# Bivariate Spline Solution of Time Dependent Nonlinear PDE for a Population Density over Irregular Domains

Juan B. Gutierrez[1,2], Ming-Jun Lai[2,3], George Slavov[2]

**Abstract**

We study a time dependent partial differential equation (PDE) which arises from classic models in ecology involving logistic growth with Allee effect by introducing a discrete weak solution. Existence, uniqueness and stability of the discrete weak solutions are discussed. We use bivariate splines to approximate the discrete weak solution of the nonlinear PDE. A computational algorithm is designed to solve this PDE numerically. A convergence analysis of the algorithm is included to establish our method. Then we present some simulations of population development over some irregular domains. Finally, we show an application in epidemiology.

*Keywords:* Bivariate splines, numerical solution, PDE, Allee effect, ecology, epidemiology
*2010 MSC:* 50-030, 70-060, 70-140

## 1. Introduction

Empirical evidence shows that the structure of environments and spatial scale can systematically influence population development and interactions in a way that can be described by mathematical models [9, 11]. The first serious attempt to model population dynamics is credited to Malthus in 1798 [24], who hypothesized that human populations grow geometrically while resources grow arithmetically, thus eventually reaching a point in which the population could not be sustained any more; this linear growth model is problematic since it allows unbounded population increase. A major refinement was introduced by Verhulst in 1838 [31] by means of a density-dependent logistic term in Malthus' model, predicting population growth if resources were available or population decay if population surpassed resources; this model takes the form $\dot{p} = r_0 p(1 - p/k)$, where $p$ represents population density, $r_0$ is the rate of growth, and $k$ represents the carrying capacity. Fisher [8] used in 1937 a diffusion operator to study the propagation of advantageous genes in population; the same year, Kolmogorov and his collaborators [14] studied the following reaction-diffusion equations in the one-dimensional setting:

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2} + kp(1-p) \text{ and } \frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2} + F(p), \tag{1.1}$$

where $F(p)$ satisfies $F(p) \geq 0$, $F(0) = F(1) = 0, F'(0) > 0, F'(1) < 0$ for $p \in [0, 1]$.

The logistic model has been central to the modern study of population dispersal in spatial domains [25, 3]. Skellam's influential paper [28] in 1951, introduced a variation in Kolmogorov's equation for the study of phytoplankton; the resulting model was $p_t = d\Delta p + c_1(x, y)p - c_2(x, y)p^2$.

[1]Institute of Bioinformatics, University of Georgia
[2]Department of Mathematics, University of Georgia
[3]Corresponding Author: mjlai@math.uga.edu, University of Georgia, Athens, GA 30602

This basic form of population dispersal is applicable in many notable cases ranging from population dispersal to recent models of information diffusion in online social networks [32]. Nevertheless, Skellam's model is too simplistic in most practical cases; it assumes a lack of interactions with other species, and that populations can grow at the same rate at low and high densities. An important refinement to Kolmogorov's model was introduced by Lewis and Kareiva in 1993 [22]. The correlation hypothesized by Allee in 1938 between population size and mean individual fitness [1], was represented in Lewis and Kareiva's model by $p_t = d\Delta p + r_0 p(1 - p/k)(p - \sigma)$, where $\sigma$ represents the population below the carrying capacity below which the population growth is negative. This is the foundation of the model we study in this paper.

More precisely, we are interested in solution of the following nonlinear time dependent partial differential equation: Letting $\Omega \subset \mathbb{R}^2$ be a polygonal domain and $\Omega_T = \Omega \times (0, T]$,

$$\begin{cases} \dfrac{\partial p(\mathbf{x},t)}{\partial t} &= \text{ div } (D(p,\mathbf{x})\nabla p(\mathbf{x},t))) + F(p(\mathbf{x},t)), \quad \mathbf{x} = (x,y) \in \Omega, t \in [0,T] \\ p(\mathbf{x},t) &\geq 0, \quad \mathbf{x} \in \Omega, t \in [0,T] \\ p(\mathbf{x},t) &= 0, \quad \mathbf{x} \in \partial\Omega, t \in [0,T] \\ p(\mathbf{x},0) &= p_0, \quad \mathbf{x} \in \Omega, \end{cases} \tag{1.2}$$

where $D(p,\mathbf{x})$ is a diffusive term, e.g. $D(p,\mathbf{x}) = D > 0$ and $F(p)$ is a growth function, e.g. $F(p) = Ap(1-p)$ which is a standard logistic growth function with $A$ being a nonnegative weighted function over $\Omega$. In this paper, we shall mainly study $F(p) = Ap(1-p)(p-\sigma)$, where $\sigma$ is a positive constant in $[0,1)$ and $A(x,y)$ are nonnegative functions on $\Omega \times [0,T)$.

Exact solutions to Kolmogorov's equation (1.1) have been found [23]. However, there does not appear to exist an exact solution to the diffusion logistic model with Allee effect; while asymptotics and speed of diffusion waves have been found analytically, the solutions to problem (1.2) over different domains remain mostly numerical. Lewis and Kareiva [22] used finite differences. Roques *et al.* [27] used a second-order finite elements method. Richter *et al.* [26] used finite elements in a model that incorporated geographic features and population dispersal.

In this paper, we present in detail a numerical solution to the diffusion logistic problem with Allee effect based on bivariate spline functions over triangulations. Bivariate spline have been studied extensively in different contexts; see [20], [2], [21], [16], [17], and [10]. An advantage of the use of bivariate splines is the ease to generate a smooth density surface over a spatial domain. The differentiability can be useful for some applications which involves the rate of changes of population along different directions at any location inside the domain, as explained in the following section.

Our numerical solution of this PDE is slightly different from the classic approach in a few ways. Instead of defining a weak solution in terms of test functions defined on domain $\Omega_T = \Omega \times (0, T]$, we define a discrete weak solution of the PDE using test functions defined on $\Omega$ together with the first order divided difference in time. See Definition 4.1. Another difference from the classic approach is that we use an optimization approach to establish the existence, uniqueness, stability and other properties of this discrete weak solution. We shall use bivariate splines to approximate the discrete weak solution using the discrete weak solution in the finite dimensional spline space. We are able to show that spline discrete weak solution converges to the discrete weak solution in $H^1(\Omega)$ as the size of underlying triangulation goes to zero.

It is clear that there are three nonlinearities in (1.2): the nonlinear diffusive term $D(p,\mathbf{x})$, the nonlinear growth function $F(p)$ and nonlinearity condition $0 \leq p(\mathbf{x},t) \leq 1$ for all $x, t$ which is essential to the theory presented. We have to design a convergent computational algorithm to find bivariate spline solutions and establish how well our bivariate spline solutions are close to the exact discrete weak solution. We implement our computational algorithm in MATLAB. With the

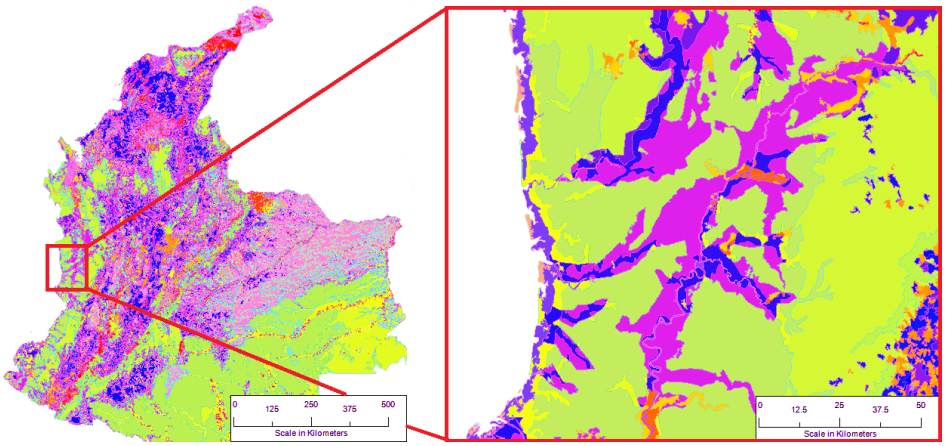



Figure 1: Map of vegetation cover in Colombia. Each color represents a specific vegetation pattern, varying from grassland to jungle. Source: Ministry of Health of Colombia, 2014.

numerical solution, we are able to simulate how a population disperses over the area $\Omega$ of interest. In particular, we are able to see how the Allee constant $\sigma$ plays a significant role in the population development.

## 2. Biological Motivation: Vector-Borne Disease

The dispersal patterns of a species in a given environment depends on the spatial scale considered, the temporal scale studied, the physical size of the model organism, and the life history the species. In most vector-borne disease dynamics, the dispersal of hosts and vectors has to be considered independently. In the case of mosquito-borne disease (e.g. malaria), human movement is a problem of a fundamentally different nature as compared to vector movement [29, 4].

We focus in this paper on the dispersal of mosquitoes, which can be characterized by the diffusion-reaction framework presented in (1.2). Note that we propose Allee dynamics. The study of Allee effect in insects is common in the context of biological invasions [12, 15]; however, the question of invasibility is not necessary to study the density dependence of a dispersal process. There is evidence that supports the assumption of Allee dynamics in this context; it has been shown that human-dependent mosquito populations (e.g. mosquito species that transmit malaria) can rapidly collapse when the population is reduced below realistic non-zero thresholds [13].

Vector control is often the tool of choice to manage mosquito-borne disease. When mosquitoes are removed from an area, or the population is brought to a level that drives local extinction, it can be repopulated by dispersal from neighboring areas. But landscapes are not even; instead, there are transitions in the patterns of vegetation cover which determine diffusivity and the suitability of local environments for the establishment and dispersal of mosquitoes. The effect of this spatial heterogeneity on population dispersal can be accounted for by (1.2), where $D(p, \mathbf{x})$ allows individual consideration of diffusivity in each segment of the spatial domain. The bivariate spline solution results in a smooth density surface over the entire spatial domain, which is very convenient to determine gradients of dispersal in all directions at all times.

Figure 1 offers an example of the type of problem that could be solved with the approach presented in this paper. It depicts a map of vegetation cover for Colombia, in the Northwest corner of South America. The region on the right side of the map shows the location where resistance to the insecticide DDT was recorded for the first time in *Anopheles darlingi* in 1990 [30]. There are two fundamental questions in public health that could be addressed using the methods presented

in this paper: (i) what is the likely pattern of dispersal of advantageous genes conferring anopheles mosquitoes resistance to insecticides and/or repellents, and (ii) what is the likely dispersal and re-population by mosquito of areas subject to vector control. In order to answer these questions with actionable recommendations, we would need to consider multiple scales, the interaction between vectors and hosts, and epidemiological data for calibration of models. The scope of this paper is limited to the numerical solution of (1.2) via bivariate splines to solve one piece of this puzzle: vector dispersal.

We present in Section 6.2 numerical simulations using the geometry of the City of Bandiagara, Mali, as a proof of concept. In this example we did not consider a space-dependent diffusivity due to local variations in the landscape. We plan to undertake the calibration of this model with actual malaria data in a posterior study using the same locality, since Bandiagara has been the subject of many modeling exercises [5, 6].

## 3. Preliminaries

For the sake of completeness, we list a number of lemmas used in this paper, which are special cases of well-known results.

**Lemma 3.1.** *For $a, b \geq 0$ and any $\alpha > 0$ we have*

$$ab \leq \frac{\alpha}{2}a^2 + \frac{1}{2\alpha}b^2$$

**Lemma 3.2** (Ladyzhenskaya's Inequality)**.** *For any $p \in H_0^1(\Omega)$ for $\Omega \subset \mathbb{R}^2$ we have the following inequality.*

$$\|p\|_{L^4} \leq C\|p\|_{L^2}^{1/2}\|\nabla p\|_{L^2}^{1/2}$$

**Theorem 3.1** (Rellich-Kondrachov)**.** *Suppose that $\Omega$ is bounded with Lipschitz boundary. Then we have the following compact injection:*

$$H^1(\Omega) \subset L^2(\Omega)$$

*That is, any bounded sequence in $H^1(\Omega)$ has a subsequence which converges to an $L^2(\Omega)$ function in $L^2$ norm.*

**Theorem 3.2** (General Sobolev Inequality)**.** *If $p \in H^2(\Omega)$, then $p \in C^{0,\gamma}$, the space of Hölder continuous functions with any exponent $0 < \gamma < 1$. Furthermore,*

$$||p||_{C^{0,\gamma}(\Omega)} \leq C\,||p||_{H^2(\Omega)}$$

*where $C$ is a constant independent of $p$.*

Preliminary on bivariate splines can be found in the Appendix of this paper. Mainly we use the theory in [20] and computational schemes in [2]. As our PDE (1.2) is nonlinear, we have to extend the MATLAB codes used in [2] to handle this nonlinear PDE discussed in this paper.

