A GENERALIZED KELLER-SEGEL MODEL

By

ERICA ZUHR

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2012
ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisors Professor Jay Gopalakrishnan and Professor Patrick De Leenheer. Despite being across the country or even out of the country for a large portion of my graduate career, they went above and beyond to continue supporting my research and professional development. Thanks also go out to Margaret for being my “third advisor” when the other two were not in the state. I would like to thank my committee members for their time and valuable feedback. I want to thank Kevin for being so supportive, and for always being able to make me smile no matter what I am dealing with in the academic world. I also want to thank my mom, dad, and stepdad for unconditional love and support in everything I do. Finally, thanks go out to all of my friends in the math department, especially my office mates. I might have been able to finish my degree without them, but it wouldn’t have been nearly as much fun.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACKNOWLEDGMENTS</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>LIST OF TABLES</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>CHEMOTAXIS ALONG A FIXED GRADIENT</td>
<td>17</td>
</tr>
<tr>
<td>2.1</td>
<td>An Example in One Dimension</td>
<td>17</td>
</tr>
<tr>
<td>2.1.1</td>
<td>Analytical Steady State Solutions</td>
<td>18</td>
</tr>
<tr>
<td>2.1.2</td>
<td>Numerical Visualization</td>
<td>24</td>
</tr>
<tr>
<td>2.2</td>
<td>The Multidimensional Case</td>
<td>24</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Existence of Classical Solutions</td>
<td>24</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Properties of Numerical Solutions</td>
<td>25</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Numerical Difficulties and Method Validation</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>A SIMPLE KELLER-SEGEL MODEL</td>
<td>32</td>
</tr>
<tr>
<td>3.1</td>
<td>Positivity of Exact Solutions</td>
<td>32</td>
</tr>
<tr>
<td>3.2</td>
<td>Preliminaries for Numerical Steady State Solutions</td>
<td>37</td>
</tr>
<tr>
<td>3.3</td>
<td>Numerical Solutions</td>
<td>49</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Solutions on the Unit Square</td>
<td>49</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Solutions on a Disk</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>AN EXTENDED KELLER-SEGEL MODEL</td>
<td>53</td>
</tr>
<tr>
<td>4.1</td>
<td>The Generalized Model Form</td>
<td>53</td>
</tr>
<tr>
<td>4.2</td>
<td>Reactions Yielding Positive Solutions</td>
<td>55</td>
</tr>
<tr>
<td>4.3</td>
<td>Existence of Homogeneous Steady State Solutions</td>
<td>61</td>
</tr>
<tr>
<td>4.4</td>
<td>Stability of Homogeneous Steady State Solutions</td>
<td>63</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Reduction to Finite Dimension</td>
<td>64</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Sufficient Conditions for Instability</td>
<td>67</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Sufficient Conditions for Stability</td>
<td>73</td>
</tr>
<tr>
<td>4.5</td>
<td>Examples</td>
<td>79</td>
</tr>
<tr>
<td>5</td>
<td>NUMERICAL SOLUTIONS</td>
<td>87</td>
</tr>
<tr>
<td>5.1</td>
<td>Spectral Bands</td>
<td>87</td>
</tr>
<tr>
<td>5.2</td>
<td>Eigenfunctions and Potential Patterns</td>
<td>91</td>
</tr>
<tr>
<td>5.3</td>
<td>Positivity of Numerical Solutions</td>
<td>94</td>
</tr>
</tbody>
</table>
5.3.1 Implementation of the Method ........................................... 97
5.3.2 Derivation of the Method .................................................. 99
5.3.3 Proof of Positivity .......................................................... 101
5.4 Vectorization .................................................................. 104
5.5 Method Validation .............................................................. 105

6 APPLICATIONS AND RESULTS OF SIMULATIONS .................. 108

6.1 Application to Tumor Growth ............................................. 108
6.2 Keller-Segel Time Simulations ............................................. 112
   6.2.1 Minimal Keller-Segel Model ........................................... 112
   6.2.2 Four Equation Keller-Segel Model .................................. 116
6.3 Parameter Fitting in the Four Equation Keller-Segel Model ....... 122

REFERENCES ........................................................................ 129

BIOGRAPHICAL SKETCH .......................................................... 132
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-1</td>
<td>Error for steady state solution on unit square</td>
<td>49</td>
</tr>
<tr>
<td>3-2</td>
<td>Error for steady state solution on a disk</td>
<td>51</td>
</tr>
<tr>
<td>6-1</td>
<td>Parameter values for four equation KS</td>
<td>124</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>2-1</td>
<td>Numerical solution to (2–1)</td>
<td>24</td>
</tr>
<tr>
<td>2-2</td>
<td>Standard FEM solutions to (2–14)</td>
<td>30</td>
</tr>
<tr>
<td>2-3</td>
<td>Exact solution to (2–14) on refined mesh</td>
<td>31</td>
</tr>
<tr>
<td>3-1</td>
<td>Solution to (3–15) on the unit square</td>
<td>50</td>
</tr>
<tr>
<td>3-2</td>
<td>Solutions to (3–15) on a disk</td>
<td>51</td>
</tr>
<tr>
<td>3-3</td>
<td>More solutions to (3–15) on a disk</td>
<td>52</td>
</tr>
<tr>
<td>5-1</td>
<td>An example of the spectral bands of the matrices $\mathcal{M}(\mu_i)$</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>corresponding to the dimerization reaction for $1 \leq i \leq 20$.</td>
<td></td>
</tr>
<tr>
<td>5-2</td>
<td>An example of the spectral bands of the matrices $\mathcal{M}(\mu_i)$</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>corresponding to the full Keller-Segel model for $1 \leq i \leq 5$.</td>
<td></td>
</tr>
<tr>
<td>5-3</td>
<td>Eigenfunction corresponding to the first positive eigenvalue of (5–4)</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>with $u_* = 170$.</td>
<td></td>
</tr>
<tr>
<td>5-4</td>
<td>Eigenfunctions corresponding to the second positive eigenvalue of (5–4)</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>with $u_* = 550$.</td>
<td></td>
</tr>
<tr>
<td>5-5</td>
<td>Eigenfunctions of (5–4)</td>
<td>95</td>
</tr>
<tr>
<td>5-6</td>
<td>More spectral bands of the full KS model.</td>
<td>96</td>
</tr>
<tr>
<td>5-7</td>
<td>Eigenfunctions of (5–4) on a disk</td>
<td>96</td>
</tr>
<tr>
<td>5-8</td>
<td>Positivity preserving FEM solutions to (2–14)</td>
<td>107</td>
</tr>
<tr>
<td>6-1</td>
<td>One dimensional FEM solutions to Equation (6–1)</td>
<td>110</td>
</tr>
<tr>
<td>6-2</td>
<td>Two dimensional FEM solutions for $u$ of Equation (6–1)</td>
<td>111</td>
</tr>
<tr>
<td>6-3</td>
<td>Spectral bands of the minimal KS model</td>
<td>114</td>
</tr>
<tr>
<td>6-4</td>
<td>Blow-up of the minimal KS model</td>
<td>115</td>
</tr>
<tr>
<td>6-5</td>
<td>Convergent ($u_* = 11$) versus blow-up ($u_* = 13$) parameter ranges of the</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>minimal KS model</td>
<td></td>
</tr>
<tr>
<td>6-6</td>
<td>Convergence to a nonhomogeneous stationary solution on a disk</td>
<td>117</td>
</tr>
<tr>
<td>6-7</td>
<td>Convergence to constant steady state for full KS model</td>
<td>118</td>
</tr>
<tr>
<td>6-8</td>
<td>Difference between standard and positivity preserving FEM for full KS time</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td>simulations</td>
<td></td>
</tr>
</tbody>
</table>
Motivated by the derivation of the standard Keller-Segel model, we introduce a generalized Keller-Segel model for chemotaxis. Our generalized system models an organism, the chemoattractant which the organism is chemotactically attracted to, and an arbitrary number of chemicals which interact with the chemoattractant. Our model accounts for the fact that some of these chemicals may be produced by the organism. We present results on the existence and linear stability of homogeneous steady states using a reduction to a finite dimensional system. These results significantly expand the class of chemical reaction networks which demonstrate chemotactically initiated pattern formation. In addition to theoretical results, we use the standard finite element method to demonstrate solutions to both the original and the generalized Keller-Segel models. In addition to demonstrating nonhomogeneous stationary solutions of the standard Keller-Segel model, we perform time simulations to demonstrate the stability of these stationary solutions. We also use a nonstandard finite element method to perform time simulations, and show that this nonstandard method reduces numerical instability. We prove that under certain conditions, this numerical method preserves positivity of solutions. Finally, we use a novel method of spectral bands to demonstrate directions of instability in the generalized model, and use these as a basis for further time simulations.
CHAPTER 1
INTRODUCTION

Chemotaxis is a fascinating biological phenomenon which occurs when the movement of a cell or organism is affected by a chemical in the environment [18]. Positive chemotaxis occurs when a chemical called the chemoattractant attracts the organism, and negative chemotaxis occurs when a so-called chemorepellent repells the organism. From fertilization of an egg in the earliest states of development to immune system function to cancer growth and metastasis, chemotaxis arises in many different biological processes [17]. The well studied two equation Keller-Segel (KS) model for chemotaxis, which we will review below, was originally developed by Keller and Segel in [21]. The strongly coupled two equation KS system models two variables, the species whose movement is affected and the chemoattractant or chemorepellent. Although generalizations of the model (reviewed below) have been introduced and studied previously, we introduce our own form of a generalized model which allows for modeling of the species whose movement is being affected, the chemoattractant, and an arbitrary number of chemicals which interact with the chemoattractant and may or may not be produced by the species. In addition to introducing the model and giving new theoretical results on existence and stability of steady states, we apply the finite element method (FEM) to find numerical solutions of both the generalized model and the simple two equation model. As an extension of results in [21, 34], we show that a destabilization mechanism which potentially leads to pattern formation extends to many chemical reaction networks (CRNs) in the framework of our generalized model.

Before continuing, we first introduce some important notation that will be used throughout. We allow the slight abuse of notation that variables such as $u$ and $v_i$ when used in an equation denote the concentrations of the species being modeled, but when the meaning is clear in context we often use the variables instead of referring to the chemical or organism by name. We denote partial derivatives by subscripts, for
\[ \partial_t u \] is the partial derivative of \( u \) with respect to time \( t \). When considering a domain \( \Omega \times [0, T] \subset \mathbb{R}^{n+1} \), the symbols \( \nabla \) and \( \Delta \) denote the gradient and the Laplacian with respect to the spatial variable \( x \in \mathbb{R}^{n} \). When used with derivatives, \( n \) or \( \vec{n} \) will denote the outward pointing normal to the boundary and \( \frac{\partial u}{\partial n} \equiv \nabla u \cdot n \) denotes the directional derivative in the outward normal direction. All integrals are taken with respect to the standard Lebesgue measure. By \( W^{k,p}(\Omega) \) we mean the Sobolev space consisting of all functions in \( L^p(\Omega) \) whose partial derivatives up to order \( k \) exist in the weak sense and are in \( L^p(\Omega) \). As is standard notation, \( H^k(\Omega) \equiv W^{k,2}(\Omega) \). We denote by \( H^k(\Omega)^N \) the product space \( H^k(\Omega) \times H^k(\Omega) \times \cdots \times H^k(\Omega) \), where the product is taken \( N \) times. By the function spaces \( C^n(\Omega) \) for any integer \( n \geq 0 \) we denote functions on \( \Omega \) whose \( n \)-th order derivatives exist and are continuous. Hence \( C^0(\Omega) = C(\Omega) \) is the space of continuous functions on \( \Omega \). Finally, we give the notation which will be used for the FEM analysis. Let \( T_h \) denote a triangulation of the given domain \( \Omega \) with global vertices typically indexed by \( x_{\ell} \). The linear FEM space is defined to be \( S_h = \{ w \in C(\Omega) : w|_T \text{ is linear on every mesh triangle } T \in T_h \} \). The function \( \phi_{\ell} \in S_h \) is defined to be the piecewise linear function which is one at vertex \( x_{\ell} \) and zero at all other vertices of the triangulation.

To better understand the motivation for our generalized model, we first review the derivation of the original Keller-Segel model. In [21], Keller and Segel base the derivation of their model on the slime mold \textit{Dictyostelium discoideum}. \textit{D. discoideum} has a unique life cycle [16, 18], and due to its interesting development is a model organism for the National Institute of Health [18]. The life cycle begins as the amoeba divide and reproduce asexually so long as a sufficient food source is present. Once the food source is depleted, the amoeba enter a “starvation mode” and, after a time, one cell will begin to emit a signal of the chemical cyclic Adenosine Monophosphate, or cAMP. This signal chemotactically attracts the other amoeba, which in turn begin to emit a cAMP signal. In addition the amoeba emit phosphodiesterase, an enzyme that degrades the
cAMP. This feedback loop causes the cells to aggregate and form a multicellular slug, which moves as one body. Once the slug has moved to a better environment, the cells differentiate into two types, and form a fruiting body which emits spores. These spores germinate into new amoeba, and the life cycle begins anew. The KS model is based on the aggregation phase of the life cycle, when the chemotactic feedback loop initiates aggregation.

In order to derive the model, Keller and Segel make the following assumptions in [21], as reviewed in [18]. Consider the model on a sufficiently smooth domain \( \Omega \subset \mathbb{R}^n \) and for time \( t > 0 \). Let \( u \) denote the concentration of \( D. discoideum \), \( v \) denote the concentration of the chemoattractant cAMP, \( \eta \) denote the concentration of the enzyme phosphodiesterase, and \( c \) the concentration of a complex formed by the enzyme and the cAMP. Assume additionally that the enzyme degrades the camp by means of a CRN represented by

\[
v + \eta \leftrightarrow c \longrightarrow \eta + \text{ degraded product.}
\]  

Let the reaction (1–1) be governed by the law of mass action with forward and backward rate constants for the first reaction given by \( k_1 \) and \( k_2 \) respectively, and the last reaction rate constant given by \( k_3 \). Assume also that the rates at which the \( D. discoideum \) produce the cAMP and the enzyme are given by the nonnegative functions \( f(v) \) and \( h(v, \eta) \) respectively. Finally, assume that all components diffuse according to Fick’s law, and assume that the flux of the cell has an additional term representing the chemotactic contribution. Hence the flux of the amoeba is given by

\[
J_u(x, t) = -D_0 \nabla u(x, t) + \kappa(u, v) \nabla v(x, t)
\]

where \( D_0 \) is the diffusion coefficient and \( \kappa(u, v) \) is called the chemotactic sensitivity function. These assumptions along with either no-flux or Neumann boundary conditions
lead to the full Keller-Segel model from \[21\],

\[
\begin{align*}
  u_t &= \nabla \cdot (D_0 \nabla u - \kappa(u,v) \nabla v), & x \in \Omega, \ t > 0 \\
  v_t &= D_1 \Delta v - k_1 v \eta + k_2 c + u f(v), & x \in \Omega, \ t > 0 \\
  \eta_t &= D_1 \Delta \eta - k_1 v \eta + (k_2 + k_3) c + u h(v, \eta), & x \in \Omega, \ t > 0 \\
  c_t &= D_3 \Delta c + k_1 v \eta - (k_2 + k_3) c, & x \in \Omega, \ t > 0 \\
  \frac{\partial u}{\partial n} = \frac{\partial v}{\partial n} = \frac{\partial \eta}{\partial n} = \frac{\partial c}{\partial n} = 0, & x \in \partial \Omega, \ t > 0.
\end{align*}
\] (1–2a) (1–2b) (1–2c) (1–2d) (1–2e)

As soon as Keller and Segel introduce (1–2), they immediately reduce it to a two equation system which models only the chemoattractant and the amoeba. This reduction uses the assumption that the total concentration of the enzyme in both free and bound forms is constant, meaning

\[\eta + c = \eta_0.\]

The second assumption is that the complex \(c\) is in steady state with regards to the reaction (1–1), meaning that

\[k_1 v \eta - (k_2 + k_3) c = 0.\]

Further assumptions reviewed in [17, 18] reduce the two equation model to the commonly studied “minimal system”,

\[
\begin{align*}
  u_t &= \nabla \cdot (D_0 \nabla u - \chi u \nabla v), & x \in \Omega, \ t > 0 \\
  v_t &= D_1 \Delta v + \alpha u - \gamma v, & x \in \Omega, \ t > 0 \\
  \frac{\partial u}{\partial n} = \frac{\partial v}{\partial n} = 0, & x \in \partial \Omega, \ t > 0 \\
  u(0,x) = u_0(x), \ v(0,x) = v_0(x), & x \in \Omega.
\end{align*}
\] (1–3a) (1–3b) (1–3c) (1–3d)

In (1–3), \(\gamma > 0\) is a constant decay rate of the chemoattractant, \(\alpha > 0\) is the constant production rate per amoeba of the chemoattractant, and the chemotactic sensitivity function \(\kappa(u,v)\) from (1–2a) is assumed to be of the form \(\chi u\) for some constant\]
The system (1–3) has been studied extensively, and results on existence and uniqueness of solutions, positivity, blow-up, and stability can be found in, for example, [17, 18, 34]. Blow-up, which is considered to occur when the $L^\infty$ norm of either $u$ or $v$ becomes unbounded in finite or infinite time, has attracted a great deal of attention but not will not be addressed in our work except for the citation of a few computationally useful results in Section 3.2 and some numerical results in Chapter 6. The unreduced system (1–2) has not been the subject of a great amount of study, and this partially motivates the generalization of the KS model which will be introduced in Chapter 4.

One of the most interesting aspects of our theoretical results on the generalized KS model is that they point to chemotactically induced pattern formation for very general classes of CRNs. Although this mechanism for pattern formation has been shown in [34] and others in the case of the minimal model, we generalize it to a much larger class of systems. To emphasize how chemotactically induced instability differs from the traditional Turing diffusion based instability [41], we review an example of diffusive instabilities as explained in [8, 32]. For this example, consider a spatially homogeneous stationary solution $(u_*, v_*)$ of the partial differential equation (PDE)

\begin{align}
\frac{\partial u}{\partial t} &= D_1 \Delta u + f(u, v), \\
\frac{\partial v}{\partial t} &= D_2 \Delta v + g(u, v), \\
\frac{\partial u}{\partial n} &= \frac{\partial v}{\partial n} = 0,
\end{align}

for some smooth functions $f$ and $g$ and diffusion coefficients $D_1 > 0$ and $D_2 > 0$. Then notice that $(u_*, v_*)$ is also a solution of the corresponding ordinary differential equation (ODE) system

\begin{align}
\frac{du}{dt} &= f(u, v), \\
\frac{dv}{dt} &= g(u, v).
\end{align}
Assume that \((u^*, v^*)\) is a linearly stable solution to (1–5), that is all eigenvalues of the Jacobian

\[
J = \begin{bmatrix}
\frac{\partial u}{\partial u} f(u^*, v^*) & \frac{\partial u}{\partial v} f(u^*, v^*) \\
\frac{\partial u}{\partial u} g(u^*, v^*) & \frac{\partial u}{\partial v} g(u^*, v^*)
\end{bmatrix}
\]

have strictly negative real part. Turing asks if it is possible that the same solution \((u^*, v^*)\) is linearly unstable in the setting of (1–4). The answer is yes, and it turns out that necessary conditions include \(d_1 \neq d_2\) and that the Jacobian matrix have one of the sign patterns

\[
\begin{bmatrix}
+ & - \\
+ & -
\end{bmatrix}, \quad \begin{bmatrix}
- & + \\
- & +
\end{bmatrix}, \quad \begin{bmatrix}
+ & + \\
- & -
\end{bmatrix}, \quad \text{or} \quad \begin{bmatrix}
- & - \\
+ & +
\end{bmatrix}.
\]

(1–6)

Our chemotactically induced stability results in Section 4.4 concerning the generalized model do not require restrictions on the diffusion coefficients. Although we do require some restrictions on the signs of the entries of a Jacobian matrix, these sign patterns are different from the ones in (1–6).

As mentioned previously, generalizations of the KS model for chemotaxis have been addressed in other recent publications [3, 9, 11, 19, 26, 43]. Most of these look at models with a specified small number of both species and chemotactic agents, for example [3, 9, 11, 26]. In [3], global existence of solutions for a two chemical version of the KS system is discussed, along with the existence of a Lyapunov functional. In [9] on the other hand, a parabolic-elliptic system with two chemotactic species but only one chemoattractant is studied. In [26] the multi-species concept is further generalized and a system with one chemoattractant but an arbitrary number of species is investigated. One additional chemical is introduced in [11] as equilibrium solutions to a two chemical, two species system are investigated. However none of the models mentioned above account for an arbitrary number of chemicals as our generalized model does.

In [19, 43] two models which do account for an arbitrary number of chemicals are introduced, and are both similar to the generalized model we introduce here. In addition
to an arbitrary number of chemical components, both models account for an arbitrary number of species which chemotactically react. However the model in [43] does not allow for nontrivial chemical reactions between the chemicals being modeled, which is one important component of our investigation. The model in [19] is quite general, and encompasses the setting of our generalized model. However after introducing the very general set up, the majority of the work in [19] focuses on specific examples which look at only one or two chemotactic agents and one or two species. Exceptions include [19, Section 5], where general conditions for the existence of a Lyapunov functional for the general model are shown, and [19, Section 6] where generalization of some results from [34] on the steady state problem are given. However [19] does not consider the generalization of the results on destabilization of homogeneous stationary solutions based on parameter ranges and the mean of the initial condition, which is the focus of one of our main results.
CHAPTER 2
CHEMOTAXIS ALONG A FIXED GRADIENT

2.1 An Example in One Dimension

In order to gain some insight into the phenomenon of chemotaxis and the mathematical modeling of it, we investigate a simple partial differential equation (PDE) similar to one derived in [25]. Let \( u(x, t) \) represent the density of a bacterial population at a given point \( x \in [0, 1] \) and time \( t \geq 0 \). It is assumed that the bacteria are attracted to areas of higher concentration of a certain chemical substance. The density of this substance at any point \( x \in [0, 1] \) is assumed to be independent of time and is denoted by \( s(x) \). As in the derivation of the original Keller-Segel (KS) model, it is assumed that the flux of the bacteria has two components, the first arising from random diffusive movement and the other from directed movement towards higher concentrations of the chemoattractant signal \( s(x) \). Given these assumptions, the flux will take the form

\[
J(x, t) = D \frac{\partial}{\partial x} u(x, t) - \chi u(x, t) \frac{d}{dx} s(x)
\]

where \( \chi > 0 \) is a constant representing the chemotactic sensitivity and \( D > 0 \) is the diffusion coefficient. Combining this flux term with the condition

\[
\int_S \frac{\partial u}{\partial t} + \nabla \cdot J \, dx = 0
\]

for every \( S \subset \Omega \) leads to the PDE

\[
\partial_t u = D \partial_{xx} u - \partial_x (\chi u s'(x)).
\]

By rescaling the time variable to allow \( D = 1 \) and applying no-flux boundary conditions, we arrive at the one-dimensional PDE

\[
\begin{align*}
\partial_t u &= \partial_{xx} u - \partial_x (\chi u(x,t)s'(x)) \quad &x \in (0, 1), \quad t > 0 \quad (2-1a) \\
\partial_x u - \chi u(x,t)s'(x) &= 0 \quad &x \in \{0, 1\}, \quad t > 0. \quad (2-1b)
\end{align*}
\]
2.1.1 Analytical Steady State Solutions

It is possible to analytically solve for steady states of (2–1). The steady state problem will be

\[ \partial_{xx} u - \partial_x (\chi u s'(x)) = 0, \]

and integrating gives \[ \partial_x u - \chi u s'(x) = k \]
for some constant \( k \). Applying the boundary conditions (2–1b) at \( x = 0 \) or \( x = 1 \) shows that for any continuous steady solution we must have \( k = 0 \), meaning that any steady state solution must satisfy

\[ \partial_x u - \chi u s'(x) = 0. \quad (2–2) \]

Since (2–2) is a linear equation in \( u \), it can always be solved analytically with the general solution being given by

\[ u(x) = C e^{\lambda x}, \]

a one dimensional set of solutions parametrized by the constant \( C \).

In order to be able to completely classify steady state solutions and their stability, we turn to the specific example where \( s(x) = ax \). We will call a stationary solution linearly stable if all eigenvalues of the linearized problem about the steady state have strictly negative real part. In this section we will show that all stationary solutions of (2–1) in the case where \( s(x) = ax \) are linearly stable with respect to perturbations \( U \) satisfying \( \int_0^1 U(x) dx = 0 \). To see the reason for the restriction on the perturbations, notice that so long as a solution \( u \) to (2–1) is sufficiently smooth we can use integration by parts and the boundary condition given by Equation (2–1a) to see

\[
\frac{\partial}{\partial t} \int_0^1 u(x, t) dx = \int_0^1 \partial_t u dx \\
= \int_0^1 \partial_x (\partial_x u - \chi u s') dx \\
= [\partial_x u - \chi u s']_0^1 \\
= 0.
\]
Since the total mass in the domain of any solution remains constant over time, to show stability we only consider perturbations $U$ away from the steady state which have zero mean and therefore do not change the mass of the solution.

We will denote the differential operator in (2–1a) with $s(x) = ax$ as

$$L := \frac{\partial^2}{\partial x^2} - \alpha \frac{\partial}{\partial x}$$

where $\alpha := a \chi$. Hence we have

$$L[u] = \partial_{xx} u - \alpha \partial_x u.$$

Then to determine stability of any steady state where $\partial_t u = L[u] = 0$, we need to solve the eigenvalue problem

$$L[u] = \lambda u.$$ \hspace{1cm} (2–4)

This becomes the ODE

$$\frac{d^2 u}{dx^2} - \alpha \frac{du}{dx} - \lambda u = 0.$$ \hspace{1cm} (2–5)

The characteristic equation for (2–5) is $r^2 - \alpha r - \lambda = 0$, which will have roots

$$r_{1,2} = \frac{\alpha \pm \sqrt{\alpha^2 + 4\lambda}}{2}.$$  

In order to determine the sign of $\lambda$ for possible solutions to (2–5), we first will show that all eigenvalues $\lambda$ must be real, and then will consider three cases based on the value $\alpha^2 + 4\lambda$, the discriminant of the characteristic equation.

We consider what happens when $\lambda$ has nonzero imaginary part. Since $\alpha^2 \neq 0$ is real and $\text{Im}(\lambda) \neq 0$, it follows that $\alpha^2 + 4\lambda \neq 0$. Hence $r_1$ and $r_2$ are distinct and solutions of (2–5) are of the form

$$U(x) = c_1 e^{r_1x} + c_2 e^{r_2x}$$

where $r_1$ and $r_2$ may be complex. Substituting this expression for $U$ and $s'(x) = a$ into the boundary condition (2–1b) and evaluating at $x = 0$ and $x = 1$ respectively gives the
system
\[ c_1(r_1 - \alpha) + c_2(r_2 - \alpha) = 0 \]  
\[ c_1 e^{r_1} (r_1 - \alpha) + c_2 e^{r_2} (r_2 - \alpha) = 0. \]

Writing this in matrix form, we see (2–6) becomes
\[ A\vec{c}: = \begin{bmatrix} r_1 - \alpha & r_2 - \alpha \\ e^{r_1} (r_1 - \alpha) & e^{r_2} (r_2 - \alpha) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \]

Hence a nonzero eigenfunction exists if and only if the matrix \( A \) is singular, so if and only if
\[ \det A = 0, \]

or
\[ (r_1 - \alpha)(r_2 - \alpha)(e^{r_2} - e^{r_1}) = 0. \] (2–7)

We consider the three cases where \( r_1 = \alpha, r_2 = \alpha, \) and \( e^{r_2} - e^{r_1} = 0. \) Consider first the case where \( r_1 = \alpha. \) Then since \( r_1 + r_2 = \alpha, r_2 = 0 \) and it follows from the definition of \( r_{1,2} \) that \( \lambda = 0, \) which is a contradiction to the fact that \( \lambda \) has nonzero imaginary part. A similar argument shows we cannot have \( r_2 = \alpha. \) It remains to consider the case where \( e^{r_1} = e^{r_2}. \) If we write the possibly complex roots as \( r_1 = a_1 + ib_1 \) and \( r_2 = a_2 + ib_2, \) then \( e^{r_1} = e^{r_2} \) holds if and only if
\[ a_1 - a_2 = 0 \text{ and } b_1 - b_2 = 2k\pi \]
for some integer \( k. \) Write the complex number \( \sqrt{\alpha^2 + 4\lambda} \) as
\[ \rho e^{i\theta} = \sqrt{\alpha^2 + 4\lambda} \]
for some real \( \rho \geq 0 \) and real \( \theta. \) Since \( \alpha^2 \) is real and \( \lambda \) has nonzero imaginary part, \( \sqrt{\alpha^2 + 4\lambda} \neq 0 \) and so in fact \( \rho > 0. \) Recalling the original definitions of \( r_{1,2}, \) we can write
\[ r_1 = \frac{1}{2} (\alpha + \rho e^{i\theta}) \text{ and } r_2 = \frac{1}{2} (\alpha - \rho e^{i\theta}). \]
Then recalling that $\alpha$ is real we can calculate
\[
0 = a_1 - a_2 = \text{Re}(r_1) - \text{Re}(r_2) = \frac{1}{2} \alpha + \frac{1}{2} \text{Re}(\rho e^{i\theta}) - \left( \frac{1}{2} \alpha - \frac{1}{2} \text{Re}(\rho e^{i\theta}) \right) = \text{Re}(\rho e^{i\theta}) = \rho \cos(\theta).
\]

Since $\rho > 0$, it must be that $\cos(\theta) = 0$ and $\theta = (2\ell + 1)\frac{\pi}{2}$ for some integer $\ell$. In other words, we know that $\rho e^{i\theta} = \sqrt{\alpha^2 + 4\lambda}$ is purely imaginary. However this implies
\[
\rho e^{i\theta} = \pm i\rho = \sqrt{\alpha^2 + 4\lambda}
\]
so that
\[
\rho^2 i^2 = -\rho^2 = \alpha^2 + 4\lambda.
\]
Recalling that $\rho$ and $\alpha$ are real numbers this implies $\text{Im}(\lambda) = 0$ which is a contradiction to the assumption that $\lambda$ had nonzero imaginary part. Hence (2–4) has only real eigenvalues.

