

Linear Reduction and Homotopy Control for Steady Drift-Diffusion Systems in Narrow Convex Domains

Joseph W. Jerome*

Key words: Solution curves, continuation, boundary value problems, drift-diffusion systems

2020 AMS classification numbers:35A01, 35B30, 35Q49, 46N20, 65J05

Abstract

This article develops and applies results, originally introduced in earlier work, for the existence of homotopy curves, terminating at a desired solution. We describe the principal hypotheses and results in section two; right inverse approximation is at the core of the theory. We apply this theory in section three to the basic drift-diffusion equations. The carrier densities are not assumed to satisfy Boltzmann statistics and the Einstein relations are not assumed. By proving the existence of the homotopy curve, we validate the underlying computational framework of a predictor/corrector scheme, where the corrector utilizes an approximate Newton method. The analysis depends on the assumption of domains of narrow width. However, no assumption is made regarding the domain diameter.

¹Department of Mathematics, Northwestern University, Evanston, IL 60208.

1 Introduction.

Potential-driven transport is now central in mathematical modeling in science and engineering. For electrostatic potentials, application areas include semiconductor transport [1], ion channel transport [2], as well as the behavior of electrochemical systems [3]. Various approaches have been employed to analyze these systems. The reference [4] established existence by the creation of an implicit composite solution mapping, essentially an abstract integral operator equation for systems. This has proven to be effective for analysis, but remains intricate, particularly as regards approximation (cf. [5]). In the present article, we employ a direct differentiable system map. This shifts the burden to one of establishing regularity and derivative mapping surjectivity. Fixed point theory is eliminated in favor of homotopy. In effect, existence is established if the desired system can be represented as the terminus of a solution curve which begins with a standard linear special case. The general result is found in section two, and is referenced from [6]. The application to the drift-diffusion model is carried out in section three. Although not explicitly considered, this result provides the rigorous infrastructure for the Euler predictor-Newton corrector approximation method. The homotopy curve emanates from the center of a ball of radius governed by the problem data, and remains within this ball until the solution is achieved.

We conclude the introduction with comments regarding the analysis literature in this subject. The first rigorous study of the drift-diffusion semiconductor model appears to be that of Mock [7]. This was followed by [8]. A framework for [4] was contained in [9]. An extensive study appeared in [10]. These studies are distinguished by the choice of boundary conditions, domain geometry, physical parameters, and recombination assumptions. For example, mixed boundary conditions, with polyhedral geometry, is apt to reduce the expected solution regularity. The reference [11] includes these technical issues.

The boundary conditions of the present article are somewhat simplified for the semiconductor application. However, they are realistic as applied to the Poisson-Nernst-Planck model for ion channel transmission and electrochemical systems. For these applications, existence is less important than system response to parameter variation and to analysis of specialized case studies. For examples of the latter, see [12, 13].

In summary, the model considered in this article is a basic scientific model of drift-diffusion, without carrier recombination, and with convex bounded domains in Euclidean space of dimension $N = 2, 3$. As formulated, the model is applicable to each of the three application areas just referenced, when Dirichlet data are appropriate.

Finally, we observe that all of the analytical investigations cited above make use of the Einstein relations, which are valid via arguments from statistical mechanics. Although we do not assume either these relations, or Boltzmann statistics, we do assume that the physical region Ω has sufficiently small width, permitting an effective application of the Poincaré inequality. Specifically, this is used both in the $\lambda = \lambda_0 = 0$ case and in the construction of inverses of the

linearized equations. The width assumption does not restrict the domain diameter, either explicitly or implicitly. The use of the basic variables is consistent with the study [14], where multi-physics effects are included to supplement the Poisson-Nernst-Planck model.

2 Prior Results

We begin with the assumptions for an existence theory.

2.1 Existence of a homotopy solution curve in Banach space

The hypotheses stated here are slightly stronger than those presented in [6], in order to facilitate verifiability. We will make a more precise statement at the conclusion of the following definition.

Definition 2.1. *Suppose that Banach spaces X and Z , a connected, open subset U of X , and a bounded, connected closed subset Λ of \mathbb{R} are given, together with a mapping,*

$$F : \bar{U} \times \Lambda \mapsto Z. \quad (1)$$

The following are assumed.

1. *The partial derivatives F_w and F_μ are Lipschitz continuous in the operator norm. Specifically, there is a constant C such that, globally,*

$$\begin{aligned} \|F_w(v, \lambda) - F_w(w, \mu)\|_{B[X, Z]} &\leq C\{\|v - w\|_X^2 + |\lambda - \mu|^2\}^{1/2}, \\ \|F_\mu(v, \lambda) - F_\mu(w, \mu)\|_{B[\mathbb{R}, Z]} &\leq C\{\|v - w\|_X^2 + |\lambda - \mu|^2\}^{1/2}. \end{aligned} \quad (2)$$

2. *$F_w(w, \mu)$ has an approximate inverse $H(w, \mu)$ satisfying, globally,*

$$\|F_w(w, \mu)H(w, \mu)z - z\|_Z \leq M\|F(w, \mu)\|_Z\|z\|_Z, \quad (3)$$

$$\|H(w, \mu)z\|_X \leq M\|z\|_Z. \quad (4)$$

Remark 2.1. *The assumptions for F_w and F_μ imply the following remainder formula and estimate:*

$$\begin{aligned} F(w, \mu) &= F(v, \lambda) + F_w(v, \lambda)(w - v) + F_\mu(v, \lambda)(\mu - \lambda) + R(v, w; \lambda, \mu), \\ \|R(v, w; \lambda, \mu)\|_Z &\leq C[\|v - w\|_X^2 + (\lambda - \mu)^2]. \end{aligned} \quad (5)$$