## 4. The Basic Properties of the Discrete Weak Solution

Let us begin with a discrete weak solution of the partial differential equation (1.2). It is a standard calculation from (1.2) to have, for any $q \in H_0^1(\Omega)$,

$$\int_\Omega \frac{\partial p(\mathbf{x},t)}{\partial t} q(\mathbf{x}) d\mathbf{x} = -\int_\Omega D(\mathbf{x}) \nabla p(\mathbf{x},t) \cdot \nabla q(\mathbf{x}) d\mathbf{x} + \int_\Omega A(\mathbf{x}) F(p(\mathbf{x},t)) q(\mathbf{x}) d\mathbf{x}. \tag{4.1}$$

Consider $t \in [0,T]$ and $0 = t_0 < t_1 < t_2 < \cdots < t_m < t_{m+1} = T$. We approximate $\frac{dp(\mathbf{x},t)}{dt}$ by its divided difference, i.e., $\frac{dp(\mathbf{x},t_i)}{dt} \approx (p(\mathbf{x},t_i) - p(\mathbf{x},t_{i-1}))/h$ with $h = t_i - t_{i-1}$. The above equation becomes

$$\begin{aligned} &\int_\Omega p(\mathbf{x},t_i) q(\mathbf{x}) d\mathbf{x} + h \int_\Omega D(\mathbf{x}) \nabla p(\mathbf{x},t_i) \cdot \nabla q(\mathbf{x}) d\mathbf{x} \\ &- h \int_\Omega A(\mathbf{x}) F(p(\mathbf{x},t_i)) q(\mathbf{x}) d\mathbf{x} = \int_\Omega p(\mathbf{x},t_{i-1}) q(\mathbf{x}) d\mathbf{x}, \quad \forall q \in H_0^1(\Omega). \end{aligned} \tag{4.2}$$

We introduce the following concept of the PDE solution:

**Definition 4.1.** *Any solution to the above equation (4.2) for a fixed $h > 0$ is called a discrete weak solution of (1.2).*

Let us first show that the discrete weak solution is a good approximation of the exact solution. Indeed, we have

**Theorem 4.1.** *Assume that $F(\cdot)$ is a Lipschitz continuous function. Let $p(\mathbf{x},t)$ be the classic solution and $p_h(\mathbf{x},t_i), i = 1, \cdots, m+1$ be the discrete weak solution dependent on $h > 0$ with $p_h(\mathbf{x},t_0) = p(\mathbf{x},t_0)$ for $\mathbf{x} \in \Omega$. Suppose that $p(\mathbf{x},t)$ is twice differentiable with respect to $t$. Then*

$$\int_\Omega |p(\mathbf{x},t_i) - p_h(\mathbf{x},t_i)|^2 d\mathbf{x} \le Ch, \quad \forall i = 0, \cdots, m+1, \tag{4.3}$$

*as $h = T/(m+1) \to 0$, where $C > 0$ is a constant independent of $h$.*

*Proof.* By Taylor expansion, we have

$$\frac{dp(\mathbf{x},t_i)}{dt} = \frac{p(\mathbf{x},t_i) - p(\mathbf{x},t_{i-1})}{h} + O(h),$$

where $O(h)$ is a quantity bounded by $Ch$ for a positive constant $C < \infty$. Using (4.1) and (4.2), we have

$$\begin{aligned} &\int_\Omega \frac{dp(\mathbf{x},t_i)}{dt} q(\mathbf{x}) d\mathbf{x} - \int_\Omega \frac{p_h(\mathbf{x},t_i) - p_h(\mathbf{x},t_{i-1})}{h} q(\mathbf{x}) d\mathbf{x} \\ =& -\int_\Omega D(\mathbf{x}) \nabla (p(\mathbf{x},t_i) - p_h(\mathbf{x},t_i)) \cdot \nabla q(\mathbf{x}) d\mathbf{x} \\ &+ \int_\Omega A(\mathbf{x}) (F(p(\mathbf{x},t_i)) - F(p_h(\mathbf{x},t_i)) q(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

That is, we have

$$\begin{aligned}
&\int_\Omega \frac{p(\mathbf{x},t_i)-p(\mathbf{x},t_{i-1})}{h}q(\mathbf{x})d\mathbf{x} - \int_\Omega \frac{p_h(\mathbf{x},t_i)-p_h(\mathbf{x},t_{i-1})}{h}q(\mathbf{x})d\mathbf{x}\\
=&O(h) - \int_\Omega D(\mathbf{x})\nabla(p(\mathbf{x},t_i)-p_h(\mathbf{x},t_i))\cdot\nabla q(\mathbf{x})d\mathbf{x}\\
&+\int_\Omega A(\mathbf{x})(F(p(\mathbf{x},t_i)-F(p_h(\mathbf{x},t_i))q(\mathbf{x})d\mathbf{x}.
\end{aligned}$$

Letting $q = p(\mathbf{x},t_i)-p_h(\mathbf{x},t_i)$ in the above inequality, we have

$$\begin{aligned}
&\int_\Omega |p(\mathbf{x},t_i)-p_h(\mathbf{x},t_i)|^2 d\mathbf{x}\\
=&O(h^2)+\int_\Omega (p(\mathbf{x},t_{i-1})-p_h(\mathbf{x},t_{i-1})(p(\mathbf{x},t_i)-p_h(\mathbf{x},t_i))\\
&-h\int_\Omega D(\mathbf{x})|\nabla(p(\mathbf{x},t_i)-p_h(\mathbf{x},t_i))|^2 d\mathbf{x}\\
&+h\int_\Omega A(\mathbf{x})(F(p(\mathbf{x},t_i))-F(p_h(\mathbf{x},t_i))(p(\mathbf{x},t_i)-p_h(\mathbf{x},t_i))d\mathbf{x}\\
\le&\frac{1}{2}\int_\Omega |p(\mathbf{x},t_i)-p_h(\mathbf{x},t_i)|^2 d\mathbf{x}+\frac{1}{2}\int_\Omega |p(\mathbf{x},t_{i-1})-p_h(\mathbf{x},t_{i-1})|^2 d\mathbf{x}+O(h^2)\\
&+h\int_\Omega A(\mathbf{x})(F(p(\mathbf{x},t_i))-F(p_h(\mathbf{x},t_i))(p(\mathbf{x},t_i)-p_h(\mathbf{x},t_i))d\mathbf{x}
\end{aligned}$$

Since $F$ is Lipschitz continuous, i.e. $|F(p)-F(q)|\le L|p-q|$, it follows that

$$\begin{aligned}
&\int_\Omega |p(\mathbf{x},t_i)-p_h(\mathbf{x},t_i)|^2 d\mathbf{x}\\
\le&\int_\Omega |p(\mathbf{x},t_{i-1})-p_h(\mathbf{x},t_{i-1})|^2 d\mathbf{x}+O(2h^2)+2hLC_A\int_\Omega |p(\mathbf{x},t_i)-p_h(\mathbf{x},t_i)|^2,
\end{aligned}$$

where $C_A = \|A\|_{L^\infty(\Omega)}$. That is, we have

$$(1-2hLC_A)\int_\Omega |p(\mathbf{x},t_i)-p_h(\mathbf{x},t_i)|^2 d\mathbf{x}\le\int_\Omega |p(\mathbf{x},t_{i-1})-p_h(\mathbf{x},t_{i-1})|^2 d\mathbf{x}+O(h^2).$$

Letting $\alpha = 1/(1-2hLC_A)$, we multiply $\alpha$ to the both sides of the inequality above and then add them together for $i=1,\cdots,k$ to have

$$\begin{aligned}
\int_\Omega |p(\mathbf{x},t_k)-p_h(\mathbf{x},t_k)|^2 d\mathbf{x}&\le\alpha\int_\Omega |p(\mathbf{x},t_{k-1})-p_h(\mathbf{x},t_{k-1})|^2 d\mathbf{x}+O(h^2)\le\cdots\cdots\\
&\le\alpha^k\int_\Omega |p(\mathbf{x},t_0)-p_h(\mathbf{x},t_0)|^2 d\mathbf{x}+O(h^2)\sum_{i=0}^{k-1}\alpha^i\\
&\le O(h^2)\frac{\alpha^m}{\alpha-1}=O(h)
\end{aligned}$$

for $k=1,\cdots,m+1$, where we have used $(m+1)h=T$ and the fact that $(1-2hL)^{-T/h}\le Ce^{2TL}$ for a positive constant $C$. These complete the proof. □

Let $\mathcal{A} = \{p \in H_0^1(\Omega), p(x,y) \geq 0 \text{ for a.e. } (x,y) \in \Omega\}$ be the set of admissible functions.Assume that $\Omega \subset \mathbb{R}^2$ is an open, bounded domain with Lipschitz boundary. That is, we look for a population density in the admissible set $p \in \mathcal{A}$ which satisfies the following equation:

$$\int_\Omega pq \, d\mathbf{x} + h \int_\Omega D(\mathbf{x}) \nabla p \cdot \nabla q \, d\mathbf{x} = \int_\Omega \hat{p} q \, d\mathbf{x} + h \int_\Omega p F_1(p) q \, d\mathbf{x} \qquad \forall q \in H_0^1(\Omega), \tag{4.4}$$

where $0 < K \leq D(\mathbf{x}) \leq K_2$ is a diffusive factor and

$$F_1(p) = A(\mathbf{x})(1-p)(p-\sigma) \tag{4.5}$$

which models population growth with an Allee effect. Here $A(\mathbf{x})$ is a given nonnegative function bounded by $M$ and $\sigma \in (0,1)$ and $\hat{p} \in \mathcal{A}$ is a given admissible function.

We would like to see that the equation (4.4) has a unique solution. In order to do that, we consider the following energy minimization problem.

$$\min_{p \in \mathcal{A}} E(p) = \min_{H_0^1(\Omega), p \geq 0} \int_\Omega p^2 \, d\mathbf{x} + h \int_\Omega D(\mathbf{x}) |\nabla p|^2 \, d\mathbf{x} - h \int_\Omega G(p) \, d\mathbf{x} - \int_\Omega \hat{p} p \, d\mathbf{x}, \tag{4.6}$$

where

$$G(p) = \int_0^p \xi F_1(\xi) \, d\xi$$

In order to show that the functional has a minimizer, we need a lower bound for its image.

**Lemma 4.1.** *Suppose we choose $h < \dfrac{1}{M}$. Then for any function $p \in \mathcal{A}$ the energy functional given in (4.6) satisfies*

$$E(p) \geq C \, ||p||^2_{H_0^1(\Omega)} - ||\hat{p}||_2^2$$

*for some constant $C > 0$. In particular, $\inf\limits_{p \in \mathcal{A}} E(p) \geq -||\hat{p}||_2^2 > -\infty$.*

*Proof.* First we will present an upper bound for one of the terms.

$$G(p) = \int_0^p \xi F_1(\xi) \, d\xi \leq M \int_0^p \xi \, d\xi = \frac{M}{2} p^2 \text{ and hence, } \int_\Omega G(p) \, d\mathbf{x} \leq \frac{M}{2} ||p||_2^2 \tag{4.7}$$

Now we prove the lower bound for the entire functional. We use the Cauchy-Schwarz inequality, the upper bound for $G(p)$ we just established and $D(\mathbf{x}) \geq K$.

$$\begin{aligned} E(p) &\geq ||p||_2^2 + hK \, ||\nabla p||_2^2 - \frac{hM}{2} \, ||p||_2^2 - ||\hat{p}||_2 \, ||p||_2 \\ &= \left(1 - \frac{hM}{2}\right) ||p||_2^2 + hK \, ||\nabla p||_2^2 - ||\hat{p}||_2 \, ||p||_2 \end{aligned}$$

Use our assumption for $h$ in this lemma and Lemma 3.1 on the last term with $\alpha = 2$.

$$\begin{aligned} E(p) &\geq \frac{1}{2} \, ||p||_2^2 + hK \, ||\nabla p||_2^2 - ||\hat{p}||_2^2 - \frac{1}{4} \, ||p||_2^2 \\ &\geq \min\left\{\frac{1}{4}, hK\right\} ||p||^2_{H_0^1(\Omega)} - ||\hat{p}||_2^2 . \end{aligned}$$

We have thus established the desired low bound. □

**Lemma 4.2.** *If $h < 1/M$, the energy functional in (4.6) is weakly lower semi-continuous on $H^1(\Omega)$. That is, if $p_k \to p^*$ weakly in $H^1(\Omega)$, then*

$$E(p^*) \leq \liminf_{k\to\infty} E(p_k)$$

*Proof.* Set $m := \liminf_{k\to\infty} E(p_k)$. By passing to a subsequence we can assume that $E(p_k) - m < 1/k$. That is, $\lim_{k\to\infty} E(p_k) = m$. Any weakly convergent sequence is bounded in $H^1(\Omega)$ norm, so by the Rellich-Kondrachov theorem (Theorem 3.1), we can pass to another subsequence which converges strongly in $L^2(\Omega)$. Taking one last subsequence, we can assume that $p_k \to p^*$ a.e. in $\Omega$.