Now that we have shown any eigenvalue $\lambda$ must be real, there are three possible cases which can be investigated separately. First we consider the case where $\alpha^2 + 4\lambda > 0$. In this case there are two distinct real roots $r_1$ and $r_2$, and the form of the solution, as in the case with $\lambda$ imaginary, is
\[
U(x) = c_1 e^{r_1 x} + c_2 e^{r_2 x}.
\]
Again applying the no-flux boundary condition, $c_1$ and $c_2$ must satisfy the system (2–7) in order for $U(x)$ to represent an eigenfunction. Again, this implies that either $r_1 = \alpha$, $r_2 = \alpha$, or $e^{r_1} = e^{r_2}$. Since $r_1$ and $r_2$ are real and distinct, we cannot have $e^{r_1} = e^{r_2}$. If $r_1 = \alpha$ then as in the previous case we conclude that $r_2 = 0$ and $\lambda = 0$. Similarly if $r_2 = \alpha$
then $r_1 = 0$ and $\lambda = 0$. We claim that $\lambda = 0$ violates the zero average condition. If this is the case, then we have shown no applicable eigenfunctions exist in the case where $\alpha^2 + 4\lambda > 0$. Indeed, note that if $\lambda = 0$ then either $r_1 = 0$ or $r_2 = 0$. If for example $r_1 = 0$, then Equation 2–6a and $r_2 = \alpha$ imply that $c_1 = 0$. However then

$$U(x) = c_2 e^{r_2 x},$$

so $U(x)$ does not have zero average unless $c_2 = 0$ as well. The same argument holds for $r_2 = 0$, and so no applicable eigenfunctions exist in this case.

The next case is $\alpha^2 + 4\lambda = 0$, which implies that the characteristic equation has one double root $r$ and that $U$ is of the form

$$U(x) = c_1 e^{rx} + c_2 xe^{rx}.$$

Application of the no-flux boundary condition at $x = 0$ and $x = 1$ for this form of $U$ leads to the system

$$c_1 r + c_2 - \alpha c_1 = 0 \quad (2–8a)$$
$$c_1 re^r + c_2 re^r + c_2 e^r - \alpha c_1 e^r - \alpha c_2 e^r = 0. \quad (2–8b)$$

Canceling $e^r \neq 0$ from Equation (2–8b) and then subtracting Equation (2–8a) gives

$$c_2 (r - \alpha) = 0.$$

Since $r = \alpha/2$ it cannot be that $r - \alpha = 0$, and so we conclude that $c_2 = 0$. Then (2–8a) becomes

$$c_1 (r - \alpha) = 0,$$

and we similarly conclude that $c_2 = 0$ and that there are no eigenfunctions in this case.
We finally turn to the case where $\alpha^2 + 4\lambda < 0$, which implies the solutions to the characteristic equation are two complex conjugate imaginary roots,

$$r_{1,2} = \frac{\alpha}{2} \pm i \frac{\sqrt{-\alpha^2 - 4\lambda}}{2}.$$ 

Letting $\gamma := \sqrt{-\alpha^2 - 4\lambda}$, $U$ is of the form

$$U(x) = e^{\frac{2\alpha}{x}} (c_1 \cos(\gamma x) + c_2 \sin(\gamma x)).$$

Applying the no-flux boundary condition at $x = 0$ and $x = 1$ respectively gives

\begin{align*}
\frac{c_2 \gamma}{\alpha} + \frac{\alpha}{2} c_1 - \alpha c_1 &= 0 \quad (2\text{–}9\text{a}) \\
\cos(\gamma) \left( c_2 \gamma + \frac{\alpha}{2} c_1 - \alpha c_1 \right) + \sin(\gamma) \left( -c_1 \gamma + \frac{\alpha}{2} c_2 - \alpha c_2 \right) &= 0. \quad (2\text{–}9\text{b})
\end{align*}

Substituting Equation (2–9a) into (2–9b) and simplifying gives

$$\sin(\gamma) \left( -c_1 \gamma - \frac{\alpha}{2} c_2 \right) = 0.$$ 

If $(-c_1 \gamma - \frac{\alpha}{2} c_2) = 0$ then together with Equation (2–9a) this leads to the linear system

$$\begin{bmatrix}
-\gamma & -\frac{\alpha}{2} \\
-\frac{\alpha}{2} & \gamma
\end{bmatrix}
\begin{bmatrix}
c_1 \\
c_2
\end{bmatrix}
= \begin{bmatrix}
0 \\
0
\end{bmatrix}$$

which has determinant $\frac{\alpha^2}{4} + \gamma^2$. However, $\frac{\alpha^2}{4} + \gamma^2 = -\lambda > 0$ in this case, so the system has no nonzero solutions for $c_1$ and $c_2$.

On the other hand if $\sin(\gamma) = 0$ then the boundary conditions are also satisfied. So, values of $\gamma = k\pi$ for some integer $k$ produce eigenvalues of the form

$$\lambda_k = -\frac{\alpha^2}{4} - k^2 \pi^2.$$ 

We have now shown all eigenvalues of $L$ are negative except for $\lambda = 0$, and so any stationary solution of (2–1) is linearly stable with respect to perturbations $U(x)$ satisfying $\int_0^1 U(x) \, dx = 0$. 

23
2.1.2 Numerical Visualization

Using MATLAB’s PDE solver pdepe, we can demonstrate that solutions quickly converge to a steady state which mimics the shape of the signal \( s(x) \) for a variety of different signals. Figure 2-1 shows the case where \( s(x) = \sin(2x) \) and the initial condition is the constant \( u(0) \equiv 1 \).

2.2 The Multidimensional Case

2.2.1 Existence of Classical Solutions

Next we consider the multidimensional version of (2–1) on some domain \( \Omega \subset \mathbb{R}^N \).

The PDE is given by

\[
\begin{align*}
\partial_t u &= \Delta u - \nabla \cdot (\chi u \nabla s(x)), & x \in \Omega, & t \in [0, T] \\
0 &= \frac{\partial u}{\partial n} - \chi u \frac{\partial s}{\partial n}, & x \in \partial \Omega, & t \in [0, T]
\end{align*}
\]

(2–10a) (2–10b)

where \( u : \mathbb{R}^N \rightarrow \mathbb{R} \) and \( s : \mathbb{R}^N \rightarrow \mathbb{R} \). If \( s \) and \( \Omega \) are sufficiently smooth, then standard theory of parabolic PDE says that a unique solution will exist for every sufficiently smooth initial condition so long as that initial condition vanishes in some \( \Omega \)-neighborhood of \( \partial \Omega \) [12, Theorem 5.3.2]. A similar result for elliptic PDE holds, which implies the existence of steady state solutions to (2–10).
2.2.2 Properties of Numerical Solutions

In dimensions higher than one it is not as trivial to find analytic solutions to (2–10) in either the time dependent or independent case. For this reason we now turn to finite element analysis to find solutions. In this section, we develop the general framework that we will use throughout to calculate solutions numerically. Consider a bounded domain $\Omega$ with Lipschitz continuous boundary $\partial \Omega$. Multiplying through by a test function $\varphi \in H^1(\Omega)$, integrating by parts, and using the boundary conditions (2–10b) gives

$$
\int_{\Omega} \left( \partial_t u \right) \varphi = - \int_{\Omega} \nabla u \cdot \nabla \varphi + \int_{\Omega} \chi u \nabla s \cdot \nabla \varphi,
$$

(2–11)

the weak formulation of (2–10). Using the notation explained in the introduction, we begin by formulating the problem using the standard finite element method in the piecewise linear finite element space $S_h$ with basis functions $\{ \phi_\ell \}_{\ell=1}^{L'}$. We first discretize only in space, and for any fixed time $t$ write the approximating function $u_h \in S_h$ of $u$ as

$$
u(x,t) \approx u_h(x,t) = \sum_{\ell=1}^{L'} U_\ell(t) \phi_\ell(x).$$

In order to give a matrix formulation of the problem, define the bilinear forms

$$a(u,v) = - \int_{\Omega} \nabla u \cdot \nabla v + \int_{\Omega} \chi u \nabla s \cdot \nabla v$$

and

$$b(u,v) = \int_{\Omega} uv.$$ 

for any $u, v \in H^1(\Omega)$. For basis functions $\phi_\ell, \phi_m \in S_h$ define the matrices $A$ and $B$ by

$$A_{\ell m} = a(\phi_m, \phi_\ell)$$

and

$$B_{\ell m} = b(\phi_m, \phi_\ell).$$
Then substitute \( u \approx u_h = \sum \ell U(t) \phi_\ell(x) \) and \( \varphi = \phi_m \) for each basis function \( \phi_m \) of \( S_h \) into (2–11) to obtain the spatially discretized weak formulation of (2–10). In matrix form this becomes the system of ordinary differential equations

\[
B \frac{d\vec{U}(t)}{dt} = A\vec{U}(t). \tag{2–12}
\]

Given initial conditions \( U_0(t) \), system (2–12) can be solved exactly since the matrices \( A \) and \( B \) are independent of \( t \) [6]. Hence piecewise linear approximating solutions are given in terms of the matrix exponential by

\[
\vec{U}(x,t) = U_0(x)e^{tB^{-1}L}. \tag{2–13}
\]

Note that \( B \) is invertible, as are all Gram matrices of inner products of a basis. Indeed, let \( y \) be in the nullspace of \( B \) and let \( \eta(x) := \sum \ell y_\ell \phi_\ell \) be the function in \( S_h \) represented by \( x \) under the current choice of basis. Then

\[
\|\eta\|^2 = < \eta, \eta > = < \sum \ell y_\ell \phi_\ell, \sum m y_m \phi_m > = \sum \ell y_\ell \sum m y_m < \phi_\ell, \phi_m > = 0
\]

where the last line follows from the \( \ell \)-th row of the expansion of \( By = 0 \). But \( \|\eta\|^2 = 0 \) if and only if \( \eta = 0 \), which implies that its vector representation \( y = 0 \), and so the nullspace of \( B \) consists of only the zero vector and \( B \) is invertible.

Once we turn to the full Keller-Segel model, the matrix \( A \) from Equation (2–12) will no longer be independent of \( t \) and so an explicit formula such as Equation (2–13) is not in general possible to find [6]. For this reason, we continue the numerical analysis and discretize (2–12) in time as well as in space. Adding indices to our notation to account for time, we write the expansion of the approximating function of \( u^*_h \in \mathbb{S}_h \) after \( n \) steps at
time $t_n = n \Delta t$ as

$$u(x, t_n) \approx u_h^n = \sum_{\ell=1}^{L'} U^n_{\ell} \phi_{\ell}.$$ 

Then approximating the time derivative by

$$\frac{du_h}{dt} \approx \frac{u_{h}^{n+1} - u_{h}^{n}}{\Delta t},$$

and using an implicit scheme, we obtain the formula

$$\vec{U}^{n+1} = (B - \Delta tA)^{-1} B\vec{U}^{n} = \vec{b}$$

to find a solution after $n + 1$ time steps when the solution after $n$ time steps is known and whenever $B - \Delta tA$ is invertible. In general it is not the case that $B - \Delta tA$ is invertible, however it can be shown in some cases under additional restrictions on the boundary conditions and, for example, $\Delta s$. Some results for the simple Keller-Segel model using a standard finite element method similar to the one just derived will be shown in Section 3.3.

### 2.2.3 Numerical Difficulties and Method Validation

It is well known that solutions to convection-diffusion problems similar to (2–10), especially those for which the convective term is relatively large, can show numerical instabilities using the standard finite element method (FEM) [20, 44]. In this section we demonstrate these potential numerical problems and at the same time validate the accuracy of the matrices used in our standard FEM schemes. Consider the convection diffusion PDE

$$D\Delta u + \beta \cdot \nabla u = f(x, y), \quad (x, y) \in \Omega \quad (2–14a)$$

$$\frac{\partial u}{\partial n} = 0, \quad (x, y) \in \partial \Omega \quad (2–14b)$$
which is motivated by an example in [20]. We will solve 2–14 for \( \Omega \in \mathbb{R}^2 \) using the standard finite element method. The weak formulation of (2–14) is

\[
- D \int_{\Omega} \nabla u \cdot \nabla \varphi + \int_{\Omega} (\beta \cdot \nabla u) \varphi = \int_{\Omega} f \varphi
\]

(2–15)

for all \( \varphi \in H^1(\Omega) \). Substituting a basis function \( \phi_\ell \) of \( S_h \) for \( \varphi \) in (2–15) as well as the FEM expansion of \( u \), we obtain

\[
\sum_{j=1}^{N} \left( -D \int_{\Omega} \nabla \phi_j \cdot \nabla \phi_i + \int_{\Omega} (\beta \cdot \nabla \phi_j) \phi_i \right) U_j = \int_{\Omega} f \phi_\ell, \quad 1 \leq i \leq N.
\]

(2–16)

Then let the stiffness matrix \( L \) be defined by

\[
L_{ij} = \int_{\Omega} \nabla \phi_j \cdot \nabla \phi_i,
\]

define the matrix \( F \) by

\[
F_{ij} = \int_{\Omega} (\beta \cdot \nabla \phi_j) \phi_i,
\]

and finally define the vector \( b \) by

\[
b_i = \int_{\Omega} f(x, y) \phi_i.
\]

Using this notation the finite element matrix approximation of Equation (2–14) is

\[
- DL\vec{U} + F\vec{U} = \vec{b}.
\]

(2–17)

Upon inspection it is easy to see that for any constant function \( u \), the left hand side of Equation 2–15 is zero. This implies that the left hand side of Equation (2–17) is also zero so if \( \vec{U} \) is equal to the vector of all ones or any constant multiple of it, then \( \vec{U} \) is in the nullspace of \( -DL + F \). Hence, as explained in the previous section, we add the restriction that any constant solution must be zero, which is equivalent to setting the value at one node equation equal to zero, or eliminating the corresponding row and column from \( DL + F \). Eliminating the first node from the matrix, we find numerically that
this new reduced matrix is invertible and may solve for a solution $\vec{U}$ using this method which is accurate up to addition of a constant value.

We test the finite element matrices from Equation 2–17 by solving the system for a function $f(x, y)$ for which (2–14) has a known solution. Consider the function $f$ defined by

$$f(x, y) = -\frac{2ae^{a(2x-1)}(3e^{a(2x-1)} - 1)}{(1 + e^{a(2x-1)})^3}$$

and the function $u$ defined by

$$u(x, y) = \frac{1}{1 + e^{a(2x-1)}}$$

on the domain $\Omega = [0, 1] \times [0, 1]$. Direct computation verifies that for the given $f$, $u$ solves the PDE (2–14a) for $D = \frac{1}{a}$ and, for example, $\beta = (1, 0)$. Clearly $\partial u / \partial y = 0$. In fact since

$$\frac{\partial u}{\partial x} = \frac{-2ae^{a(2x-1)}}{(1 + 2ae^{a(2x-1)})^2}$$

we see that

$$\left. \frac{\partial u}{\partial x} \right|_{x=0} = \frac{-2a}{e^a(2e^{-2a} + 1)}$$

and

$$\left. \frac{\partial u}{\partial x} \right|_{x=1} = \frac{-2ae^a}{e^{2a} + 1}$$

so that clearly

$$\lim_{a \to \infty} \left. \frac{\partial u}{\partial x} \right|_{x=0} = \lim_{a \to \infty} \left. \frac{\partial u}{\partial x} \right|_{x=1} = 0.$$}

Hence for large enough values of $a$, $u$ has approximately Neumann boundary conditions.

Using the standard finite element matrix formulation to solve for $u$, we see good convergence to the solution as the mesh size gets very fine, but for a rough mesh the solution shows oscillations near the steep gradient seen in the exact solution in Figure 2-3 around $x = 0.5$. The FEM solutions for two different mesh sizes can be seen in Figure 2-2. For comparison, the exact solution $u$ to (2–14) is shown in Figure 2-3.

The oscillatory behavior seen when using the standard FEM to solve this test problem is
Figure 2-2. Standard FEM solutions to (2–14)

part of the motivation for a nonstandard finite element method which we will introduce in Section 5.3.
Figure 2-3. Exact solution to (2–14) on refined mesh
CHAPTER 3
A SIMPLE KELLER-SEGEL MODEL

3.1 Positivity of Exact Solutions

In this section we will investigate some properties of the two equation Keller-Segel (KS) model derived in Chapter 1. For ease of reference, we restate the system here.

\[ \begin{align*}
\partial_t u &= D \nabla \cdot (\nabla u - \chi u \nabla v), \quad x \in \Omega, \quad t > 0 \\
\partial_t v &= \tilde{D} \Delta v - \gamma v + \alpha u, \quad x \in \Omega, \quad t > 0 \\
\frac{\partial u}{\partial n} = \frac{\partial v}{\partial n} = 0, \quad x \in \partial \Omega, \quad t > 0 \\
u(x, 0) &= u_0, \quad v(x, 0) = v_0 \quad x \in \Omega
\end{align*} \] (3–1)

In (3–1d), \( u_0 \) and \( v_0 \) are non-negative smooth functions on \( \Omega \) which are assumed to not be identically zero. For the remainder of the chapter, let \( \Omega \) be a bounded, connected open subset of \( \mathbb{R}^n \) with smooth boundary \( \partial \Omega \). For any fixed \( T > 0 \), we let \( \Omega_T := \Omega \times (0, T] \) be the parabolic cylinder.

Strict positivity of exact solutions is considered standard, but as it is difficult to find a careful, complete proof, we include the details for completeness. The main result on positivity is Theorem 3.3, but in order to prove it we first need some preliminary theorem statements and arguments. We begin with statements of Hopf’s Lemma and a strong maximum principle for parabolic partial differential equations (PDE). Throughout this chapter we assume that \( L \) is a uniformly elliptic differential operator of the form

\[ Lu = -\sum_{i,j} a^{ij}(x, t) \partial_{x_i x_j} u(x, t) + \sum_{i} b^i(x, t) \partial_{x_i} u(x, t) + c(x, t) u(x, t) \] (3–2)

with \( a^{ij}, b^i \) and \( c \) all continuous and bounded. Part (i) of the following version of Hopf’s Lemma for parabolic PDE is stated and proved in [27], and part (ii) follows by applying part (i) to \(-z\). Theorem 3.1 is a version of the strong maximum principle from Section 7.1 in [10].
Lemma 3.1. Let $\eta > 0$ and $R > 0$ be some fixed constants, and fix $(y, s) \in \mathbb{R}^{N+1}$.
Suppose the coefficients $a^{ij}, b^j$ and $c$ are all bounded and that the operator $\partial_t + L$ is
uniformly parabolic in the lower parabolic frustum

$$PF = \{(x, t) \in \mathbb{R}^{n+1} : |x - y|^2 + \eta^2(s - t) < R^2, t < s\}.$$ 

Suppose also that the derivatives $\partial_{x_1}z, \partial_{x_1,x_j}z$ and $\partial_t z$ exist and are continuous and let
$(x_1, s) \in \mathbb{R}^{N+1}$ with $|x_1 - y| = R$.

(i) Suppose $\partial_t z + Lz \leq 0$ on $PF$ and $x_1$ satisfies

$$z(x_1, s) \geq z(x, t), \quad (x, t) \in PF \quad (3-3a)$$

$$z(x_1, s) > z(x, t), \quad (x, t) \in PF \quad \text{with} \quad |x - y| \leq R/2 \quad (3-3b)$$

$$c(x, t)z(x_1, s) \geq 0, \quad (x, t) \in PF. \quad (3-3c)$$

If $n = (x_1 - y)/|x_1 - y|$ is the outward unit normal of $PF$ at $x_1$, then

$$\frac{\partial z}{\partial n} > 0 \quad (3-4)$$

and $z < z(x_1, s)$ in $PF$.

(ii) If on the other hand $\partial_t z + Lz \geq 0$ on $PF$ and $x_1$ satisfies

$$z(x_1, s) \leq z(x, t), \quad (x, t) \in PF \quad (3-5a)$$

$$z(x_1, s) < z(x, t), \quad (x, t) \in PF \quad \text{with} \quad |x - y| \leq R/2 \quad (3-5b)$$

$$c(x, t)z(x_1, s) \leq 0, \quad (x, t) \in PF. \quad (3-5c)$$

If $n = (x_1 - y)/|x_1 - y|$ is the outward unit normal of $PF$ at $x_1$, then

$$\frac{\partial z}{\partial n} < 0 \quad (3-6)$$

and $z > z(x_1, s)$ in $PF$.

Theorem 3.1. (Strong Maximum Principle for $c \geq 0$) Assume the function $u(x, t)$ is
continuous on $\overline{\Omega}_T$ and that the derivatives $\partial_{x_1}u, \partial_{x_1,x_j}u$ and $\partial_t u$ exist and are continuous
on $\Omega_T$. Suppose also that $\partial_t + L$ is uniformly parabolic with $c \geq 0$ in $\Omega_T$ and that $\Omega$ is connected.
(i) If \( \partial_t u + Lu \leq 0 \) in \( \Omega_T \) and \( u \) attains a nonnegative maximum over \( \bar{\Omega}_T \) at point \((x_0, t_0)\) \( \in \Omega_T \) then \( u \equiv \text{constant on } \Omega_{t_0} \).

(ii) Similarly if \( \partial_t u + Lu \geq 0 \) in \( \Omega_T \) and \( u \) attains a nonpositive minimum over \( \bar{\Omega}_T \) at point \((x_0, t_0)\) \( \in \Omega_T \) then \( u \equiv \text{constant on } \Omega_{t_0} \).

The following lemma is in Chapter 7, corollary 2.3 in [36], but here we significantly simplify the proof from [36] by using Lemma 3.1.

**Lemma 3.2.** Let \( \frac{\partial}{\partial t} + L \) be a uniformly parabolic operator with \( c \geq 0 \). If \( v \) is continuous on \( \bar{\Omega} \times [0,T] \) and satisfies

\[
\begin{align*}
\partial_t v + Lv & \geq 0 \quad \text{in} \quad \Omega \times (0,T) \\
\beta(x)v + \frac{\partial v}{\partial n} & = 0 \quad \text{on} \quad \partial \Omega \times (0,T) \\
v & = g \quad \text{on} \quad \Omega \times \{t = 0\}
\end{align*}
\]

where \( g \geq 0 \) but \( g \) not identically zero and \( \beta(x) \geq 0 \) is continuous on \( \Omega \), then \( v(x,t) > 0 \) for all \((x,t) \in \bar{\Omega} \times (0,T)\).

**Proof.** We first show that \( v \geq 0 \) on \( \bar{\Omega} \times (0,T) \). By way of contradiction, assume not. Then there exists some point \((x_0, t_0) \in \bar{\Omega} \times (0,T)\) such that \( v(x_0, t_0) < 0 \). Then \( \min_{\Omega \times [0,T]} v < 0 \) and must be achieved at some point \((\bar{x}, \bar{t}) \in \bar{\Omega} \times [0,T] \). We must have \( \bar{t} > 0 \) because of the initial conditions \( g \geq 0 \). Now, if \( \bar{x} \in \Omega \) with \( 0 < \bar{t} \leq T \) then by the strong maximum principle \( v \equiv \text{constant} < 0 \) on \( \Omega \times (0, \bar{t}) \) which is a contradiction to \( v = g \geq 0 \) at \( t = 0 \) and \( v \) continuous on \( \bar{\Omega} \times [0,T] \).

Next we find a contradiction for the case where \( \bar{x} \in \partial \Omega \). We wish to use part (ii) of Hopf’s Lemma (Lemma 3.1) and so must verify the hypotheses of (3–5). Clearly (3–5a) holds since \((\bar{x}, \bar{t})\) is a minimum. If \( v(\bar{x}, \bar{t}) = \min_{\Omega \times [0,T]} v \) is achieved anywhere in the interior \( \Omega \) then the argument in the previous paragraph applies, so without loss of generality we may assume that the minimum is achieved only on the boundary. Hence (3–5b) holds. Since \( c \geq 0 \) and \( v(\bar{x}, \bar{t}) < 0 \), (3–5c) holds as well, so we may apply part (ii)
of Lemma 3.1 to conclude \( \frac{\partial v(\bar{x}, \bar{t})}{\partial n} < 0 \). But then because \( \beta \geq 0 \) and \( v(\bar{x}, \bar{t}) < 0 \) we have

\[
\beta(\bar{x})v(\bar{x}, \bar{t}) + \frac{\partial v(\bar{x}, \bar{t})}{\partial n} < 0
\]

which is a contradiction to the fact that \( v \) satisfies (3–8). We have now established a contradiction in all cases, and so we have shown \( v(x, t) \geq 0 \) for all \( (x, t) \in \bar{\Omega} \times [0, T] \).

It only remains to show \( v > 0 \) on \( \bar{\Omega} \times (0, T] \). By way of contradiction, again suppose not. Since we have already shown \( v \geq 0 \), then there must exist some point \( (x_0, t_0) \in \bar{\Omega} \times (0, T] \) such that \( v(x_0, t_0) = 0 \). Then the minimum over \( \bar{\Omega} \times [0, T] \) is achieved at \( (x_0, t_0) \), so if \( x_0 \in \Omega \) the strong maximum principle (Theorem 3.1) implies \( v \equiv 0 \) on \( \bar{\Omega} \times [0, T] \) which is again a contradiction to \( v(x, 0) = g(x) \) and the assumption that \( g \) is not identically zero. On the other hand if \( x_0 \in \partial \Omega \) then as in the previous argument to show \( v \geq 0 \), we can assume that the minimum is attained only on the boundary of \( \Omega \) and apply Lemma 3.1 to obtain a contradiction to the boundary conditions (3–7). \( \square \)

Next we will prove a lemma which is similar to problem 7.7 in [10].

**Lemma 3.3.** Suppose \( u \) is a smooth solution of

\[
\begin{align*}
\partial_t u + Lu &= 0 & \text{in } \Omega \times (0, T] \tag{3–10a} \\
\frac{\partial u}{\partial n} &= 0 & \text{on } \partial \Omega \times (0, T] \tag{3–10b} \\
u &= g & \text{on } \Omega \times \{t = 0\}. \tag{3–10c}
\end{align*}
\]

where \( \partial_t + L \) is a uniformly parabolic operator. If \( g \geq 0 \), \( g \) is not identically zero, and the function \( c \) is bounded below, then \( u > 0 \) on \( \bar{\Omega} \times (0, T] \).