This follows from a direct estimate of the following representation for the remainder.

$$R(v, w; \lambda, \mu) = \int_0^1 [F_w(v + s(w - v), \lambda) - F_w(v, \lambda)](w - v) ds +$$

$$\int_0^1 [F_\mu(w, \lambda + t(\mu - \lambda)) - F_\mu(w, \lambda)](\mu - \lambda) + [F_\mu(w, \lambda) - F_\mu(v, \lambda)](\mu - \lambda) dt. \quad (6)$$

In [6], (5) is included in the hypotheses, whereas the Lipschitz continuity of F_μ is not. Here, it is explicitly included.

We have the following.

Theorem 2.1. *Suppose that Banach spaces X and Z , a closed interval $[\lambda_0, \lambda_1]$, a closed ball $B_r \subset X$, with interior B_r° , and a mapping F are given, satisfying Definition 2.1. We suppose that (2, 3, 4) of Definition 2.1 hold and also the following:*

- (i) $F(u_0, \lambda_0) = 0$, where u_0 is the center of B_r .
- (ii) r is sufficiently large relative to M and $\lambda_1 - \lambda_0$:

$$r \geq \max(C, M)(\lambda_1 - \lambda_0) \left[1 + \sup_{v, \lambda} \|F_\mu(v, \lambda)\|_{B[\mathbb{R}, Z]} \right]. \quad (7)$$

Then there exists a solution set $u = u(\lambda)$ such that

$$u(\lambda) \in B_r^\circ, F(u(\lambda), \lambda) = 0, \lambda \in [\lambda_0, \lambda_1]. \quad (8)$$

Remark 2.2. *This result follows from [6, Th. 4.1] and its proof. The proof is based upon a predictor/corrector method which ensures that the iterates remain in the radius of convergence of Newton's method; here, this means the interior of B_r . The argument uses the chain of inequalities given by (3.10) of [6]. However, for the purposes here, two of the inequalities in the chain need to be replaced by strict inequalities: that preceding the term $\rho^{-1}/4$ and the final inequality preceding the term $3r/4$. When modified in this way, each predictor is seen to lie in the interior of B_r .*

Note that the viability of inequality (7) requires the decoupling of the norm of F_μ from the choice of r . This is the most serious challenge to the application of this theory to systems of partial differential equations.

2.2 Surjectivity of partial derivatives

An essential hypothesis of the theory is a parameter dependent approximate right inverse for the derivative map. Here, we provide a result which guarantees, for the application considered, the existence of a right inverse. The following result is a concatenation of [15, Th. 27.A, Prop. 27.6]

Proposition 2.1. *Suppose that X is a reflexive Banach space and that $L : X \mapsto X^*$ is a given operator with decomposition $L = L_1 + L_2$, satisfying the following conditions.*

1. L_1 is monotone and continuous.
2. L_2 is strongly continuous.

3. L is bounded.

4. L is coercive.

Then L is surjective.

3 Existence of a Homotopy Curve for Drift-Diffusion Systems

We apply the results of the preceding section to construct a solution curve. We begin by writing the system in strong form with parametric dependence. The dependent variables are electrostatic potential, charge density for negative carriers, and charge density for positive carriers, denoted by u, n, p , resp. Carrier transport is not assumed to satisfy the Einstein relations. Boltzmann statistics are not assumed. Moreover, additional carriers are easily accommodated if desired. The system is given as follows. In the following section, we will provide a precise definition of solution, which in this article is defined as a strong solution for u and a weak solution for n and p .

$$\begin{aligned} -\nabla^2 u &= \lambda(p - n) + D, \\ -\nabla \cdot (d_n \nabla n - c_n n \nabla u) &= 0, \\ -\nabla \cdot (d_p \nabla p + c_p p \nabla u) &= 0. \end{aligned} \tag{9}$$

Here, d_n, c_n, d_p, c_p are positive constants, related to diffusion and carrier drift, resp. In the first equation, D represents so-called permanent charge, and is an arbitrary L^∞ function. Also, units in this equation are chosen so that the ratio of charge modulus to dielectric constant is one. Dirichlet boundary conditions are imposed as follows, via the boundary trace operator Γ , for given functions a_u, a_n, a_p .

$$\Gamma u = \Gamma a_u, \Gamma n = \Gamma a_n, \Gamma p = \Gamma a_p. \tag{10}$$

The domain $\Omega \subset \mathbb{R}^N$ is an open convex bounded set. Here, $N = 2, 3$. The parameter λ ranges between 0, 1.

3.1 The fundamental mapping F

We begin with the component definitions of F .