Fix $\epsilon > 0$. By Egoroff's theorem there exists a measurable set $U_\epsilon$ such that $p_k \to p^*$ uniformly on $U_\epsilon$ and $|\Omega - U_\epsilon| < \epsilon$. Also write

$$V_\epsilon = \left\{ x \in \Omega \middle| |p^*(\mathbf{x})| + |\nabla p^*(\mathbf{x})| < \frac{1}{\epsilon} \right\} \tag{4.8}$$

Then $|\Omega - V_\epsilon| \to 0$ as $\epsilon \to 0$. Let $O_\epsilon = U_\epsilon \cap V_\epsilon$ and note that

$$|\Omega - O_\epsilon| = |(\Omega - U_\epsilon) \cup (\Omega - V_\epsilon)| \leq |\Omega - U_\epsilon| + |\Omega - V_\epsilon| \to 0 \quad \text{as } \epsilon \to 0$$

Now

$$E(p_k) + \int_\Omega \hat{p} p_k \, d\mathbf{x} = \int_\Omega p_k^2 + hD(\mathbf{x})|\nabla p_k|^2 - hG(p_k) \, d\mathbf{x}.$$

From the proof of Lemma 4.1, i.e. (4.7), we know that the right-hand side is nonnegative. Thus, we have

$$E(p_k) + \int_\Omega \hat{p} p_k \, d\mathbf{x} \geq \int_{O_\epsilon} p_k^2 + hD(\mathbf{x})|\nabla p_k|^2 - hG(p_k) \, d\mathbf{x}.$$

Since the function $\eta : \mathbb{R}^n \to \mathbb{R}$ given by $\eta(\mathbf{x}) = |x|^2$ is convex, it follows that

$$\begin{aligned} E(p_k) + \int_\Omega \hat{p} p_k \, d\mathbf{x} &\geq \int_{O_\epsilon} p_k^2 + hD(\mathbf{x}) \left( |\nabla p^*|^2 + 2\nabla p^* \cdot (\nabla p_k - \nabla p^*) \right) - hG(p_k) \, d\mathbf{x} \\ &= \int_{O_\epsilon} p_k^2 + hD(\mathbf{x})|\nabla p^*|^2 - hG(p_k) \, d\mathbf{x} + \int_{O_\epsilon} 2hD(\mathbf{x}) \nabla p^* \cdot (\nabla p_k - \nabla p^*) \, d\mathbf{x}. \end{aligned} \tag{4.9}$$

Recall equation (4.8) and note that in the first integral every term is bounded above. In addition, $p_k \to p^*$ uniformly on $O_\epsilon$ and $G$ is an absolutely continuous function, so $G(p_k) \to G(p^*)$ uniformly on $O_\epsilon$. Thus,

$$\lim_{k\to\infty} \int_{O_\epsilon} p_k^2 + hD(\mathbf{x})|\nabla p^*|^2 - hG(p_k) \, d\mathbf{x} = \int_{O_\epsilon} (p^*)^2 + hD(\mathbf{x})|\nabla p^*|^2 - hG(p^*) \, d\mathbf{x} \tag{4.10}$$

As for the second integral, note that $\nabla p_k \to \nabla p^*$ weakly in $L^2(\Omega; \mathbb{R}^n)$. Since $hD(\mathbf{x})\nabla p^* \in L^2(\Omega; \mathbb{R}^n)$ it follows that

$$\lim_{k\to\infty} \int_{O_\epsilon} 2hD(\mathbf{x}) \nabla p^* \cdot (\nabla p_k - \nabla p^*) \, d\mathbf{x} = 0 \tag{4.11}$$

We then take limits as $k \to \infty$ on both sides of (4.9) and as a result of ( 4.10) and (4.11), we have

$$m + \int_\Omega \hat{p}p^* \, d\mathbf{x} \geq \int_{O_\epsilon} (p^*)^2 + hD(\mathbf{x})|\nabla p^*|^2 - hG(p^*) \, d\mathbf{x}$$
$$m \geq \int_{O_\epsilon} (p^*)^2 + hD(\mathbf{x})|\nabla p^*|^2 - hG(p^*) \, d\mathbf{x} - \int_\Omega \hat{p}p^* \, d\mathbf{x}$$

Now we take the limit as $\epsilon \to 0$. Since the integrand is nonnegative and $O_\epsilon \uparrow \Omega$, the monotone convergence theorem guarantees that

$$m \geq \int_\Omega (p^*)^2 + hD(\mathbf{x})|\nabla p^*|^2 - hG(p^*) - \hat{p}p^* \, d\mathbf{x} = E(p^*)$$

This completes the proof. □

**Theorem 4.2.** *There exists a function $p^* \in \mathcal{A}$ which minimizes the energy functional $E(p)$ defined in (4.6).*

*Proof.* Set $m := \inf_{p \in \mathcal{A}} E(p)$ and choose a minimizing sequence $\{p_k\}$. Then $E(p_k) \to m$. As a result of Lemma 4.1 we know that

$$||p_k||_{H_0^1(\Omega)} \leq E(p_k) + ||\hat{p}||_2^2$$

$E(p_k) \to m$, so $\sup_k E(p_k) < \infty$. Thus, the minimizing sequence is bounded in $H_0^1(\Omega)$. Since $H_0^1(\Omega)$ is weakly compact, there exists a subsequence $p_k$ which converges weakly to some function $p^* \in H_0^1(\Omega)$. We'd like to know that $p^*$ is also in the admissible set $\mathcal{A}$. By the Rellich-Kondrachov theorem (Theorem 3.1), we can pass to a subsequence which converges strongly in $L^2(\Omega)$. By taking another subsequence, we can assume that $p_k \to p^*$ a.e., so we conclude that $p^* \geq 0$ a.e. That is, $p^*$ is in the admissible set $\mathcal{A}$.

It remains to show that $p^*$ is a minimizer of $E(p)$. Lemma 4.2 assures us that

$$E(p^*) \leq \liminf_{k \to \infty} E(p_k) = m \tag{4.12}$$

Since $p^* \in \mathcal{A}$, we have $m \leq E(p)$. Together with (4.12), this implies that $E(p^*) = m = \min_{p \in \mathcal{A}} E(p)$. □

**Lemma 4.3.** *Recall that $M = \max_{\mathbf{x} \in \Omega, p \geq 0} F_1(p)$. Let $M' = \max_{\mathbf{x} \in \Omega, p \geq 0} F_1'(p)$. If $h$ is small enough so that*

$$2 - hM - hM' p_{max} > 0$$

*then the functional $E(p)$ defined in (4.6) is $\mu$-strongly convex. That is, $\exists \mu > 0$ such that*

$$E(\mathbf{y}) \geq E(\mathbf{x}) + \langle \nabla E(\mathbf{x}), \mathbf{x} - \mathbf{y} \rangle + \frac{\mu}{2} ||\mathbf{x} - \mathbf{y}||_2^2$$

*where $\langle \nabla E(\mathbf{x}), \mathbf{x} - \mathbf{y} \rangle$ is the Gâteaux derivative of $E$ at the point $\mathbf{x}$ in the direction $\mathbf{x} - \mathbf{y}$.*

*Proof.* We use an equivalent formulation of $\mu$-strong convexity. It is enough to show that $\forall q \in \mathcal{A}$ we have

$$\partial^2 E(p, q) \geq \mu \, ||q||_2^2$$

We compute the second Gâteaux derivative. Let $q \in \mathcal{A}$. Then the second derivative is given by $\mathcal{F}''(0)$ which can be calculated as follows.

$$\begin{aligned}
\mathcal{F}(t) =& E(p+tq) \\
\mathcal{F}'(t) =& \int_\Omega 2(p+tq)q \, d\mathbf{x} + 2h \int_\Omega D(\mathbf{x}) \nabla(p+tq) \cdot \nabla q \, d\mathbf{x} \\
& - h \int_\Omega (p+tq) F_1(p+tq) q \, d\mathbf{x} - \int_\Omega \hat{p} q \\
\mathcal{F}''(t) =& 2 \int_\Omega q^2 \, d\mathbf{x} + 2h \int_\Omega D(\mathbf{x}) |\nabla q|^2 \, d\mathbf{x} - h \int_\Omega F_1(p+qt) q^2 d\mathbf{x} \\
& + \int_\Omega (p+tq) F_1'(p+tq) q^2 \, d\mathbf{x} \\
\partial^2 E(p,q) = \mathcal{F}''(0) =& 2 \left|\left|q\right|\right|_2^2 + 2h \int_\Omega D(\mathbf{x}) |\nabla q|^2 \, d\mathbf{x} - h \int_\Omega F_1(p) q^2 d\mathbf{x} - h \int_\Omega p F_1'(p) q^2 \, d\mathbf{x} \\
\geq& 2 \left|\left|q\right|\right|_2^2 - hM \left|\left|q\right|\right|_2^2 - hM' p_{\max} \left|\left|q\right|\right|_2^2 \\
=& (2 - hM - hM' p_{\max}) \left|\left|q\right|\right|_2^2
\end{aligned}$$

as desired. □

Note that we shall use the following relation throughout the rest of the paper.

$$\mathcal{F}'(0) = \langle \nabla E(p), q \rangle = \int_\Omega 2pq \, d\mathbf{x} + 2h \int_\Omega D(\mathbf{x}) \nabla p \cdot \nabla q d\mathbf{x} - h \int_\Omega p F_1(p) q d\mathbf{x} - \int_\Omega \hat{p} q d\mathbf{x}$$

**Theorem 4.3.** *The minimization problem in (4.6) has a unique minimizer.*

*Proof.* Suppose $p$ and $\tilde{p}$ are both minimizers of $E(p)$. Then for any $q \in H_0^1(\Omega)$ we have

$$\langle \nabla E(p), q - p \rangle \geq 0 \text{ and } \langle \nabla E(\tilde{p}), q - \tilde{p}) \rangle \geq 0, \quad \forall q \in \mathcal{A}.$$

By Lemma 4.3 the following two inequalities hold.

$$E(p) \geq E(\tilde{p}) + \frac{\mu}{2} \left|\left|p - \tilde{p}\right|\right|_2^2 \text{ and } E(\tilde{p}) \geq E(p) + \frac{\mu}{2} \left|\left|p - \tilde{p}\right|\right|_2^2$$

Adding the two inequalities yields $0 \geq \mu \left|\left|p - \tilde{p}\right|\right|_2^2$. Thus, $p = \tilde{p}$ a.e. and hence, the minimizer is unique. □

**Theorem 4.4.** *A function $p \in \mathcal{A}$ is the minimizer of (4.6) if and only if $p$ is a discrete weak solution to (4.4) and $p \in \mathcal{A}$.*

*Proof.* It is clear that when $p \in \mathcal{A}$ is a discrete weak solution (4.4), we have $\langle \nabla E(p), q \rangle = 0$. Hence, we have

$$\langle \nabla E(p), q - p \rangle \geq 0, \quad \forall q \in \mathcal{A}. \tag{4.13}$$

Thus, $p$ is a minimizer of (4.6).

One the other hand, as seen from Theorem 4.3, there is a unique minimizer $p^*$ of (4.6). We can use the standard projected gradient method to find $p^*$ which is given as follows.

**Algorithm 4.1.** *Starting with $P^1 = \hat{p} \in \mathcal{A}$, we iteratively compute $\tilde{P}^{k+1}$*

$$\int_\Omega \tilde{P}^{k+1} q d\mathbf{x} = \int_\Omega P^k q d\mathbf{x} - \tau \langle \nabla E(P^k), q \rangle, \quad q \in H_0^1(\Omega) \tag{4.14}$$

*and find $P^{k+1} = \mathcal{P}_+(\tilde{P}^{k+1})$ for $k = 1, \cdots$, until the consecutive error $\|P^{k+1} - P^k\|_2$ is within a tolerance, where $\mathcal{P}_+$ is the nonnegative projection operator and $\tau > 0$ is a step size.*

It is standard that $\mathcal{P}_+(p^*) = \mathcal{P}_+(p^* - \tau \nabla E(p^*))$ and

$$\int_\Omega |\mathcal{P}_+(p) - \mathcal{P}_+(q)|^2 d\mathbf{x} \leq \int_\Omega |p - q|^2 d\mathbf{x}, \quad \forall p, q \in L^2(\Omega). \tag{4.15}$$

Thus, we have

$$\begin{aligned}
&\|P^{k+1} - p^*\|^2_{L^2(\Omega)} = \int_\Omega |\mathcal{P}_+(\tilde{P}^{k+1}) - \mathcal{P}_+(p^*)|^2 d\mathbf{x} \\
&\leq \int_\Omega |\tilde{P}^{k+1} - (p^* - \tau \nabla E(p^*))|^2 d\mathbf{x} \\
&= \int_\Omega |P^k - p^* - \tau(\nabla E(P^k) - \nabla E(p^*))|^2 d\mathbf{x} \\
&= \|P^k - p^*\|^2_{L^2(\Omega)} - 2\tau \int_\Omega (P^k - p^*)(\nabla E(P^k) - \nabla E(p^*)) d\mathbf{x} + \tau^2 \|\nabla E(P^k) - \nabla E(p^*)\|^2_{L^2(\Omega)} \\
&\leq \|P^k - p^*\|^2_{L^2(\Omega)} (1 - 2\tau\mu + \tau^2 L^2)
\end{aligned}$$

where $\mu$ is the constant in Lemma 4.3 and $L$ is the Lipschitz differentiability of the energy functional $E$. As long as $\tau < 2\mu/(L^2)$, we have $\rho = 1 - 2\tau\mu + \tau^2 L^2 < 1$. For example, $\tau = \mu/L^2$ is a good choice. Thus, it follows that $P^k, k \geq 1$ are a Cauchy sequence and converge to $p^*$ in $L^2(\Omega)$ norm.

Furthermore, we can consider $\|\tilde{P}^{k+1} - \tilde{P}^{\ell+1}\|^2_{L^2(\Omega)}$ and use the above analysis to conclude that $\tilde{P}^k, k \geq 1$ are a Cauchy sequence in $L^2(\Omega)$ norm and hence, converge to a limit by $\tilde{P}^*$. It follows that $p^* = \mathcal{P}_+(\tilde{P}^*)$ almost every where in $\Omega$ since

$$\begin{aligned}
\|p^* - \mathcal{P}_+(\tilde{P}^*)\|^2_{L^2(\Omega)} &\leq \|p^* - P^k\|^2_{L^2(\Omega)} + \|P^k - \mathcal{P}_+(\tilde{P}^*)\|^2_{L^2(\Omega)} \\
&= \|p^* - P^k\|^2_{L^2(\Omega)} + \|\mathcal{P}_+(\tilde{P}^k) - \mathcal{P}_+(\tilde{P}^*)\|^2_{L^2(\Omega)} \\
&\leq \|p^* - P^k\|^2_{L^2(\Omega)} + \|\tilde{P}^k - \tilde{P}^*\|^2_{L^2(\Omega)} \to 0
\end{aligned}$$

when $k \to \infty$.

