaces*, Pergamon Press, New York, 1964.

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

[30] T. Kerkhoven and Y. Saad, *On acceleration methods for coupled nonlinear elliptic systems*, Numer. Math., 60 (1992), pp. 525–548.

[31] A. Kolmogorov, *Über die beste Annähergung von Funktionen einer gegebenen Funktionklasse*, Ann. Math., 37 (1936), pp. 107–111.

[32] M. Krasnosel'skii, G. Vainikko, P. Zabreiko, Y. Rititskii, and V. Stetsenko, *Approximate Solution of Operator Equations*, Wolters-Noordhoff, Groningen, 1972.

[33] G. Lorentz, *Approximation Theory*, Holt, New York, 1966.

[34] J. Moser, *A rapidly convergent iteration method and nonlinear partial differential equations I*, Ann. Scuola Norm. Sup. Pisa Cl. Sci., XX (1966), pp. 265–315.

[35] M. Murthy and G. Stampacchia, *A variational inequality with mixed boundary conditions*, Israel J. Math., 13 (1972), pp. 188–224.

[36] J. Nitsche, *Ein Kriterium für die Quasi-Optimalität des Ritzschen Verfahrens*, Numer. Math., 11 (1968), pp. 346–348.

[37] J. Ortega and W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, 1970.

[38] A. Pinkus, *n-Widths in Approximation Theory*, Springer-Verlag, Berlin-New York, 1985.

[39] W. Roosbroeck, *Theory of flow of electrons and holes in germanium and other semiconductors*, Bell System Tech. J., 29 (1950), pp. 560–607.

[40] I. Rubinstein, *Electro-Diffusion of Ions*, SIAM Studies in Applied Mathematics, 1990.

[41] M. Rudan and F. Odeh, *Multi-dimensional discretization scheme for the hydrodynamic model of semiconductor devices*, COMPEL, 5 (1986), pp. 149–183.

[42] Y. Saad and M. Schultz, *GMRES: A generalized minimal residual method for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869.

[43] M. H. Schultz, *Multivariate spline functions and elliptic problems*, in Approximations with Special Emphasis on Spline Functions, L.B. Rall, ed., Academic Press, Orlando, 1969, pp. 279–347.

[44] S. Sobolev, *Applications of Functional Analysis in Mathematical Physics*, vol. 7 of Translations of Math. Monographs, American Mathematical Society, Providence, Rhode Island, 1963.

[45] E. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, Princeton, New Jersey, 1970.

[46] G. Strang, *Approximation in the finite element method*, Numerische Mathematik, 19 (1972), pp. 81–98.

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

[48] G.-L. Tan, X.-L.Yuan, Q.-M. Zhang, W. Tu, and A.-J.Shey, *Two-dimensional semiconductor device analysis based on new finite-element discretization employing the S-G scheme*, IEEE Trans. on C.A.D., 8 (1989), pp. 468–478.

[49] R. Varga, *Matrix Iterative Analysis*, Prentice-Hall, 1962.

[50] M. Ward, L. Reyna, and F. Odeh, *Multiple steady-state solutions in a multijunction semiconductor device*, SIAM J. Appl. Math., 51 (1991), pp. 90–123.