Definition 3.1. Set $\mathcal{H} = (H^2(\Omega) \cap H_0^1(\Omega)) \times H_0^1(\Omega) \times H_0^1(\Omega)$ and $\mathcal{G} = L^2(\Omega) \times H^{-1}(\Omega) \times H^{-1}(\Omega)$. Arbitrary elements of \mathcal{H} are designated as (ρ, σ, τ) . Standard norms are employed, except for the equivalent norm in H_0^1 : $\|f\|_{H_0^1} = \|\nabla f\|_{L^2}$. Then $F : \mathcal{H} \times [0, 1] \mapsto \mathcal{G}$ is defined via its components as follows. Here, $(u, n, p) := (\rho, \sigma, \tau) + (a_u, a_n, a_p)$ for a fixed specified function $(a_u, a_n, a_p) \in H^2 \times H^1 \times H^1$.

$$\begin{aligned} F_1(\rho, \sigma, \tau; \lambda) &= -\nabla^2 u + \lambda(n - p) - D, \\ F_2(\rho, \sigma, \tau; \lambda) &= -\nabla \cdot (d_n \nabla n - c_n n \nabla u), \\ F_3(\rho, \sigma, \tau; \lambda) &= -\nabla \cdot (d_p \nabla p + c_p p \nabla u). \end{aligned} \tag{11}$$

F_1 has natural range in L^2 . The precise (standard) meaning of the action of F_2, F_3 is given by the following. In describing this action, we write an arbitrary vector test function for the final two components of \mathcal{G} as (ϕ, ψ) . We have:

$$\begin{aligned} F_2(\rho, \sigma, \tau; \lambda)[\phi] &= \int_{\Omega} [(d_n \nabla n - c_n n \nabla u) \cdot \nabla \phi] dx, \\ F_3(\rho, \sigma, \tau; \lambda)[\psi] &= \int_{\Omega} [(d_p \nabla n + c_p n \nabla u) \cdot \nabla \psi] dx. \end{aligned} \quad (12)$$

By a homotopy solution curve is meant a set $\{h_\lambda = (\rho_\lambda, \sigma_\lambda, \tau_\lambda)\} \subset \mathcal{H}, 0 \leq \lambda \leq 1$, such that $F(h_\lambda, \lambda) = 0, \forall \lambda$.

Remark 3.1. As noted earlier, we will place restrictions on the width of the convex domain Ω . We use the terminology of [16, p. 359]. A domain Ω has finite width d if the latter is the smallest positive number such that Ω lies between two parallel planes separated by a distance d .

Lemma 3.1. For each fixed $\lambda \in [0, 1]$, the mapping F is well-defined from \mathcal{H} to \mathcal{G} .

Proof. The analysis for F_1 is clear. We consider F_2, F_3 . Each integral in the definition has a product integrand which is L^1 , as seen by the Hölder inequality. Moreover, the corresponding estimate includes a factor dominated by the H_0^1 norm of the test function. Note that in Euclidean dimension not exceeding four, an H^1 function is in L^4 [17]. This permits the conclusion. \square

3.2 Solution at $\lambda = 0$

We analyze now the $\lambda = 0$ case. The system decouples into three independent linear equations. We have the following.

Proposition 3.1. There is a number d_0 such that for domain width $d < d_0$, the system for $\lambda = 0$ has a solution. More precisely, there is an element $h_0 \in \mathcal{H}$ satisfying $F(h_0, 0) = 0$. Moreover, we have the explicit formula,

$$d_0^2 = 4 \frac{1}{\|D\|_{L^\infty} + \|a_u\|_{L^\infty}} \min \left(\frac{d_n}{c_n}, \frac{d_p}{c_p} \right). \quad (13)$$

Finally, the solution is unique if $d < d_0$.

Proof. We notice that (13) makes use of the fact that, for $N \leq 3$, H^2 functions are bounded. Now the first equation of (9) is a Poisson equation:

$$-\nabla^2 \rho = D + \nabla^2 a_u, \quad (14)$$

subject to homogeneous boundary values. The results of [18] imply the existence of a solution $\rho \in H^2$. Standard methods imply uniqueness.

The second equation of (9) can be rewritten as

$$-\nabla \cdot (d_n \nabla \sigma - c_n \sigma \nabla u) = \nabla \cdot (d_n \nabla a_n - c_n a_n \nabla u) \quad (15)$$

The rhs of this equation is identified with a member of H^{-1} . It follows that the equation is satisfied if we can prove surjectivity of the mapping defined by the lhs. To achieve this, we use Proposition 2.1, with

$$L_1 = -\nabla \cdot d_n \nabla \sigma, L_2 = \nabla \cdot c_n \sigma \nabla u. \quad (16)$$

The norm properties imply that L_1 is monotone and continuous. We next consider the operator L_2 . Suppose that σ_k converges weakly to σ in H_0^1 . By the Sobolev compact embedding theorem, σ_k converges to σ in L^p , $1 \leq p < 6$. We claim that the vector sequence $\sigma_k \nabla u$ converges to $\sigma \nabla u$ in L^2 . This is demonstrated directly by the Schwarz inequality. Once established, we use the following characterization of the H^{-1} norm:

$$\|L_2(\sigma_k - \sigma)\|_{H^{-1}} = \sup_{\|\phi\|_{H_0^1} \leq 1} \int_{\Omega} \nabla u (\sigma_k - \sigma) \cdot \nabla \phi \, dx. \quad (17)$$

If the Schwarz inequality is applied to this expression, then it follows from the preceding that the sequence $L_2 \sigma_k$ converges to $L_2 \sigma$. Strong continuity follows.