We now claim that $\tilde{P}^* \in \mathcal{A}$. Otherwise, let $\omega = \{(x, y) \in \Omega, \tilde{P}^* < 0\}$. If $\omega$ is an open set with a positive measure, we can choose a function $q \in H_0^1(\Omega)$ such that $q = 1$ in an interior of $\omega$ and 0 outside of $\omega$ such that

$$\begin{aligned}
\int_\Omega \frac{p^* - \tilde{P}^*}{\tau} q d\mathbf{x} &= \int_\Omega \nabla E(p^*) q d\mathbf{x} \\
&= \int_\Omega p^* q \mathbf{x} + h \int_\Omega D(\mathbf{x}) \nabla p^* \cdot \nabla q d\mathbf{x} - h \int_\Omega p^* F_1(p^*) q d\mathbf{x} - \int_\Omega \hat{p} q d\mathbf{x} \\
&= \int_\omega p^* q \mathbf{x} + h \int_\omega D(\mathbf{x}) \nabla p^* \cdot \nabla q d\mathbf{x} - h \int_\omega p^* F_1(p^*) q d\mathbf{x} - \int_\omega \hat{p} q d\mathbf{x}
\end{aligned} \tag{4.16}$$

It follows that

$$0 < \int_\omega \frac{-\tilde{P}^*}{\tau} d\mathbf{x} = -\int_\omega \hat{p} q d\mathbf{x} \leq 0 \tag{4.17}$$

which is a contradiction as $\hat{p} \geq 0$. If $\omega$ is not an open set, we can choose an open set $\widetilde{\omega}$ containing $\omega$ such that the measure of $\widetilde{\omega} \backslash \omega$ arbitrarily close to zero. We shall have the similar equality of (4.16) with $\widetilde{\omega}$ replacing $\omega$ and use a $q \in H_0^1(\Omega)$ which is 1 on an interior of $\widetilde{\omega}$ while zero out of $\widetilde{\omega}$. As $p^* \in H_0^1(\Omega)$, the terms on the right-hand side of the modified version of (4.16) can be arbitrarily small except for the last term, i.e. $-\int_{\widetilde{\omega}} \hat{p} q d\mathbf{x}$ while the left-hand side term is $\int_{\widetilde{\omega}} \frac{-\tilde{P}^*}{\tau} q d\mathbf{x} > 0$. These show that $\omega$ has to be of zero measure. Hence, $\tilde{P}^* \geq 0$ almost everywhere in $\Omega$. That is, $p^* = \mathcal{P}_+(\tilde{P}^*) = \tilde{P}^*$. From (4.14), it follows that $\langle \nabla E(p^*), q \rangle = 0, \quad \forall q \in H_0^1(\Omega)$. □

**Remark 4.1.** *Theorem 4.4 implies that there exists a unique discrete weak solution to (4.4).*

**Lemma 4.4.** *The minimizer $p^*$ of the energy functional (4.6), hereby denoted by $E_{\hat{p}}$, is stable with respect to perturbations in $\hat{p}$. In particular, if we let $q^*$ be the minimizer associated with the energy functional*

$$E_{\hat{q}}(q) = \int_\Omega q^2 \, d\mathbf{x} + h \int_\Omega D(\mathbf{x}) |\nabla q|^2 \, d\mathbf{x} - h \int_\Omega G(q) \, d\mathbf{x} - \int_\Omega \hat{q} q \, d\mathbf{x}$$

*then we are assured that*

$$||p^* - q^*||_2 \leq \frac{1}{\mu} ||\hat{p} - \hat{q}||_2$$

*Proof.* Since $p^*$ is the minimizer, we know $\partial E_{\hat{p}}(p^*, \nu) = 0$ for all $\nu$. Similarly, $\partial E_{\hat{q}}(q^*, \nu) = 0$ for all $\nu$. As a result of Lemma 4.3 we get the following two inequalities.

$$E_{\hat{p}}(q^*) \geq E_{\hat{p}}(p^*) + \frac{\mu}{2} \|p^* - q^*\|_2^2$$
$$E_{\hat{q}}(p^*) \geq E_{\hat{q}}(q^*) + \frac{\mu}{2} \|p^* - q^*\|_2^2$$

We add the two inequalities. After some cancellation we obtain the following inequality.

$$\begin{aligned} -\langle \hat{p}, q^* \rangle - \langle \hat{q}, p^* \rangle &\geq -\langle \hat{p}, p^* \rangle - \langle \hat{q}, q^* \rangle + \mu \|p^* - q^*\|_2^2 \\ \langle \hat{p}, p^* - q^* \rangle - \langle \hat{q}, p^* - q^* \rangle &\geq \mu \|p^* - q^*\|_2^2 \\ \langle \hat{p} - \hat{q}, p^* - q^* \rangle &\geq \mu \|p^* - q^*\|_2^2 \end{aligned}$$

We use the Cauchy-Schwarz's inequality to conclude

$$\|p^* - q^*\|_2 \leq \frac{1}{\mu} \|\hat{p} - \hat{q}\|_2$$

which is the desired inequality. □

## 5. Bivariate Spline Approximation of the Discrete Weak Solution

### *5.1. The Discrete Weak Solution in Finite Dimensional Space*

So far we have established that there exists a unique discrete weak solution to the problem posed in (4.4). Our next goal is to find an approximate solution in a finite-dimensional spline space. That is, we will approximate $p$ and $\hat{p}$ by using the spline space $S_d^r(\triangle)$ defined as follows.

**Definition 5.1** (Spline Space)**.** *Let $\triangle$ be a given triangulation of the domain $\Omega$ of interest. Then we define the spline space of smoothness $r \geq 1$ and degree $d \geq 3r+2$ over $\triangle$ by,*

$$S_d^r(\triangle) = \{s \in C^r(\Omega) \mid s|_T \in \mathcal{P}_d, \ \forall\ T \in \triangle\},$$

*where $\mathcal{P}_d$ is the space of polynomials of degree at most $d$.*

It is known from [20] that $S_d^1(\triangle)$ with $d \geq 5$ can be dense in $H_0^1(\Omega)$ when the size $|\triangle|$ of triangulation $\triangle$ goes to zero. See Theorem 8.1 in Appendix. We now set out to find $p^* \in \mathcal{A} \cap S_d^1(\triangle), d \geq 5$ which satisfies the following minimization problem:

$$\min_{p \in \mathcal{A} \cap S_d^1(\triangle)} E(p). \tag{5.1}$$

Similar to the previous section, we can show that the minimization above has a unique minimizer and is stable. It is known that the minimizer $p^*$ satisfies

$$\langle \nabla E(p^*), q - p^* \rangle \geq 0, \quad \forall q \in \mathcal{A} \cap S_d^1(\triangle). \tag{5.2}$$

Based on Theorem 4.4, one approach to find $p^*$ is to solve the following

$$\int_\Omega pq \, d\mathbf{x} + h \int_\Omega D(\mathbf{x}) \nabla p \cdot \nabla q \, d\mathbf{x} = \int_\Omega \hat{p} q \, d\mathbf{x} + h \int_\Omega p F_1(p) q \, d\mathbf{x}, \qquad \forall q \in S_d^r(\triangle). \tag{5.3}$$

If $p^*$ satisfies the above equations, i.e. $\nabla E(p^*) = 0$ and $p^* \in \mathcal{A}$, we have (5.2) and hence, $p^*$ is the minimizer.

**Theorem 5.1.** *If $h$ is small enough, then there is $p^* \in S_d^r(\triangle)$ satisfies (5.3).*

*Proof.* The proof of this theorem is constructive and we only give an overview of the construction here. The detail is contained in the rest of this subsection and the next subsection. We first devise an iterative computational scheme. Each iteration requires solving a simple linear equation, for which we can guarantee the existence of such iterative solution. We then show that this sequence of iterative solutions actually forms a Cauchy sequence. Thus, the sequence converges to a spline in $S_d^r(\triangle)$ which is a finite dimensional, and hence a complete space. Finally, by simply taking limits as the number of iteration goes to infinity, we demonstrate that we get a discrete weak spline solution satisfying (5.3). □

We shall need the following

**Theorem 5.2.** *The weak solution of (5.3) is unique.*

*Proof.* The proof is analogous to the one in Theorem 4.3. Detail is omitted here. □

*5.2. Our Computational Scheme*

At each time step $t_i$, we have to solve the nonlinear problem (5.3). We shall linearize the equation and use a fixed-point method as explained below.

**Algorithm 5.1.** *Writing $\hat{p} = p(\mathbf{x}, t_{i-1})$ or $\hat{p} = p_0(\mathbf{x}, t_0)$, the initial value, find $p^{(k)} := p^{(i,k)}, k \geq 1$ such that*

$$\int_\Omega p^{(k)} q + hD \int_\Omega \nabla p^{(k)} \cdot \nabla q = \langle \hat{p}, q \rangle + h \int_\Omega p^{(k)} F_1\left(p^{(k-1)}\right) q \, d\mathbf{x} \quad \forall q \in S_d^r(\triangle) \tag{5.4}$$

*for $k = 1, 2, \cdots$, until a given accuracy for $\|p^{(k)} - p^{(k-1)}\|$ is met.*

**Remark 5.1.** *We stated in the outline of the proof for Theorem 5.1 that we will show the sequence of $p^{(k)}$ is Cauchy and hence converges to a limit $p^* \in S_d^r(\triangle)$. Note that in (5.4), we can take the limit as $k \to \infty$ of both sides and obtain precisely (5.3). This requires the use of the Dominated Convergence Theorem and so we prove boundedness of all the iterates in Theorem 5.3.*

**Remark 5.2.** *Although one can use the standard projected gradient method to find the minimizer $p^*$ of (5.1), we use Algorithm 5.1 to do our computation.*

**Lemma 5.1.** *Given splines $p^{(k-1)}$ and $\hat{p} \geq 0$, there exists a unique spline solution for $p^{(k)}$ in equation (5.4).*

*Proof.* We shall denote by $\{\phi_j\}_{1 \leq j \leq n}$ the basis of this space $S_d^1(\triangle)$. Any spline function in $S_d^r(\triangle)$ can be written as $\sum_{i=1}^n c_i \phi_i$. Let $\vec{c}$ be the vector of coefficients for $p^{(k)}$ and $\vec{p}$ be the vector of coefficients for $\hat{p}$. Define the following matrices.

$$\begin{aligned} M(i,j) &:= \int_\Omega \phi_i \phi_j \, d\mathbf{x} \\ K_D(i,j) &:= \int_\Omega D(\mathbf{x}) \nabla \phi_i \cdot \nabla \phi_j \, d\mathbf{x} \\ M_{F(p^{(k-1)})}(i,j) &:= \int_\Omega F_1(p^{(k-1)}) \phi_i \phi_j \, d\mathbf{x} \end{aligned}$$

Note that all these matrices are symmetric. In addition, $M$ is positive-definite.

We have to solve (5.4) for each $q \in S_d^r(\triangle)$, but it's sufficient to solve for each basis spline $\phi_j$. Thus, we have $n$ equations and $n$ unknowns in the coefficient vector, which is equivalent to the following linear system.

$$\begin{aligned} M\vec{c} + hK_D\vec{c} &= M\vec{p} + hM_{F(p^{(k-1)})}\vec{c} \\ \left(M + hK_D - hM_{F(p^{(k-1)})}\right)\vec{c} &= M\vec{p} \end{aligned}$$

Let $L = M + hK_D - hM_{F(p^{(k-1)})}$. $M$ is positive-definite and invertible. If $h$ is small enough, $L$ is also invertible. Thus, we can solve for $\vec{c}$, the spline coefficients of $p^{(k)}$. □

**Theorem 5.3.** *If $h < 1/M$, then the successive solutions $p^{(k)}$ of the equation (5.4) satisfy*

$$\left|\left|p^{(k)}\right|\right|_2 \leq \frac{1}{1 - hM} \left|\left|\hat{p}\right|\right|_2 \tag{5.5}$$

$$\left|\left|\nabla p^{(k)}\right|\right|_2 \leq \frac{1}{\sqrt{hK}} \sqrt{\left|\left|p^{(k)}\right|\right|_2 \left(\left|\left|\hat{p}\right|\right|_2 - (1 - hM)\left|\left|p^{(k)}\right|\right|_2\right)} \tag{5.6}$$

*If we substitute the estimate from (5.5) into (5.7), we obtain a bound which is less sharp but is independent of $k$.*

$$\left|\left|\nabla p^{(k)}\right|\right|_2 \leq \frac{1}{\sqrt{hK}} \sqrt{\left|\left|p^{(k)}\right|\right|_2 \left|\left|\hat{p}\right|\right|_2} \leq \frac{1}{\sqrt{hK(1 - hM)}} \left|\left|\hat{p}\right|\right|_2$$

*Proof.* Substitute $q = p$ into (5.4). Then

$$\left|\left|p^{(k)}\right|\right|_2^2 + h \underbrace{\int_\Omega D(\mathbf{x}) |\nabla p^{(k)}|^2 \, d\mathbf{x}}_{\geq 0} = \langle \hat{p}, p^{(k)} \rangle + h \int_\Omega F(p^{(k-1)}) (p^{(k)})^2 \, d\mathbf{x}$$

Use the Cauchy-Schwarz inequality and the fact that $F(p) \leq M$ for any $p$.

$$\begin{aligned}
\left|\left|p^{(k)}\right|\right|_2^2 &\leq ||\hat{p}||_2 \left|\left|p^{(k)}\right|\right|_2 + hM \left|\left|p^{(k)}\right|\right|_2^2 \\
\left|\left|p^{(k)}\right|\right|_2 &\leq ||\hat{p}||_2 + hM \left|\left|p^{(k)}\right|\right|_2 \\
\left|\left|p^{(k)}\right|\right|_2 &\leq \frac{1}{1-hM} ||\hat{p}||_2
\end{aligned}$$

Now we prove the bound for $\nabla p^{(k)}$ by substituting $q = p$ once more into (5.4).