**Remark 3.2.** For our application we only need to see that this Lemma 3.3 holds for solutions on \( \Omega \times [0, T] \) for \( T \) finite and \( \Omega \) bounded, in which case the boundedness of \( c \) is clear from the continuity assumption on all coefficients of the operator. However the proof will hold exactly as written for a solution \( u \) for (3–10) on \( \Omega \times [0, \infty) \) if \( c \) is bounded.
Proof. Let \( m = \inf_{\Omega \times [0,T]} c(x,t) > -\infty \) since \( c \) is assumed to be bounded below. Define \( v(x,t) := e^{mt}u(x,t) \). Then notice that for any partial derivatives, we have

\[
\partial_{x_i} v = e^{mt} \partial_{x_i} u \quad (3\text{-}11)
\]

and

\[
\partial_{x_i x_j} v = e^{mt} \partial_{x_i x_j} u. \quad (3\text{-}12)
\]

Then (3–11) and (3–12) imply that for our operator \( L \) we have \( Lv = e^{mt}Lu \), or rewriting that

\[
Lu = e^{-mt}Lv. \quad (3\text{-}13)
\]

Then using (3–10a), (3–13) and the definition of \( v \) we calculate

\[
\partial_t v(x,t) = e^{mt} u_t(x,t) + me^{mt}u(x,t)
\]

\[
= -e^{mt} Lu(x,t) + mv(x,t)
\]

\[
= -e^{mt} e^{-mt}Lv(x,t) + mv(x,t)
\]

to see that \( v \) satisfies the PDE \( \partial_t v + L v - mv = 0 \). Notice also that by the boundary and initial conditions for \( u \) and the definition of \( v \), we have that \( v \) satisfies (3–10b) on \( \partial\Omega \times [0,T] \) and \( v = g \) on \( \Omega \times \{ t = 0 \} \). Defining a new operator \( K v := L v - mv \), we see that \( \partial_t + K \) is also a uniformly parabolic operator and the “c” term of this operator is \( c(x,t) - m \geq 0 \) by the definition of \( m \). Then we can apply Lemma 3.2 (with \( \alpha \equiv 0 \)) to say that \( v > 0 \) on \( \overline{\Omega} \times (0,T] \). Then we have that \( u(x,t) = e^{-mt} v(x,t) > 0 \) as well.

We now have the tools to prove strict positivity of exact solutions to the minimal KS system.

**Theorem 3.3.** Let \((u,v)\) be a smooth solution to (3–1) on the interval \([0,T]\) with \(u_0,v_0\) not identically zero, \( \alpha \geq 0 \), and \( \gamma \geq 0 \). Then \( u(x,t) > 0 \) and \( v(x,t) > 0 \) on \( \overline{\Omega} \times (0,T] \).
Proof. We first look at the equation for $u$. From (3–1a) and the product rule we have

$$\partial_t u - \Delta u + \chi \nabla u \cdot \nabla v + \chi u \Delta v = 0.$$ 

Then for $Lu := -\Delta u + \chi \nabla u \cdot \nabla v + \chi u \Delta v, \frac{\partial}{\partial t} + L$ is a uniformly parabolic operator with smooth $a^{ij}, b^i$ and $c$, and since $c = \chi \Delta v$ is a continuous function over the compact set $\bar{\Omega} \times [0, T]$, it is bounded below. Then since $\partial_t u + Lu = 0$ on $\Omega \times (0, T]$ we can apply Lemma 3.3 to say that $u > 0$ on $\bar{\Omega} \times (0, T]$. Next look at $v$. Equation (3–1b) implies

$$\partial_t v - \Delta v + \gamma v = \alpha u.$$ 

Again defining a differential operator based on this, if $Lv := -\Delta v + \gamma v$ then $\frac{\partial}{\partial t} + L$ is a uniformly parabolic operator. We just showed $u \geq 0$, so since $\gamma \geq 0$ and $\alpha \geq 0$, we have that $v_t + Lv = \alpha u \geq 0$. We also have $c = \gamma \geq 0$ so that Lemma 3.2 with $\beta = 0$ may be applied to conclude $v > 0$ on $\bar{\Omega} \times (0, T]$ as well.

3.2 Preliminaries for Numerical Steady State Solutions

Our first numerical experiments are carried out on (3–1) with all constants and the linear sensitivity function assumed to be one. This reduces the time dependent problem to a system from [18, 33] which is

$$\begin{align*}
\partial_t u &= \nabla \cdot (\nabla u - u \nabla v), & x \in \Omega, \ t > 0 \quad (3–14a) \\
\partial_t v &= \Delta v - v + u, & x \in \Omega, \ t > 0 \quad (3–14b) \\
0 &= \frac{\partial u}{\partial n} = \frac{\partial v}{\partial n}, & x \in \partial \Omega, \ t > 0. \quad (3–14c)
\end{align*}$$

The steady state problem then becomes

$$\begin{align*}
0 &= \nabla \cdot (\nabla u - u \nabla v), & x \in \Omega, \ t > 0 \quad (3–15a) \\
0 &= \Delta v - v + u, & x \in \Omega, \ t > 0 \quad (3–15b) \\
0 &= \frac{\partial u}{\partial n} = \frac{\partial v}{\partial n}, & x \in \partial \Omega, \ t > 0. \quad (3–15c)
\end{align*}$$
In order to compute solutions to (3–15), we will first reduce the system to one partial differential equation which is equivalent to the original system. This approach is summarized in [18], and we fill in the details here for completeness. We first need statements of Hopf’s Lemma and the strong maximum principle for elliptic PDE. The following two theorem statements are the elliptic analogues of Lemma 3.1 and Theorem 3.1, and with the exception of part (ii) of Lemma 3.4, are from [10]. Part (ii) of Lemma 3.4 follows from applying part (i) to the function $-u$.

**Lemma 3.4.** Assume $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega})$ and the coefficient $c$ from Equation (3–2) satisfies $c = 0$ in $\Omega$. Suppose also that $\Omega$ is open, connected, and bounded.

(i) Suppose $Lu \leq 0$ in $\Omega$ and there exists a point $x_0 \in \partial \Omega$ such that $u(x_0) > u(x)$ for all $x \in U$.

Finally, assume $\Omega$ satisfies the interior ball condition at $x_0$, meaning there exists an open ball $B \subset \Omega$ with $x_0 \in \partial B$. Then

$$\frac{\partial u}{\partial n}(x_0) > 0,$$

where $n$ denotes the outer unit normal to $B$ at $x_0$.

(ii) Similarly suppose $Lu \geq 0$ in $\Omega$ and there exists a point $x_0 \in \partial \Omega$ such that $u(x_0) < u(x)$ for all $x \in U$.

If $\Omega$ satisfies the interior ball condition at $x_0$, then

$$\frac{\partial u}{\partial n}(x_0) < 0.$$

**Theorem 3.4.** Assume $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega})$ and the function $c$ from (3–2) satisfies $c \geq 0$ in $\Omega$. Suppose also that $\Omega$ is connected.

(i) Suppose $Lu \leq 0$ in $\Omega$ and $u$ attains a nonnegative maximum over $\overline{\Omega}$ at an interior point. Then $u$ is constant within $\Omega$.

(ii) Similarly, if $Lu \leq 0$ in $\Omega$ and $u$ attains a nonpositive minimum over $\overline{\Omega}$ at an interior point, then $u$ is constant within $\Omega$. 

38
Theorem 3.5. Let $\Omega$ be bounded and $\partial \Omega$ be $C^2$. Assume $u, v \in C^2(\Omega) \cap C^1(\bar{\Omega})$ and are finite everywhere. Then $(u, v)$ satisfies (3–15) if and only if $v$ satisfies

\begin{align*}
0 &= \Delta v + ke^v - v, \quad x \in \Omega \\
\frac{\partial v}{\partial n} &= 0, \quad x \in \partial \Omega, \quad t > 0
\end{align*}

(3–16a) (3–16b)

and $u = ke^v$ for some constant $k$.

Proof. For the forward implication, assume that $u, v \in C^2(\Omega) \cap C^1(\bar{\Omega})$ satisfy (3–15). Consider a new function $\psi(x)$ defined by $\psi(x) = u(x)e^{-v(x)}$. Notice that $\psi$ also has Neumann boundary conditions. Then since $u$ and $v$ are both finite everywhere, we see $u = \psi e^v$. Substituting this into (3–15a), applying the product rule twice and canceling out the non-zero function $e^v$ then gives

$$\Delta \psi + \nabla v \cdot \nabla \psi = 0.$$ 

Consider the elliptic operator $L$ defined by $L\psi := \Delta \psi + \nabla v \cdot \nabla u$, so that coefficients $b^j$ of the first derivative part of the operator are given by the first order partials of the known function $v$. Then since $v \in C^2(\Omega) \cap C^1(\bar{\Omega})$, the coefficients of $L$ are bounded by $\Omega$ bounded and are also continuous. Since $L\psi \equiv 0$ in $\Omega$, we may apply the strong maximum principle of Theorem 3.4 and Hopf’s Lemma 3.4. So, either $\psi$ attains its maximum over $\bar{\Omega}$ at an interior point, which by the strong maximum principle implies $\psi \equiv k$ for some constant $k$, or $\psi$ achieves it maximum only on the boundary of the domain which implies by Hopf’s Lemma that $\frac{\partial \psi}{\partial n} > 0$. However this last inequality is a contradiction to the Neumann Boundary conditions of $\psi$. Thus it must be that $\psi \equiv k$ in $\bar{\Omega}$, and so $u = ke^v$. We now substitute $u = ke^v$ into (3–15b) to obtain (3–16a). For the reverse direction, we notice that if $v$ is a solution of (3–16) and we define $u = ke^v$, then Neumann boundary conditions for $u$ follow from the boundary conditions on $v$. Substituting $u$ and $v$ into (3–15a)-(3–15b), the equations hold. \qed
Notice that for any fixed constant $u_*>0$, the system (3–15) has exactly one spatially constant solution $(u_*, v_*)$ with $v_* = u_*$. As expected, the same result follows if one considers a homogeneous solution $v_*$ to (3–16). Then (3–16a) reduces to $v_* = ke^{v_*}$, and since $u_* = ke^{v_*}$ is an assumed, we see that indeed $u_* = v_*$ is the only stationary solution, and additionally we can solve to see that $k = \frac{v_*}{e^{v_*}}$. The stability of such homogeneous stationary solutions to Equation 3–14 is fully characterized as a specific example of a more general case analyzed in [34]. The stability result uses a technique which we will explain and generalize in Section 4.4. The technique involves $\mu_1 < 0$, the nonzero eigenvalue of the Laplacian on $\Omega$ with Neumann boundary conditions which is closest to zero. Then it is shown in [34] that if $u_* < 1 - \mu_1$ the homogeneous stationary solution $(u_*, v_*)$ is linearly stable and if $u_* > 1 - \mu_1$ it is linearly unstable. As that analysis fully characterizes constant solutions to (3–15), we turn to numerical methods to search for nonconstant solutions. In order to find such solutions to (3–16), we first prove the following lemma which is explained in [18]. From here on we denote the mean of a function over the domain $\Omega$ by $m(z) := \frac{1}{|\Omega|} \int_\Omega z(x)dx$.

**Lemma 3.5.** Let $v \in C^2(\Omega)$. Then $v$ is a solution to (3–16) with $m(v) = \beta$ if and only if $z = v - \beta$ is a solution of

\[
\Delta z - z + \beta \left( \frac{e^z}{m(e^z)} - 1 \right) = 0, \quad x \in \Omega \tag{3–17a}
\]

\[
\frac{\partial z}{\partial n} = 0, \quad x \in \partial \Omega \tag{3–17b}
\]

with $m(z) = 0$.

**Proof.** We first assume that $v \in C^2(\Omega)$ is a solution to (3–16). Then define $z = v - m(v)$. Clearly $m(z) = 0$, and we claim that $z$ solves (3–17). It is easy to see that (3–17b) holds by the Neumann boundary conditions on $v$. To see that (3–17a) holds, substitute $v = z + m(v)$ into (3–16a) to obtain

\[
\Delta z - (z + m(v)) + ke^{z + m(v)} = 0. \tag{3–18}
\]
Integrating both sides of (3–18) over $\Omega$ and using integration by parts and Neumann boundary conditions, along with the fact that $\int_{\Omega} z = 0$ gives

$$m(v) = ke^{m(v)} m(e^z). \quad (3–19)$$

Rewriting (3–18) using (3–19) then gives

$$0 = \Delta z - z - ke^{m(v)} m(e^z) + ke^{m(v)} e^z$$

$$= \Delta z - z + ke^{m(v)} m(e^z) \left( \frac{e^z}{m(e^z)} - 1 \right)$$

$$= \Delta z - z + m(v) \left( \frac{e^z}{m(e^z)} - 1 \right)$$

which is (3–17a) with $\beta = m(v)$.

For the reverse direction, assume that $z$ solves (3–17) with $m(z) = 0$. Let $v := z + \beta$ and notice that $v$ still has Neumann boundary conditions. Then substituting $z = v - \beta$ into (3–17a) and simplifying yields

$$\Delta v - v + \beta \frac{e^z}{m(e^z)e^\beta} e^v = 0,$$

which is (3–16a) with $k = \frac{\beta}{m(e^z)e^\beta}$. \hfill \square

Hence to find a solution $(u, v)$ to (3–15) it is sufficient to find a solution $z$ to (3–17) with $m(z) = 0$. We may choose $\beta$ arbitrarily, and once we have solved for $z$ with a fixed $\beta$ then

$$v = z + \beta, \quad u = \frac{\beta e^z}{m(e^z)}$$

solve (3–15). Notice that $\beta = m(v) = m(u)$. We want to look for a solution in the subspace of $H^1(\Omega)$ consisting of functions with zero average. Motivated by this we define

$$\widehat{H}^1(\Omega) := \{w \in H^1(\Omega) : \int_{\Omega} w(x) dx = 0\}.$$
As in [18], we try to show the existence of a nontrivial solution using the functional

\[
J(z) = \frac{1}{2} \left( \int_\Omega |\nabla z|^2 + |z|^2 \right) dx - \beta |\Omega| \log \left( \frac{1}{|\Omega|} \int e^z dx \right).
\]

(3–20)

Let \( \mathcal{L}(H^1(\Omega), \mathbb{R}) \) denote the function space of bounded linear operators mapping \( H^1(\Omega) \) to \( \mathbb{R} \). Then we wish to show that \( J \) is a well defined functional, and calculate the Frechet derivative of \( J \), which is a function \( J' : H^1(\Omega) \to \mathcal{L}(H^1(\Omega), \mathbb{R}) \).

We first note that for the remainder of this section we abuse notation and allow values of the constant \( C \) to vary depending on the context being used. In order to do this, we first need a preliminary result. It requires that \( \Omega \) is bounded and satisfies the cone condition, which according to [14, Theorem 1.2.2.2] is equivalent to \( \partial \Omega \) locally Lipschitz and compact. If \( \Omega \) satisfies these conditions, then we have from [40, Theorem 2] that the embedding

\[
H^1(\Omega) \hookrightarrow L_\Phi(\Omega)
\]

(3–21)

is continuous where \( \Phi(t) = e^{t^2} - 1 \) and \( L_\Phi(\Omega) \) is the corresponding Orlicz space normed by

\[
\|u\|_{L_\Phi(\Omega)} = \inf \{ k > 0 | \Phi(u/k) \leq 1 \}.
\]

The continuous embedding (3–21) implies that there exists some constant \( C > 0 \) with

\[
\|u\|_{L_\Phi(\Omega)} \leq C\|u\|_{H^1(\Omega)}
\]

for all \( u \in H^1(\Omega) \). We use the existence of such a constant to prove the following result.

**Lemma 3.6.** There exists constant \( C > 0 \) such that for all \( u \in H^1(\Omega) \) and \( \alpha \in \mathbb{R} \),

\[
\int_\Omega e^{\alpha u} \leq e^{C\|u\|_{H^1(\Omega)}^2/4} (1 + |\Omega|).
\]

**Proof.** Consider a minimizing sequence \( k_i \to \|u\|_{L_\Phi(\Omega)} \). Then for every \( \epsilon > 0 \) there exists \( k \equiv k_i \) such that

\[
\|u\|_{L_\Phi(\Omega)} \leq k \leq \|u\|_{L_\Phi(\Omega)} + \epsilon
\]
and by the definition of $\| \cdot \|_{L^\Phi(\Omega)}$
\[
\int_{\Omega} \Phi(u/k) = \int_{\Omega} e^{(u/k)^2} - 1 \, dx \leq 1
\]
which implies
\[
\int_{\Omega} e^{(u/k)^2} \leq 1 + |\Omega|.
\]
Now, for any real $\alpha$ the arithmetic-geometric mean inequality implies that
\[
\sqrt{\frac{u}{k}(k\alpha)} = \frac{u}{2k} + \frac{k}{2\alpha}
\]
or
\[
\frac{u}{\alpha} \leq \frac{1}{4} \left( \frac{u}{k} \right)^2 + \frac{k^2\alpha^2}{4} \leq \left( \frac{u}{k} \right)^2 + \frac{k^2\alpha^2}{4}.
\]
Hence $e^{\alpha u} \leq e^{(u/k)^2 + (\alpha k/2)^2}$ and so
\[
\int_{\Omega} e^{\alpha u} \leq e^{(\alpha k/2)^2} \int_{\Omega} e^{(u/k)^2} \leq e^{(\alpha k/2)^2} (1 + |\Omega|) \leq e^{\frac{\alpha^2}{4} (\|u\|_{L^\Phi(\Omega)} + \epsilon)^2} (1 + |\Omega|) \leq e^{\frac{\alpha^2}{4} (C \|u\|_{H^1(\Omega)} + \epsilon)^2} (1 + |\Omega|)
\]
for every $\epsilon > 0$. Then letting $\epsilon \to 0$, the result follows.

Notice 3.6 immediately shows that $J$ is well defined. We now continue to calculate $J'(z)$ by finding a bounded linear operator $A_z = J'(z)$ such that
\[
\lim_{h \to 0} \frac{|J(z + h) - J(z) - A_z(h)|}{\|h\|_{H^1}}
\]
exists and is equal to zero. Using this limit definition, for any $z \in H^1(\Omega)$ we can calculate that the derivative evaluated at a given $\delta \in H^1(\Omega)$ is
\[
J'(z)(\delta) = (\nabla z, \nabla \delta) + (z, \delta) - \frac{\beta}{m(e^z)} (e^z, \delta).
\]
(3–22)
To verify the derivative of the first two terms of $J$, we let $J_1 := \frac{1}{2} \left( \int_\Omega |\nabla z|^2 + |z|^2 \right) dx$ and calculate

$$J_1(z + h) - J_1(z) = \frac{1}{2} \int_\Omega |\nabla z + \nabla h|^2 - |\nabla z|^2 - |z - h|^2 - |z|^2 \, dx$$

$$= \frac{1}{2} \int_\Omega |\nabla z|^2 + 2\nabla z \cdot \nabla h + |\nabla h|^2 - |\nabla z|^2 - |z|^2 - 2zh + |h|^2 - |z|^2 \, dx$$

$$= \int (\nabla z \cdot \nabla h) - zh \, dx + \|h\|_{H^1(\Omega)}^2.$$  

for a function $h \in H^1(\Omega)$ with $\|h\|_{H^1(\Omega)}$ small. Motivated by this calculation, define for any fixed $z \in H^1(\Omega)$ the linear functional $A_1^1 : H^1(\Omega) \to \mathbb{R}$ by

$$A_1^1(\delta) = (\nabla z, \nabla \delta) + (z, \delta)$$

for any $\delta \in H^1(\Omega)$. Then

$$|J_1(z + h) - J_1(z) - A_1^1(h)| = \|h\|_{H^1(\Omega)}^2$$

so that clearly

$$\lim_{h \to 0} \frac{|J_1(z + h) - J_1(z) - A_1^1(h)|}{\|h\|_{H^1}} = 0.$$  

This verifies the derivative of the first two terms of $J$. To compute the derivative of the third and final term, we will use the chain rule for Frechet derivatives [15] on the functions $g(\cdot) = \log(\cdot) : \mathbb{R} \to \mathbb{R}$ and $f(\cdot) = \frac{1}{|\Omega|} \int_\Omega e^\cdot : H^1(\Omega) \to \mathbb{R}$. It is easy to see that the Frechet derivative of functions mapping $\mathbb{R} \to \mathbb{R}$ is just the linear map taking $t \in \mathbb{R}$ to $t$ times the derivative of the function, meaning that if prime denotes the Frechet derivative, then

$$g'(x)(t) = \frac{1}{x} t.$$  

Then to completely verify the derivative of the functional $J$ it only remains to compute the derivative of functionals of the form $f(z) = \int_\Omega e^{az}$. This leads to the following lemma.
Lemma 3.7. For any constant $\alpha$ if $f(z) = \int_{\Omega} e^{\alpha z}$ then the Frechet derivative of $f$ at any function $z \in H^1(\Omega)$ is

$$B_z(\delta) := \alpha \int_{\Omega} e^{\alpha z} \delta dx.$$  

Proof. We calculate

$$|f(z + h) - f(h) - B_z(\delta)| = \left| \int_{\Omega} e^{\alpha z} (e^{\alpha h} - 1 - \alpha h) \right|$$

$$\leq \|e^{\alpha z}\|_{L^2(\Omega)} \|e^{\alpha h} - 1 - \alpha h\|_{L^2(\Omega)}$$

$$= \|e^{\alpha z}\|_{L^2(\Omega)} \left| \sum_{m \geq 2} \frac{(\alpha h)^m}{m!} \right|_{L^2(\Omega)}$$

$$\leq \|e^{\alpha h}\|_{L^2(\Omega)} \sum_{m \geq 2} \frac{|\alpha|^m}{m!} \|h^m\|_{L^2(\Omega)}$$

$$= \|e^{\alpha z}\|_{L^2(\Omega)} \sum_{m \geq 2} \frac{|\alpha|^m}{m!} \|h\|^m_{L^2(\Omega)}.$$  \hspace{1cm} (3–23)

Now, by Lemma 3.6, $\|e^{\alpha z}\|_{L^1(\Omega)} \leq C_1(\|z\|_{H^1(\Omega)})$ is bounded by a constant dependent on $\|z\|_{H^1(\Omega)}$ but independent of $\|h\|_{H^1(\Omega)}$. We next turn to estimating the sum in 3–23. By a specific version of the Sobolev embedding theorem seen in [37], there exists a constant $C > 0$ such that for all $v \in L^p(\Omega)$

$$\|v\|_{L^p(\Omega)} \leq C p^{1/2} \|v\|_{H^1(\Omega)}$$

for $p \geq 2$ on Lipschitz domains. Hence for each $m \geq 0$ we have

$$\frac{|\alpha|^m}{m!} \|h\|^m_{L^2(\Omega)} \leq \frac{|\alpha|^m C^m (2m)^{m/2}}{m!} \|h\|^m_{H^1(\Omega)}.$$  \hspace{1cm} (3–24)

Then as $\|h\|_{H^1(\Omega)} \to 0$, fix $E \geq \|h\|_{H^1(\Omega)}$ so that we may write

$$\sum_{m \geq 2} \frac{|\alpha|^m}{m!} \|h\|^m_{L^2(\Omega)} \leq \|h\|^2_{H^1(\Omega)} \sum_{m \geq 2} \frac{|\alpha|^m C^m (2m)^{m/2} E^{m-2}}{m!}.$$  \hspace{1cm} (3–25)

We need to show

$$\frac{|\alpha|^m C^m (2m)^{m/2} E^{m-2}}{m!}$$

45
is finite, and then by the limit definition of the derivative we are done. To see this, let
\[
\gamma := \frac{\alpha |C\sqrt{2eE}}{\sqrt{m}}
\]
and note that \(\gamma < 1\) for large enough \(m\). Then recall Stirling’s approximation for \(m!\), that is
\[
\frac{1}{m!} < \frac{e^m}{\sqrt{2\pi m^{m-\frac{1}{2}}}}.
\]  
(3–26)
Using (3–26) we obtain
\[
\frac{|\alpha|^m C^m (2m)^{m/2} E^{m/2}}{m!} \leq \frac{1}{E^2 \sqrt{2\pi m}} \gamma^m,
\]
where the latter term corresponds to a convergent series. Hence combining (3–23) with (3–24) and (3–25), we have
\[
|f(z + h) - f(z) - B_z h| \leq C \|h\|^2_{H^1(\Omega)}
\]
for some constant \(C < \infty\) independent of \(\|h\|_{H^1(\Omega)}\) which concludes the proof. \(\square\)

Now that we have established the validity of the functional \(J\), we want to look for critical points of \(J\), or points where \(J'(z) \equiv 0\). However since \(\int_{\Omega} \delta dx = 0\) the addition of a constant to the last inner product does nothing to change the value of the functional, so we instead search for zeros of a modified \(J'(z)\),
\[
J'(z)(\delta) = (\nabla z, \nabla \delta) + (z, \delta) - \frac{\beta}{m(e^z)} (e^z - m(e^z), \delta),
\]  
(3–27)
which is the variational form of (3–17a). We were unable to find zeros of this functional in the linear finite element subspace \(\tilde{S}_h \subset \tilde{H}^1(\Omega)\) using a fixed point iteration, so we implemented Newton’s method using the second derivative \(J'' : H^1(\Omega) \rightarrow \mathcal{L}(H^1(\Omega), \mathcal{L}(H^1(\Omega), \mathbb{R}))\). Similarly to the case of the calculation of \(J'\), the second derivative evaluated at  \(z, v, w \in \tilde{H}^1(\Omega)\) is calculated to be
\[
J''(z)(v)(w) = (\nabla w, \nabla v) + (w, v) - \frac{\beta}{m(e^z)} \left( (e^z, w)(e^z, v) - (e^z, 1) \right)
\]  
(3–28)
where \((u, v) = \int_{\Omega} uv\) and \((\nabla u, \nabla v) = \int_{\Omega} \nabla u \cdot \nabla v\).

In [18] Horstmann explains the idea of a proof of existence of a solution to (3–16) in the case where \(k > 4\pi\) based on the Mountain Pass Theorem. Although we will not end up using the Mountain Pass Theorem because the parameter ranges we are interested are not in the required range, we state a version of the theorem from [10] below for reference.

**Theorem 3.6.** Consider an abstract functional \(I[\cdot] \in C^1(H, \mathbb{R})\) for some real Hilbert space \(H\) such that \(I'[u]\) exists for each \(u \in H\) and is continuous. Additionally assume that \(I\) satisfies the Palais-Smale compactness condition. Suppose also that

1. \(I[0]=0\),
2. there exist constants \(r, a > 0\) such that \(I[u] \geq a\) if \(\|u\| = r\), and
3. there exists an element \(v \in H\) with \(\|v\| > r\) and \(I[v] \leq 0\).

Define
\[
\Gamma := \{ g \in C([0, 1], H) | g(0) = 0, g(1) = v \}.
\]

Then
\[
c := \inf_{g \in \Gamma} \max_{0 \leq t \leq 1} I[g(t)]
\]
is a critical value of \(I\).

In [18], the sequence of functions
\[
u_0^\epsilon := \log \left( \frac{\epsilon^2}{(\epsilon^2 + \pi |x - x_0|^2)^2} \right) - m \left( \log \left( \frac{\epsilon^2}{(\epsilon^2 + \pi |x - x_0|^2)^2} \right) \right) \in \widetilde{H}^1(\Omega) \tag{3–29}
\]
where \(x_0\) is any fixed point in \(\partial \Omega\) is used to find an appropriate function \(v\) for condition 3 of Theorem 3.6. Although we did not use parameters within the range for application of this proof, we still used the sequence \(u_0^\epsilon\) to find initial guesses for the numerical simulations as discussed in Section 3.3.

In order to determine for which values of \(\beta\) we are likely to find nonhomogeneous solutions to (3–17), we recall the following previously mentioned result from [34] that
spatially constant stationary solutions \((u_*, v_*)\) to (3–14) will be linearly stable with respect to perturbations \(U\) satisfying \(\int_{\Omega} U \, dx = 0\) whenever \(u_* < 1 - \mu_1\) and linearly unstable if \(u_* > 1 - \mu_1\) where \(\mu_1 < 0\) is the nonzero eigenvalue of smallest norm satisfying \(\Delta \omega_1 = \mu_1 \omega_1\) on \(\Omega\) with zero Neumann boundary conditions. To see the reason for the restriction \(\int_{\Omega} U \, dx = 0\), notice that for sufficiently smooth \(u\) we may calculate using Equation (3–14a)

\[
\partial_t \int_{\Omega} u(x, t) \, dx = \int_{\Omega} \partial_t u(x, t) \, dx \\
= \int_{\Omega} \nabla \cdot (\nabla u - u \nabla v) \, dx \\
= \int_{\partial \Omega} (\nabla u - u \nabla v) \cdot n = 0
\]

by Neumann boundary conditions. Hence the mean of \(u(x, t)\) is a constant \(u_0\) which is independent of time. It therefore makes sense to choose only perturbations with zero mean and, additionally, to choose a starting point for the iteration for which \(m(u_0) = \beta\) falls into the range where the constant steady state is unstable; this allows for a higher possibility of convergence to a more interesting nonconstant steady state.