The third condition, coerciveness of L , follows directly from the following inequality [15, p. 501]:

$$\langle L\sigma, \sigma \rangle \geq C(\sigma, \sigma)_{H_0^1} \quad (18)$$

To derive (18), we reason as follows.

$$\begin{aligned} \langle L\sigma, \sigma \rangle &= d_n \|\sigma\|_{H_0^1}^2 + c_n \frac{1}{2} \int_{\Omega} \nabla u \nabla \sigma^2 \, dx \\ &\geq d_n \|\sigma\|_{H_0^1}^2 - \frac{c_n}{2} \|D + \nabla^2 a_u\|_{L^\infty} \|\sigma\|_{L^2}^2 \\ &\geq d_n \|\sigma\|_{H_0^1}^2 - (d^2/2) \frac{c_n}{2} \|D + \nabla^2 a_u\|_{L^\infty} \|\nabla \sigma\|_{L^2}^2, \end{aligned} \quad (19)$$

where we have used the Poincaré inequality. Inequality (18) now follows for any $d < d_0$.

Finally, we establish the boundedness of L . Suppose that \mathcal{B} is bounded in H_0^1 , with $\|\sigma\|_{H_0^1} \leq \beta$. We obtain an upper bound for the norm of $L\sigma$ in H^{-1} . The first term satisfies $\|L_1 \sigma\|_{H^{-1}} \leq d_n \beta$. The estimation of $\|L_2 \sigma\|_{H^{-1}}$ involves two terms, resulting from the application of the divergence operator.

- The term involving $\nabla^2 u$, which is an L^∞ function, is estimated as

$$c_n \left\{ \sup_{\|\phi\|_{H_0^1} \leq 1} \int_{\Omega} \nabla^2 u \sigma \phi \, dx \right\} \leq c_n \|\nabla^2 u\|_{L^\infty} \|\sigma\|_{L^2} \|\phi\|_{L^2}. \quad (20)$$

An application of the Poincaré inequality to each of the terms involving σ and ϕ completes the analysis of this term.

- The term including $\nabla \sigma$ follows a similar pattern. We begin with

$$c_n \left\{ \sup_{\|\phi\|_{H_0^1} \leq 1} \int_{\Omega} \nabla u \cdot \nabla \sigma \phi \, dx \right\} \leq c_n \|\nabla u\|_{L^4} \|\nabla \sigma\|_{L^2} \|\phi\|_{L^4}, \quad (21)$$

followed by Sobolev's inequality. We obtain a similar upper bound for this term.

The term involving the positive ion current is handled identically. This concludes the proof of existence.

To establish uniqueness, suppose that two solutions n_1, n_2 exist for the second equation of the system, and set $\sigma = n_1 - n_2$. By the above arguments, (18) is shown to hold for $d < d_0$, so that $\sigma = 0$. A similar argument holds for the third component. \square

3.3 Fréchet differentiability

In this section, we establish, in the sense of Fréchet differentiability, the existence of the partial derivatives F_h and F_λ . Moreover, F_h and F_λ are continuous so that F' exists and is continuous. We employ the notation $h = (\rho, \sigma, \tau)$ for an arbitrary element of \mathcal{H} at which the Fréchet derivative is computed. The variables acted upon by F_h are designated by $h' = (\rho', \sigma', \tau')$. In the process, we use some general results, documented in [19]. The structure of the mapping F is such that one may begin with the Gâteaux derivative. Because of directional uniformity, it will be seen that these partial derivatives are Fréchet derivatives [19, Prop. 4.8]. Further, it will be shown that the partial derivatives are continuous in the joint variable (h, λ) , so that $F'(h, \lambda)$ is continuously Fréchet differentiable [19, Prop. 4.14].

Lemma 3.2. *The partial derivative F_h exists as a Fréchet derivative which is continuous in the variable (h, λ) . Its evaluation, as a linear operator applied to the variables ρ', σ', τ' , is given as follows:*

$$\begin{aligned} F_{1,h}(\rho', \sigma', \tau') &= -\nabla^2 \rho' + \lambda(\sigma' - \tau'), \\ F_{2,h}(\rho', \sigma', \tau')[\phi] &= \int_{\Omega} [(d_n \nabla \sigma' - c_n \sigma' \nabla u - c_n n \nabla \rho') \cdot \nabla \phi] dx, \\ F_{3,h}(\rho', \sigma', \tau')[\psi] &= \int_{\Omega} [(d_p \nabla \tau' + c_p \tau' \nabla u + c_p p \nabla \rho') \cdot \nabla \psi] dx. \end{aligned} \quad (22)$$

As previously, $(u, n, p) := (\rho, \sigma, \tau) + (a_u, a_n, a_p)$.

Proof. The most efficient way to verify the validity of the system (22) is to compute the Gâteaux derivatives with respect to the components of h ; then, verify sufficient conditions for Fréchet differentiability. Finally, the partial derivatives with respect to the elements of h can be summed to obtain (22) because of the continuity of the associated component derivatives.

The Gâteaux partial derivatives are directional derivatives in the directions ρ', σ' , and τ' . Because of the linear structure of the variables, the difference quotients reduce to the derivatives straightforwardly. Summation yields (22). Continuity of F_h is a direct consequence of the linearity of the variables associated with (h, λ) . Indeed, the duality norm, coupled with Hölder's inequality, yields the result. \square

There is a parallel result for F_λ . The proof is similar to the preceding lemma so that we omit the proof.