$$\begin{aligned}
\left|\left|p^{(k)}\right|\right|_2^2 + h\int_\Omega D(\mathbf{x})|\nabla p^{(k)}|^2 \, d\mathbf{x} &= \langle \hat{p}, p^{(k)} \rangle + h\int_\Omega F(p^{(k-1)})(p^{(k)})^2 \, d\mathbf{x} \\
\left|\left|p^{(k)}\right|\right|_2^2 + hK\left|\left|\nabla p^{(k)}\right|\right|_2^2 &\leq ||\hat{p}||_2 \left|\left|p^{(k)}\right|\right|_2 + hM\left|\left|p^{(k)}\right|\right|_2^2 \\
hK\left|\left|\nabla p^{(k)}\right|\right|_2^2 &\leq ||\hat{p}||_2 \left|\left|p^{(k)}\right|\right|_2 - \left|\left|p^{(k)}\right|\right|_2^2 + hM\left|\left|p^{(k)}\right|\right|_2^2 \\
hK\left|\left|\nabla p^{(k)}\right|\right|_2^2 &\leq \left|\left|p^{(k)}\right|\right|_2 \left(||\hat{p}||_2 - (1-hM)\left|\left|p^{(k)}\right|\right|_2\right) \\
\left|\left|\nabla p^{(k)}\right|\right|_2 &\leq \frac{1}{\sqrt{hK}} \sqrt{||p^{(k)}||_2 \left(||\hat{p}||_2 - (1-hM)\,||p^{(k)}||_2\right)}.
\end{aligned}$$

These complete the proof. □

**Remark 5.3.** *The constant in the bound for $\nabla p^{(k)}$, which can be found under the square root, is non-negative as a result of the bound for $p^{(k)}$. In fact, it can be very close to zero.*

**Remark 5.4.** *Since we are now working within a finite-dimensional space, all norms are equivalent. As a result, we have just established that $p$ and its derivatives are bounded functions. That is,*

$$\left|\left|p^{(k)}\right|\right|_\infty \leq \frac{C}{1-hM} ||\hat{p}||_2$$

*for a positive constant $C$.*

**Theorem 5.4.** *If $h$ is small enough so that*

$$hL\frac{C}{(1-hM)^2} ||\hat{p}||_2 < 1$$

*where $C$ is the constant from Remark 5.4, then successive iterates of (5.4) are a Cauchy sequence in $L^2(\Omega)$. That is,*

$$\left|\left|p^{(k)} - p^{(k-1)}\right|\right|_2 \leq \alpha \left|\left|p^{(k-1)} - p^{(k-2)}\right|\right|_2,$$

*where $0 < \alpha < 1$. Furthermore, when $D(\mathbf{x}) \geq K > 0$ over $\Omega$, $p^k, k \geq 1$ are a Cauchy sequence in $H^1(\Omega)$.*

*Proof.* Take two successive solutions which satisfy the following equations.

$$\begin{aligned}
\int_\Omega p^{(k)} q \, d\mathbf{x} + h\int_\Omega D(\mathbf{x})\nabla p^{(k)} \cdot \nabla q \, d\mathbf{x} &= \int_\Omega \hat{p} q \, d\mathbf{x} + h\int_\Omega p^{(k)} F(p^{(k-1)}) q \, d\mathbf{x} \\
\int_\Omega p^{(k-1)} q \, d\mathbf{x} + h\int_\Omega D(\mathbf{x})\nabla p^{(k-1)} \cdot \nabla q \, d\mathbf{x} &= \int_\Omega \hat{p} q \, d\mathbf{x} + h\int_\Omega p^{(k-1)} F(p^{(k-2)}) q \, d\mathbf{x}
\end{aligned}$$

Subtract the two equations and substitute $q = p^{(k)} - p^{(k-1)}$.

$$\left|\left|p^{(k)} - p^{(k-1)}\right|\right|_2^2 + h \underbrace{\int_\Omega D(\mathbf{x})|\nabla p^{(k)} - \nabla p^{(k-1)}|^2 \, d\mathbf{x}}_{\geq 0} \tag{5.7}$$

$$= h \int_\Omega \left(F(p^{(k-1)})p^{(k)} - F(p^{(k-2)})p^{(k-1)}\right)\left(p^{(k)} - p^{(k-1)}\right) d\mathbf{x} \tag{5.8}$$

Add and subtract $F(p^{(k-1)})$ and rearrange on the right-hand side. We have

$$\left|\left|p^{(k)} - p^{(k-1)}\right|\right|_2^2 \leq h \int_\Omega F(p^{(k-1)}) \left(p^{(k)} - p^{(k-1)}\right)^2 \tag{5.9}$$

$$+ h \int_\Omega \left(F(p^{(k-1)}) - F(p^{(k-2)})\right) p^{(k-1)}(p^{(k)} - p^{(k-1)}) \, d\mathbf{x}. \tag{5.10}$$

Use Remark 5.4 to bound $\|p^{(k-1)}\|_\infty$ and hence,

$$\left|\left|p^{(k)} - p^{(k-1)}\right|\right|_2^2 \leq hM \left|\left|p^{(k)} - p^{(k-1)}\right|\right|_2^2 + h\frac{C}{1-hM} \left|\left|\hat{p}\right|\right|_2 \int_\Omega \left|F(p^{(k-1)}) - F(p^{(k-2)})\right| \left|p^{(k)} - p^{(k-1)}\right| \, d\mathbf{x}$$

Group like terms, we have

$$(1-hM)\left|\left|p^{(k)} - p^{(k-1)}\right|\right|_2^2 \leq h\frac{C}{1-hM} \left|\left|\hat{p}\right|\right|_2 \int_\Omega \left|F(p^{(k-1)}) - F(p^{(k-2)})\right| \left|p^{(k)} - p^{(k-1)}\right| \, d\mathbf{x}.$$

As $F(p)$ is a differentiable function and by Remark 5.4, it has a bounded derivative on the compact interval $\left[0, \sup_k \left|\left|p^{(k)}\right|\right|_\infty\right]$. Thus, $F(p)$ is Lipschitz continuous with constant $L_F$.

$$(1-hM)\left|\left|p^{(k)} - p^{(k-1)}\right|\right|_2^2 \leq hL_F\frac{C}{1-hM} \left|\left|\hat{p}\right|\right|_2 \int_\Omega \left|p^{(k-1)} - p^{(k-2)}\right| \left|p^{(k)} - p^{(k-1)}\right| \, d\mathbf{x}.$$

Apply the Cauchy-Schwartz inequality to get

$$(1-hM)\left|\left|p^{(k)} - p^{(k-1)}\right|\right|_2^2 \leq hL_F\frac{C}{1-hM} \left|\left|\hat{p}\right|\right|_2 \left|\left|p^{(k-1)} - p^{(k-2)}\right|\right|_2 \left|\left|p^{(k)} - p^{(k-1)}\right|\right|_2$$

$$\left|\left|p^{(k)} - p^{(k-1)}\right|\right|_2 \leq hL_F\frac{C}{(1-hM)^2} \left|\left|\hat{p}\right|\right| \left|\left|p^{(k-1)} - p^{(k-2)}\right|\right|_2$$

We can choose a small enough $h$ so that $\alpha = hL\dfrac{C}{(1-hM)^2}$ satisfies $0 < \alpha < 1$.

Furthermore, we use (5.10) again with the property $D(\mathbf{x}) \geq K > 0$ together with the fact that $p^k, k \geq 1$ are a Cauchy sequence in $L^2(\Omega)$ to conclude that $p^k, k \geq 1$ is a Cauchy sequence in $H^1(\Omega)$. □

*5.3. Bivariate Spline Approximation to the Discrete Weak Solution in Sobolev Space*

In this subsection, we show that the spline solutions obtained above are a good approximation to the weak solution in (4.4). Let $p^*$ be the weak solution of (4.4) and let $S^*$ be the spline solution

which is the limit of the iterative solutions from Algorithm 5.1. By using Lemma 4.3 and noting that $\langle \nabla E(p^*), q - p^* \rangle \geq 0$ for any $q \in H_0^1(\Omega)$, we have

$$E(S^*) - E(p^*) \geq \frac{\mu}{2} ||S^* - p^*||_2^2 \tag{5.11}$$

Let $S_{p^*}$ be the nonnegative preserving interpolation of $p^*$ in the spline space $S_d^1(\triangle)$ with $d \geq 5$ as in [18]. Since $S^*$ is the minimizer of (4.6) with respect to all $q \in S_d^r(\triangle)$, we conclude that $E(S_{p^*}) \geq E(S^*)$. Together with (5.11) we can write

$$\frac{\mu}{2} ||S^* - p^*||_2^2 \leq E(S_{p^*}) - E(p^*) \tag{5.12}$$

**Theorem 5.5.** *Suppose that $h > 0$ is small enough and $p^* > 0$, the weak solution of (4.4), is in $H^2(\Omega)$ and continuously twice differentiable over the closure of $\Omega$. Then $S^*$, the limit of the iterative solutions from Algorithm 5.1, approximates $p^*$ in the following sense:*

$$||S^* - p^*||_2^2 \leq C|\triangle||p^*|_{2,2,\Omega}, \tag{5.13}$$

*where $C$ is a positive constant and $|\triangle|$ small enough.*

*Proof.* We rewrite equation (5.12)

$$\begin{aligned}
\frac{\mu}{2} ||S^* - p^*||_2^2 &\leq \int_\Omega S_{p^*}^2 - (p^*)^2 \, d\mathbf{x} + h \int_\Omega D(\mathbf{x}) \left( |\nabla S_{p^*}|^2 - |\nabla p^*|^2 \right) \, d\mathbf{x} + h \int_\Omega G(p^*) - G(S_{p^*}) \, d\mathbf{x} \\
&= \int_\Omega (S_{p^*} - p^*)(S_{p^*} + p^*) \, d\mathbf{x} \\
&\quad + h \int_\Omega D(\mathbf{x}) (\nabla S_{p^*} - \nabla p^*) \cdot (\nabla S_{p^*} + \nabla p^*) \, d\mathbf{x} + h \int_\Omega G(p^*) - G(S_{p^*}) \, d\mathbf{x}
\end{aligned}$$

Note that $G$ is a differentiable function by construction. Since $p^* \in H^2(\Omega)$, by Theorem 3.2 we conclude that $p^*$ is Hölder continuous and hence it has some maximal value $M^*$ on the compact set $\overline{\Omega}$. Analogously, we can conclude the same for $S_{p^*}$. As a result, $G'(p)$ has a maximum value on the compact set $[0, M^*]$ and so $G$ is Lipschitz continuous with some constant $L_G$. Continuing where we left off above, we use the Cauchy-Schwarz inequality and $L_G$:

$$\begin{aligned}
\frac{\mu}{2} ||S^* - p^*||_2^2 \leq & ||S_{p^*} - p^*||_2 \, ||S_{p^*} + p^*||_2 \\
& + hK_2 \, ||\nabla S_{p^*} - \nabla p^*||_2 \, ||\nabla S_{p^*} + \nabla p^*||_2 + hL_G \int_\Omega |p^* - S_{p^*}| \, d\mathbf{x} \\
\leq & C_1 \, ||S_{p^*} - p^*||_2 + hK_2 C_2 \, ||\nabla S_{p^*} - \nabla p^*||_2 + hL_G |\Omega|^{1/2} \, ||p^* - S_{p^*}||_2 ,
\end{aligned}$$

where $C_1 = ||S_{p^*}||_2 + ||p^*||_2$, $C_2 = ||\nabla S_{p^*}||_2 + ||\nabla p^*||_2$.

By the approximation property of nonnegative preserving interpolatory splines, Theorem 2.3 in [18] and the standard approximation property of spline spaces, i.e. Theorem 8.1 together with the Markov inequality, i.e. Theorem 8.2 as in Appendix, we then write

$$\begin{aligned}
||S_{p^*} - p^*||_2 &\leq C_3 |\triangle|^2 |p^*|_{2,\infty,\Omega} \\
||\nabla S_{p^*} - \nabla p^*||_2 &\leq C_4 |\triangle| |p^*|_{2,\infty,\Omega}
\end{aligned}$$

where $|\triangle|$ is the length of the longest edge in the triangulation which is small enough and $C_3$ and $C_4$ are positive constants independent of $p^*$. In addition, we have assumed that $\triangle$ is a triangulation with all interior edge active, e.g., $\triangle$ is an acute triangulation. □

As a corollary, we have that $E(S_{p^*}) - E(p^*) \to 0$ as $|\triangle| \to 0$ and hence, $S^* \to p^*$ by using (5.12).

## 6. Computational Results

We have implemented the computational scheme discussed in the previous section in MATLAB 2014a. In this section we will show some of our computational results. Since no exact solutions to this PDE were known, we modify the equation by adding an appropriate forcing term. By doing so we can force any twice differentiable function $p(\mathbf{x}, t)$ to be a solution. Then we can make sure our algorithm recovers it for any given just $p(\mathbf{x}, 0)$, the initial condition. In this way, we are able to verify that our MATLAB code works. Then we remove the forcing term and use the resulting MATLAB code to numerically solve (1.2) for various initial conditions, various diffusive factor $D(\mathbf{x})$, and Allee constant $\sigma$.

Although in this paper we focused on the theory for Dirichlet zero boundary condition, the theory holds equally well for Neumann boundary conditions. We tested both boundary conditions numerically. Numerical results are explained in the following subsequences.