In addition to choosing \(\beta\) in a range where the constant steady state is unstable, we would also like to ensure that solutions do not blow up. Horstmann summarizes results on blow-up and time asymptotic behavior to the minimal system in Table 4.4 of [18], and among the results is that solutions to the system (3–14) exist globally in time, have bounded \(L^\infty\) norm, and converge to a stationary solution as \(t \to \infty\) if

\[
\int_{\Omega} u_0 < 4\pi.
\]

Hence if we chose \(\beta = \frac{1}{|\Omega|} \int_{\Omega} u_0 = m(u)\) so that

\[
1 - \mu_1 < \beta < \frac{4\pi}{|\Omega|},
\]
we have a good chance of finding a nonconstant steady state solution of (3–14) or equivalently a constant solution to (3–15).

3.3 Numerical Solutions

3.3.1 Solutions on the Unit Square

Before moving on to introduce our generalized Keller-Segel model, we show some numerical solutions to (3–15) found by calculating solutions to (3–17) using the standard FEM introduced in Section 2.2.2. All numerical solutions in the current section were found using a Newton iteration to find zeros of the discretized form of (3–27). We begin with solutions on the unit square $\Omega = [0, 1] \times [0, 1]$. Since the eigenvalues of the Laplacian on the unit square are $\mu_{n,m} = -(n+m)^2\pi^2$ for $n, m = 0, 1, 2, \ldots$ we are likely to find nonconstant solutions to (3–17) for $\beta$ values which satisfy

$$1 + \pi^2 < \beta < 4\pi.$$

Indeed, only when $\beta > 1 + \pi^2$ was chosen were we able to demonstrate convergence to a nontrivial function $z$, and in many cases we also needed $\beta < 4\pi$.

We experimented with initial guess functions of the form $cu_0^\epsilon$ from equation (3–29) for different values of $x_0 \in \partial\Omega$, $\epsilon > 0$ small, and constants $0 < c < 1$. A solution to (3–15) for $m(u_0) = \beta = 18$ is shown in Figure 3-1. Convergence of the solution for different mesh sizes is given in Table 3-1, where $h$ corresponds to maximum triangle size. Meshes for this solution and many later solutions were generated using the freely available program Triangle [35].

<table>
<thead>
<tr>
<th>Number of nodes</th>
<th>$h$ value</th>
<th>$|z_h - z_{h/2}|_{H^1(\Omega)}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>0.5</td>
<td>4.76</td>
</tr>
<tr>
<td>25</td>
<td>0.25</td>
<td>2.30</td>
</tr>
<tr>
<td>81</td>
<td>0.125</td>
<td>0.698</td>
</tr>
<tr>
<td>289</td>
<td>0.0625</td>
<td>0.100</td>
</tr>
<tr>
<td>1089</td>
<td>0.03125</td>
<td>0.020</td>
</tr>
<tr>
<td>4225</td>
<td>0.0156</td>
<td>–</td>
</tr>
</tbody>
</table>
We next look at solutions on a disk. Our original computations were performed on
the unit disk, but we had difficulty finding initial conditions of the form (3–29) that did not
either blow up or converge to the constant solution. Indeed, notice that on a disk of
radius one, the smallest nonzero eigenvalue of the Laplacian is $\mu_1 \approx -3.38996$. Hence
there are no values of $\beta$ which satisfy

$$1 - \mu_1 \approx 4.38996 < \beta < 4 = \frac{4\pi}{|\Omega|}$$

and therefore no values for which we are assured of having instability of constant steady
states as well as no blow up.

Therefore we turn to computation on a domain $\Omega$ equal to a disk of some radius $R$,
for which the nonzero eigenvalue closest to zero satisfies $\mu_1 \approx -3.38996/R^2$. In order
to ensure a nonzero range of initial conditions satisfying the required inequality, it is
necessary to have

$$1 - \mu_1 \approx 1 + \frac{3.38996}{R^2} < \frac{4}{R^2},$$

that is

$$R < 0.7811.$$
Figure 3-2. Solutions to (3–15) on a disk

shown in Figure 3-2, and convergence of the solution for various mesh sizes is seen in

Table 3-2. Varying the initial condition (3–29) by changing the value of $x_0 \in \partial \Omega$ we can in

fact demonstrate an infinite number of solutions with peaks anywhere on the boundary

of $\partial \Omega$, for example as in Figure 3-3.

Table 3-2. Error for steady state solution on a disk

<table>
<thead>
<tr>
<th>Number of nodes</th>
<th>$h$ value</th>
<th>$|z_h - z_{h/2}|_{H^1(\Omega)}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>145</td>
<td>0.1</td>
<td>4.17</td>
</tr>
<tr>
<td>545</td>
<td>0.05</td>
<td>1.34</td>
</tr>
<tr>
<td>2113</td>
<td>0.025</td>
<td>0.305</td>
</tr>
<tr>
<td>8321</td>
<td>0.0125</td>
<td>0.0475</td>
</tr>
<tr>
<td>33,025</td>
<td>0.00625</td>
<td>–</td>
</tr>
</tbody>
</table>
Figure 3-3. More solutions to (3–15) on a disk
We would like to extend the results of the previous chapters to more general systems similar to the Keller-Segel (KS) model. To do this, again consider a population of some species of density \( u \) climbing the gradient of a chemical of concentration \( v_N \). Biologically, it is likely that this chemical will be involved in reactions with other compounds in the environment, some of which may also be produced by the species. So in this more general setting, we also model the concentrations of \( N - 1 \) other chemicals \( v_1, v_2, ..., v_{N-1} \) which interact with each other. Then writing \( \vec{v} = [v_1, ..., v_N]^t \), it can be assumed that the chemical reaction network’s contribution to the change of \( \vec{v} \) with respect to time is modeled by \( \vec{g}(\vec{v}) \) for some function \( \vec{g} : \mathbb{R}^{N+} \rightarrow \mathbb{R}^N \). In our applications this function \( \vec{g} \) will likely be determined by mass action kinetics. Note that \( g \) may include decay of the chemicals. Assume also that for some subset \( A = \{i_1, i_2, ..., i_k\} \subset \{1, 2, ..., N\} \) the population produces the chemical \( v_{i_j} \) at a rate \( \alpha_{i_j} \geq 0 \). Let \( \vec{\alpha} \in \mathbb{R}^N \) be the vector which has \( \alpha_i \) in the \( i \)-th component if \( i \in A \) and 0 otherwise. Assume that the diffusion coefficient for \( u \) is given by the constant \( D > 0 \), and the diffusion coefficients for each \( v_i \) are given by constants \( \tilde{D}_i > 0 \). Let \( \tilde{\vec{D}} \) be the diagonal matrix with elements \( \tilde{D}_i \) along the diagonal and denote the chemotactic coefficient by \( \chi \). Then applying Neumann boundary conditions and nonnegative initial conditions, we obtain the following system of equations.

\[
\begin{align*}
\frac{\partial u(x,t)}{\partial n} &= 0, & x \in \partial \Omega, & t \geq 0, & 1 \leq i \leq N \\
\end{align*}
\]
Now that the form of the generalized model has been explained, we give some examples of situations which can be modeled using \((4–1)\).

**Example 4.1.** We begin with a simple theoretical case with two chemical species in order to better understand the framework of the model. Dimerization is the chemical reaction

\[ 2v_1 \leftrightarrow v_2, \]

where the reaction may or may not be reversible. If the forward and backwards rate constants for the dimerization reaction are given by \(k_1 > 0\) and \(k_2 \geq 0\) respectively and the decay rates of each chemical are denote by \(\gamma_i\), then using the law of mass action the function

\[
\vec{g}(\vec{v}) = \begin{pmatrix}
- k_1 v_1^2 + k_2 v_2 - \gamma_1 v_1 \\
 k_1 v_1^2 - k_2 v_2 - \gamma_2 v_2
\end{pmatrix}
\]

\((4–2)\)

describes the kinetics of the CRN. Hence the theoretical scenario where an organism emits a chemical \(v_1\) which undergoes dimerization to form the chemical \(v_2\), which in turn is the chemoattractant of \(u\) can be modeled by \((4–1)\) for \(\vec{g}\) as in \((4–2)\) and

\[
\vec{\alpha} = \begin{pmatrix}
\alpha_1 \\
0
\end{pmatrix}
\]

for some constant \(\alpha_1 > 0\).

**Example 4.2.** Consider the unsimplified four equation KS system \((1–2)\). Note that assuming mass action kinetics on the chemicals in question, constant rates of production, and a chemotactic sensitivity function of the form \(\kappa(u, v) = \chi u\) for some constant \(\chi\), \((1–2)\) with possible decay of all chemicals can be expressed in the framework of \((4–1)\) by setting

\[
\vec{\alpha} = \begin{pmatrix}
\alpha_1 \\
0 \\
\alpha_3
\end{pmatrix}
\]
\[ \vec{g}(\vec{v}) = \begin{pmatrix} -k_1 v_1 v_3 + (k_2 + k_3)v_2 - \gamma_1 v_1 \\ k_1 v_1 v_3 - (k_2 + k_3)v_2 - \gamma_2 v_2 \\ -k_1 v_1 v_3 + k_2 v_2 - \gamma_3 v_3 \end{pmatrix}. \]

In this new framework, \( u \) denotes the density of the amoeba \( Dictyostelium discoideum \), \( v_1 \) denotes the density of the phosphodiesterase enzyme emitted by the amoeba, \( v_3 \) denote the density of the chemoattractant cAMP, and \( v_2 \) the density of the complex formed from the enzyme and the cAMP. All \( k_i, \alpha_1 \) and \( \alpha_3 \) are assumed to be positive constants, while the \( \gamma_i \) are assumed to be nonnegative.

### 4.2 Reactions Yielding Positive Solutions

In this section we will show conditions under which sufficiently smooth solutions to (4–1) remain positive, or at least nonnegative, if the initial conditions \( u_0 \) and \( v_0 \) are nonnegative and not identically zero. The fact that \( u > 0 \) will follow without any further assumptions exactly as in Theorem 3.3 for the case of the simple Keller-Segel model. In order to prove nonnegativity for \( \vec{v} \) a certain inward pointing condition is required of the function \( \vec{g} \) on the boundary of the positive orthant. As we will see, this assumption arises naturally if \( \vec{g} \) is derived using the law of mass action kinetics. The main result of this section is Theorem 4.4, but we need a preliminary technical result before it can proven.

**Lemma 4.1.** Suppose there is an \( \varepsilon_0 > 0 \) such that for all \( 0 < \varepsilon < \varepsilon_0 \), the function \( f : \mathbb{R} \times \bar{\Omega} \times [0, T] \to \mathbb{R} \) satisfies

\[ f(\varepsilon, x, t) > 0, \quad (x, t) \in \bar{\Omega} \times [0, T] \] (4–3)
Let \( w : \bar{\Omega} \times [0, T] \to \mathbb{R} \) be continuously differentiable with respect to \( t \), twice continuously differentiable with respect to \( x \), and satisfy

\[
\begin{align*}
\partial_t w &= \tilde{D} \Delta w + f(w, x, t), \\
w(x, t) &\geq 0
\end{align*}
\]

\( x \in \bar{\Omega}, \quad t \in (0, T] \) \( \tag{4–4a} \)

\[ w(x, t) \geq 0 \]

\( x \in \bar{\Omega}, \quad t = 0, \) \( \tag{4–4b} \)

**together with the following boundary condition:**

**Either**

\[ w(x, t) \geq 0, \quad x \in \partial \Omega, \quad t \in (0, T], \] \( \tag{4–4c} \)

**or**

\[ \frac{\partial w}{\partial n} \geq 0, \quad x \in \partial \Omega, \quad t \in (0, T]. \] \( \tag{4–4d} \)

Then \( w(x, t) \geq 0 \) for all \( x \in \bar{\Omega} \) and \( t \in [0, T] \).

**Proof.** If not, there is an \( (\tilde{x}_0, \tilde{t}_0) \in \bar{\Omega} \times [0, T] \) such that \( w(\tilde{x}_0, \tilde{t}_0) < 0 \). Then \( w(\tilde{x}_0, \tau) \) is then negative at \( \tau = \tilde{t}_0 \) but nonnegative at \( \tau = 0 \) due to \( (4–4b) \). Hence, there are (small enough) values of \( \varepsilon \in (0, \varepsilon_0) \) such that \( w \) attains the value of \(-\varepsilon\) at one or more \( \tau \in (0, \tilde{t}_0) \). Fix such an \( \varepsilon \), and let \( t_0 \) be the first time when \( w \) attains the value of \(-\varepsilon\).

Let \( x_0 \) be any point in \( \bar{\Omega} \) such that \( w(x_0, t_0) = -\varepsilon \). Then, as \( \tau \) increases to \( t_0 \) and is sufficiently close to \( t_0 \), the function \( w(x_0, \tau) \) cannot increase, so

\[
\partial_t w(x_0, t_0) \leq 0.
\] \( \tag{4–5} \)

Note also that

\[
\min_{(x, t) \in \bar{\Omega} \times [0, t_0]} w(x, t) = w(x_0, t_0) = -\varepsilon
\] \( \tag{4–6} \)

(because if \( w \) took a value lesser than \(-\varepsilon\) in \( \bar{\Omega} \times [0, t_0] \), then \( t_0 \) would not be the first time when \( w \) attains \(-\varepsilon\)).

The remainder of the proof is split in two parts. First, suppose \( (4–4c) \) holds. Then, \( x_0 \notin \partial \Omega \). In view of \( (4–6) \), we therefore have

\[
\Delta w(x_0, t_0) \geq 0.
\] (4–7)
Combining (4–5) and (4–7), we find that $0 \geq [\partial_t w - \bar{D}\Delta w](x_0, t_0) = f(-\epsilon, x_0, t_0)$. This contradicts (4–3) and finishes the proof in the case of the boundary condition (4–4c).

To complete the proof for the case of the boundary condition (4–4d), first note that if $x_0$ is an interior point of $\Omega$, then we obtain a contradiction using (4–5) and (4–7) as above, so we need only consider $x_0 \in \partial \Omega$. By (4–4a) and (4–3), the inequality $\partial_t w - \bar{D} \Delta w = f > 0$ holds at $(x_0, t_0)$, and so by continuity, it holds in a closed neighborhood $\bar{\Omega}_0 \times [t_1, t_0] \subseteq \Omega \times (0, t_0]$, where $\Omega_0$ is a ball whose boundary contains $x_0$. By (4–6), we know that the minimum of $w$ in $\Omega_0 \times [t_1, t_0]$ is attained at $(x_0, t_0)$. If this minimum is also attained at another point $(x'_0, t'_0)$ in the same neighborhood, then $x'_0$ is an interior point of $\Omega$, so (4–5) and (4–7) finish the proof as before. Hence it only remains to consider the situation when $x_0 \in \partial \Omega_0$ is the sole point in $\Omega_0 \times [t_1, t_0]$ where the maximum is attained. But in this situation, all conditions of part (ii) of Lemma 3.1 are satisfied in a sufficiently small parabolic frustum contained in $\Omega_0 \times (t_1, t_0]$. Hence, (3–4) implies that $\partial(-w)/\partial n > 0$ at $x_0$ which contradicts (4–4d).

In order to strengthen this result to a more useful form, we need a result on existence and uniqueness of a PDE of the form (4–4). The following is a combination of versions of [12, Theorems 6 and 10 in Section 7.4], modified to fit our application. Before stating the theorem, we recall that a function $f(w, x, t)$ is said to be “locally Hölder continuous” in $(x, t)$ if there is a $C$ and $0 < \alpha < 1$ such that

$$|f(w, x_1, t_1) - f(w, x_2, t_2)| \leq C|(x_1, t_1) - (x_2, t_2)|^\alpha$$

for all $(x_1, t_1)$ and $(x_2, t_2)$ in every closed bounded subset $B$ of $\bar{\Omega}_T$.

**Theorem 4.3.** Consider (4–4a) together with the initial and boundary condition

$$w(x, t) = w_0(x, t), \quad (x, t) \in (\bar{\Omega} \times \{t = 0\}) \cup (\partial \Omega \times [0, T])$$

(4–8)

for some smooth $w_0$. Assume that $f(w, x, t)$ is Lipschitz continuous in $w$, uniformly with respect to bounded subsets of $\mathbb{R} \times \bar{\Omega} \times [0, T]$, and also locally Hölder continuous
with respect to \((x,t)\). Suppose \(\partial_t w_0 = \tilde{D}\Delta w_0 + f(w_0, x, 0)\) on \(\partial\Omega\). Then there is some \(0 < T_0 \leq T\) such that a unique solution to (4–4a) satisfying the initial and boundary condition (4–8) exists in \(\bar{\Omega} \times [0, T_0]\).

Before applying the previous theorem to prove our desired result, introduce the following terminology. We say that \(\vec{f} : \mathbb{R}^N \times \bar{\Omega} \times [0, T] \) is “locally Lipschitz continuous in \(\vec{y} \equiv (y_1, \ldots, y_N)\), uniformly in \((x,t)\), if for every bounded subset \(D \subset \mathbb{R}^N\), there is a constant \(M_f > 0\) such that

\[
\max_i |f_i(\vec{y}, x, t) - f_i(\vec{z}, x, t)| \leq M_f \max_i |y_i - z_i| \quad (4–9)
\]

for all \(\vec{y}, \vec{z} \in D\) and all \((x,t) \in \bar{\Omega} \times [0, T]\). Using this terminology we prove one more lemma before proving the desired final result. Lemma 4.2 was inspired by a statement in lecture notes from Chris Cosner (Analogues of Maximum Principles for Systems, private communication) about strengthening the results of [42], but the result was not explicitly proven in the notes.

**Lemma 4.2.** Suppose that \(\vec{f} : \mathbb{R}^N \times \bar{\Omega} \times [0, T] \) is locally Lipschitz continuous in \(\vec{y}\), uniformly in \((x,t)\), and uniformly on bounded subsets of \(\mathbb{R}^N\). Suppose \(\vec{f}\) is also locally Hölder continuous in \((x,t)\). Finally, assume that for all \((x,t) \in \bar{\Omega} \times [0, T]\),

\[
f_i(y_1, y_2, \ldots, y_{i-1}, 0, y_{i+1}, \ldots, y_N, x, t) \geq 0, \quad i = 1, 2, \ldots, N \quad (4–10)
\]

whenever \(y_j \geq 0\) for all \(j \neq i\). Let \(\vec{w} : \mathbb{R}^N \times \bar{\Omega} \times [0, T] \rightarrow \mathbb{R}^N\) be a smooth solution of

\[
\begin{align*}
\partial_t w_i &= \tilde{D}_i \Delta w_i + f_i(\vec{w}, x, t), \quad x \in \bar{\Omega}, \quad t \in (0, T], \quad i = 1, 2, \ldots, N, \quad (4–11a) \\
\vec{w}(x,t) &\geq 0 \quad x \in \bar{\Omega}, \quad t = 0, \quad (4–11b)
\end{align*}
\]

together with the following boundary condition:

Either \(\vec{w}(x,t) \geq 0\), \(x \in \partial\Omega, \ t \in (0, T]\), \(4–11c)\)

or \(\frac{\partial \vec{w}}{\partial n} \geq 0\), \(x \in \partial\Omega, \ t \in (0, T]\). \(4–11d)\)
Then $\bar{w}(x,t) \geq 0$ for all $x \in \bar{\Omega}$ and $t \in [0,T]$.

Proof. The idea is to construct a new function $F_i$ from $f_i$ such that Lemma 4.1 can be applied. To this end, first let $a^+ = \max(a,0)$. It can be easily verified, case by case, that for any two numbers $a$ and $b$,

$$|a^+ - b^+| \leq |a - b|.$$  

(4–12)

Next, define $F_i : \mathbb{R} \times \bar{\Omega} \times [0,T] \rightarrow \mathbb{R}$ by

$$F_i(v,x,t) = f_i( w_1(x,t), \ldots, w_{i-1}(x,t), v^+, w_{i+1}(x,t), \ldots, w_N(x,t), x, t ) + v^+ - v.$$

By (4–12) and (4–9), $F_i$ is locally Lipschitz continuous in $v$, uniformly in $(x,t)$.

We now prove that, for an arbitrary $\varepsilon > 0$,

$$F_i(-\varepsilon,x,t) > 0, \quad \forall (x,t) \in K_{i}^{-\varepsilon_1}$$  

(4–13)

for any $\varepsilon_1 < \varepsilon/M_f$. Here $K_{i}^\alpha = \{(x,t) \in \bar{\Omega} \times [0,T] : w_j(x,t) \geq \alpha \text{ for all } j \neq i\}$. Before proving this, note that although (4–10) and the definition of $F$ immediately imply the inequality $F_i(-\varepsilon,x,t) \geq \varepsilon$, this inequality holds in general only for $(x,t) \in K_{i}^\alpha$. To obtain a similar inequality in a larger set, we use (4–9). So, for any $\varepsilon_1 > 0$ and any $(x,t) \in K_{i}^{-\varepsilon_1}$,

$$f_i(\bar{\Omega},x,t) - f_i(w_1, \ldots, w_{i-1}, 0, w_{i+1}, \ldots, w_N, x, t) \leq M_f \varepsilon_1.$$  

Since the first term above is nonnegative due to (4–10), this implies that

$$F_i(-\varepsilon,x,t) = f_i(w_1, \ldots, w_{i-1}, 0, w_{i+1}, \ldots, w_N, x, t) + \varepsilon \geq -M_f \varepsilon_1 + \varepsilon.$$  

and (4–13) follows. Accordingly, we fix $\varepsilon_1 < \varepsilon/M_f$ and proceed.

To prove the lemma by way of contradiction, suppose $\bar{w} \not\geq 0$. Then there exists some $\varepsilon_2$, sufficiently small, and chosen so that $0 < \varepsilon_2 \leq \varepsilon_1$, such that at least one of the components of $\bar{w}$ attain the value $-\varepsilon_2$. Let $t_1 > 0$ be the first time that any of the components of $\bar{w}$ attain the value $-\varepsilon_2$, and let $i^*$ and $x_1 \in \bar{\Omega}$ be such that
\( w_i^*(x_1, t_1) = -\varepsilon_2. \) Then (cf. (4–6))

\[
\min_i \min_{(x,t) \in \Omega \times [0,t_1]} w_i(x,t) = w_i^*(x_1, t_1) = -\varepsilon_2.
\]  

(4–14)

Clearly, this implies that

\[
\bar{\Omega} \times [0, t_1] \subseteq K^{-\varepsilon_2}_i \subseteq K^{-\varepsilon_1}_i.
\]  

(4–15)

Now, let \( v_i^* \) be the solution to

\[
\partial_t v_i^* = \bar{D}_i \Delta v_i^* + F_i^*(v_i^*, x, t), \quad x \in \bar{\Omega}, \quad t \in (0, t_1],
\]

(4–16a)

\[
v_i^* = w_i^*, \quad (x, t) \in (\bar{\Omega} \times \{t = 0\}) \cup (\partial \Omega \times [0, t_1)).
\]

(4–16b)

By Theorem 4.3 and the aforementioned continuity properties of each \( F_i, v_i \) exists in an interval \([0, t_2] \), where \( t_2 > 0 \) is the maximal time of existence of the solution. The remainder of the proof is split into two cases: \( t_2 \geq t_1 \) and \( t_2 < t_1 \).

Consider the first case \( t_2 \geq t_1 \). Because of (4–15), we find that the inequality in (4–13) holds for all \( x \in \bar{\Omega} \) and all \( 0 \leq t \leq t_1 \). So we can apply Lemma 4.1 to conclude that \( v_i^* \geq 0 \) on \([0, t_1]\), which in turn implies that

\[
F_i^*(v_i^*, x, t) \equiv f_i^*(w_1, \ldots, w_{i-1}, v_i^*, w_{i+1}, \ldots, w_N, x, t).
\]

But then, the \((i^*)th\) equations in (4–11a)–(4–11c) show that \( w_i^* \) also solves (4–16) on \([0, t_1]\). By uniqueness (Theorem 4.3) we conclude that \( w_i^* = v_i^* \geq 0 \) on \([0, t_1]\), a contradiction to (4–14).

The other case, \( t_2 < t_1 \), also leads to a contradiction, as we now show. As above, we conclude that \( w_i^* = v_i^* \geq 0 \) on \([0, t_1]\). Now, due to the smoothness assumptions on \( w_i^* \) on \([0, T]\) (and noting that \( t_2 < t_1 \leq T \)) the solution \( v_i^* \) is smooth at \( t = t_2 \). However then we can extend the solution \( v_i^* \) to some interval \([0, t_3]\) with \( t_3 > t_2 \) by again invoking Theorem 4.3. This is a contradiction to the maximality of \( t_2 \) and finishes the proof. \( \square \)
Theorem 4.4. Suppose a smooth solution \((u, \vec{v})\) to (4–1) exists on the interval \([0, T]\) for some \(T > 0\) for smooth initial conditions \(u_0\) and \(\vec{v}_0\) which are nonnegative and not identically zero on \(\Omega\). Assume \(\partial \Omega\) is smooth, \(\vec{g} : \mathbb{R}^N \to \mathbb{R}^N\) is uniformly Lipschitz on compact subsets of \(\mathbb{R}^N\), and that \(\vec{g}\) satisfies
\[
g_i(v_1, v_2, \ldots, v_{i-1}, 0, v_{i+1}, \ldots, v_N) \geq 0, \quad i = 1, 2, \ldots, N \tag{4–17}
\]
whenever \(v_j \geq 0\) for all \(j \neq i\). Then on \(\bar{\Omega} \times [0, T]\), \(u > 0\) and each component of \(\vec{v}\) is nonnegative.

Proof. The argument that \(u > 0\) on \(\bar{\Omega} \times [0, T]\) is exactly as in Theorem 3.3. To show that \(\vec{v}\) is nonnegative, write (4–1) in the form of (4–11a) to see that
\[
f_i(\vec{v}, x, t) = \alpha_i u(x, t) + g_i(\vec{v}(x, t)).
\]
Since \(u_i\) is smooth on the compact set \(\bar{\Omega} \times [0, T]\) and \(g_i\) was assumed to be uniformly Lipschitz in \(\vec{v}\) on compact sets, it follows that the function \(f_i\) is locally Lipschitz continuous in \(\vec{v}\), uniformly in \((x, t)\). By (4–17), \(\alpha_i \geq 0\), and \(u(x, t) > 0\) we see that (4–10) holds on \(\Omega \times [0, T]\) whenever \(v_j \geq 0\). Therefore we may apply Lemma 4.2 to conclude \(\vec{v}\) is nonnegative on \(\bar{\Omega} \times [0, T]\). \(\square\)

Remark 4.5. Notice that if \(\vec{g}\) is derived according to the law of mass action kinetics, then \(\vec{g}\) will be a smooth function which satisfies (4–17).

4.3 Existence of Homogeneous Steady State Solutions

Again consider the generalized KS model (4–1). A steady state \((u_*, \vec{v}_*)\) of the system will satisfy
\[
D \Delta u_* - \nabla \cdot (\chi u_* \nabla v_{N,*}) = 0 \tag{4–18a}
\]
\[
\bar{D} \Delta \vec{v}_* + \bar{\alpha} u_* + \bar{g}(\vec{v}_*) = 0. \tag{4–18b}
\]
If such a steady state is constant then (4–18) will reduce to

\[ \vec{\alpha} u^* + \vec{g}(\vec{v}^*) = 0. \]  

(4–19)

Given a non-zero constant \( u^* \), there is a question as to whether or not a non-zero constant steady state \( (u^*, \vec{v}^*) \) exists for some constant \( \vec{v}^* \in \mathbb{R}^N \), and if it does exist if the homogeneous steady state is stable.

Before continuing with the analysis on this subject, some terminology on matrices needs to be introduced and one result will be stated. The following definitions and subsequent result are all from [2]. We note that in the following analysis when we write \( \vec{v} \geq 0 \) and \( A \geq 0 \) for a vector or matrix we mean componentwise nonnegativity.

**Definition 4.1.** A Metzler matrix \( A \in \mathbb{R}^N \times \mathbb{R}^N \) is a matrix where all off diagonal entries are nonnegative, i.e. for all \( 1 \leq i, j \leq N \)

\[ A_{ij} \geq 0, \quad i \neq j. \]

**Definition 4.2.** A matrix \( A \in \mathbb{R}^N \times \mathbb{R}^N \) is an M-matrix if for all \( 1 \leq i, j \leq N \) it satisfies

1. \( A_{ij} \leq 0 \), \( i \neq j \)
2. \( A_{ii} \geq 0 \)
3. When \( A \) is expressed in the form \( A = sI - B \) for \( s > 0 \) and \( B \) nonnegative, then \( \rho(B) \leq s \)

The following characterization of a nonsingular M-matrix is from [2, Theorem 6.2.3, condition N_{38}].