Lemma 3.3. *The partial derivative F_λ exists as a Fréchet derivative which is continuous in the variable (h, λ) . Its evaluation, as a linear operator applied to the real variable θ , is given as follows:*

$$\begin{aligned} F_{1,\lambda}(\theta) &= \theta(n-p), \\ F_{2,\lambda}(\theta) &= 0 \\ F_{3,\lambda}(\theta) &= 0. \end{aligned} \tag{23}$$

3.4 Lipschitz continuity for derivatives

We begin with the statement of Lipschitz continuity.

Proposition 3.2. *The derivative mapping F_h is globally Lipschitz continuous in the variable (h, λ) .*

Proof. It suffices to show that $F_h(\cdot, \lambda)$ is Lipschitz continuous for each fixed λ and that $F_h(h, \cdot)$ Lipschitz continuous for each fixed h .

We begin with the first case. Because of the structure of the system (22), only differences in the second and third components are relevant. These are quite similar in their analysis, so that we will consider differences in $F_{2,h}$. The following estimates hold for $\|\phi\|_{H_0^1} \leq 1$. They are required for H^{-1} estimation.

$$\begin{aligned} c_n \|\sigma' \nabla(\rho_2 - \rho_1) \cdot \nabla \phi\|_{L^1} &\leq c_n \|\sigma'\|_{L^4} \|\rho_2 - \rho_1\|_{W^{1,4}} \leq c_n C \|\sigma'\|_{H_0^1} \|\rho_2 - \rho_1\|_{H^2} \\ c_n \|(\sigma_2 - \sigma_1) \nabla \rho' \cdot \nabla \phi\|_{L^1} &\leq c_n \|\nabla \rho'\|_{L^4} \|\sigma_2 - \sigma_1\|_{L^4} \leq c_n C \|\nabla \rho'\|_{H_0^1} \|\sigma_2 - \sigma_1\|_{H_0^1} \end{aligned}$$

In these inequalities, C is a generic constant obtained from the product of Sobolev embedding constants. As mentioned, the third component analysis is exactly parallel. Since the duality norm for \mathcal{G} involves the supremum over test functions with \mathcal{H} norm not exceeding one, we may obtain an upper bound for

$$\|(F_h(h_2, \lambda) - F_h(h_1, \lambda))h'\|_{\mathcal{G}} \tag{24}$$

from the estimates above. Indeed, each estimate is bounded above by a constant times $\|h_2 - h_1\|_{\mathcal{H}} \|h'\|_{\mathcal{H}}$. The result that $F_h(\cdot, \lambda)$ is Lipschitz continuous follows from an application of basic ℓ^2 inequalities.

The second case, with fixed h , is immediate. This concludes the proof. \square

Proposition 3.3. *The partial derivative $F_\lambda(h, \lambda)$ is globally Lipschitz continuous.*

Proof. The partial derivative $F_\lambda(h, \cdot)$ is invariant for each fixed h . It is sufficient then to consider $F_\lambda(\cdot, \lambda)$ for each fixed λ . In this case, the linearity of the components of h implies the result. \square

Remark 3.2. *The hypothesis of (2) has now been verified. The constant C is obtained directly from the current analysis.*

3.5 A bounded inverse for F_h

Recall that h_0 denotes the solution of the system for $\lambda = 0$. In the previous subsections, the action of $F(\cdot, \lambda)$ has included all of \mathcal{H} . Here, we restrict the domain of $F(\cdot, \lambda)$ to a ball $S = \{h \in \mathcal{H} : \|h\|_{\mathcal{H}} \leq R\}$, where $R = \|h_0\|_{\mathcal{H}} + r$, with r to be specified. Observe that the general theory guarantees that the homotopy curve lies in an open ball of radius r , centered at h_0 . We note here that the restriction of F to S guarantees the logical consistency of the choice of the domain width d , so that it remains bounded away from 0. This enters decisively into the contraction mapping theorem discussed below. We now construct inverses for F_w along the curve.

For each fixed λ , and each fixed h , we will show that there is a bounded linear operator $H(h, \lambda)$ from \mathcal{G} to \mathcal{H} which is an inverse for $F_h(h, \lambda)$.

Proposition 3.4. *There is a positive real number $d_1 = d_1(R)$ such that, if the width d of Ω satisfies $d \leq d_1$ and $d < d_0$, for each pair (h, λ) , the linear operator $F_h(h, \lambda)$ is bijective. More precisely, for each $g \in \mathcal{G}$, there is a unique element $h' \in \mathcal{H}$ such that $F_h(h, \lambda)h' = g$. In particular, $F_h(h, \lambda)$ is invertible. The norm M of the inverse may be selected so that $M = 2$.*

Proof. There are two stages to the proof.

- (a) Let $g \in \mathcal{G}$ be given, and consider the system $F_h(h, \lambda)h' = g$. We show that this system has a unique solution h' under assumptions on the width d .
- (b) There is a uniform bound for the norms of the inverses of (a). This is achieved by an appropriate uniform selection of $d(R)$.

We consider these in turn.