### *6.1. Dirichlet Boundary with Forcing*

In order to make sure that our implementation works, we use a few test functions over a rectangular domain $\Omega = [0, 1] \times [0, 1]$. These test functions are not weak solutions to the PDE in (4.4), but instead they are the exact solutions to the following modified PDE with forcing term.

$$\frac{\partial p(\mathbf{x}, t)}{\partial t} = \text{ div } (D(\mathbf{x})\nabla p(\mathbf{x}, t))) + p(\mathbf{x}, t)F(p(\mathbf{x}, t)) + f(\mathbf{x}, t), \tag{6.1}$$

where $\mathbf{x} = (x, y) \in \Omega \subset \mathbb{R}^2, t \geq 0$, and $f(\mathbf{x}, t)$ can be computed with a CAS. The weak solution $p$ satisfies

$$\begin{aligned}\int_\Omega p(\mathbf{x}, t_i)q(\mathbf{x})d\mathbf{x} + h\int_\Omega D(\mathbf{x})\nabla p(\mathbf{x}, t_i) \cdot \nabla q(\mathbf{x})d\mathbf{x} \\ = \int_\Omega p(\mathbf{x}, t_{i-1})q(\mathbf{x})d\mathbf{x} + \int_\Omega p(\mathbf{x}, t_i)F(p(\mathbf{x}, t_i)d\mathbf{x} \\ + \int_\Omega f(\mathbf{x}, t_i)q(\mathbf{x})d\mathbf{x}.\end{aligned} \tag{6.2}$$

We then ran our MATALB code to recover the function $p$ using bivariate splines of degree $d$ and recorded the maximal error in population density at some fixed time on a 100 by 100 grid. The numerical results are given in Tables 1, 2, where $d$ is the spline degree and $|\triangle|$ is the number of triangles in the triangulation, $h$ is the size of the time step, T refers to how far in time we have evolved. In all cases $A(\mathbf{x}) = 1$ and the domain is $\Omega = [0, 1] \times [0, 1]$.

Table 1: $d = 5$, $T = 5$, $p(\mathbf{x}, t) = \dfrac{13x(x-1)y(y-1)}{1+t}$, $D(\mathbf{x}) = 1/200$.

| $h\backslash\lvert\triangle\rvert$ | 2 | 8 | 32 | 128 | 512 |
|---|---|---|---|---|---|
| $5 \times 10^{-2}$ | 0.039429 | 0.032977 | 0.034431 | 0.034433 | 0.034433 |
| $5 \times 10^{-3}$ | 0.054059 | 0.0041368 | 0.0033432 | 0.0033453 | 0.0033454 |
| $5 \times 10^{-4}$ | 0.055708 | 0.0055911 | 3.3120e-04 | 3.3343e-04 | 3.3353e-04 |
| $5 \times 10^{-5}$ | 0.055873 | 0.0057463 | 3.1034e-05 | 3.3341e-05 | 3.3341e-05 |

Table 2: $d = 5$, $T = 1$, $p(\mathbf{x}, t) = \dfrac{13x(x-1)y(y-1)}{1+t}$, $D(x, y) = \dfrac{1}{200}e^{-(x-.5)^2-(y-.5)^2}$.

| $h\backslash\lvert\triangle\rvert$ | 2 | 8 | 32 | 128 |
|---|---|---|---|---|
| $5 \times 10^{-2}$ | 0.019157 | 0.016883 | 0.016599 | 0.016599 |
| $5 \times 10^{-3}$ | 0.0046455 | 0.0019749 | 0.0016833 | 0.0016832 |
| $5 \times 10^{-4}$ | 0.0043546 | 4.7506e-04 | 1.6861e-04 | 1.6852e-04 |

Table 3: $d = 5$, $T = 1$, $p(\mathbf{x}, t) = \dfrac{\sin(\pi x)\sin(\pi y)}{1+t}$, $D(\mathbf{x}) = 1/200$.

| $h\backslash\lvert\triangle\rvert$ | 2 | 8 | 32 | 128 |
|---|---|---|---|---|
| $5 \times 10^{-2}$ | 0.01982 | 0.019187 | 0.018405 | 0.018398 |
| $5 \times 10^{-3}$ | 0.005989 | 0.0026479 | 0.0018789 | 0.0018710 |
| $5 \times 10^{-4}$ | 0.005609 | 9.6351e-04 | 1.962e-04 | 1.8732e-04 |

In Table 1 we see that in order to reduce the error, it is necessary to reduce both $h$ and the size of the triangulation. A refinement in just one of these parameters, usually has diminishing returns. The error decreases roughly like $O(h)$. In Table 2 we complicate the model further by using diffusion which varies inside $\Omega$. In Table 3 we use a solution which is not a polynomial and hence is not exactly representable in spline space.

*6.2. Several Simulations of Population Development*

From the previous subsection, we have seen that our MATLAB code works well. Thus we removed the forcing terms and ran simulations of the solution of (1.2) for various initial conditions and parameters. We shall use the following two domains shown with a triangulation in Fig. 2

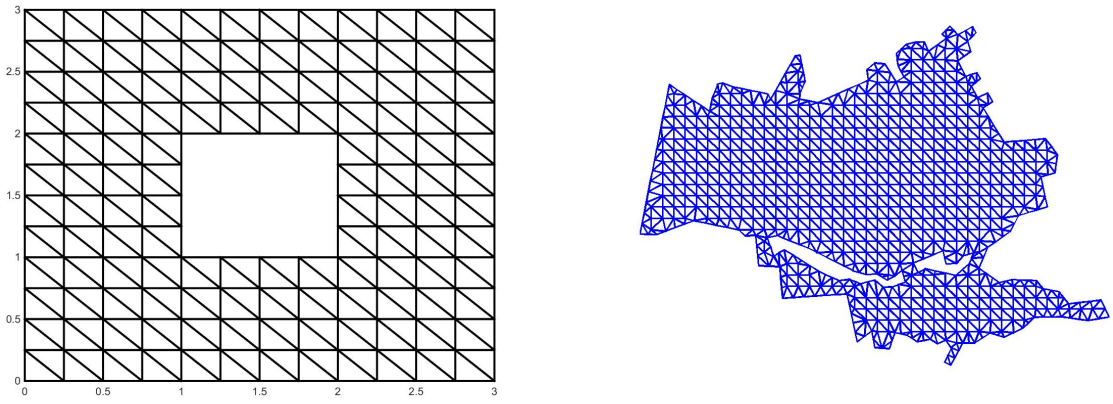



Figure 2: Two domains with triangulation for numerical simulations

We provide several examples to show how various growth functions affect the rate at which the solution reaches the asymptotically stable constant solution of $p(x, y) = 1$ or $p(x, y) = 0$.

Figures 3 through 6 show several 3D renders of how solutions grow over time over two domains indicated in Fig 2. Each subfigure shows four equally-spaced time slices, plotted on the same $xy$-

axes, one on top of each other, allowing the reader to observe how the solution grows over time. Initial time slice appears as the bottommost surface and the final state is the topmost surface. In addition, each figure shows the effect of varying the Allee threshold $\sigma$. With low $\sigma$ we see a very quick spread since any amount of infection will expand to infect all individuals. Higher $\sigma$ corresponds to a need for a critical mass before infection can permanently establish itself in a region. A high value for $\sigma$ causes the average population to grow more slowly as seen in Figure 7. It can introduce sharp rises in population density between regions where $p(x) < \sigma$ and regions where $p(x) > \sigma$ as seen in Figure 8e. In order to make the difference in the behavior of the solution clearer, the value of $t$ for each time slice is indicated in the caption of each figure.

We can observe some expected behavior from the solutions presented in Figure 3. The initial condition is uniformly $p = 0.1$ on a large portion of $\Omega$ with an isolated bump function in one corner. In Figure 3b the second time slice shows the population has become extinct on the area where $p = 0.1$. At the same time the bump grows to population capacity and eventually spreads life into formerly dead areas. We observe similar results in Figure 3c, but the rate at which the population grows has been severely diminished. In Figure 3d, the threshold $\sigma$ is so high that the population becomes extinct everywhere and very quickly.

*6.3. Simulations of Malaria Study*

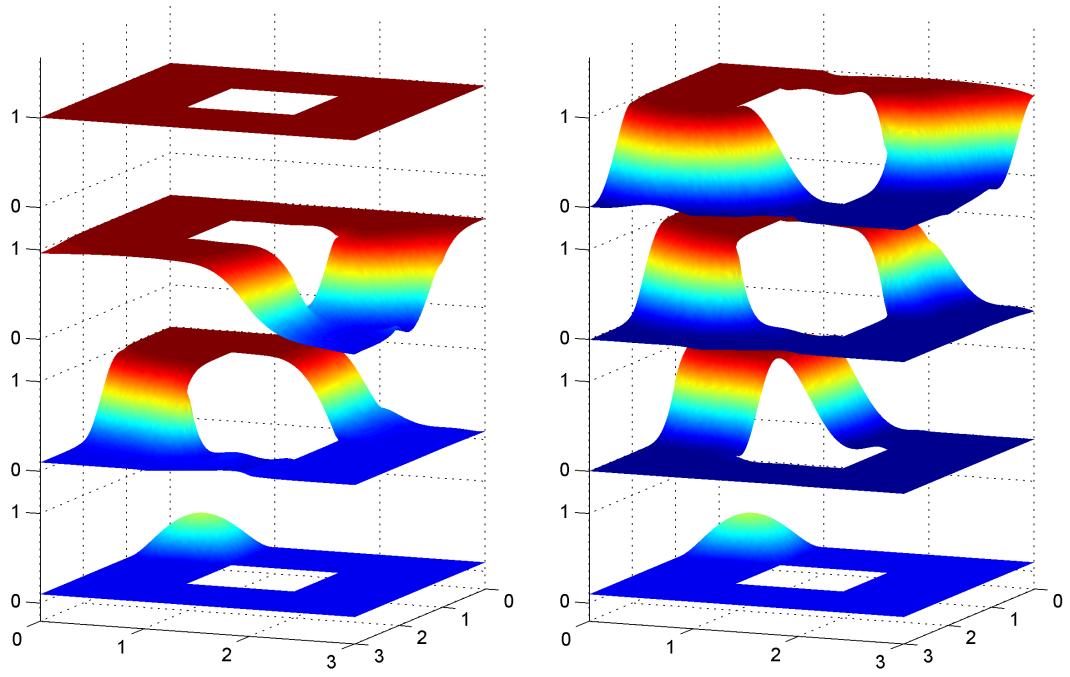
It is well-known that malaria is one of the leading causes of mortality in the world and an estimated 3.3 billion people are at risk of malaria (cf. [33]). WHO is interested in spatial models which can identify high-risk zones of infection on a fine geographical scale as indicated in the 18th and 20th WHO reports (cf. [34] and [35]). An example of such a study can be found in [6] and in [5] where Coulibaly et al. provide data samples of individuals infected with malaria in Bandiagara, Mali [5]. We used this data to form an initial value for our PDE model using a bivariate spline data fitting technique (cf. [2]). Then we use our MATLAB program to simulate the development of malarial infection over a period of time using various Allee parameters. Our results are presented in Figure 8 of this paper. $\sigma$ plays a vital role in the growth rate of the infected region. When $\sigma$ is small, the initial population of infected individuals is sufficiently large to cover a majority of the region by the end of the simulation. When $\sigma$ is a bit larger we see that some regions become free from infection for a while since the local population density is less than $\sigma$. In Figure 9 we see that with a high enough $\sigma$ it is possible for average infected population to decrease at first yet ultimately return to growth.

An appropriate calibration of this constant based on real data would be an important achievement, as high-risk zones can be identified using our model by examining a time-horizon of one year and analyzing regions where infection has taken hold.

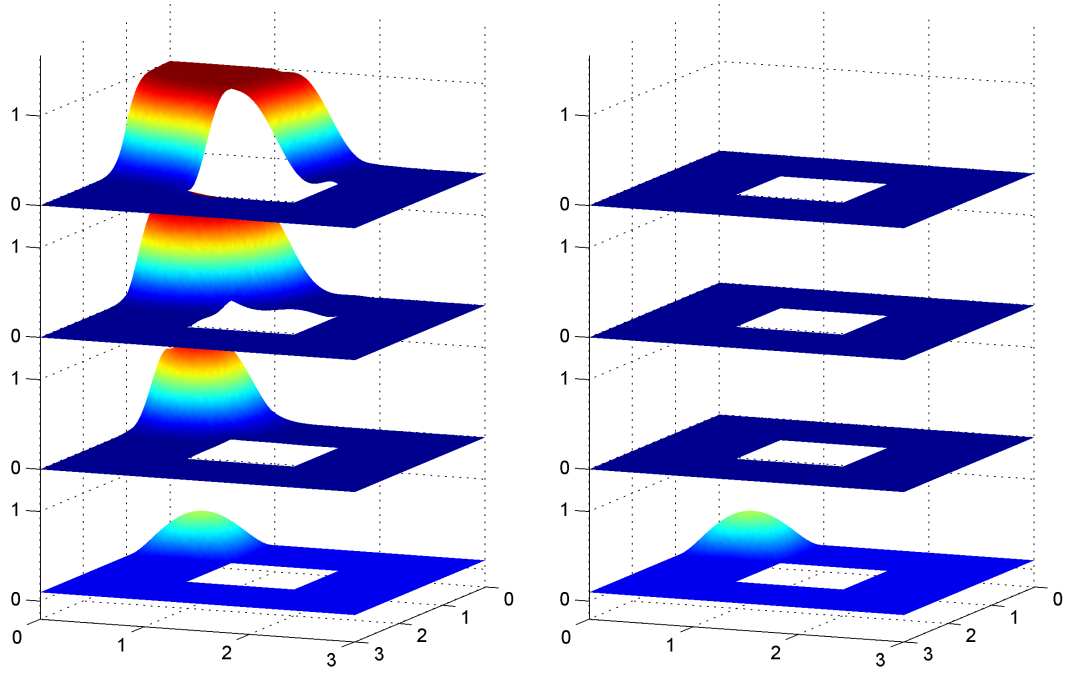
Figures 7 through show average population over time over the city of Bandiagara, Mali. Each subfigure corresponds to a certain set of initial conditions for the PDE, while separating the cases by the choice for $\sigma$, emphasizing the effect $\sigma$ has on the rate at which the population reaches an asymptotically stable solution.