**Lemma 4.3.** \( A \in \mathbb{R}^N \times \mathbb{R}^N \) is a nonsingular M-matrix if and only if \( A \) is inverse positive, that is, \( A^{-1} \) exists and \( A^{-1} \geq 0 \).

Using the previous results, we can now give a sufficient condition for existence of nonnegative homogeneous steady states in the case where \( g \) is a linear function.
Lemma 4.4. Assume that $\bar{g}(\bar{v}) = -A\bar{v}$ for some matrix $A \in \mathbb{R}^N \times \mathbb{R}^N$. Then $A$ is a nonsingular $M$-matrix if and only if (4–1) has a nonnegative homogeneous steady state solution for each constant $u_*>0$ and vector $\bar{\alpha} \geq 0$.

Proof. First assume that $A$ is a nonsingular $M$-matrix, and fix any $u_*>0$ and $\bar{\alpha} \geq 0$. Then defining

$$\bar{v}_* = u_*A^{-1}\bar{\alpha}, \quad (4–20)$$

$\bar{v}_*$ satisfies (4–19). By Lemma 4.3, $\bar{v}_* \geq 0$ and therefore $(u_*, \bar{v}_*)$ is a nonnegative homogeneous steady state solution to (4–1). Furthermore, notice that $\bar{v}_*$ is unique. Indeed, if $\bar{w}_*$ satisfies (4–19) then $A\bar{w}_* = u_*\bar{\alpha}$ which implies that $\bar{w}_* = u_*A^{-1}\bar{\alpha} = \bar{v}_*$ by (4–20).

To prove the converse, let $(u_*, \bar{v}_*^{(i)})$ be a nonnegative homogeneous steady state obtained by the hypothesis using $u_*=1$ and $\bar{v}_* = \bar{e}_i$, the $i$th coordinate vector. Then by (4–19), we have $A\bar{v}_*^{(i)} = \bar{e}_i$. Defining the matrix

$$C = [\bar{v}_*^{(1)} | \bar{v}_*^{(2)} | \cdots | \bar{v}_*^{(N)}],$$

we have that $AC = I$. It follows that $A$ is invertible, and since $A^{-1} \equiv C \geq 0$, by Lemma 4.3, $A$ is inverse nonnegative and therefore a nonsingular $M$-matrix.

4.4 Stability of Homogeneous Steady State Solutions

Having proven the existence of homogeneous stationary solutions under certain conditions, we now assume that they exist and turn to determining their stability. In order to analyze the stability of a given homogeneous solution to (4–18), we use the methods of Schaaf in [34] to obtain some general results. The general method involves linearizing about a constant steady state and using the eigenvalues of the linearized problem to determine stability. Then for example as in [22] we give the following definitions for a steady state $u_*$ of the generic PDE

$$u_t = F(u) \quad \text{on} \quad \Omega \times (0,T) \quad (4–21)$$
where \( u(t) \) takes values in some Banach space \((X, \| \cdot \|)\).

**Definition 4.3.** We say that \( u_* \) is linearly stable if all eigenvalues \( \lambda \) of the linearized eigenvalue problem

\[
DF|_{u_*}(U) = \lambda U
\]

have negative real part and do not have zero as a limit point. It is said to be linearly unstable if there exists an eigenvalue with positive real part.

Assume that a constant steady state \((u^*, \vec{v}^*)\) of (4–1), or equivalently a solution of (4–18), exists. We wish to find the eigenvalues of the linearized system about that steady state in order to determine linear stability. Suppose that \( \lambda \) is an eigenvalue of the linearized system, meaning that \( \lambda \) satisfies

\[
\begin{align*}
D \Delta U - \chi u_* \Delta \vec{V}_N &= \lambda U, & x \in \Omega, & t > 0 \\
\dot{D} \Delta \vec{V} + \vec{a} U + J \vec{V} &= \lambda \vec{V}, & x \in \Omega, & t > 0 \\
\frac{\partial U}{\partial n} = \frac{\partial \vec{V}_i}{\partial n} &= 0, & 1 \leq i \leq N, & x \in \partial \Omega, & t > 0
\end{align*}
\]

(4–22a)

(4–22b)

(4–22c)

for some nontrivial functions \((U(x), \vec{V}(x))\) on \( \Omega \). In (4–22b), \( J \) is the \( N \times N \) Jacobian matrix of \( \vec{g} \) with respect to \( \vec{v} \) evaluated at the steady state \( \vec{v}^* \), meaning

\[
J_{ij} = \frac{\partial g_i}{\partial v_j} \bigg|_{\vec{v}^*}.
\]

We wish to study the eigenvalue problem (4–22) in order to determine conditions under which spatially homogeneous stationary solutions are unstable.

### 4.4.1 Reduction to Finite Dimension

To simplify the stability analysis, we will reduce the infinite dimensional eigenproblem (4–22) to a countable number of finite dimensional matrix eigenproblems. In order to allow our results to be as general as possible, we look at the weak formulation of the
eigenproblem and search for \((u, \vec{v}) \in H^1(\Omega)^{N+1}\) which satisfy

\[
- \int_{\Omega} D \nabla U \cdot \nabla \varphi + \int_{\Omega} \chi u_s \nabla V_N \cdot \nabla \varphi = \lambda \int_{\Omega} U \varphi, \\
- \int_{\Omega} \tilde{D}_k \nabla V_k \cdot \nabla \varphi + \alpha_k \int_{\Omega} U \varphi + \sum_{l=1}^{N} J_{kl} \int_{\Omega} V_l \varphi = \lambda \int_{\Omega} V_k \varphi, \quad k = 1, 2, \ldots, N,
\]

for all \(\varphi \in H^1(\Omega)\). Note that in the case of sufficiently smooth domain boundary \(\partial \Omega\) and eigenfunctions \((u, \vec{v})\), \((4–23)\) can be recovered from \((4–22)\) by multiplying through by a test function, integrating by parts, and using boundary conditions.

We first reduce \((4–23)\) to a series of finite dimensional eigenproblems by generalizing a method due to [34]. To proceed, we let \(\{\omega_i\}_{i \geq 1}\) be a set of functions in \(H^1(\Omega)\) normalized to have \(L^2(\Omega)\) norm equal to one and which are weak eigenfunctions of the Laplacian operator with Neumann boundary conditions. Write the eigenvalues of each \(\omega_i\) as \(\mu_i\) and order so that \(0 = \mu_0 > \mu_1 \geq \mu_2 \geq \mu_3, \geq \cdots\). Hence the \(\omega_i\) satisfy

\[
- \int_{\Omega} \nabla \omega_i \cdot \nabla \varphi = \mu_i \int_{\Omega} \omega_i \varphi, \quad \forall \varphi \in H^1(\Omega), \quad i = 0, 1, 2, \ldots.
\]

Then substituting any \(\omega_i\) in for \(\varphi\) in \((4–23)\) and using \((4–24)\) to simplify, we obtain the system

\[
D \mu_i \int_{\Omega} U \bar{\omega}_i - \chi u_s \mu_i \int_{\Omega} V_N \bar{\omega}_i = \lambda \int_{\Omega} u \bar{\omega}_i \\
\tilde{D}_k \mu_i \int_{\Omega} V_k \bar{\omega}_i + \alpha_k \int_{\Omega} U \bar{\omega}_i + \sum_{l=1}^{N} J_{kl} \int_{\Omega} V_l \bar{\omega}_i = \lambda \int_{\Omega} V_k \bar{\omega}_i
\]

To express the left hand side of \((4–25)\) more concisely in matrix form, fix \(i \geq 0\) and let \(x_0 = \int_{\Omega} U \bar{\omega}_i\) and \(x_k = \int_{\Omega} V_k \bar{\omega}_i\) for \(1 \leq k \leq N\). Then if \(\vec{x} = [x_0, x_1, \ldots, x_N]^t\), \((4–25)\) reduces
to the \((N + 1) \times (N + 1)\) matrix system

\[
\begin{bmatrix}
D_{\mu} & 0 & 0 & \cdots & 0 & -\chi_{\mu}u^* \\
\alpha_1 & J_{11} + \mu_1\tilde{D}_1 & J_{12} & \cdots & J_{1,N-1} & J_{1N} \\
\alpha_2 & J_{21} & J_{22} + \mu_2\tilde{D}_2 & \cdots & J_{2,N-1} & J_{2N} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\alpha_N & J_{N1} & J_{N2} & \cdots & J_{N,N-1} & J_{NN} + \mu_N\tilde{D}_N
\end{bmatrix}
\begin{bmatrix}
\vec{x} \\
\end{bmatrix}
= \lambda \begin{bmatrix}
\vec{x}
\end{bmatrix}
\tag{4–26}
\]

Denote the matrix on the left hand side of (4–26) by \(A(\mu_i)\). Then the stability of the steady state depends on the eigenvalues of \(A(\mu_i)\) according to the following theorem.

**Theorem 4.6.** The number \(\lambda\) solves the weak eigenproblem (4–23) if and only if it solves (4–26) for some \(\mu_i, i \geq 0\).

**Proof.** First assume \(\lambda\) solves (4–23) for nontrivial eigenfunctions

\[
\begin{bmatrix}
U \\
V
\end{bmatrix} \in L^2(\Omega)
\]

Then since the \(\{\omega_i\}_{i \geq 0}\) form a complete orthonormal set in \(L^2(\Omega)\), it is possible to fix \(i\) so that either \(\int_\Omega U\omega_i \neq 0\) or \(\int_\Omega V_k\omega_i \neq 0\) for some \(1 \leq k \leq N\). Then for this \(\mu_i\), the previous calculations show that \(\lambda\) solves (4–26) with \(x \neq 0\).

To show the converse, assume that \(\lambda\) is an eigenvalue of (4–26) with eigenvector \(x \neq 0\) for some fixed eigenvalue \(\mu_i\) of the Laplacian. Let \(\omega_i\) be the eigenfunction of the Laplacian which corresponds to \(\mu_i\). Define the nontrivial function

\[
\begin{bmatrix}
U \\
V
\end{bmatrix} \in L^2(\Omega)^{N+1}
\]

by

\[
U = x_0\omega_i
\]

and

\[
\begin{bmatrix}
x_1\omega_i \\
x_2\omega_i \\
\vdots \\
x_N\omega_i
\end{bmatrix}
\]
Substituting these definitions into the left hand side of (4–23) and using (4–24) and (4–26) shows that \( u \) and \( v \) as defined above indeed satisfy (4–23). For example for (4–23a) we use (4–24) to calculate

\[
- \int_{\Omega} D \nabla U \cdot \nabla \varphi + \int_{\Omega} \chi u_* \nabla V_N \cdot \nabla \varphi = -D \int_{\Omega} \nabla (x_0 \omega_i) \cdot \nabla \varphi + \int_{\Omega} \chi u_* \nabla (x_N \omega_i) \cdot \nabla \varphi
\]

\[
= Dx_0 \mu_i \int_{\Omega} \omega_i \varphi - \chi u_* x_N \mu_i \int_{\Omega} \omega_i \varphi
\]

\[
= (Dx_0 \mu_i - \chi u_* x_N \mu_i) \int_{\Omega} \omega_i \varphi.
\]

Then noticing that the first row of (4–26) gives

\[
Dx_0 \mu_i - \chi u_* x_N \mu_i = \lambda x_0,
\]

it follows that (4–23a) holds. Equation (4–23b) follows similarly substituting in each \( V_k = x_k \omega_i \) into the left hand side of equation (4–23) and then using the \( k \)th row of (4–26) to simplify.

\[\Box\]

### 4.4.2 Sufficient Conditions for Instability

In order to analyze the eigenproblem (4–26) and determine sufficient conditions for instability of homogeneous steady states, we need some standard results from matrix theory and linear algebra. The following definitions and results are stated and proved in [2]. We also note that throughout this section, a positive vector will refer to one in which each component is positive, while a nonnegative vector means that each component is nonnegative.

**Definition 4.4.** The associated directed graph \( G(A) \) of an \( N \times N \) matrix \( A \), consists of \( N \) vertices \( P_1, P_2, \ldots, P_N \) where and edge leads from \( P_i \) to \( P_j \) if and only of \( a_{ij} \neq 0 \).

**Definition 4.5.** A directed graph \( G \) is strongly connected if for any ordered pair \( (P_i, P_j) \) of vertices of \( G \), there exists a sequence of edges (a path) which leads from \( P_i \) to \( P_j \).

**Definition 4.6.** A matrix \( A \) is called irreducible if \( G(A) \) is strongly connected.

**Theorem 4.7** (Perron-Frobenius). Let \( A \) be a square, nonnegative matrix.
(i) If $A$ is positive, meaning each element of $A$ is strictly greater than zero, then the spectral radius $\rho(A)$ is a simple eigenvalue of $A$, greater than the magnitude of any other eigenvalue.

(ii) If $A$ is nonnegative and irreducible, then $\rho(A)$ is a simple eigenvalue, any eigenvalue of $A$ of the same modulus is also simple, $A$ has a positive eigenvector $x$ corresponding to $\rho(A)$, and any nonnegative eigenvector of $A$ is a multiple of $x$.

**Theorem 4.8** (Spectral radius bounds). Let $A$ be a non-negative irreducible $N \times N$ matrix. Let $s_i$ denote the sum of the elements of the $i$th row of $A$. Let $S = \max_{1 \leq i \leq N} s_i$ and $s = \min_{1 \leq i \leq N} s_i$. Then the spectral radius $\rho(A)$ satisfies

$$s \leq \rho(A) \leq S.$$

Using these standard results, we now can state and prove a theorem giving sufficient conditions for instability of a homogeneous steady state.

**Theorem 4.9.** Assume the system (4–1) has a positive homogeneous steady state solution $(u^*, \vec{v}^*)$. Let $J$ denote the Jacobian of $\vec{g}$ evaluated at $(u^*, v^*)$, and assume also that the matrix $J$ and the vector $\vec{\alpha}$ satisfy the following conditions.

1. There exists some $1 \leq i \leq N$ such that $\alpha_i > 0$.

2. $J$ is irreducible.

3. $J$ is Metzler.

Then if the product $\chi u^*_i$ or $\alpha_i$ is sufficiently large, $(u^*, \vec{v}^*)$ will be unstable.

**Remark 4.10.** Biologically the first condition means that the species in question must produce one of the chemicals being modeled, which is what creates the interesting feedback mechanism as well as the coupling of the system. The second condition is a technical assumption which will be relaxed in the next theorem. The third condition is the most restrictive of the three, however there are large classes of chemical reaction networks for which the assumption holds. These will be explored in Section 4.5, along with an example in which condition three does not hold, but the conclusion of the theorem remains true.
Proof. Using the technique of Shaaf in [34] that was outlined in the previous Section 4.4.1, recall that the linearized eigenvalue problem about a steady state solution of (4–1) can be reduced to the countable number of \((N + 1) \times (N + 1)\) matrix eigenproblems \(A(\mu)x = \lambda x\) from (4–26). In the following calculations we fix \(i > 0\) so that \(\mu_i < 0\) and denote \(\mu\) simply as \(\mu\). Denote the matrix in (4–26) as \(A = A(\mu)\) and let \(K := -\mu u^* \chi > 0\) be the entry in the upper right hand corner of \(A\). We can choose \(r > 0\), dependent on \(\mu\) but independent of \(K\) and \(\vec{\alpha}\), large enough so that \(B := A + rI\) has positive diagonal entries. Then since \(J\) was assumed to be Metzler, \(K > 0\) and \(\vec{\alpha}\) is a nonnegative vector, \(B\) is a non-negative matrix.

In fact, \(B\) is a nonnegative irreducible matrix. To see this, consider the associated directed graph \(G(B)\) on \(N + 1\) vertices \(v_0, v_1, ..., v_N\). The graph can be constructed by augmenting \(G(J)\), the associated directed graph of \(J\), with the vertex \(v_0\) corresponding to the first row and column of \(B\) and assuming that each vertex has a loop to account for the positive diagonal entries. Since \(J\) is irreducible, \(G(J)\) is strongly connected. Then since \(\vec{\alpha} \neq 0\) there exists a path from \(v_j\) to \(v_0\) for any \(1 \leq j \leq N\), and since \(K > 0\) there exists a path from \(v_0\) to \(v_j\) for any \(1 \leq j \leq N\). Hence \(G(B)\) is strongly connected.

We now claim that

\[
\lim_{K \to \infty} \rho(B) = \lim_{\alpha_i \to \infty} \rho(B) = \infty \tag{4–27}
\]

If (4–27) holds then the theorem holds by the following argument. Since \(B\) is nonnegative, by Theorem 4.7 \(\rho(B)\) is an eigenvalue of \(B\). Then if (4–27) holds, we may choose \(K\) (or \(\alpha_i\)) large enough so that \(\rho(B) > r\). Then \(\rho(B) - r > 0\) is an eigenvalue of \(A = B - rI\) and so \((u_s, \vec{v}_s)\) is linearly unstable.

So, we conclude the proof by verifying equation (4–27). First consider the case where \(u^* \chi\) is arbitrarily large. Before continuing, we note that for any integer \(m \geq 1\)

\[
\rho((B^t)^m) = (\rho(B^t))^m, \tag{4–28}
\]
so it is enough to show that
\[
\lim_{K \to \infty} \rho((B^t)^m) = \infty
\]  
(4–29)
for some \(m \geq 1\), take the \(m\)th root, and use \(\rho(B) = \rho(B^t)\).

Denote the entries of the transpose of the matrix \(B\) by
\[
B^t = (b^t)_{ij}.
\]
To show (4–29), we first observe the formula for matrix multiplication
\[
(B^t)^n_{ij} = \sum_I b^t_{i_1 j_1} b^t_{i_2 j_2} ... b^t_{i_{n-1} j}
\]  
(4–30)
where \(I\) is the set of all multi-indices \(i_1, i_2, ..., i_{n-1}\) with each index \(i_k\) such that \(0 \leq i_k \leq N\). Note that here we continue indexing the \(N + 1\) entries of \(B^t\) by \(0, 1, ..., N\) instead of the standard \(1, 2, ..., N + 1\) so the indices of the vertices match with their corresponding variable.

Next, recall that \(G(B)\) and hence \(G(B^t)\) is strongly connected. Therefore there is a path between every pair of vertices and, more specifically, for each \(1 \leq i \leq N\) there is a path from \(v_i\) to \(v_0\). By construction there is an edge from vertex \(i\) to vertex \(j\) if and only if \(b^t_{ij} > 0\), so also denote this edge, if it exists, by \(b^t_{ij}\). Finally observe that since every vertex of \(G(B^t)\) has a loop, if there is a path of length \(m\) from one vertex to another, there is a path of any length longer than \(m\) as well.

Using irreducibility of \(B^t\) along with the observations above, we can find a fixed integer \(m > 0\) such that there is a path of length \(m - 1\) from \(v_i\) to \(v_N\) for every \(0 \leq i \leq N\). Then adding the edge \(b^t_{N0}\) from \(v_N\) to \(v_0\) to each of these paths, we have a path of length \(m\) from \(v_i\) to \(v_N\) which is of the form
\[
b^t_{i_0 i_1} b^t_{i_1 i_2} ... b^t_{i_{m-1} i_m} > 0
\]
where $i_0 = i$ and $b_{i_{m-1}i_m}^t = b_{N0}^t = b_{0N} = K$. The product is positive by the definition of the graph and hypothesis 4. Let

$$C_i(K) = b_{i_0 i_1}^t b_{i_1 i_2}^t b_{i_2 i_3}^t \ldots b_{i_{m-2} i_{m-1}}^t, \quad 0 \leq i \leq N$$

be a constant dependent only on $K$, and notice that $C_i$ will only increase as $K$ increases, since its dependence on $K$ must be linear or higher. Then by formula (4–30) for matrix multiplication, there exists a term $C_i(K)$ in the $(i, 0)$ entry of $(B^t)^m$ for $0 \leq i \leq N$.

As in Theorem 4.8, let $s_i$ be the sum of the entries in the $i$-th row of $(B^t)^m$. Define

$$s = \min_{0 \leq i \leq N} s_i$$

and let $C(K) = \min_{0 \leq i \leq N} C_i(K)$. $C(K)$ also increases with $K$ since each $C_i(K)$ does. Because $B^t$ is a nonnegative matrix, all of the terms in each entry of $(B^t)^m$ will nonnegative, so $s$ satisfies

$$s = \min_{0 \leq i \leq N} s_i \geq \min_{0 \leq i \leq N} ((B^t)^m)_{i0} \geq \min_{0 \leq i \leq N} C_i(K)K = C(K)K.$$  

Then by Theorem 4.8, the spectral radius of $(B^t)^m$ satisfies

$$\rho((B^t)^m) \geq C(K)K.$$  

Letting $K \rightarrow \infty$, (4–27) and (4–29) follow for the case of $\chi u_+$. In the case where there is some fixed $i^*$ with $1 \leq i^* \leq N$ for which $\alpha_{i^*}$ is arbitrarily large, the proof follows in the exact same manner by considering the graph $G(B^t)$ and finding for each $0 \leq i \leq N$ a path of length $m$ from $v_i$ to $v_{i^*}$, which ends in the edge $\alpha_{i^*}$. \qed

In order to make one of the hypotheses of Theorem 4.9 less restrictive, we need to introduce some more terminology. Recall that a strongly connected component of a graph is a maximal strongly connected subgraph. Then consider the following equivalence relation on vertices of a graph. Let two vertices $v_i$ and $v_j$ be equivalent
if there is a directed path from \( v_i \) to \( v_j \) and a path from \( v_j \) to \( v_i \). Then notice the equivalence classes created by this equivalence relation consist exactly of the vertices of the strongly connected components of the graph. Using this language, we state the following definition from [2]

**Definition 4.7.** The classes of a nonnegative \( N \times N \) matrix \( A \) are the disjoint subsets

\[
\{i_1, i_2, ..., i_k\} \subset \{1, 2, ..., N\}
\]

corresponding to vertices of the equivalence classes of \( G(A) \). We also identify a class with its strongly connected component.

**Definition 4.8.** The classes of a nonnegative matrix \( A \) are the equivalence classes of its associated directed graph \( G(A) \) as described above.

As described in Section 2.3 of [2], note that by a reordering of vertices a permutation matrix \( P \) can be found so that

\[
T = PAP^t
\]

is a lower block triangular matrix with the blocks corresponding to the classes of \( A \).

Since the classes of \( G(A) \) are strongly connected, it follows that the diagonal blocks of \( T \) are irreducible. Using these facts we can drop the irreducibility assumption in Theorem 4.9 and replace it with a weaker assumption on paths in the chemical reaction network as follows.

**Theorem 4.11.** Assume the system (4–1) has a positive homogeneous steady state solution \((u^*, \vec{v}^*)\), and that the matrix \( J \) the vector \( \vec{\alpha} \) satisfy the following conditions.

1. There exists some \( 1 \leq i^* \leq N \) such that \( \alpha_{i^*} > 0 \) and there is a directed path from \( v_{i^*} \) to \( v_N \) in \( G(J^t) \).

2. \( J \) is Metzler

Then if either the product \( \chi u^* \) or \( \alpha_{i^*} \) is sufficiently large, \((u^*, \vec{v}^*)\) is linearly unstable.

**Proof.** As in the proof of Theorem 4.9, the problem is reduced to the series of eigenvalue problems in (4–26). Again fixing \( \mu_i < 0 \), denoting the matrix in (4–26) as
$A \equiv A(\mu_i)$ and choosing $r > 0$ large enough so that $B := A + rI$ has positive diagonal entries, we can consider the graph $G(B^t)$ as described in the proof of Theorem 4.9. Fix $i^*$ for which $\alpha_{i^*} > 0$ and there is a path from $v_{i^*}$ to $v_N$. Then in $G(B^t)$ it follows that there is an edge from $v_0$ to $v_{i^*}$ since $B^{t}_{0i^*} = \alpha_{i^*} > 0$. Also since $B^{t}_{N0} = K > 0$ there is an edge from $v_N$ to $v_0$. The path from $v_i$ to $v_N$ along with the two edges mentioned previously creates a cycle and shows that $v_0$, $v_{i^*}$, and $v_N$ are part of the same strongly connected component. Therefore there exists a permutation matrix $P$ such that

$$T = P B^t P^t$$

where $T$ is an upper block triangular matrix with some block $T_j$ corresponding to the strongly connected component containing $v_N$ and $v_i$. Since this component is strongly connected, it follows that $T_j$ is an irreducible matrix. Then viewing the vertices in $G(T_j)$ as their own chemical reaction network, the matrix $T_j$ falls into the case of the proof of Theorem 9 and hence has an arbitrarily large positive eigenvalue as $K \to \infty$ and as $\alpha_{i^*} \to \infty$, specifically one larger than $r$. However since the eigenvalues of a block triangular matrix are simply the eigenvalues of the diagonal blocks, this implies that $T$ has an eigenvalue larger than $r$ which implies $B^t$ and $B$ do as well. Hence $A = B - rI$ has a positive eigenvalue.

4.4.3 Sufficient Conditions for Stability

As mentioned in the introduction, the idea of initiation of pattern formation by destabilization of a homogeneous steady state dates back to Turing’s classic paper [41]. We have shown necessary conditions for a constant steady state to be linearly unstable, so what would make this result of even more interest with regards to pattern formation is if the same homogeneous steady state is unstable without the chemotactic feedback term. In this section we show necessary conditions for a steady state to be linearly stable with respect to certain reasonable perturbations. A similar analysis was explained in [34] for the case of the simple KS model of Chapter 3.
The restriction of perturbations follows from the fact that if \( u \) satisfies (4–1a) with either Neumann boundary conditions on \( u \) and \( v_N \) or no flux boundary conditions on \( u \), then it must be that \( \int_\Omega u(x,t)dx = \int_\Omega u_0(x)dx \) is constant in time. This can be seen using (4–1a), integration by parts and boundary conditions as follows.

\[
\frac{d}{dt} \int_\Omega u(x,t)dx = \int_\Omega u_t(x,t)dx
\]

\[
= \int_\Omega D\Delta u - \nabla \cdot (\chi u \nabla v_N)dx
\]

\[
= D \int_{\partial \Omega} \left( \frac{\partial u}{\partial n} - \chi u \frac{\partial v_N}{\partial n} \right) dx
\]

\[
= 0
\]

So since \( \int_\Omega u dx \) is constant for any solution, when determining linear stability it makes sense to look only at perturbations \( U \) which satisfy \( \int_\Omega U(x,t)dx = 0 \).

We consider stability in the case where \( \vec{g}(\vec{v}) \) in (4–1b) is a linear function, and let \( A \in \mathbb{R}^N \times \mathbb{R}^N \) be the matrix such that \( \vec{g}(\vec{v}) = A\vec{v} \). This limits the chemical reaction networks to reactions of the form \( v_i \to v_j \) or \( v_i \to \emptyset \), but these component reactions may be combined arbitrarily and still generate a linear function under the law of mass action. Consider the previous two reaction types in the setting of mass action kinetics, and let the rate constant for the reaction going to chemical \( i \) from chemical \( j \) be \( k_{ij} \geq 0 \). Similarly let the rate constant for the decay reaction \( v_i \to \emptyset \) be denoted by \( \gamma_i \geq 0 \). Then for every \( 1 \leq i \leq N \), the function \( \vec{g}(\vec{v}) \) will be given by

\[
g_i(\vec{v}) = \sum_{j=1,j\neq i}^{N} k_{ij}v_j - \sum_{j=1,j\neq i}^{N} k_{ji}v_i - \gamma_i v_i.
\]
Then the matrix formulation of \( \vec{g} \) will be

\[
\vec{g}(\vec{v}) = \begin{pmatrix}
-\gamma_1 - \sum_{j \neq 1} k_{j1} & k_{12} & k_{13} & \cdots & k_{1N} \\
k_{21} & -\gamma_2 - \sum_{j \neq 2} k_{j2} & k_{23} & \cdots & k_{2N} \\
k_{31} & k_{32} & -\gamma_3 - \sum_{j \neq 3} k_{j3} & \cdots & k_{3N} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
k_{N1} & k_{N2} & k_{N3} & \cdots & k_{NN}
\end{pmatrix}
\begin{pmatrix}
v_1 \\
v_2 \\
v_3 \\
\vdots \\
v_N
\end{pmatrix}
\]

which implies that the Jacobian of \( \vec{g} \) with respect to \( \vec{v} \) is

\[
J = \begin{pmatrix}
-\gamma_1 - \sum_{j \neq 1} k_{j1} & k_{12} & k_{13} & \cdots & k_{1N} \\
k_{21} & -\gamma_2 - \sum_{j \neq 2} k_{j2} & k_{23} & \cdots & k_{2N} \\
k_{31} & k_{32} & -\gamma_3 - \sum_{j \neq 3} k_{j3} & \cdots & k_{3N} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
k_{N1} & k_{N2} & k_{N3} & \cdots & k_{NN}
\end{pmatrix}
\]

Before proving a theorem on stability in the linear case, we need two preliminary results.