(a) Although the system is linear, we employ the contraction mapping theorem for an appropriately defined operator \mathcal{T} . For the remainder of the proof, we use the equivalent norm on \mathcal{H} given by the Laplacian norm for the first component and, as previously mentioned, the derivative part of the H_0^1 norm for the remaining components. Also, for convenience of notation, we write $\Delta = \nabla^2$. In order to reformulate the system for an application of the contraction mapping theorem, we define $-\Delta^{-1} : \mathcal{G} \mapsto \mathcal{H}$ by the componentwise definition:

$$-\Delta^{-1}g_i = f_i, \quad (25)$$

where

$$(\nabla f_i, \nabla \phi)_{L^2} = g_i(\phi), i = 1, 2, 3. \quad (26)$$

Results of [18] for convex domains allow us to conclude that the Laplacian defines an isomorphism between $H^2 \cap H_0^1$ and L^2 .

The surjectivity will follow from a fixed point argument applied to an equivalent system, given by

$$h' = \mathcal{T}h' - \Delta^{-1}\tilde{g}. \quad (27)$$

Here, \tilde{g} is obtained from g by dividing the second and third components by d_n, d_p , respectively. The same operations are applied to the components of F_h in the formation of \mathcal{T} .

Since \mathcal{T} is a strict contraction if and only if $\mathcal{T} - \Delta^{-1}\tilde{g} := \mathcal{S}$ is, we consider \mathcal{T} . For reference, we write the form explicitly.

$$\begin{aligned}\mathcal{T}_1(\rho', \sigma', \tau') &= \lambda \Delta^{-1}(\sigma' - \tau'), \\ \mathcal{T}_2(\rho', \sigma', \tau') &= \frac{c_n}{d_n} \Delta^{-1} \nabla \cdot (\sigma' \nabla u + n \nabla \rho') \\ \mathcal{T}_3(\rho', \sigma', \tau') &= -\frac{c_p}{d_p} \Delta^{-1} \nabla \cdot (\tau' \nabla u + p \nabla \rho')\end{aligned}\tag{28}$$

The proof will show that \mathcal{T} maps \mathcal{H} into itself. This is inherited by \mathcal{S} . We verify the strict contractive property. Since \mathcal{T} is linear, it suffices to estimate $\|\mathcal{T}h'\|_{\mathcal{H}}$. A fundamental tool in this estimation is the use of the version of the Poincaré inequality in [16, Th. 12.17]. This inequality permits L^p estimation in terms of a $W_0^{1,p}$ estimation with a constant proportional to the finite width of Ω . Recall that this property was used effectively to derive the solution h_0 . Now, for sufficiently small width, the contraction constant is less than 1. We present the argument as follows, with emphasis on the case $N = 3$, the most delicate case. The case $N = 2$ is implied by the arguments.

- $H^2 \cap H_0^1$ estimation of \mathcal{T}_1

$$\|\Delta^{-1}\sigma'\|_{H^2} \leq \|\Delta^{-1}\|\|\sigma'\|_{L^2} \leq (d/\sqrt{2})\|\Delta^{-1}\|\|\nabla\sigma'\|_{L^2}.\tag{29}$$

A similar result holds for the term involving τ' .

- H_0^1 estimation of \mathcal{T}_2

When the divergence operator is applied, each of the four terms is of the form $-\Delta^{-1}\chi$, where it will be shown that $\chi \in L^{3/2}$. In order to estimate the H_0^1 norm of $-\Delta^{-1}\chi$, we observe that, by definition, this term is numerically equal to the H^{-1} norm of χ . A representation of this norm on H^{-1} will be chosen, allowing for a direct application of the Poincaré inequality. This form was used in [20], in the study of the Stefan problem and is used for L^p functions which define continuous linear functionals, provided $-\Delta^{-1}\chi \in H_0^1$, as is the case for $\chi \in L^{3/2}$. In particular, for χ described above, we have

$$\|\chi\|_{H^{-1}} = \sqrt{\int_{\Omega} \mathcal{R}\chi \chi \, dx}.\tag{30}$$

Here, we have written \mathcal{R} for the map $-\Delta^{-1}$.

For χ in (30), we have the following inequalities.

$$\int_{\Omega} \mathcal{R}\chi \chi \, dx \leq \|\mathcal{R}\chi\|_{L^3} \|\chi\|_{L^{3/2}}\tag{31}$$

$$\leq (d/3^{1/3}) \|\nabla \mathcal{R}\chi\|_{L^3} \|\chi\|_{L^{3/2}}\tag{32}$$

$$\leq C(d/3^{1/3}) \|\mathcal{R}\chi\|_{W^{2,3/2}} \|\chi\|_{L^{3/2}}\tag{33}$$

$$\leq C(d/3^{1/3}) \|\mathcal{R}\| \|\chi\|_{L^{3/2}}^2.\tag{34}$$

The first inequality follows from Hölder's inequality; the second from the Poincaré inequality [16]; the third from Sobolev's embedding theorem [17]; and the fourth from [18].