An additional benefit to our method is that the use of splines allows us to produce smooth population density surfaces. Fisher's [8] traveling waves travel in the direction of steepest-descent on the surface and thus can visualized quite well by a contour graph of $\nabla p$, which is helpful in identifying the direction of the spread of infection. See an example of this visualization in Figure 10.

Finally, let us use the code developed in this paper to simulate the real life situation studied in [5]. Although we were not able to obtain the data set from the authors in [5], we picked some values from Figure 11 which is a photo copy of Figure 2 in [5] and use the scattered data interpolation

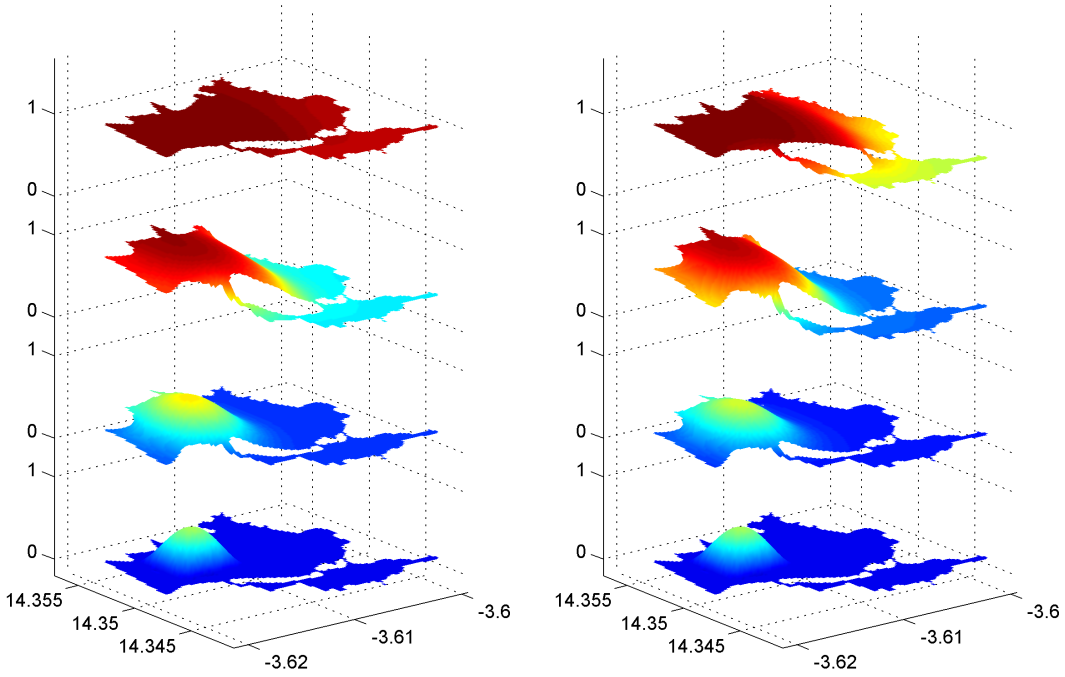


(a) $\sigma = 0.1$ (b) $\sigma = 0.2$



(c) $\sigma = 0.3$ (d) $\sigma = 0.5$

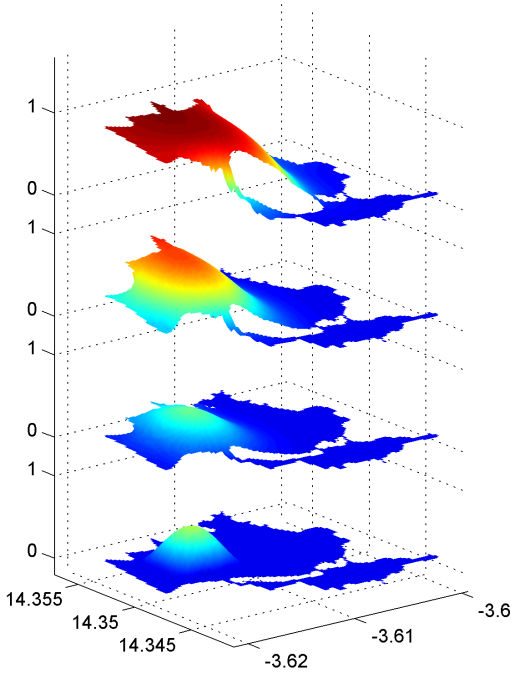
Figure 3: Donut-shape domain. Constant growth and diffusion. Various Allee effect thresholds $\sigma$. The vertical axis shows population density $p \in [0, 1]$ at 4 points in time: $t \in \{0, 30, 60, 90\}$, where the bottom manifold represents $t = 0$, and the top manifold represents $t = 90$ in each case.



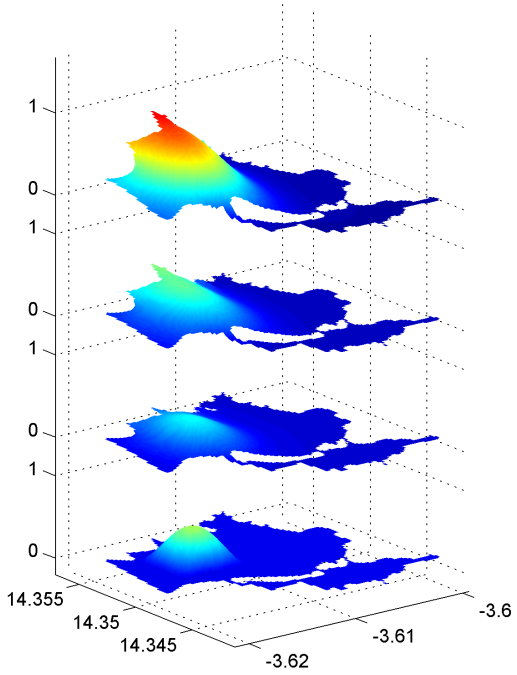

(a) $\sigma = 0.02$




(b) $\sigma = 0.05$

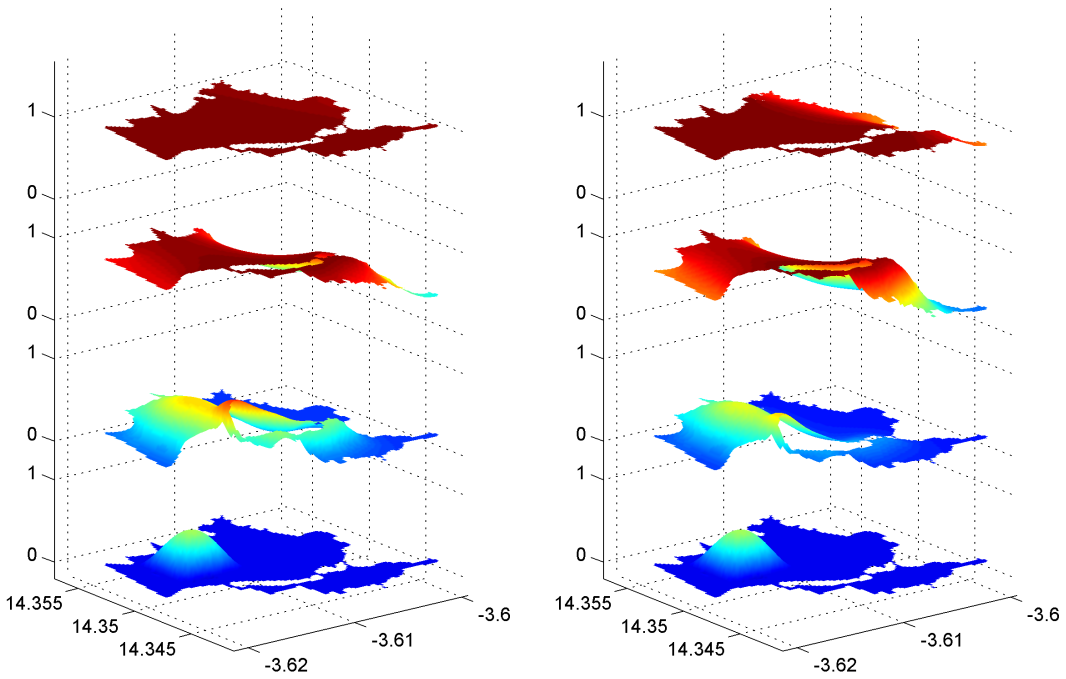



(c) $\sigma = 0.1$




(d) $\sigma = 0.15$

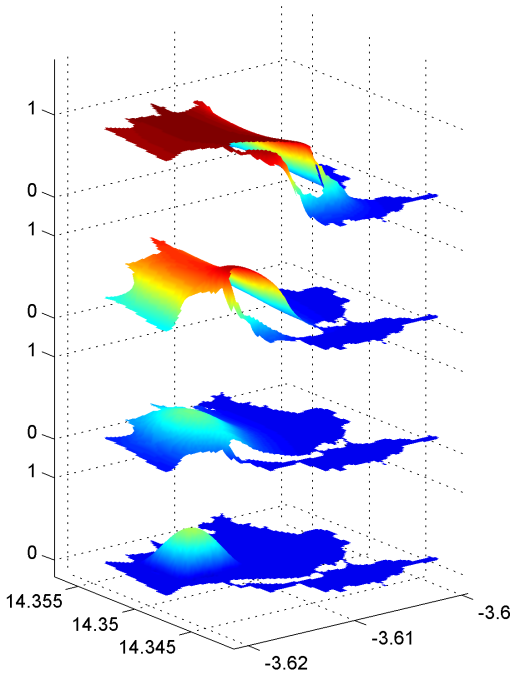
Figure 4: City of Bandiagara, Mali. Constant growth and diffusion. Various Allee effect thresholds $\sigma$. The vertical axis shows population density $p \in [0, 1]$ at 4 points in time: $t \in \{0, 5, 15, 20\}$, where the bottom manifold represents $t = 0$, and the top manifold represents $t = 20$ in each case.



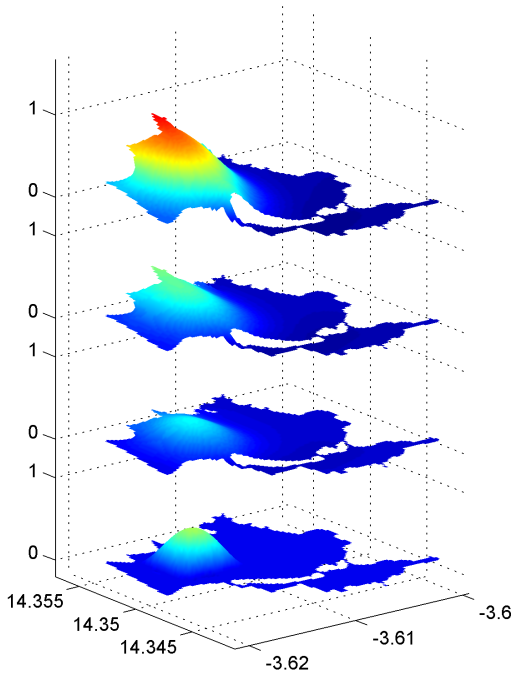

(a) $\sigma = 0.02$




(b) $\sigma = 0.05$

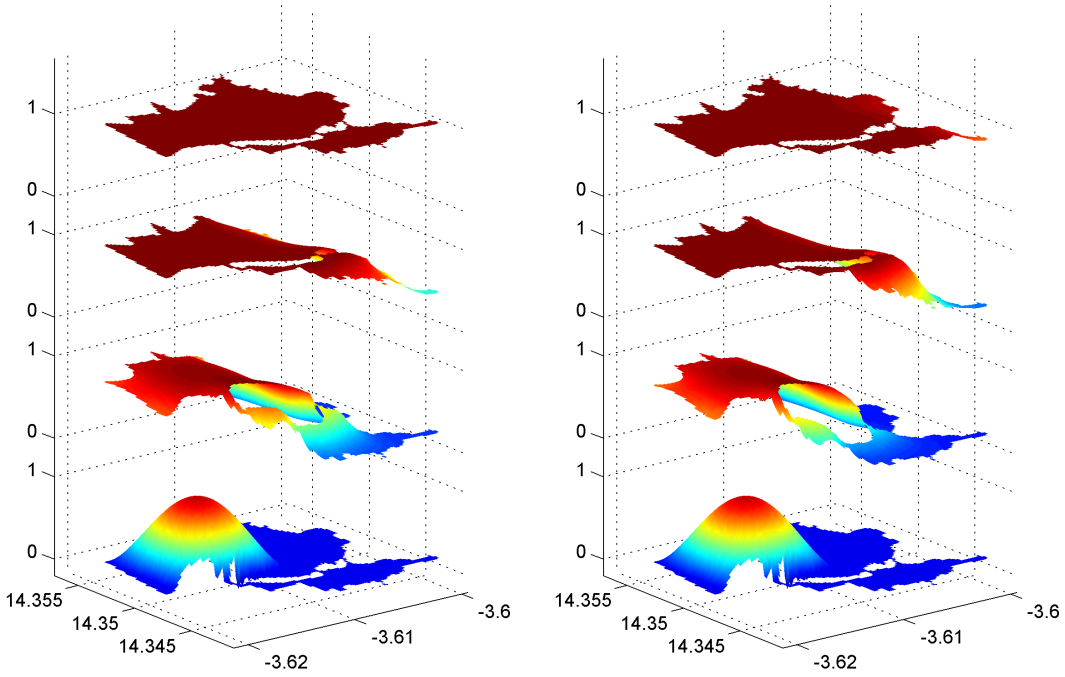



(c) $\sigma = 0.1$




(d) $\sigma = 0.15$

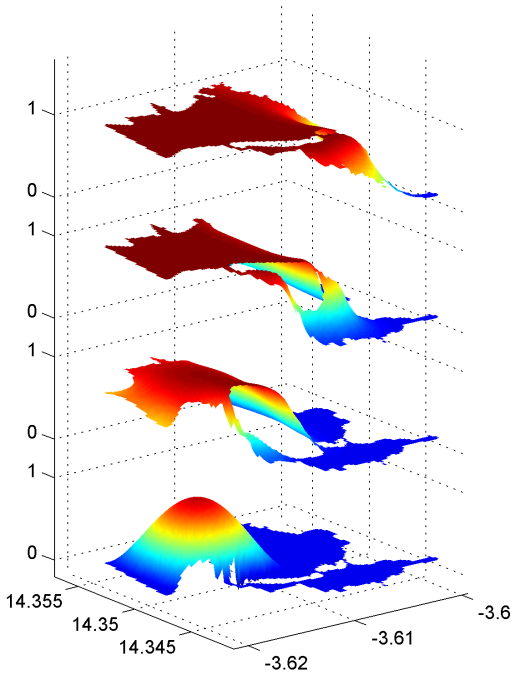
Figure 5: City of Bandiagara, Mali. Constant diffusion. Various Allee effect thresholds. Growth function is piecewise-constant with triple magnitude for patches near the city's river. The vertical axis shows population density $p \in [0, 1]$ at 4 points in time: $t \in \{0, 5, 15, 20\}$, where the bottom manifold represents $t = 0$, and the top manifold represents $t = 20$ in each case.