**Lemma 4.5.** Assume that the kinetics of a system of chemicals of densities \( v_1, v_2, \ldots, v_N \) are represented by a linear function \( \vec{g} : \mathbb{R}^N \to \mathbb{R}^N \) which is derived using the law of mass action kinetics. Assume also that the matrix \( J \), the Jacobian of \( \vec{g} \) with respect to \( \vec{v} \), is irreducible, and that at least one chemical decays at a rate \( \gamma_i > 0 \). Then the eigenvalues of \( J \) will all have strictly negative real parts.

**Proof.** If the chemical reaction network is as described, then the Jacobian \( J \) will be of the form (4–31). This is a Metzler matrix since each \( k_{ij} \geq 0 \), and so there exists some real number \( s > 0 \) such that \( J + sI \) is a nonnegative irreducible matrix. Then we can apply the Perron Frobenius Theorem to conclude the existence of an eigenvalue \( \lambda_p > 0 \) of \( J + sI \) corresponding to a positive eigenvector \( x_p \) such that \( \rho(J + sI) = \lambda_p \). Since \( \lambda_p > 0 \) is the dominant eigenvalue, notice this implies all eigenvalues \( \lambda \) of \( J + sI \) are in a
disc of radius $\lambda_p$ centered at 0 in the complex plane and so will satisfy
\[
\text{Re}(\lambda) < \lambda_p. \tag{4–32}
\]

Now for any eigenvalue $\lambda$ of $J + sI$,
\[
(J + sI)x = \lambda x
\]
holds, which is true if and only if
\[
Jx = (\lambda - s)x. \tag{4–33}
\]
Hence $\lambda$ is an eigenvalue of $J + sI$ with eigenvector $x$ if and only if $\lambda - s$ is an eigenvalue of $J$ with eigenvector $x$. Therefore if $\nu$ is any eigenvalue of $J$, then $\nu + s$ is an eigenvalue of $J + sI$ which implies by (4–32) that
\[
\text{Re}(\nu + s) = \text{Re}(\nu) + s < \lambda_p. \tag{4–34}
\]

Now, noticing that since $J$ is of the form (4–31), it is easy to see that
\[
\begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} J = \begin{pmatrix} -\gamma_1 & -\gamma_2 & \cdots & -\gamma_N \end{pmatrix}.
\]
Hence multiplying (4–33) with $x = x_p$ and $\lambda = \lambda_p$ by a row vector of ones shows that
\[
0 > \begin{pmatrix} -\gamma_1 & -\gamma_2 & \cdots & -\gamma_N \end{pmatrix} x_p = (\lambda_p - s) \sum_{i=1}^{N} x^i
\]
which implies that $\lambda_p - s < 0$ since $\sum x_i > 0$. Combining this fact with (4–34), we see
\[
\text{Re}(\nu) < \lambda_p - s < 0
\]
for any eigenvalue $\nu$ of $J$.

\[\blacklozenge\]

**Lemma 4.6.** If $J$ is an irreducible Metzler matrix of the form (4–31) with $\gamma_i > 0$ for some $1 \leq i \leq N$, $\mu < 0$ is fixed, and $\tilde{D}$ is a nonnegative diagonal matrix, then $J + \mu \tilde{D}$ is also a Metzler matrix with all negative eigenvalues. Furthermore, every eigenvalue $\lambda$ of $J + \mu \tilde{D}$
satisfies $\text{Re}(\lambda) < m < 0$ where

$$m := -\max_{1 \leq j \leq N} \gamma_j + \mu \min_{1 \leq j \leq N} \tilde{D}_j$$

**Proof.** Clearly $J + \mu \tilde{D}$ is still a Metzler matrix since $\tilde{D}$ is diagonal by definition and hence the off diagonal components of $J$ are not affected. Then as in the proof of Lemma 4.5, $J + \mu \tilde{D}$ has a positive eigenvector $x_p$ corresponding to some eigenvalue $\lambda_p$ so that

$$\text{Re}(\lambda) < \lambda_p$$

for all eigenvalues $\lambda$ of $J + \mu \tilde{D}$. Since $x_p$ is positive it has positive sum, and so without loss of generality we may “normalize” $x_p$ so that $\sum_{i=1}^{N} x_{p,i} = 1$. Then multiplying

$$(J + \mu \tilde{D})x_p = \lambda_p x.$$ 

through by the row vector of all ones and recalling that

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} J = \begin{pmatrix} -\gamma_1 & -\gamma_2 & \cdots & -\gamma_N \end{pmatrix},$$

we have

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} (J + \mu \tilde{D})x_p = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \lambda_p x$$

which implies that

$$\lambda = -\sum_{i=1}^{N} \gamma_i x_{p,i} + \mu \sum_{i=1}^{N} \tilde{D}_i x_{p,i} < 0.$$ 

Since $\text{Re}(\lambda) < \lambda_p$ for all eigenvalues of $J + \mu \tilde{D}$ we conclude that all eigenvalues of $J + \mu \tilde{D}$ have negative real part. To show the bound on the eigenvalues, calculate for any...
\( \lambda \) an eigenvalue of \( J + \mu \tilde{D} \)

\[
\text{Re}(\lambda) < \lambda_p = -\sum_{i=1}^{N} \gamma_i x_{i,p} + \mu \sum_{i=1}^{N} \tilde{D}_i x_{i,p} \\
\leq -\max_{1 \leq j \leq N} \gamma_j \sum_{i=1}^{N} x_{p,i} + \mu \min_{1 \leq j \leq N} \tilde{D}_j \sum_{i=1}^{N} x_{p,i} \\
= m.
\]

\[\square\]

**Theorem 4.12.** Assume that \((u_*, \vec{v}_*)\) is a nonnegative, homogeneous steady state of (4–1), and that \(\vec{g}\) in (4–1b) is a linear function derived using the law of mass action such that the Jacobian matrix \(J\) evaluated at the steady state \(\vec{v}_*\) is irreducible. Assume also that \(\gamma_i > 0\) for some \(1 \leq i \leq N\). If \(\chi = 0\), then \((u_*, v_*)\) is linearly stable with respect to perturbations satisfying \(\int_{\Omega} u(x, t)dx = 0\).

**Proof.** Using Theorem 4.6 and the technique of Schaaf in [34] outlined at the beginning of this section, recall that we can reduce the problem of finding eigenvalues of the linearized general KS system (4–23) about a nonnegative homogeneous steady state \((u_*, v_*)\) to the family of eigenvalue problems (4–26). In this specific case, we see from (4–31) that we are searching for eigenvalues of the family of matrices

\[
A(\mu_i) = \begin{bmatrix}
D_{\mu_1} & 0 & 0 & \cdots & 0 & 0 \\
\alpha_1 & J_{11} + \mu_i \tilde{D}_1 & J_{12} & \cdots & J_{1,N-1} & J_{1N} \\
\alpha_2 & J_{21} & J_{22} + \mu_i \tilde{D}_2 & \cdots & J_{2,N-1} & J_{2N} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\alpha_N & J_{N1} & J_{N2} & \cdots & J_{N,N-1} & J_{NN} + \mu_i \tilde{D}_N
\end{bmatrix}, \quad 1 \leq i \leq N.
\]

Since each \(A(\mu_i)\) is a block lower triangular matrix, the eigenvalues of \(A_i\) will be \(D_{\mu_i} \leq 0\) and the eigenvalues of \(J + \mu_i \tilde{D}\). We see from (4–31) that \(J\) is Metzler. Hence by Lemma 4.6 for \(i = 0, 1, 2, \ldots\) the eigenvalues of \(J + \mu_i \tilde{D}\) all have strictly negative real part. Also by Lemma 4.6 any eigenvalue \(\lambda\) of \(J + \mu_i \tilde{D}\) is bounded away from zero uniformly in
Recall that $0 = \mu_0 < \mu_1 \leq \mu_2 \leq \cdots$ so that $D\mu_i \leq \mu_1 < 0$ for all $i \geq 1$. So with the exception the eigenvalue $D\mu_0 = 0$, all eigenvalues of every $A(\mu_i)$ have strictly negative real part and do not have zero as a limit point.

So, it only remains to show that any eigenfunction $\begin{pmatrix} u \\ v \end{pmatrix}$ corresponding to the eigenvalue $\lambda = 0$ violates the condition $\int_{\Omega} u(x,t)dx = 0$. To show this, let $\vec{x} = [x_0 x_1 \cdots x_N]^t \neq 0$ be the eigenvector of $A(\mu_0)$ corresponding to the eigenvalue $\lambda = 0$. Notice that if $x_0 = 0$, then it must be that $[x_1 x_2 \cdots x_N]^t \neq 0$ and so

$$J \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = 0.$$ 

However this implies that 0 is an eigenvalue of $J$, which is a contradiction based on Lemma 4.5, so we conclude $x_0 \neq 0$. By Theorem 4.6, the eigenfunction corresponding to $\lambda = 0$ is $\begin{pmatrix} u \\ v \end{pmatrix}$ where $u = x_0 \omega_1$

is a nonzero function. However

$$\int_\Omega u(x,t)dx = \int_\Omega x_0 \omega_1(x,t)dx = \int_\Omega x_0 = |\Omega| x_0 \neq 0.$$ 

\[\Box\]

### 4.5 Examples

In this section we review some examples which we explored in [8]. The first example is a large class of CRNs to which we may apply Theorem 4.11, while the second is a specific case in which the same theorem may be applied. The final
example shows a case where one of the hypotheses of Theorem 4.11 is violated, but the conclusion still remains true.

**Example 4.13.** Again consider the case where \( \vec{g}(\vec{v}) \) in (4–1b) is a linear function and let \( A \in \mathbb{R}^N \times \mathbb{R}^N \) be the matrix such that \( \vec{g}(\vec{v}) = A\vec{v} \). With terminology as in Section 4.4.3, recall that the Jacobian of \( \vec{g} \) will be given by (4–31). We want to look at the stability of positive steady states under these conditions. Now, since \( A \) is Metzler, it follows that \(-A\) is of the form of an \( M \)-matrix, so if it is in fact a nonsingular \( M \)-matrix then Lemma 4.4 shows the existence of nonnegative homogeneous steady states. In this setting we have the following corollary on the stability of positive homogeneous steady states.

**Corollary 4.1.** Assume that (4–1) has a positive homogeneous steady state solution \((u_*, \vec{v}_*)\), and that the function \( \vec{g} \) is linear and obtained by the law of mass action kinetics. If for some chemical \( v_i^* \) there exists a path in the chemical reaction network from \( v_i^* \to v_N \) and \( \alpha_{i^*} > 0 \), then \((u_*, \vec{v}_*)\) is linearly unstable whenever \( \chi u_* \) or \( \alpha_{i^*} \) is sufficiently large.

**Proof.** We must verify the conditions of Theorem 4.11. The path in the CRN from \( v_i^* \) to \( v_N \) implies that there exists positive reaction rate constants

\[
k_{i_1i_1}, \ k_{i_2i_1}, \ k_{i_3i_2}, \ \cdots, \ k_{Ni_{i-1}}
\]

for some positive integer \( n \). By (4–31), this corresponds exactly to a path in \( G(J^t) \) from \( v_i^* \) to \( v_N \). Furthermore, by hypothesis \( \alpha_{i^*} > 0 \), so the first condition is satisfied. Since each \( \gamma_i \geq 0 \) and \( k_{ij} \geq 0 \), by (4–31) \( J \) is Metzler and the second condition is satisfied. \( \square \)

Furthermore, notice that in the case where \( J \) is irreducible Theorem 4.12 will apply. Hence positive homogeneous steady states in the irreducible case are always stable when \( \chi = 0 \), and can always be destabilized by strong enough chemotactic feedback according to Corollary 4.1.

**Example 4.14 (Dimerization with decay).** We return again to Example 4.1 to explore the existence and stability of positive homogeneous steady states. Recall that the
dimerization reaction is given by
\[ 2v_1 \leftrightarrow v_2. \]
Additionally, assume decay of chemicals \( v_1 \) and \( v_2 \) at rates \( \gamma_1 \geq 0 \) and \( \gamma_2 > 0 \) respectively. Denote the rate constant for the forward reaction by \( k_1 \) and the reverse reaction by \( k_2 \). Then according to the law of mass action kinetics, the function \( \bar{g}(\vec{v}) \) describing the CRN kinetics will be
\[
\bar{g}(\vec{v}) = \begin{pmatrix}
- k_1 v_1^2 + k_2 v_2 - \gamma_1 v_1 \\
 k_1 v_1^2 - k_2 v_2 - \gamma_2 v_2 
\end{pmatrix}.
\]
We also assume that the species of density \( u \) produces \( v_1 \) at a rate \( \alpha_1 > 0 \) and that the chemical \( v_2 \) is not produced so that \( \alpha_1 = 0 \). Considering (4–1) in this setting, we know by (4–19) that homogeneous steady states are solutions to the system
\[
\begin{align*}
\alpha_1 u - k_1 v_1^2 + k_2 v_2 - \gamma_1 v_1 &= 0, \quad (4–35a) \\
k_1 v_1^2 - (k_2 + \gamma_2) v_2 &= 0. \quad (4–35b)
\end{align*}
\]
We solve (4–35) to verify that positive steady states exist. Adding (4–35a) to (4–35b) to obtain a new first equation gives the equivalent system
\[
\begin{align*}
\alpha_1 u &= \gamma_1 v_1 + \gamma_2 v_2, \quad (4–36a) \\
v_2 &= \frac{k_1}{k_2 + \gamma_2} v_1^2. \quad (4–36b)
\end{align*}
\]
Then substituting (4–36b) into (4–35a) shows that \( v_1 \) must be a solution to the quadratic equation
\[
\frac{k_1 \gamma_2}{k_2 + \gamma_2} v_1^2 + \gamma_1 v_1 - \alpha_1 u = 0. \quad (4–37)
\]
It can be directly verified using the quadratic equation that (4–37) has a unique positive root whenever \( u > 0 \). Denoting this root by \( r(u) \), we see that there is a family of positive homogeneous steady state solutions that can be indexed by \( u \). So given any constant
$u > 0$, a positive steady state solution to (4–1) is given by

$$u > 0, \quad v_1 = r(u) > 0, \quad v_2 = \frac{k_1}{k_2 + \gamma_2} r^2(u) > 0.$$ 

Hence we see that positive homogeneous stationary solutions exist, and we will now show that Theorem 4.11 applies to this example. The Jacobian of $\vec{g}$ will be given by

$$J = \begin{bmatrix}
-k_1 v_1 v_2 + k_2 v_3 - \gamma_1 v_1 & 2k_1 v_1 v_2 - k_2 v_3 \\
-2k_1 v_1 v_2 + k_2 v_3 & -k_2 - \gamma_2
\end{bmatrix},$$

which is a Metzler matrix. Also since $k_2 > 0$ and $2k_1 v_1 > 0$, it follows that there is a path in $G(J^t)$ from $v_1$ to $v_2$ and from $v_2$ to $v_1$. Hence the hypotheses of Theorem 4.11 are verified, and we conclude that the positive homogeneous steady states in this example will be linearly unstable for large enough values of $\chi u$ or $\alpha_1$.

**Example 4.15.** As our last example in this section, we demonstrate a case where the hypotheses of Theorem 4.11 are not satisfied but the conclusion still remains valid. This shows that while the conditions of the theorem are sufficient, they are not necessary.

Consider the reaction

$$v_1 + v_2 \leftrightarrow v_3.$$ 

We again assume that the species $u$ produces the first chemical $v_1$ at a rate $\alpha_1 > 0$, but does not produce the other two chemicals so that $\alpha_1 = \alpha_2 = 0$. Also assume the decay rate of $v_1$ is $\gamma_1 > 0$ while the decay rates of $v_2$ and $v_3$ satisfy $\gamma_2 = \gamma_3 = 0$.

In this case the function $\vec{g}(\vec{v})$ will be given by

$$\vec{g}(\vec{v}) = \begin{pmatrix}
-k_1 v_1 v_2 + k_2 v_3 - \gamma_1 v_1 \\
-k_1 v_1 v_2 + k_2 v_3 \\
k_1 v_1 v_2 - k_2 v_3
\end{pmatrix},$$
Again we first verify that positive homogeneous steady states exist. By (4–19), this means we must find positive solutions to the system

\(-k_1 v_1 v_2 + k_2 v_3 - \gamma_1 v_1 + \alpha u = 0, \tag{4–38a}\)

\(-k_1 v_1 v_2 + k_2 v_3 = 0 \tag{4–38b}\)

\(k_1 v_1 v_2 - k_2 v_3 = 0. \tag{4–38c}\)

Notice that (4–38b) and (4–38c) are equivalent, and so we only need to search for solutions to (4–38a)–(4–38b). Then subtracting (4–38b) from (4–38a) we obtain the equivalent system

\(\gamma_1 v_1 = \alpha u, \tag{4–39a}\)

\(k_1 v_1 v_2 = k_2 v_3. \tag{4–39b}\)

Hence we see that there are an infinite number of positive homogeneous steady solutions, indexed by arbitrary positive constants \(u > 0\) and \(v_2 > 0\). These stationary solutions are given by the formulas

\(u > 0, \quad v_1 = \frac{\alpha}{\gamma_1} u, \quad v_2 > 0, \quad v_3 = \frac{\alpha k_1}{\gamma_1 k_2} uv_2.\)

We now turn to determining the stability of these positive homogeneous stationary solutions. Since the Jacobian evaluated at some positive homogeneous steady state \((\bar{u}_s, \bar{v}_s)\) is

\(J = \begin{bmatrix}
-k_1 v_{s,2} - \gamma_1 & -k_1 v_{s,1} & k_2 \\
-k_1 v_{s,2} & -k_1 v_{s,1} & k_2 \\
k_1 v_{s,2} & k_1 v_{s,1} & -k_2
\end{bmatrix},\)

which is not in general a Metzler matrix, we cannot directly apply any of our stability results. However, we may still apply Theorem 4.6 to reduce the linearized eigenvalue
problem to finding the eigenvalues of the matrices

$$A(\mu) = \begin{bmatrix}
\mu D & 0 & 0 & -\chi u_* \mu \\
\alpha_1 & -k_1 v_{*,2} + \mu \tilde{D}_1 - \gamma_1 & -k_1 v_{*,1} & k_2 \\
0 & -k_1 v_{*,2} & -k_1 v_{*,1} + \mu \tilde{D}_2 & k_2 \\
0 & k_1 v_{*,2} & k_1 v_{*,1} & -k_2 + \mu \tilde{D}_3
\end{bmatrix}$$

for eigenvalues $\mu$ of the Laplacian with Neumann boundary conditions. As in previous calculations, we define $K = \chi \mu u_* < 0$. We will first consider the case when $K = 0$, find conditions on the eigenvalues, and then show that these conditions will change as $K$ becomes arbitrarily large.

Fix some eigenvalue $\mu < 0$ of the Laplacian as in (4–24). For the $K = 0$ case, the matrix $A(\mu)$ is block diagonal with eigenvalues $\mu D < 0$ and the eigenvalues of the lower right $3 \times 3$ submatrix. This submatrix can be written as

$$\begin{bmatrix}
-a - d_1 & -b & c \\
-a & -b - d_2 & c \\
a & b & -c - d_3
\end{bmatrix}$$

using the positive numbers $a = k_1 v_{*,2}, b = k_1 v_{*,1}, c = k_2, d_1 = \gamma_1 - \mu \tilde{D}_1, d_2 = -\mu \tilde{D}_2$, and $d_3 = -\mu \tilde{D}_3$. Then characteristic equation of (4–41) is

$$\lambda^3 + b_2 \lambda^2 + b_1 \lambda + b_0 := \lambda^3 + (a + b + c + d_1 + d_2 + d_3) \lambda^2 \\
+ (a(d_2 + d_3) + b(d_1 + d_3) + c(d_1 + d_2) + d_1 d_2 + d_1 d_3 + d_2 d_3) \lambda \\
+(a d_2 d_3 + b d_1 d_3 + c d_1 d_2 + d_1 d_2 d_3) = 0.$$
We apply the Routh-Hurwitz criteria (see for example [13]), which says that all
eigenvalues of a matrix have negative real part if and only if the inequalities

\[ b_2 > 0 \] (4–42a)
\[ b_0 > 0 \] (4–42b)
\[ b_1 b_2 - b_0 > 0 \] (4–42c)

hold. The inequalities (4–42a) and (4–42b) are immediate, and (4–42c) follows by direct
calculation. Thus whenever \( \mu < 0 \) and \( K = 0 \) all eigenvalues of \( A(\mu) \) have negative real
part. This will not be true for sufficiently large values of \( K > 0 \), as we now show.

Considering the case where \( K > 0 \) and \( \mu < 0 \) is still fixed, we use cofactor
expansion on the first row of (4–40) to compute

\[
\det (A(\mu)) = C - K \det \begin{bmatrix}
\alpha_1 & -k_1 v_{*,2} + \mu \tilde{D}_1 - \gamma_1 & -k_1 v_{*,1} \\
0 & -k_1 v_{*,2} & -k_1 v_{*,1} + \mu \tilde{D}_2 \\
0 & k_1 v_{*,2} & k_1 v_{*,1}
\end{bmatrix}
\]

where \( C > 0 \) is the determinant of the matrix \( A \equiv A(\mu) \) in the case \( K = 0 \), and is thus
independent of \( K \). Then expanding determinant of the submatrix to investigate the \( K \)
dependence, we see the determinant is some function of \( K \) given by

\[
f(K) := \det(A(\mu, K)) = C + K \alpha_1 k_1 v_{*,2} \mu \tilde{D}_2
\]

where \( K \alpha_1 k_1 v_{*,2} \mu \tilde{D}_2 < 0 \). Hence for some sufficiently large value of \( K_0 \) of \( K \), we
have \( f(K_0) < 0 \) while for \( K = 0 \), \( F(0) > 0 \). We claim this implies that for some \( K_1 \)
with \( 0 < K_1 \leq K_0 \), at least one eigenvector must have positive real part. Since the
determinant is the product of the eigenvalues of the matrix, it is clear if all eigenvalues
are real at least one must change signs if the product is to change signs. Indeed, since
all eigenvalues are negative in the \( K = 0 \) case, this means one must become positive
over the interval 0 to \( K_0 \). In the case that some eigenvalue has nonzero imaginary
parts, then because the eigenvalues will appear in conjugate pairs which have positive product, it follows that exactly two of the eigenvalues must be real and have opposite signs in order for the determinant to ever be negative. We have now shown that for sufficiently large values of $K$, the initially stable positive homogeneous steady state will become linearly unstable.

Along with the case where $\vec{g}$ is a linear function and the Jacobian of $\vec{g}$ is irreducible at the steady state, we now have two examples of how chemotactic feedback in a general CRN can destabilize homogeneous steady states. So, both of these examples are cases where chemotactically induced instability could potentially lead to pattern formation.
As in the two equation case, we employ the finite element method (FEM) to search for nonhomogeneous stationary solutions of Equation 4–1 and to implement time simulations. In this chapter we give a generalization of the methods used in Section 3.3 to find nonhomogeneous stationary solutions. We also present time simulations using both a standard FEM and a method using a non-standard discretization in order to preserve positivity and reduce numerical instability. We will prove numerical nonnegativity in the case of the simple two equation Keller-Segel (KS) model, and also in the generalized model under certain conditions on the function $\vec{g}$ representing the chemical reaction network (CRN). Throughout this chapter we assume that $\Omega$ is a polygonal domain so that it can be meshed exactly.

## 5.1 Spectral Bands

Recall that in Section 3.2 we used results from [34] and [18] to determine the stability of spatially homogeneous solutions $(u_*, \vec{v}_*)$ to (3–15). There exists a range of values for which $u_*$ is large enough so that the homogeneous steady state of (3–15) is unstable, but small enough to avoid blow-up solutions. Only within this range of parameters were we able to find convergence to nonhomogeneous steady states. Inspired by Theorem 4.11, we hope to generalize this method to solutions of the system (4–1) to determine for which values of the constant steady state $u_*$ eigenvalues of the linearized problem become positive. We then search for solutions in the neighborhood where the sign change takes place.

In order to visualize the spectrum of linearized operators of the form 4–22, we once again turn to Theorem 4.6 to reduce the infinite dimensional eigenproblem to an indexed series of $(N + 1) \times (N + 1)$ matrices where $N$ is the number of chemicals being modeled. We may calculate all eigenvalues of 4–22 about a given homogeneous steady state $(u_*, \vec{v}_*)$ by numerically finding for $i = 1, 2, 3, ...$ eigenvalues of
the matrices $A(\mu_i)$ in (4–26). Recall that in many of our examples, we found that positive or nonnegative steady states could be indexed by the value of the constant $u_*$. For any of these examples we can produce graphs visualizing the spectrum using the following procedure.

1. Fix some $u_* \geq 0$ and determine the positive homogeneous steady state corresponding to this $u_*$. 

2. For some finite number $I$, calculate all eigenvalues $\{\lambda_j\}_{j=1}^{N+1}$ of the matrix $A(\mu_i, u_*)$ for $i = 0, 1, \ldots I$. 

3. Let $m(u_*, \mu_i) = \max_{1 \leq j \leq N+1} \text{Re}(\lambda_j)$. 

4. For each $i = 0, 1, \ldots, I$, plot the point $(u_*, m(u_*, \mu_i))$. 

5. Repeat for some finite number of indices $i$ for $u_*$ in a range $[u_{\text{start}}, u_{\text{stop}}]$. 

By plotting the first 10, 20 or more bands over various ranges $[u_{\text{start}}, u_{\text{stop}}]$, we can see where bands cross the $x$ axis and determine possible values of $u_*$ for which the linearized eigenproblem may produce eigenfunctions which lead to spatially nonhomogeneous stationary solutions.

We now return to some previous examples in order to see the results of the spectral band computations. Consider again Example 4.1 where $\vec{g}$ arises from a dimerization reaction. In Example 4.14 we further explored this example to show the existence of an infinite family of steady states which can be indexed by $u > 0$ and are given by

$$u > 0, \quad v_1 = r(u) > 0, \quad v_2 = \frac{k_1}{k_2 + \gamma_2} r^2(u) > 0$$

where $r(u)$ is the unique positive root of a specific quadratic equation. Following the algorithm above for $i = 1, \ldots, 20$, performing all calculations on the unit square and setting all nonzero constants equal to unity, Figure 5-1 is obtained. Upon closer inspection, it can be seen that the line corresponding to the maximum real parts of eigenvalues of $m(u_*, \mu_1)$ crosses the $x$-axis at around $u_{\text{crit}} = 22.14$. 

88
Figure 5-1. An example of the spectral bands of the matrices $M(\mu_i)$ corresponding to the dimerization reaction for $1 \leq i \leq 20$.

Turning next to the full four equation KS model discussed in Example 4.2, we can obtain a similar spectral band plot given a few assumptions on the parameters. The validity of these assumptions, as well as computations done with parameter values approximated from the literature, will be further explored in Section 6.3. For the time being we assume $\gamma_2 = \gamma_3 = 0$ and so consider (4–1) with $\vec{g}$ given by

$$
\vec{g}(\vec{v}) = \begin{pmatrix}
-k_1v_1v_3 + (k_2 + k_3)v_2 - \gamma_1v_1 \\
k_1v_1v_3 - (k_2 + k_3)v_2 \\
-k_1v_1v_3 + k_2v_2
\end{pmatrix}
$$

and $\vec{\alpha}$ given by

$$
\vec{\alpha} = \begin{pmatrix}
\alpha_1 \\
0 \\
\alpha_3
\end{pmatrix}.
$$

We first want to show that the full, four equation KS model has positive homogeneous steady state solutions. By (4–19) this problem reduces to showing that positive solutions
exist. Adding (5–1a) and (5–1b), we immediately see that given a fixed constant \( u > 0 \), we have
\[ v_1 = \frac{\alpha_1}{\gamma} u > 0. \]
Similarly (5–1b) and (5–1c) imply that
\[ v_2 = \frac{\alpha_3}{k_3} u > 0. \]
Then substituting these two values for \( v_1 \) and \( v_2 \) into (5–1b) and assuming that \( \gamma > 0 \) and \( k_3 > 0 \), we see that so long as our fixed \( u > 0 \),
\[ v_3 = \frac{(k_2 + k_3)\alpha_3\gamma}{\alpha_1 k_1 k_3} > 0, \]
a constant which is independent of the value chosen for \( u \). Hence we see that for any given constant value \( u_* > 0 \), there exists a unique positive vector \( \bar{v} \in \mathbb{R}^3 \) such that \((u, \bar{v})\) is a homogeneous steady state solution of the four equation KS model.