In this estimation, we will choose $\chi = \chi_1\chi_2$, where $\chi_1 \in L^2$ and $\chi_2 \in H^1$. Typically, these functions are vector-valued. Hölder's inequality implies that $\chi \in L^{3/2}$. Indeed, when the inequality is applied with $p = 4/3$ and $q = 4$ to $\|\chi_1\chi_2\|_{L^{3/2}}$, one obtains the respective product of the L^2 and L^6 norms; Sobolev's inequality gives the desired result:

$$\|\chi\|_{H^{-1}}^2 \leq C'd\|\chi_1\|_{L^2}^2\|\chi_2\|_{H^1}^2. \quad (35)$$

Here, C' depends only on operator and embedding constants. These general remarks will now be applied to the terms above obtained by the application of the divergence operator as follows. Consider \mathcal{T}_2 .

$$\frac{c_n}{d_n}(\nabla\sigma' \cdot \nabla u + \sigma'\nabla^2 u + \nabla n \cdot \nabla\rho' + n\nabla^2\rho'). \quad (36)$$

Each of these four terms is an example of the generic function $\chi = \chi_1\chi_2$ discussed previously. In each case, the terms involving u and n and their derivatives have norm bounds depending on a constant $C(R)$. A similar analysis holds for the third component of \mathcal{T} . In summary, by standard inequalities, we obtain an inequality of the form:

$$\|\mathcal{T}h'\|_{\mathcal{H}} \leq dC(R)\|h'\|_{\mathcal{H}}. \quad (37)$$

We thus obtain a contraction if d is sufficiently small. In particular, we may choose $d \leq d_2$ so that the contraction constant does not exceed $1/2$. Note that this choice of d depends only upon R . Thus, \mathcal{S} has a unique fixed point; the existence of an inverse follows.

(b) To obtain a uniform estimate for the inverse norm, we begin with the fixed point relation,

$$h' = \mathcal{T}h' + \Delta^{-1}\tilde{g}, \quad (38)$$

and we estimate the \mathcal{H} norm. We obtain,

$$\|h'\|_{\mathcal{H}} \leq \|\mathcal{T}h'\|_{\mathcal{H}} + \|\tilde{g}\|_{\mathcal{G}}. \quad (39)$$

Since we have selected d to ensure an upper bound of $1/2$ for the contraction constant, we conclude that 2 serves as an upper bound for the inverse mapping. This is the assertion $M = 2$ in the statement of the proposition. \square

3.6 The existence theorem

We begin with an essential lemma.

Lemma 3.4. *A uniform estimate for F_λ is given as follows.*

$$\|F_\lambda\| \leq \|a_n\|_{L^2} + \|a_p\|_{L^2} + (2d/\sqrt{2})R. \quad (40)$$

Here, d is the width of Ω and $R = r + \|h_0\|_{\mathcal{H}}$. Also, h_0 is a solution of the system for $\lambda = 0$ and r , to be specified below, is the radius of a ball centered at h_0 .

Proof. We proceed from the explicit representation of $\|F_\lambda\|$, and calculate the operator norm. It is bounded above by the L^2 norm. The factor $2d/\sqrt{2}$ follows from the Poincaré inequality applied to the elements σ, τ . \square

Remark 3.3. *One sees that the choices, for $0 < \alpha_0 \leq 1$,*

$$\begin{aligned} d &= \frac{\alpha_0}{2\sqrt{2} \max(C, 2)}, \\ r &= \frac{1}{1 - \alpha_0/2} \max(C, 2)(1 + \|a_n\|_{L^2} + \|a_p\|_{L^2}) + \frac{\alpha_0}{2 - \alpha_0} \|h_0\|_{\mathcal{H}}, \end{aligned}$$

lead to the satisfaction of the requirement for the existence of the homotopy curve, with the proper choice of α_0 . This is made specific as follows.

The following theorem is now a direct consequence of Theorem 2.1.

Theorem 3.1. *Consider the drift-diffusion system $F(u, \lambda) = 0$, where F is defined in Definition 3.1. If the width d of Ω satisfies $d < d_0$, where d_0 is defined in (13), then there is a solution to the system for $\lambda = 0$. If r is defined as in the remark above, and d also satisfies $d \leq d_1, d \leq d_2$, then there exists a homotopy solution curve within the open ball of radius r , centered at h_0 .*

3.7 Nonnegativity for ion densities

Thus far, we have studied the system independently of the sign properties associated with the ion densities. In the process, we have not made assumptions on the signs of a_n, a_p . We now assume their nonnegative boundary traces. Note that charge voids are not eliminated, so that a situation can arise which is excluded when Boltzmann statistics are assumed.

Theorem 3.2. *Suppose a homotopy solution curve $h_\lambda = (\rho_\lambda, \sigma_\lambda, \tau_\lambda)$, exists as described in Definition 3.1. We write*

$$(u_\lambda, n_\lambda, p_\lambda) = (a_u, a_n, a_p) + h_\lambda$$

for the basic variables. There is a number D_0 such that, if the width d satisfies $d < D_0$, and if $a_n \geq 0, a_p \geq 0$ on $\partial\Omega$, then $n_\lambda \geq 0, p_\lambda \geq 0$ on Ω for each fixed λ .

Proof. We observe that $\|h_\lambda\|_{\mathcal{H}} \leq R, \forall \lambda$. Consider the system for a fixed $\lambda, 0 \leq \lambda \leq 1$. For clarity, we suppress the subscript λ for the variables. We consider the second equation of the system, and we employ n^- as a test function, where n^- is the negative part of n . Here, n^- satisfies $n = n^+ + n^-$. The assumption that $a_n \geq 0$ on $\partial\omega$ implies that the boundary trace of n^- is zero, so that $n^- \in H_0^1$. We now use the second equation to show that n^- is the zero function.

We have:

$$0 = d_n \int_{\Omega} \nabla n \cdot \nabla n^- \, dx - c_n \int_{\Omega} n \nabla u \cdot \nabla n^- \, dx. \quad (41)$$

It follows that

$$\begin{aligned}
0 &= d_n \int_{\Omega} |\nabla n^-|^2 dx - c_n/2 \int_{\Omega} \nabla u \cdot \nabla (n^-)^2 dx \\
&= d_n \int_{\Omega} |\nabla n^-|^2 dx + c_n/2 \int_{\Omega} \nabla^2 u (n^-)^2 dx \\
&\geq d_n \|n^-\|_{H_0^1}^2 - c_n/2 \|\nabla^2 u\|_{L^4} \|n^-\|_{L^4} \|n^-\|_{L^2} \\
&\geq d_n \|n^-\|_{H_0^1}^2 - c_n/2 d C(D, R) \|\nabla n^-\|_{L^2}^2.
\end{aligned}$$

The constant $C(D, R)$ arises from the estimation of the L^4 norm of $\nabla^2 u$ as determined from (9), and from Sobolev embedding constants. It is now immediate that, if d is sufficiently small, then

$$\|n^-\|_{H_0^1} = 0, \quad (42)$$

which implies that $n^- = 0$. A similar analysis yields the result that $p^- = 0$ if d is sufficiently small. The number D_0 in the statement of the theorem is chosen as the largest number such that the final expression above, as well as the corresponding expression for p^- , is nonnegative. \square

3.8 Final remarks

We have analyzed a basic scientific model of electrodiffusion and drift in an alternative way, through the basic variables. Although we avoid the Einstein relations and Boltzmann statistics, our analysis is restricted to narrow convex domains. We establish the existence of a homotopy solution curve, starting from the uniquely defined center of a specifically located sphere in function space. The result then allows the application of [6, Th. 3.1], which describes a predictor/corrector algorithm, terminating at the starting iterate of a convergent Newton sequence. The conditions of the general theory, verified here for the drift-diffusion model, allow the implementation of the algorithm. Although the results are theoretical, they provide a basic strategy for the solution of the system.

As a concluding observation, we note that the arguments are not sign sensitive, so that the electrostatic potential can be replaced by potentials which induce drift.

References

- [1] M. Lundstrom, *Fundamentals of Carrier Transport*. Addison-Wesley, Reading MA, 1990.
- [2] J. Keener and J. Sneyd, *Mathematical Physiology*. Springer-Verlag, New York, 1998.
- [3] I. Rubinstein, *Electrodiffusion of Ions*. SIAM, Philadelphia, 1990.

- [4] J.W. Jerome, Consistency in semiconductor modeling: existence/stability analysis for the stationary Van Roosbroeck system. *SIAM J. Appl. Math.* **45** (1985), 565-590.
- [5] J.W. Jerome and T. Kerkhoven, A finite element approximation theory for the drift-diffusion semiconductor model. *SIAM J. Numer. Anal.* **28** (1991), 403-422.
- [6] J.W. Jerome, Approximate Newton methods and homotopy for stationary operator equations. *Constr. App.* **1** (1985), 271-285.
- [7] M. Mock, On equations describing steady-state carrier distributions in a semiconductor device. *Comm. Pure Appl. Math.* **25** (1972), 781-792.
- [8] T. Seidman, Steady state solutions of diffusion reaction systems with electrostatic convection. *Nonlinear Anal.* **4** (1980), 623-637.
- [9] R.E. Bank, J.W. Jerome, and D.J. Rose, Analytical and numerical aspects of semiconductor device modeling. In, R. Glowinski and J. Lions (eds.) *Proc. Fifth International Conference on Computing Methods in Science and Engineering*. North Holland, Amsterdam, 1982, pp. 593-597.
- [10] P.A. Markowich, *The Stationary Semiconductor Device Equations*. Springer, Vienna and New York, 1986.
- [11] J.W. Jerome, *Analysis of Charge Transport: A Mathematical Study of Semiconductor Devices*. Springer, Berlin, 1996.
- [12] J. Park and J.W. Jerome, Qualitative properties of solutions of steady-state Poisson-Nernst-Planck systems: Mathematical study, *SIAM J. Appl. Math.* **57** (1997), 609-630.
- [13] V. Barcion, D.P. Chen, R.S. Eisenberg, and J.W. Jerome, Qualitative properties of solutions of steady-state Poisson-Nernst-Planck systems: Perturbation and simulation study, *SIAM J. Appl. Math.* **57** (1997), 631-648.
- [14] J.W. Jerome, The steady boundary-value problem for charged incompressible fluids: PNP Navier-Stokes systems. *Nonlinear Anal.* **74** (2011), 7486-7498.
- [15] E. Zeidler, *Nonlinear Functional Analysis and its Applications II/B*. Springer, 1990.
- [16] G. Leoni, *A First Course in Sobolev Spaces*. Graduate Texts in Mathematics 105, Amer. Math. Soc. , Providence, 2009.
- [17] R. Adams, *Sobolev Spaces*, Academic Press, 1976.
- [18] V. Adolfsson, L^p integrability of the second order derivatives of Green potentials in convex domains. *Pacific J. Math.* **159** (1993), no. 2, 201-225.

- [19] E. Zeidler, *Nonlinear Functional Analysis and its Applications, I: Fixed Point Theorems* (English translation). Springer-Verlag, New York, 1986..
- [20] J.W. Jerome and M.E. Rose, Error estimates for the multidimensional two-phase Stefan problem. *Math. Comp.* **39** (1982).