In this particular case the \( 4 \times 4 \) matrix \( A(\mu_i) \) from (4–26) will be given by
\[
\begin{bmatrix}
D\mu_i & 0 & 0 & -\chi u_* \mu_i \\
\alpha_1 & -k_1 v_3^* - \gamma + \mu_i \bar{D}_1 & k_2 + k_3 & -k_1 v_1^* \\
0 & k_1 v_3^* & -(k_2 + k_3) + \mu_i \bar{D}_2 & k_1 v_1^* \\
\alpha_3 & -k_1 v_3^* & k_2 & -k_1 v_1^* + \mu_i \bar{D}_3 \\
\end{bmatrix}.
\]
(5–2)

Setting all nonzero constants equal to unity and performing the procedure described above on the matrix (5–2), we obtain Figure 5-2.
5.2 Eigenfunctions and Potential Patterns

Using the method described in Section 5.1, we can find exact values for positive eigenvalues of the linearized system (4–23) about a homogeneous steady state \((u^*, \vec{v}^*)\). We next use the standard finite element discretization of (4–23), as explained in Section 3.3, to solve for approximate eigenvectors corresponding to these positive eigenvalues. Since the positive eigenvalues represent directions of instability, these eigenfunctions will represent potential patterns formed by the chemotactic instability. We also hope that time simulations will show some of these may indicate patterns of nonconstant steady states. We explore this last idea further in Section 6.2.2, but for now we simply demonstrate eigenfunctions arising from the positive eigenvalues seen in Figures 5-1 and 5-2. For now we will assume that all constants are equal to unity in order to qualitatively understand patterns in a setting similar to that of the minimal system in Equation (3–15). In Section 6.3, we further investigate biologically realistic parameter values and explore the effects of changing parameter values.

To show how to compute these eigenfunctions, we again return to the full KS model introduced in Example 4.2 with \(\gamma_2 = \gamma_3 = 0\). Considering this specific case, the weak
eigenvalue problem in the form of 4–23 is given by

\[- \int D \nabla U \cdot \nabla \phi + \chi u_\ast \int \nabla V_3 \cdot \nabla \phi = \lambda \int U \phi \]

(5–3a)

\[- \tilde{D}_1 \int \nabla V_1 \cdot \nabla \phi + \alpha_1 \int U \phi - (k_1 v_3^* + \gamma) \int V_1 \phi - (k_2 + k_3) \int V_2 \phi - k_1 v_1^* \int V_3 \phi = \lambda \int V_1 \phi \]

(5–3b)

\[- \tilde{D}_2 \int \nabla V_2 \cdot \nabla \phi + k_1 v_3^* \int V_1 \phi - (k_2 + k_3) \int V_2 \phi + k_1 v_1^* \int V_3 \phi = \lambda \int V_2 \phi \]

(5–3c)

\[- \tilde{D}_3 \int \nabla V_3 \cdot \nabla \phi + \alpha_3 \int U \phi - k_1 v_3^* \int V_1 \phi + k_2 \int V_2 \phi - k_1 v_1^* \int V_3 \phi = \lambda \int V_3 \phi. \]

(5–3d)

Using notation as in the introduction and Section 3.3, consider the linear finite element subspace \( S_h \) with basis \( \{ \phi_\ell \}_{\ell=1}^{L'} \), and let \( L \) be the \( L' \times L' \) stiffness matrix defined by

\[ L_{ij} = \int \nabla \phi_j \cdot \nabla \phi_i. \]

Similarly let \( M \) be the \( L' \times L' \) mass matrix defined by \( M_{ij} = \int \phi_j \phi_i. \)

Define the approximations \( u_h \) and \( v_h \) in \( S_h \) by

\[ u \approx u_h = \sum_\ell U^n_\ell \phi_\ell, \quad v_i \approx v^n_{h,i} = \sum_\ell V^n_{i,\ell} \phi_\ell. \]

Then the matrix formulation of the system (5–3) for a solution \((u_h, v_h)\) is

\[
C \begin{bmatrix} \bar{U} \\ \bar{V}_1 \\ \bar{V}_2 \\ \bar{V}_3 \end{bmatrix} = \lambda E \begin{bmatrix} \bar{U} \\ \bar{V}_1 \\ \bar{V}_2 \\ \bar{V}_3 \end{bmatrix}
\]

(5–4)
where the $4L' \times 4L'$ matrices $C$ and $E$ are defined by

$$C = \begin{bmatrix}
-DL & 0 & 0 & \chi u \cdot L \\
\alpha_1 M & -\tilde{D}_1 L - (k_1 v_3^* + \gamma) M & (k_2 + k_3) M & -k_1 v_1^* M \\
0 & k_1 v_3^* M & -\tilde{D}_2 L - (k_2 + k_3) M & k_1 v_1^* M \\
\alpha_3 M & -k_1 v_3^* M & k_2 M & -\tilde{D}_3 L - k_1 v_1^* M
\end{bmatrix}$$

(5–5)

and

$$E = \begin{bmatrix}
M & 0 & 0 & 0 \\
0 & M & 0 & 0 \\
0 & 0 & M & 0 \\
0 & 0 & 0 & M
\end{bmatrix}.$$  

(5–6)

With knowledge from the spectral band computations of the value of positive eigenvalues we can use the MATLAB command eigs to find the eigenvectors of the generalized eigenproblem (5–4) corresponding to positive eigenvalues. We see that, as we hoped, values of $(u^*, \vec{v}^*)$ with $u^*$ slightly larger than $u_{\text{crit}}$ do indeed return positive eigenvalues in the spatially discretized setting approximating the exact eigenvalues calculated previously. If we use the eigs command option to return eigenvectors as well, then we have the finite element representation of the eigenfunction, and hence the direction of instability, corresponding to an eigenvalue with positive real part.

In the case of the four equation Keller-Segel model eigenproblem (5–4) with all parameters set to unity, Figure 5-2 shows that the first eigenvalue becomes positive around $u_{\text{crit}} = 165.6$. Then fixing $u^* = 170$ slightly larger than $u_{\text{crit}}$ and solving for eigenvalues near the one obtained in the exact calculation, we obtained a positive eigenvalue $\lambda_1 \approx 0.016$ with the corresponding eigenfunction seen in Figure 5-3. This eigenvalue has multiplicity two, with a second eigenfunction similar to the one seen in Figure 5-3 but with the peaks and valleys in opposite corners.

Another spectral band becomes positive around $u^* = 525$, so linearizing around the steady state corresponding to $u^* = 550$, we have two positive eigenvalues of (5–4), $\lambda_1 \approx$
Figure 5-3. Eigenfunction corresponding to the first positive eigenvalue of (5–4) with $u_* = 170$.

0.0463 and $\lambda_2 \approx 0.474$. The new, smallest eigenvalue corresponds to the eigenfunction pictured in Figure 5-4. Notice that qualitatively, this new pattern is the same as the steady state solution to the minimal model shown in Figure 3-1. Further increasing $u_*$ so that two more bands cross the $x$-axis, eigenfunctions of the $u$ component only corresponding to the newly positive eigenvalues are seen in Figure 5-5. Also of note is that the bands of the first two eigenvalues cross, so that the eigenvalue corresponding to the eigenfunction of Figure 5-4 becomes the eigenvalue with largest positive real part. The bands for this higher range of $u_*$ values can be seen in Figure 5-6. Results of similar computations can done on a disk can be seen in Figure 5-7, which shows the eigenfunctions of $u$ only corresponding to the first four spectral bands to cross the $x$-axis.

5.3 Positivity of Numerical Solutions

Having found potential initial conditions for which pattern formation may occur, it is of interest to next perform time simulations to see if patterns do develop and if
Figure 5-4. Eigenfunctions corresponding to the second positive eigenvalue of (5–4) with $u_\ast = 550$.

Figure 5-5. Eigenfunctions of (5–4)
nonhomogeneous steady states are found. In the case of the simple KS model, we have already demonstrated nonhomogeneous stationary solutions so time simulations will demonstrate whether such stationary solutions are stable, and therefore potentially observed in nature. As we have shown that classical solutions of (3–1) and (4–1) remain at least nonnegative considering nonnegative, nonzero initial conditions, we wish to know if our numerical schemes maintain nonnegativity as well. For the standard finite element method we have no such guarantee, but in this section we introduce
a nonstandard discretization which in many cases ensures that solutions remain nonnegative and additionally minimizes the numerical instability for convection diffusion problems shown in Section 2.2.3.

5.3.1 Implementation of the Method

As before we let \( T_h \) denote a triangulation of the bounded domain \( \Omega \subset \mathbb{R}^2 \). We again work in the Lagrange finite element space \( S_h \). Let \( \{ \vec{x}_\ell \}_{\ell=1}^{L'} \) denote the global vertices of \( T_h \) and \( \phi_\ell \in S_h \) the corresponding basis functions. We write the expansions of the approximating functions after \( n \) steps at time \( t_n = n \Delta t \) as \( u^n_h, v^n_{h,i} \in S_h \) of \( u \) and \( v_i \) as

\[
\begin{align*}
    u(x, t_n) &\approx u^n_h = \sum_{\ell=1}^{L'} U^n_\ell \phi_\ell, \\
    v_i(x, t_n) &\approx v^n_{h,i} = \sum_{\ell=1}^{L'} V^n_{i,\ell} \phi_\ell.
\end{align*}
\]

Then if the solutions \( \vec{U}^n \) and \( \vec{V}^n_i \) are known, we may solve for solutions at time \( t_{n+1} \) by

\[
\begin{align*}
    \frac{1}{k} M (\vec{U}^{n+1} - \vec{U}^n) &= -DA(v^n_{h,N})\vec{U}^{n+1} \quad (5–7a) \\
    \frac{1}{k} M (\vec{V}^{n+1}_{i} - \vec{V}^n_{i}) &= -\tilde{D}_i L \vec{V}^{n+1}_{i} + \alpha_i M \vec{U}^n + \vec{G}_i(\vec{v}^n_h) \quad (5–7b)
\end{align*}
\]

where \( M \) is a diagonal matrix obtained after lumping the masses [38], meaning

\[
M_{\ell \ell} = \sum_m \bar{M}_{\ell m}, \quad \bar{M}_{\ell m} = \int_\Omega \phi_\ell \phi_m,
\]

and the stiffness matrices are

\[
\begin{align*}
    L_{lm} &= \int_\Omega \nabla \phi_m \cdot \nabla \phi_l, \\
    A_{lm}(z_h) &= \sum_K A_{j(l),j(m)}^{K,z_h}
\end{align*}
\]

where \( j(l) \in \{0, 1, 2\} \) denotes the local vertex number of the \( l \)th global vertex. \( A_{j(l),j(m)}^{K,z_h} \) and \( \vec{G}_i(\vec{v}_h^n) \) will be defined shortly.
We now describe the local stiffness matrix $A^K$. For the local indices $j$ use addition modulo three so that if, for example $j = 2$ then $A^K_{j,j+1} = A^K_{2,0}$. For any $z_h \in S_h$, we set

$$c_j = \frac{1}{|E_j|} \int_{E_j} e^{-(\chi/D)z_h}$$

where $E_j$ is the edge opposite to the $j$th vertex, and define the entries of the matrix $A^K_{j,z} \equiv A^K_{jk}$ by

$$(5–8a)\quad A^K_{jj} = w_{j+1} \frac{p_j}{c_{j+1}} + w_{j+2} \frac{p_j}{c_{j+2}}$$

$$(5–8b)\quad A^K_{j,j+1} = -w_{j+2} \frac{p_j+1}{c_{j+2}}$$

$$(5–8c)\quad A^K_{j,j+2} = -w_{j+1} \frac{p_j+2}{c_{j+1}}$$

where $p_j = e^{-(\chi/D)z_j}$, $z_j$ equals the value of $z_h$ at the $j$th vertex, and $w_j = (1/2) \cot \theta_j$ where $\theta_j$ is the interior angle of $K$ at the $j$th vertex. As mentioned previously, the discretization leading to the above defined $A^K_{jk}$ was proposed in [44] and will be derived for our particular case in Section 5.3.2.

The $\ell$th component of vector $\vec{G}_i(v^n_h)$ should be an approximation of $\int_{\Omega} g_i(\vec{v}_h^n) \phi_\ell$, so calculate

$$G^n_{i,\ell} \equiv [\vec{G}_i(v^n_h)]_\ell = \int_{\Omega} g_i(\vec{v}_h^n) \phi_\ell$$

$$\approx \int_{\Omega} \left[ \sum_{m=1}^{L'} g_i(\vec{v}(\vec{x}_m, t_n)) \phi_m \right] \phi_\ell$$

$$= \sum_{m=1}^{L'} g_i(V_{1,m}^n, V_{2,m}^n, ..., V_{N,m}^n) \int_{\Omega} \phi_m \phi_\ell$$

$$= [\tilde{M}\vec{h}_i]_\ell$$

where $\vec{h}_i$ is the constant vector defined for each $1 \leq i \leq N$ by

$$(5–9)\quad [\vec{h}_i]_m = g_i(V_{1,m}^n, V_{2,m}^n, ..., V_{N,m}^n).$$
In order to maintain some positivity preserving properties we again use the mass lumping technique as in [38] and in general define the vector \( \vec{G}_i \), \( 1 \leq i \leq N \), by
\[
\vec{G}_i = M \vec{h}_i.
\] (5–10)

### 5.3.2 Derivation of the Method

As in Section 5.3.1, let \( \mathcal{T}_h \) be a triangulation of a bounded domain \( \Omega \subset \mathbb{R}^2 \) and let \( S_h \subset H^1(\Omega) \) be the subspace of functions which are continuous and piecewise linear with respect to \( \mathcal{T}_h \). Let \( q_j, j \in \{0, 1, 2\} \) denote the locally indexed vertices of a triangle \( T \in \mathcal{T}_h \), and consider addition over the local indices \( j \) to be mod 3 for the remainder of this section. Let \( a_{ij}^K \) denote the local stiffness matrix on a given element \( K \), ie
\[
a_{ij}^K = \int_T \nabla \lambda_j \cdot \nabla \lambda_i
\]
where the \( \lambda_i \) are the local linear basis functions on \( K \). Let \( E = E_{ij} \) denote the edge between the \( i \)th and \( j \)th vertices and with endpoints \( q_i \) and \( q_j \). Define
\[
\delta_E f = f(q_i) - f(q_j)
\]
for any function \( f \) defined on \( E \). Finally, let \( \theta^K_E \) denote the interior angle of the vertex opposite of edge \( E = E_{ij} \).

A derivation of the following important fact is given in the appendix of [44]. Given a local bilinear form
\[
a(u_h, v_h) = \int_K \nabla u_h \cdot \nabla v_h
\] (5–11)
defined for \( u_h, v_h \in S_h \), we can use the symmetry of \( a_{ij} \) and the fact that \( a_{ii}^K = -\sum_{j \neq i}^3 a_{ij}^K \) along with a geometric derivation to rewrite 5–11 as
\[
a(u_h, v_h) = \sum_{E \subset K} \omega^K_E \delta_E u_h \delta_E v_h
\] (5–12)
where
\[
\omega^K_E = \frac{1}{2} \cot(\theta^K_E).
\]
Note that for $i \neq j$, $\omega_{E_{ij}}^K = -a_{ij}^K$. We will return to Equation (5–12) later.

We now turn to the specific case of (4–1a). We wish to approximate the variational formulation of (4–1a) in such a way that we may use (5–12) to construct the local stiffness matrices and then use restrictions on angle size in the triangulation to ensure positivity. In [29] and later in [44], this is done by approximating the flux of (4–1a) by a constant over each triangle $K$. Recall that the flux of (4–1a) is

$$\vec{J}(u) = D \nabla u - \chi u \nabla v_N,$$

and so we may write the variational formulation of (3–1a), assuming Neumann boundary conditions on both $u$ and $v$, as

$$\int_{\Omega} u \phi = -\int_{\Omega} \vec{J}(u) \cdot \nabla \phi$$

for all test functions $\phi \in H^1(\Omega)$.

In order to discretize (5–13), we perform the following calculations from [44], modified to fit our boundary conditions and specific equation. Define the function $\psi_E$ on an edge $E$ by

$$\psi_E = (-\chi/D)v_N$$

for $v_N \in S_h$ and perform the following calculations. Here we abuse notation and denote the derivative over an edge $E$ in the tangential direction $\tau_E$ by $\frac{\partial \psi_E}{\partial \tau_E}$. We calculate

$$e^{-\psi_E} \frac{\partial (e^{\psi_E} u)}{\partial \tau_E} = e^{-\psi_E} \left[ \frac{\partial \psi_E}{\partial \tau_E} u + u e^{\psi_E} \frac{\partial \psi_E}{\partial \tau_E} \right]$$

$$= \frac{1}{|\tau_E|} \left[ \nabla u \cdot \tau_E - \frac{\chi}{D} u \nabla v_N \cdot \tau_E \right]$$

$$= \frac{1}{D|\tau_E|} \vec{J}(u) \cdot \tau_E.$$
Then multiplying through by $e^{\psi_E}$, integrating over the edge $E$, and substituting in the definition of $\psi_E$ gives

$$\delta_E(e^{-(\chi/D)v_N}u) = \frac{1}{D|\tau_E|} \int_E e^{-(\chi/D)v_N}(\vec{J}(u) \cdot \tau_E)ds.$$ (5–14)

Now, approximating the flux on any given triangle $K$ with some constant vector $\vec{J}_c(u)$, we can obtain from (5–14) the equality

$$\vec{J}_c(u) \cdot \tau_E = D\delta_E(e^{-(\chi/D)v_N}u) \left[ \frac{1}{|\tau_E|} \int_E e^{-(\chi/D)v_N}ds \right]^{-1}.$$ (5–15)

Using (5–12) along with

$$\vec{J}_c(u) = \nabla(\vec{J}_c(u) \cdot \vec{x}) \quad \text{and} \quad \delta_E(\vec{J}_c(u) \cdot \vec{x}) = \vec{J}_c(u) \cdot \tau_E,$$

we can obtain the approximation

$$-\int_K \vec{J}(u) \cdot \nabla \varphi \approx -\int_K \vec{J}_c(u) \cdot \nabla \varphi = -\sum_{E \subset K} \int_E \omega_T^E \delta_E(\vec{J}_c(u) \cdot \tau_E) \delta_E \varphi.$$ (5–16)

Finally, by combining (5–16) with (5–15) we see

$$-\int_K \vec{J}(u) \cdot \nabla \varphi \approx -D \sum_{E \subset K} \int_E \omega_T^E \delta_E(e^{-(\chi/D)v_N}u) \left[ \frac{1}{|\tau_E|} \int_E e^{-(\chi/D)v_N}ds \right]^{-1} \delta_E \varphi.$$ (5–17)

Notice that (5–17) reduces to (5–8) with $u = \lambda_j$, $\varphi = \lambda_i$, and $|\tau_E| = |E|$.

### 5.3.3 Proof of Positivity

We first show positivity of $u_h$ for numerical solutions of (4–1) calculated using (5–7).

**Theorem 5.1.** Suppose that $T_h$ is a triangulation of $\Omega$ such that all mesh triangles have non-obtuse interior angles. If $u_0 \geq 0$ then for any step size $\Delta t > 0$ we have $u_h^n \geq 0$ on $\Omega$ for all time $t = n'(\Delta t)$ such that solutions $v^n_h$ exist for all $n < n'$.

**Proof.** For simplicity of notation, fix any $n \in \mathbb{N}$ and let $A \equiv A(v^n_h, N)$ for the approximating function $v^n_h, N \in S_h$. Then by (5–7a),

$$\vec{U}^{n+1} = (M + (\Delta t)DA)^{-1}M\vec{U}^n.$$
Since by definition $M$ is a diagonal matrix with positive diagonal entries, it follows that if $\mathbf{U}^n \geq 0$ then $M\mathbf{U}^n \geq 0$. Hence to prove that $\mathbf{U}^{n+1} \geq 0$ one only need to show that $B \equiv M + (\Delta t)DA$ is inverse positive, one of the equivalent conditions to being an invertible $M$-matrix. By [2, the equivalence $M_{35} \iff N_{38}$], it is enough to show that $B$ has positive diagonal entries and that $B$ is strictly diagonally dominant, meaning

$$B_{ll} > \sum_{m \neq l} |B_{ml}|.$$  

To show the necessary conditions, we first observe that the diagonal elements of $B$ are given by

$$B_{ll} = M_{ll} + (\Delta t)DA_{ll} > 0.$$  

Clearly $M_{ll} > 0$, $D > 0$ and $\Delta t > 0$ by assumption. Furthermore $A_{ll} > 0$ follows from (5–8a) and the fact that $\omega_j \geq 0$ since $0 < \theta_j \leq \pi/2$ for every angle. To show the diagonal dominance condition, first note that the column sums of each local stiffness matrix $A^K$ on a given triangle $K$ are zero. Indeed, for any $k \in \{0, 1, 2\}$

$$A_{kk}^K = w_{k+1} \frac{p_k}{c_{k+1}} + w_{k+2} \frac{p_k}{c_{k+2}}, \quad A_{k+1,k}^K = -w_{k+2} \frac{p_k}{c_{k+2}}, \quad A_{k+2,k}^K = -w_{k+1} \frac{p_k}{c_{k+1}},$$

so that the column sums are zero. Therefore the column sums of the global matrix $A$ are zero, so that

$$A_{ll} = -\sum_{m \neq l} A_{ml}.$$  

Again by the assumption on each $\theta_j$ we observe that each $w_j \geq 0$ so that all off diagonal entries of the local matrices $A^K$, and therefore the global matrix $A$, are nonpositive.
by (5–8b)–(5–8c). Using the above observations, calculate

\[ B_{\ell \ell} = M_{\ell \ell} + kD \sum_{m \neq \ell} (-A_{m \ell}) \]

\[ = M_{\ell \ell} + kD \sum_{m \neq \ell} |A_{m \ell}| \]

\[ > kD \sum_{m \neq \ell} |A_{m \ell}| \quad \text{(as } M_{\ell \ell} > 0) \]

\[ = \sum_{m \neq \ell} |B_{m \ell}| \quad \text{(as } M_{m \ell} = 0). \]

Hence \( B \) is a nonsingular \( M \)-matrix, so \( B^{-1} \geq 0 \). Since \( u_0 \) implies \( \vec{U}_0 \), the result holds by induction. \( \square \)

We next prove positivity of the method defined by (5–7) in the particular case where \( \vec{g} \) is a linear function.

**Theorem 5.2.** Let the assumptions on \( T_h, \Delta t, \text{ and } t_n = n\Delta t \) be as in Theorem 5.1. Additionally assume that \( \vec{g} = R\vec{v} \) is a linear function, with the matrix \( R \) such that \( I + (\Delta t)R \) is a nonnegative matrix. If \( u_0 \) and every component of \( \vec{v}_0 \) are nonnegative functions on \( \Omega \), then \( v_{h,i}^n \geq 0 \) for all \( 1 \leq i \leq N \) so long as \( v_{h,i}^n \) exists for all \( n < n' \).

**Proof.** From (5–9)–(5–10), in this case

\[ [\vec{h}_i]_m = \sum_{j=1}^{N} R_{ij} V_{j,m}^n \]

so that

\[ \vec{G}_i^n = M \sum_{j=1}^{N} R_{ij} \vec{V}_j^n. \]

Then from (5–7b),

\[ (M + \Delta t \vec{D}_i L)\vec{V}_i^{n+1} = MV_i^n + \alpha_i(\Delta t)MU_i^n + \vec{G}_i^n \Delta t \]

\[ = \alpha_i \Delta t MU_i^n + \sum_{j=1}^{N} [I + \Delta t R]_{ij} M \vec{V}_j^n. \quad (5–18) \]
Since $U^n \geq 0$ by Theorem 5.1 and $M$ is a nonnegative diagonal matrix, the first term of (5–18) is clearly a nonnegative vector. Using the standard fact that for any local basis elements $\varphi_i$ and $\varphi_j$ on a triangle $K$ we have $\int_K \nabla \varphi_i \cdot \nabla \varphi_j = -(1/2) \cot(\theta_{E_{ij}})$ and the restriction $\theta \leq \pi/2$ it is clear that off-diagonal entries of $L$ are nonpositive. Also obviously diagonal entries of $L$ are positive. Therefore we may argue just as in the proof of Theorem 5.1 that $(M + \Delta t \tilde{D}_i L)$ is an $M$-matrix and hence inverse positive. It follows that $\vec{V}_{n+1}$ is nonnegative in all components whenever $\vec{V}^n$ is nonnegative, and hence the result follows by induction.

**Remark 5.3.** Notice that we can consider the simple KS model (3–1) as a specific case of (4–1) with $N = 1$ and $g(v) = -\gamma v$ a linear function so that Theorems 5.1–5.2 show positivity of $u_h$ and $v_h$ for approximate solutions of (3–1) using the method in (5–7).

### 5.4 Vectorization

We make some final comments about implementation of finite element method time simulations to find approximate solutions of (4–1). The difficulty arises from the term

$$- \nabla \cdot (\chi u \nabla v_N)$$

(5–19)

in (4–1a) which is highly nonlinear and also contributes to the strong coupling of the equations for $u$ and $\vec{v}$. Consider the linear finite element space $S_h$ with a basis of $L'$ functions. If the standard finite element formulation is used, the term (5–19) leads to a $L' \times L'$ submatrix of the full $(N + 1)L' \times (N + 1)L'$ matrix used to solve the discretized form of (4–1). By the nonlinearity of the term, this matrix must be dependent on either the value of $u_h$ or $v_{h,N}$ at the previous time step, and so must be calculated anew at each time iteration. For example, one possible discretization of the term (5–19) is to define the matrix $D^{(u)}$ by

$$D^{(u)}_{ij} = \int_{\Omega} u \nabla \phi_j \cdot \nabla \phi_i.$$

Then if $u_h \approx \sum \ell U_{\ell} \phi_{\ell}$ and $v_{h,N} \approx \sum \ell V_{N,\ell}$ are the approximations of $u$ and $v_N$ in the finite element space at any fixed time, (5–19) is approximated in matrix form by $D^{(u,)} \vec{V}_N$. 

104
Even when the positivity preserving method is used, notice that (5–8) still depends on the values of \( v_N \) at the previous time step and so must also be calculated within the time step loop at each iteration. For this reason it is very important to be as efficient as possible not only when performing the matrix inversion to solve for the next step but also when constructing the matrices. As our programming was primarily done using MATLAB, we used a vectorization technique from [5] to ensure efficiency in matrix construction. The idea behind the vectorization technique is to use MATLAB’s built in structures for sparse matrices along with efficient calculations on vectors. For example, consider two vectors \( \vec{a} \) and \( \vec{b} \) and the operation \( \vec{c} = \vec{a} \cdot \vec{b} \), where the \( \cdot \) is a built in MATLAB function that performs elementwise vector or matrix multiplication. It is much faster to run the line of code \( \vec{c} = \vec{a} \cdot \vec{b} \) than to run the loop

```plaintext
>> for i = 1:length(a)
    >> c(i) = a(i) * b(i)
    >> end,
```
even when storage for \( \vec{c} \) is preallocated [5]. In our code, we apply this idea to the construction of stiffness matrices as in [5] so that only two nested loops of three iterations each are run to build the local stiffness matrices. Within each nested loop a vectorized calculation builds the local stiffness matrix in one line. These local matrices are then placed into the global matrix using MATLAB’s sparse matrix notation.

### 5.5 Method Validation

In order to validate the accuracy of the matrix \( A \) constructed as described in Section 5.3.1, we again return to the test problem (2–14) from Section 2.2.3. In addition to validating the accuracy of the positivity preserving matrix, the test problem will also exemplify how solutions obtained using this new method are better behaved when exact solutions have sharp gradients.

Recall that \( A^{(z)} \) was be derived so that the FEM discretization of the term

\[
D\Delta u - \nabla \cdot \chi u \nabla v,
\]
with Neumann boundary conditions and weak formulation

\[-Dt \int_{\Omega} \nabla u \cdot \nabla \phi_i + \chi \int_{\Omega} u \nabla v \cdot \nabla \phi_i \tag{5–20}\]

is given by

\[-DA^{(\nabla v)} U.\]

We will use the matrix \( A \equiv A^{(v)} \) for \( v(x, y) = x \). Then we have \( \nabla v = (1, 0) = \beta \).

Substituting this \( v \) and the FEM approximation \( u = \sum_i u_i \phi_i \) into the weak formulation given by Equation (5–20) becomes

\[\sum_j \left( -D \int \nabla \phi_j \cdot \nabla \phi_i + \chi \int \phi_j \beta \cdot \nabla \phi_i \right) u_j \quad 1 \leq i \leq N.\]

Thus we must have that

\[A_{ij} \approx \int \nabla \phi_j \cdot \nabla \phi_i - \frac{\chi}{D} \int \phi_j (\beta \cdot \nabla \phi_i)\]

for \( v(x, t) = x \), or in other words

\[-DA_{ij} \approx -D \int \nabla \phi_j \cdot \nabla \phi_i + \chi \int \phi_j (\beta \cdot \nabla \phi_i),\]

which is exactly the transpose of the matrix formulation given by (2–16) for \( \chi = 1 \).

When using this method to solve for \( U \), the non-standard finite element solution gives a good approximation of the actual solution, even for rough meshes where the standard method showed oscillation. Solutions using the method described above are shown in Figure 5-8. By comparing these solutions to the solutions from the standard FEM in Figure 2-2 from Section 2.2.3, we see that the positivity preserving method does indeed reduce the oscillatory behavior, even for the coarse mesh.
Figure 5-8. Positivity preserving FEM solutions to (2–14)
6.1 Application to Tumor Growth

Although its effects are not yet fully understood, chemotaxis is believed to be a factor in tumor growth and metastasis. In this section we look at a tumor growth model proposed in [4] which is based on haptotaxis, or chemotaxis in which the chemoattractant is bound and therefore does not diffuse. The model accounts for tumor cell production or activation of matrix degrading enzymes (MDEs), which cause decay of components of the extracellular matrix (ECM). The degradation of this matrix gives tumor cells room to grow, and at the same time the tumor cells climb gradients of the ECM in order to support their growth. For the remainder of this section we let \( u \) represent the cell density, \( v_1 \) represent the concentration of the MDE and \( v_2 \) represent the ECM concentration. It [4] it is assumed that the MDE is produced at a constant rate \( \alpha_1 > 0 \) per cancer cell, and that it also degrades with rate constant \( \beta \geq 0 \). Then using the same chemotactic flux term as the Keller-Segel model, the following system of equations is given in [4].

\[
\begin{align*}
\partial_t u &= D_0 \Delta u - \chi \nabla \cdot (u \nabla v_2), & x \in \Omega, \ t > 0 \quad \text{(6–1a)} \\
\partial_t v_1 &= \bar{D}_1 \Delta v_1 + \alpha_1 u - \beta v_1, & x \in \Omega, \ t > 0 \quad \text{(6–1b)} \\
\partial_t v_2 &= - \eta v_1 v_2, & x \in \Omega, \ t > 0. \quad \text{(6–1c)}
\end{align*}
\]

The authors of [4] also assume that the cancer cells and the MDE remain inside the domain and therefore impose no-flux boundary conditions on \( u \) and \( v_1 \), which results in the boundary conditions

\[
\begin{align*}
-D_0 \frac{\partial u}{\partial n} + \chi u \frac{\partial v_2}{\partial n} &= 0, & x \in \partial \Omega, \ t > 0 \quad \text{(6–2a)} \\
\frac{\partial v_1}{\partial n} &= 0, & x \in \partial \Omega, \ t > 0. \quad \text{(6–2b)}
\end{align*}
\]
Notice that up to boundary conditions this model fits the framework of the generalized KS model \((4–1)\) with the function \(\vec{g}\) given by

\[
\vec{g} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} -\beta v_1 \\ -\eta v_1 v_2 \end{pmatrix},
\]

the vector \(\alpha = [\alpha_1, 0]\) and the diffusion matrix

\[
D = \begin{bmatrix} \tilde{D}_1 & 0 \\ 0 & 0 \end{bmatrix}.
\]

If we assume the MDE degrades the ECM by means of a reaction of the form

\[ v_1 + v_2 \longrightarrow v_1 + \text{degraded product} \]

with reaction rate constant \(\eta\) and that

\[ v_1 \longrightarrow \emptyset \]

with rate constant \(\beta\), then \(\vec{g}\) can be recovered using the law of mass action kinetics.

In addition to providing a real biological example which fits the framework of our generalized model, we are interested in \((6–1)\) because we may use it to validate our time simulations by reproducing simulated results from \([4]\). The initial conditions used in \([4]\) are based on the idea of an already formed tumor with cell density given by

\[
u(x, 0) = \begin{cases} 
\exp\left(-\frac{x^2}{\epsilon}\right), & x \in [0, 0.25] \\
0, & x \in (0.25, 1].
\end{cases}
\]

They then assume that the initial MDE concentration, as it is emitted by the tumor cells, will be proportional to the tumor cell density, and define \(v_1(x, 0) = 0.5u(x, 0)\).

Assuming the MDE from the previously formed tumor has already degraded the ECM, they consider \(v_2(x, 0) = 1 - 0.5u(x)\). Using these initial conditions and a standard finite element method (FEM) with piecewise linear basis functions and Neumann boundary
conditions, we solved the system (6–1) with boundary conditions (6–2) in one dimension until time $t = 20$. The results are shown in Figure 6-1, and appear to closely match the solutions from [4].

For two dimensions the initial conditions again consider a previously formed tumor, this time centered in the middle of the unit square domain. The initial condition for the tumor cells is given by

$$u(x, 0) = \begin{cases} \exp \left( \frac{-r^2}{\epsilon} \right), & r \in [0, 0.1] \\ 0, & x \in (0.1, 1) \end{cases}$$

where $r^2 = (x - 0.5)^2 + (y - 0.5)^2$. The initial conditions for the MDE and ECM are given as in the one dimensional case, that is $v_1(x, 0) = 0.5u(x, 0)$ and $v_2(x, 0) = 1 - 0.5u(x)$. 

Figure 6-1. One dimensional FEM solutions to Equation (6–1)
Our standard FEM solution in two dimensions also appears to match the solutions in [4], including the higher density ring of cells which develops near the edge of the tumor and continues to invade the ECM. The tumor densities for various times are shown in Figure 6-2. As an additional verification, it was checked that the total cell density over the domain is conserved, which should be implied by the no-flux boundary conditions and lack of birth or death terms in Equation (6–1a). This held up to machine precision, or $10^{-17}$. The following parameters were used for both the one-dimensional and two-dimensional simulations, as stipulated in [4]: $D_0 = 0.001$, $\bar{D}_1 = 0.001$, $\chi = 0.005$, $\eta = 10$, $\alpha_1 = 0.1$, and $\beta = 0$. For the one dimensional simulation, $\epsilon = 0.01$ while for the two dimensional simulation $\epsilon = 0.0025$. 

Figure 6-2. Two dimensional FEM solutions for $u$ of Equation (6–1)
6.2 Keller-Segel Time Simulations

6.2.1 Minimal Keller-Segel Model

In this section we numerically reproduce some well known results on the minimal Keller-Segel (KS) model in order to further validate our numerical methods and have a basis for comparison for the results of time simulations of the full KS model. We first review some results mentioned in Section 3.2 on the stability and global behavior of solutions to Equation (3–15). Recall that as in [18], we define blow up of a solution \((u, v)\) to (3–15) as having occurred when either \(\|u(x, t)\|_{L^{\infty}(\Omega)}\) or \(\|v(x, t)\|_{L^{\infty}(\Omega)}\) becomes unbounded in either finite or infinite time. In [18], it is stated that solutions to (3–15) exist globally in time and have bounded \(L^{\infty}(\Omega)\) norm if the initial condition for \(u\) satisfies

\[
\int_{\Omega} u_0(x) \, dx < 4\pi.
\]

If on the other hand

\[
4\pi < \int_{\Omega} u_0(x) \, dx < 8\pi,
\]

there exist initial conditions \(u_0\) and \(v_0\) such that solutions to (3–15) blow up on the boundary of \(\Omega\) in either finite or infinite time. This type of boundary blow-up is referred to as chemotactic collapse, and is exhibited by the formation of one or more \(\delta\)-singularities at the boundary of \(\Omega\) [18, 33]. Additionally if

\[
\int_{\Omega} u_0(x) \, dx > 8\pi
\]

then there still exist initial conditions for which blow-up occurs, it is just not necessarily on the boundary of \(\Omega\)[18]. The next result, from [34], involves the eigenvalue \(\mu_1 < 0\) of the Laplacian operator on the domain \(\Omega\) which is closest to zero. If

\[
|\Omega|(1 - \mu_1) > \int_{\Omega} u_0(x) \, dx
\]

then the spatially homogeneous stationary solution to (3–15) is linearly stable.
We performed time simulations of solutions to (3–15) using both the standard FEM described in Section 2.2.2 and the positivity preserving method described in Section 5.3. The following discussion is for results using the domain $\Omega$ equal to the unit square. Unless otherwise noted, simulations in this section were computed using a mesh size of around $h \approx 0.03$ and time steps $\Delta t = 0.01$. The choice of $\Delta t$ was based on what time step size was needed for convergence to the known spatially constant steady states from randomly perturbed initial conditions. For any constant $u_* > 0$, notice that if $v_* = u_*$, then $(u_*, v_*)$ is a spatially homogeneous stationary solution to the system (3–15). Then for $u_* < (1 - \mu_1)|\Omega| = 1 + \pi^2 \approx 10.87$ [34] says that $(u_*, v_*)$ is linearly stable. This stability result was confirmed in numerical time simulation experiments using initial conditions of the constant solution $(u_*, v_*)$ perturbed both randomly and by perturbations in the direction of eigenfunctions corresponding to positive eigenvalues when $u_* > 1 + \pi^2$. These eigenfunctions were determined using the spectral bands method described in Section 5.1. The spectral bands for the minimal Keller-Segel (KS) system can be seen in Figure 6-3. Notice this plot also numerically and visually confirms the theoretical result from [34] that the stationary solution to (3–15) on the unit square becomes unstable around $u_* = 11$.

We next investigate time simulation of solutions to (3–15) when $u_* > 4\pi \approx 12.57$, in particular for $u_* = 13$. Recall that for this parameter range the theory predicts that the spatially constant steady state is unstable and that solutions which exhibit blow-up behavior on the boundary exist. We again used a small perturbation away from the spatially homogeneous stationary solution $(u_*, v_*)$ as the initial condition for the time simulation. For the standard finite element method, the solution propagates away from the stationary solution when perturbed slightly, just as predicted by the theory. However after a period of growth away from the stationary state, the solution begins to blow-up in a numerically unstable way. It shows large oscillations and attains nonphysical negative values as seen in Figure 6-4. This numerical instability motivate the need
Figure 6-3. Spectral bands of the minimal KS model

for the positivity preserving method. Figure 6-4 also shows the positivity preserving method does indeed preserve positivity, and the solution obtained using the positivity preserving method continues to converge towards what appears to be a $\delta$-function singularity in a corner of the domain even after the standard method begins to show numerical instability. As mentioned in [18], a $\delta$-function in $u$ is the expected behavior of solutions when blow up occurs. The solutions shown in Figure 6-4 are all for the function $u$, but solutions of $v$ are qualitatively similar. Both solutions were for time steps of size $\Delta t = 0.01$ and mesh size $h \approx 0.03$.

We finally turn to values of $u_*$ on the unit square for which $1 + \pi^2 < u_* < 4\pi$, or values where the homogeneous stationary solution is unstable but the theory ensures that solutions remain bounded for all time $t > 0$. In fact results cited in [18] state that solutions should converge to a stationary solution as $t \to \infty$. Since the unique spatially constant stationary solution is unstable for $u_* > 1 + \pi^2$, this implies a spatially nonhomogeneous stationary solution exists. Although we have found some such solutions, pictured in Section 3.3, using a Newton iteration, we have not been able to
Figure 6-4. Blow-up of the minimal KS model

show convergence to the stationary solution through time simulations. However as seen in Figure 6-5, the time before blow-up and numerical instability occurs is longer when $1 + \pi^2 < u^* < 4\pi$ as opposed to when $u^* > 4\pi$. This gives hope for future work that if mesh size or time steps are sufficiently small, the convergence behavior can be demonstrated numerically.

We note that for computations on the unit circle, similar results were obtained. Convergence to the stationary solution was shown numerically in the appropriate parameter range, while blow-up and numerical instability was observed in the parameter range where such solutions are theoretically guaranteed to exist. However we in this case we were able to show convergence to a nonhomogeneous stationary solution within the range where the spatially constant stationary solution is unstable but solutions are theoretically guaranteed to exist for all time. The error for convergence in this range is shown in Figure 6-6.
6.2.2 Four Equation Keller-Segel Model

Having verified some standard theoretical results in the case of the minimal KS model, we now turn to comparing numerical simulations to our new theoretical results from Chapter 4. We will refer to the two equation KS model studied in the previous section as the “minimal model”, and to the four equation KS model which we numerically investigate in this section as the “full” or “four equation” KS model. This specific full KS model should not be confused with the “general KS model” introduced in Chapter 4 as it is in fact a special case of that model. As with the simulations of the minimal model, we study the case where all constant parameters are set to unity. While these ranges
Figure 6-6. Convergence to a nonhomogeneous stationary solution on a disk

are clearly not biologically realistic, they should still allow for a comparison between the qualitative behavior of the minimal and full KS models. Also as in the last section, unless otherwise noted all simulations were performed using a FEM on a mesh size of \( h \approx 0.03 \) with \( \Delta t = 0.01 \).

Recall that in Section 5.1 we found that the first eigenvalue of the linearized eigenproblem about the spatially homogeneous stationary solution to the full KS model became positive around \( u_{\text{crit}} = 165.6 \). Hence we first tested time simulations in the case of the full KS model for perturbation around the spatially constant stationary solution corresponding to \( u_* = 150 < u_{\text{crit}} \). The spectral bands plot seen in Figure 5-2 predicts that this spatially homogeneous stationary solution should be linearly stable for \( u_* = 150 \). Numerical simulations confirm stability of the stationary solution, and imply that in this case the linearized stability does imply stability. Figure 6-7 shows the \( H^1(\Omega) \) norm of the error between the solution obtained using the standard FEM at various time steps and the spatially homogeneous solution corresponding to \( u_* = 150 \). The initial conditions used were of the constant steady state perturbed by the eigenfunction that corresponds
to an eigenvalue that will later become positive. This is the same perturbation that will be used when we show divergence from the stationary solution for values of $u^* > u_{\text{crit}}$, and was determined using the method described in Section 5.2.

As our theoretical results and spectral bands plots do not give any information about blow-up or non blow-up, we can only use numerical time simulations to test behavior for values of $u^* > u_{\text{crit}}$. We performed time simulations with initial conditions equal to small perturbations from the spatially constant stationary solution. We begin with results obtained from simulations of initial conditions corresponding to perturbations in the direction of the first unstable eigenfunction, obtained as described in Section 5.2. Unlike in the case of $u^* < u_{\text{crit}}$, time iterates continued to grow in the direction of the eigenfunction and away from the spatially homogeneous stationary solution. This was true for both the standard FEM and the positivity preserving FEM. As a further verification of both methods, the error between solutions of the two methods during this period of initial growth is shown in Figure 6-8. After a period of growth in the direction of the eigenfunction, both methods started to demonstrate blow-up behavior. As in the case of the minimal model, the standard FEM began to show numerical instability and solutions become meaningless. We then turn to the results of the positivity preserving
method. As these results accurately showed the behavior of blow up in the two equation case, they should give an accurate portrayal of blow-up in this case as well. Similarly to the case of the minimal model, the positivity preserving method for the full model shows blow-up of $u(x, t)$ in the form of a $\delta$-function on the boundary. The solutions for all four functions when blow-up occurs in $u$ can be seen in Figure 6-9. So, as in the case of the minimal KS model, it appears that when blow-up occurs in the full KS model it is, at least in these simulations, of a $\delta$-function singularity type.

Although we have not been able to find convergence in time to a spatially nonhomogeneous stationary solution in the case of either the minimal or full KS model, we do have calculations that show the patterns arising from the eigenfunctions corresponding to positive eigenvalues are stable in some sense. To see this stability, we used time simulations using an initial condition of the spatially constant stationary solution perturbed by a small random function instead of perturbing in the direction of the unstable eigenfunction of the stationary solution. Under these time simulations using the constant stationary solution corresponding to $u_* = 180$, both the standard and
positivity preserving methods quickly “converge” to a pattern similar to the one of the dominant eigenvalue before slowly starting to increase in the direction of this eigenvalue and eventually demonstrating either numerical instability (in the case of the standard method) or $\delta$-function blow-up (in the case of the positivity preserving method). Solutions simulated using the standard FEM from these randomly perturbed initial conditions at various times can be seen in Figure 6-10. The arbitrary initial condition at time $t = 0$ is the first image, and very quickly after that small aggregation centers begin to appear as at time $t = 0.02$. Of interest is that around time $t = 0.15$, for a time the pattern resembles one of the eigenfunctions corresponding to an eigenvalue which is negative for linearization about the constant steady state corresponding to $u_* = 180$ but becomes positive, and therefore a direction of instability, for larger values of $u_*$. The form of $u$ for this eigenfunction can be seen in Figure 6-11, and the eigenfunctions of $v_1$, $v_2$ and $v_3$ are qualitatively similar.
In summary, the numerical experiments in this section have shown that immediately before and after destabilization of the spatially constant stationary solution, the minimal KS model and the full KS model have very similar behavior in the case where all constants are set to unity. Both systems show convergence to the spatially constant steady state when $u_*$ is below a critical value, and propagate in the direction of a qualitatively similar eigenfunction when $u_*$ is slightly above a critical value. An open
question for further experimentation is if the behavior of the two models remains
the same for larger values of $u_*$ when different eigenvalues, and hence different
eigenfunctions corresponding to them, become dominant. The numerical technique
of finding a direction of instability from a homogeneous stationary solution using a
spectral bands plot can also be generalized to address any number of other applications
which may fit the generalized KS model, which is another open direction for future
research.

6.3 Parameter Fitting in the Four Equation Keller-Segel Model

In all of the previous sections we have taken the approach of [33] by setting all
nonzero constants equal to unity and observing the qualitative behavior of the system.
In this section we will apply the spectral band and time simulation methods to the full
KS model with more realistic parameter values. A summary of the parameters used for
the calculations in this section is given in Table 6-1. We do not claim that these values
are exactly accurate, and in fact many are very rough estimates, but we hope that they
provide some idea of the correct biological scale in order to see relevant results.

Some brief comments on the values are necessary. Diffusion coefficients for three
of the four components were found in the literature, but a value could not be found for a
diffusion coefficient of the complex consisting of the enzyme bound with the cAMP. For
this reason, the parameter $D_2$ was estimated using the Stokes-Einstein equation (see for
example [31]),

\[ D = \frac{k_B T}{6\pi \eta r}. \] (6–3)

In (6–3) $k_B$ is the Boltzmann constant, $T$ is the absolute temperature, $\eta$ is the viscosity,
and $r$ is the radius of the particle, which in this case is a molecule of the chemical
compound. If all other constants are assumed to be the same for the chemicals in
question and the effective radius of the complex is assumed to be sum of the radii of the
enzyme and the cAMP, then using (6–3) $D_2$ may be estimated by

$$D_2 = \left( \frac{1}{D_1} + \frac{1}{D_3} \right)^{-1}.$$  

We also note that the more commonly studied parameter in the literature for chemical reactions seems to be the ratio $K_d = \frac{k_2}{k_1}$, known as the dissociation constant. It was assumed that $k_2$ should be on the scale of $k_3 = 9.3 \times 10^{-4}$, the other rate constant with units of $s^{-1}$. Hence we produced spectral bands plots for values of $k_2$ varying from $7.5 \times 10^{-6}$ to $7.5 \times 10^2$ and the corresponding $k_1$ values $10^{-14}$ to $10^{-6}$ as calculated from $K_d$. Within this range the spectral bands did not change visibly, and so it seems reasonable to assume, as was stated in the literature, that the value of $K_d$ is the important ratio and that so long as it is constant the results will not be significantly affected. In order to keep all values within a computable range, we chose the value of $k_2 = 0.75$. Then using $K_d$ and the chosen value of $k_2$, $k_1$ was calculated to be $10^{-9}$.

Finally, in order to guarantee the existence of a homogeneous steady state solution which may be calculated exactly, we found that we need $\gamma_2 = \gamma_3 = 0$ and $\gamma_1 > 0$. In [24], it is mentioned that “starved cells excrete a protein which inhibits the activity of their extracellular phosphodiesterase [enzyme]”. If this effect is estimated by decay of the enzyme at a positive rate constant, then $\gamma_1 > 0$ would be a plausible biological assumption. Additionally, if no decay of the enzyme is assumed then no spatially homogeneous stationary solutions exist for any value of $u_*>0$. This would be a definite qualitative difference between the behavior of solutions to the full KS system and the minimal KS system, and so would place further importance on the study of the general model we develop in this work, if not necessarily the case studied in this section where $\gamma_1 > 0$. Since no value could be found in the literature, we experimented with values of $\gamma_1$ between $\gamma_1 = 10^{-6}$ and $\gamma_1 = 10$ and found very few visible differences between the spectral bands plots. Hence we estimate $\gamma_1$ to be on the order of the decay rate constant $k_3 = 9.3 \times 10^{-4}$ and choose $\gamma_1 = 10^{-4}$.
Using the values in Table 6-1 and the method of spectral bands in Section 5.1, we obtain Figure 6-12. It at first appeared that, unlike in the case where all parameters were set to unity, as soon as the steady state value $u^*$ becomes positive, positive eigenvalues occur. However experimentation with parameter values showed that in fact the same behavior is exhibited, we simply needed to zoom in much closer to zero. A spectral band plot in which the destabilization behavior is visible is shown in Figure 6-13.

Figure 6-13 shows that the constant stationary solution becomes unstable around $u_* = 0.002$. As the units of $u$ are in cells/cm$^2$, this implies that any realistic homogeneous steady state would be unstable. For this reason, we try a different approach to the parameter values. Other than $\gamma_1$, which according to numerical experiments does not appear to greatly affect the critical value at which destabilization occurs, the parameter value for which the estimate is the most rough is $\chi$. The value for $\chi$ was approximated from literature that views the chemotactic sensitivity function $\chi u$ as either a step function in $u$ or a constant function [28, 30], not a linear function, and no literature could be found on what an appropriate linear chemotactic sensitivity function would be. As the linear form of the chemotactic sensitivity function is one of the most commonly studied forms of the KS model [18], it is of interest to obtain a more accurate approximate value for $\chi$. 

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_0 = D_{\text{Dicty}}$</td>
<td>$4 \times 10^{-8}$ cm$^2$/sec</td>
<td>[28]</td>
</tr>
<tr>
<td>$D_1 = D_{\text{enzyme}}$</td>
<td>$6.1 \times 10^{-7}$ cm$^2$/sec</td>
<td>[39]</td>
</tr>
<tr>
<td>$D_2 = D_{\text{complex}}$</td>
<td>$5.8 \times 10^{-7}$ cm$^2$/sec</td>
<td>Estimated with Stokes-Einstein</td>
</tr>
<tr>
<td>$D_3 = D_{cAMP}$</td>
<td>$9.7 \times 10^{-6}$ cm$^2$/sec</td>
<td>[28]</td>
</tr>
<tr>
<td>$k_1$</td>
<td>$10^{-9}$ cm$^2$/ (molecule - sec)</td>
<td>Calculated from $K_d$ and $k_1$</td>
</tr>
<tr>
<td>$k_2$</td>
<td>0.75 sec$^{-1}$</td>
<td>Chosen for scale</td>
</tr>
<tr>
<td>$k_3$</td>
<td>$9.3 \times 10^{-4}$ s$^{-1}$</td>
<td>[23]</td>
</tr>
<tr>
<td>$\chi$</td>
<td>$1.2 \times 10^{-14}$ cm$^4$/ (molecule - sec)</td>
<td>[28, 30]</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>$10^{-4}$ s$^{-1}$</td>
<td>On the order of $k_3$ [24]</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>3000 molecules/sec per cell</td>
<td>[1]</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>$4 \times 10^5$ molecules/sec per cell</td>
<td>[28]</td>
</tr>
<tr>
<td>$K_d = k_2/k_1$</td>
<td>$7.5 \times 10^8$ molecules/cm$^2$</td>
<td>[16]</td>
</tr>
</tbody>
</table>
Figure 6-12. Spectral bands using parameters from Table 6-1

Figure 6-13. Behavior near $u_* = 0$ of spectral bands using parameters from Table 6-1
The information we do have is that aggregation fails to occur when the concentration of the amoeba is less than 4000-5000 cells/cm\(^2\) [7]. If chemotactic feedback is indeed what destabilizes the amoeba and initiates aggregation, we may adjust the value of \(\chi\) so that the critical value of \(u_*\) is between 4000 and 5000 cells/cm\(^2\). Using the parameter fitting concept, we see that a value of \(\chi = 5.4 \times 10^{-19}\) with all other parameter values as in Table 6-1 gives the value \(u_{\text{crit}} \approx 4270\). Using the methods of Section 5.1 with \(\chi = 5.4 \times 10^{-19}\) we obtain the spectral bands plot from Figure 6-14. As the order of magnitude \(10^{-19}\) is below machine precision, it should also be noted that in order to have computable values, we converted all of the parameters for the \(v_i\) equations to units of \(10^6\) molecules/cm\(^2\) instead of molecules/cm\(^2\). Although this scaling does not affect, for example, any of the diffusion coefficients, it will affect parameter values which include molecules in the units. The values that are affected become \(k_1 = 10^{-3}\) cm\(^2\)/ \((10^6\) molecules-sec\), \(\chi = 5.4 \times 10^{-13}\) cm\(^4\)/\((10^6\) molecules-sec\), \(\alpha_1 = 0.003\) cm\(^4\)/\((10^6\) molecules-sec\), and \(\alpha_3 = 0.4\) cm\(^4\)/\((10^6\) molecules-sec\) after the scaling. As would be expected, this parameter rescaling did not affect the computations in any way other than scaling of the solutions.
Figure 6-15. The linearized eigenfunction corresponding to the largest positive eigenvalue

We can now use the methods of Section 5.2 to find eigenfunctions corresponding to the positive eigenvalues and therefore visualize potential patterns formed just after destabilization by chemotaxis. We found the eigenfunctions of the eigenproblem linearized about the constant stationary solution corresponding to \( u^* = 4500 \) molecules/cm\(^2\). The largest positive eigenvalue was \( \lambda \approx 1.13 \times 10^{-7} \), and the eigenfunctions for \( u \) and all \( v_i \) corresponding to this eigenvalue can be seen in Figure 6-15. In this parameter range, unlike in the case where all parameters were set to unity, the shape of the eigenfunction for \( v_1 \) is exactly the opposite of the shapes of the other three eigenfunctions. In other words, the eigenfunction for \( v_1 \) has peaks when the other three eigenfunctions have valleys, and valleys when the others have peaks. Biologically the chemical \( v_1 \) reduces the amount of chemoattractant in the environment, so this mirroring behavior makes sense. Other eigenfunctions corresponding to positive eigenvalues can be seen in Figure 6-16. These eigenfunctions all demonstrate the mirroring behavior of the eigenfunction corresponding to the dominant eigenvalue.
We conclude by emphasizing that in the case of the full KS model we have just shown that parameter values can qualitatively change the nature of the stationary solutions. An open question for future research is if time simulations also differ in this case. As solutions to the four equation KS model did tend to converge to positive eigenfunctions, it is likely that parameter values affect the nature of time dependent solutions as well. So, although behavior of the minimal KS system and the full KS system is qualitatively similar for the case where all constants are set to unity, the results of this final section show that further investigation is needed to show that in all cases the behavior of solutions to the two systems is similar.
REFERENCES


[16] E.J. HENDERSON, The cyclic adenosine 3': 5'-monophosphate receptor of Dictyostelium discoideum. binding characteristics of aggregation-competent


BIOGRAPHICAL SKETCH

Erica Zuhr was born and raised in Oak Ridge, Tennessee. She graduated from Oak Ridge High School in 2003, and went on to attend the University of North Carolina at Chapel Hill and major in mathematics. She graduated from the University of North Carolina with honors in 2007, and at that time was awarded the Archibald Henderson Prize in Mathematics. She was then offered full support in the form of an Alumni Fellowship to pursue a doctoral degree in mathematics at the University of Florida.