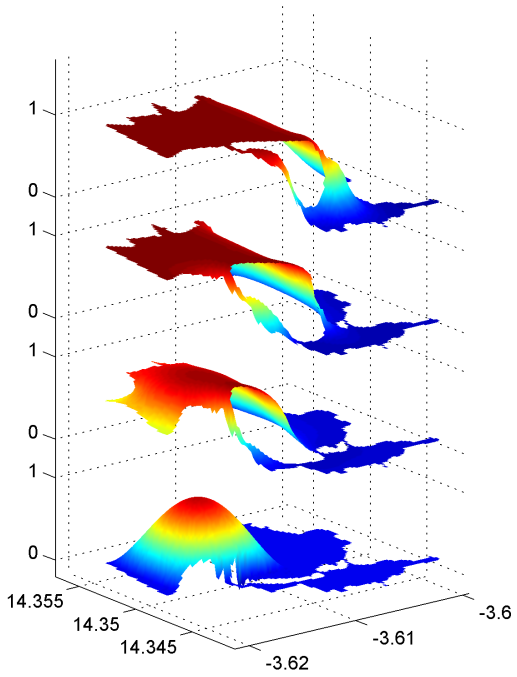

(a) $\sigma = 0.02$




(b) $\sigma = 0.05$

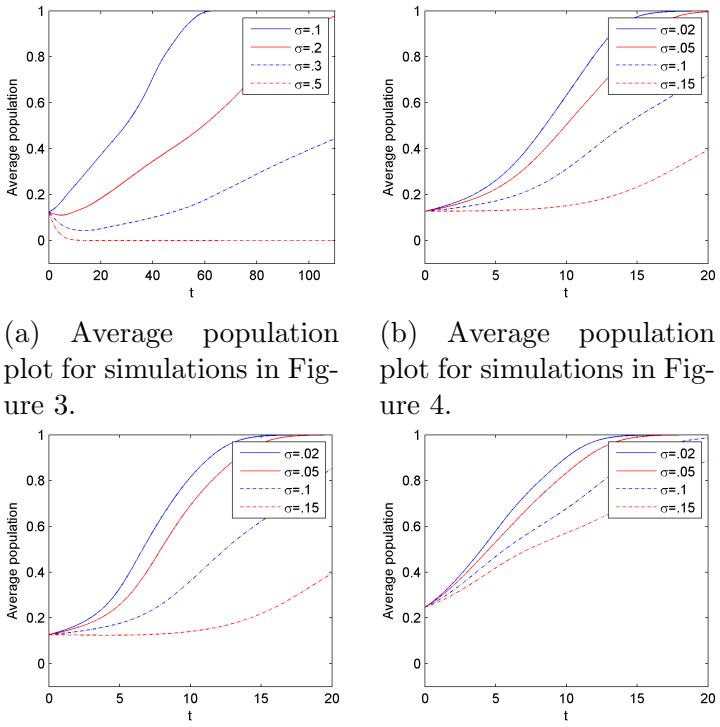



(c) $\sigma = 0.1$




(d) $\sigma = 0.15$

Figure 6: City of Bandiagara, Mali. Same as Figure 5 but the initial condition has a much higher total population. The vertical axis shows population density $p \in [0, 1]$ at 4 points in time: $t \in \{0, 5, 15, 20\}$, where the bottom manifold represents $t = 0$, and the top manifold represents $t = 20$ in each case.




(a) Average population plot for simulations in Figure 3.

(b) Average population plot for simulations in Figure 4.

(c) Average population plot for simulations in Figure 5.

(d) Average population plot for simulations in Figure 6.

Figure 7: Average population density in $\Omega$ plotted over time for each of the four preceding figures.

with nonnegative preservation method discussed in [18] to find an initial value which is shown on the top row of Figure 8. Then we chose $\sigma = 0.15$ and obtained the evolution of the sick children from the $\nabla p$ showed in Figure 10.

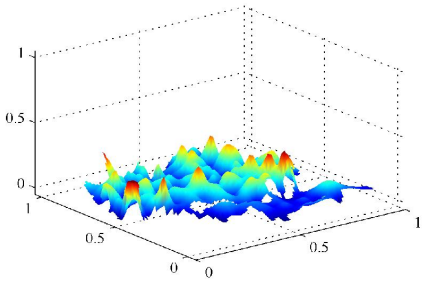
It is interesting to see that the direction of the spread of malaria which matches with the pattern in Fig. 5 A. B. C in [5]. For convenience, we include a snapshot of these graphs in [5] for the comparison. The area C2 in red indicated in graph C shows the spread of the disease which is the same as the heavy arrows shown in Figure 10.

## 7. Discussion

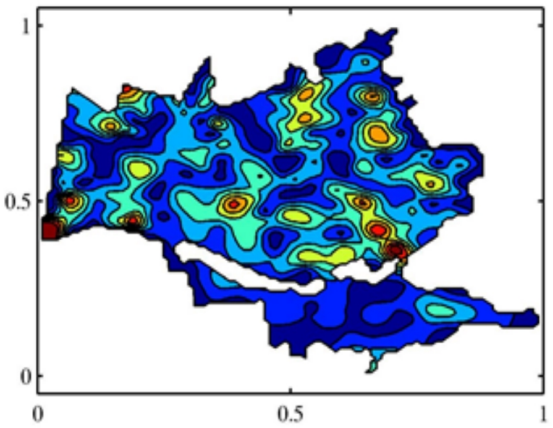
The results presented in this paper show that a problem of diffusive population growth with Allee effect has a unique bivariate spline solution, based on a discrete formulation using smooth functions defined on $\Omega$ with the first order divided difference in time. Domains of arbitrary shape can be easily handled with this approach; while the method of finite-elements has been traditionally used to study arbitrary domains, here we propose an advanced alternative which can now be easily applied due to the computer power.

The study of diffusion over arbitrary domains is the necessary link between theoretical and applied population dynamics. As geographic data sets become increasingly available, it has become easier to automate the segmentation of spatial domains into subdomains with specific properties. This type of approach makes it possible to conduct numerical studies on the global behavior of complex ecological systems, something extremely hard to accomplish with purely theoretical approaches.

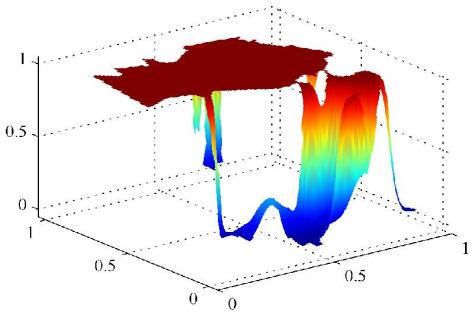
A direction of future work is related to the interaction between within-host dynamics and between-host dynamics. Due to a host's immune response, exposure to low levels of a pathogen may



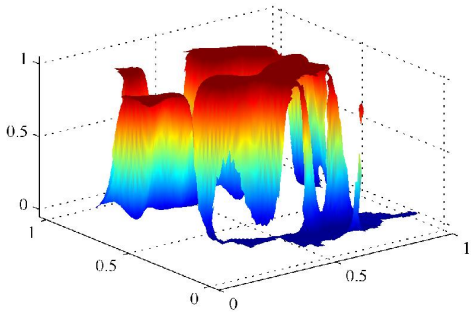
(a) Initial population density



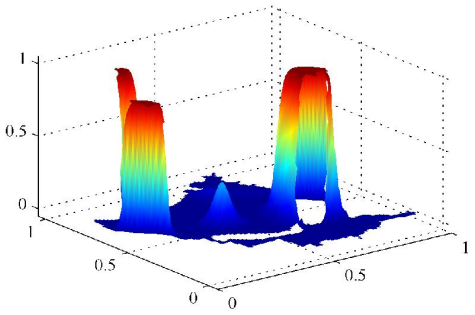
(b) Contour Plot of Initial population density



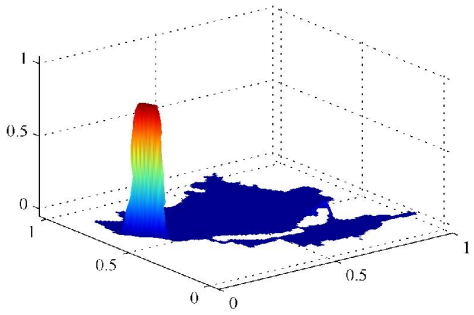
(c) $\sigma = 0.05$



(d) $\sigma = 0.1$



(e) $\sigma = 0.15$



(f) $\sigma = 0.2$

Figure 8: We used a spline data fitting technique on artificial data of infected population which mimics the one presented in [5] and then applied our model to examine future development. Figures 8c through 8f correspond to the same time $t = 27$.

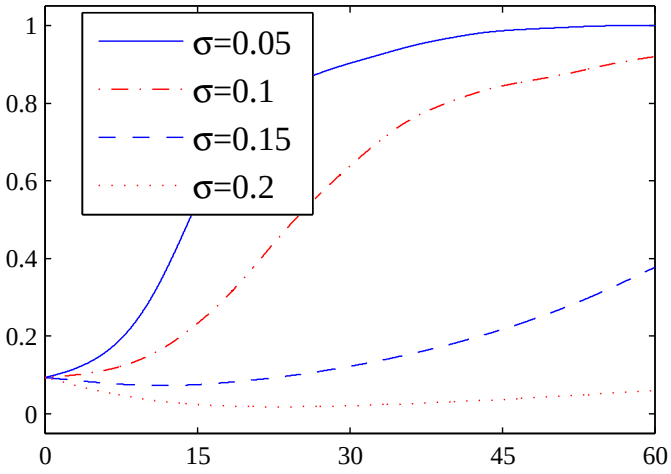



Figure 9: Average population over time corresponding to each choice of $\sigma$ in the adjacent subfigures.

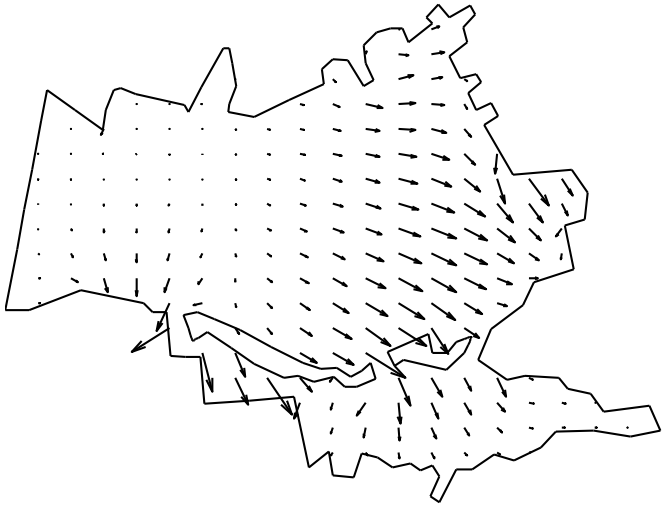


Figure 10: A contour plot of $\nabla p$ which corresponds to Figure 4a at T=15, indicating the direction in which infection spreads.

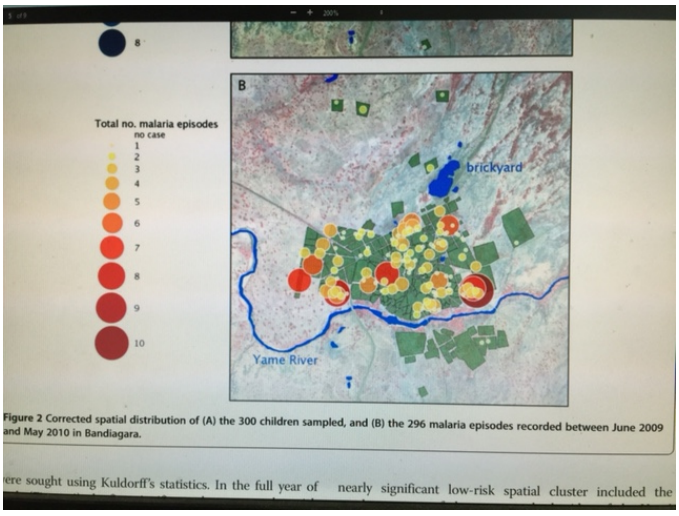



Figure 11: Photo copy of Figure 2 from [5] show the number of sick children over the area.

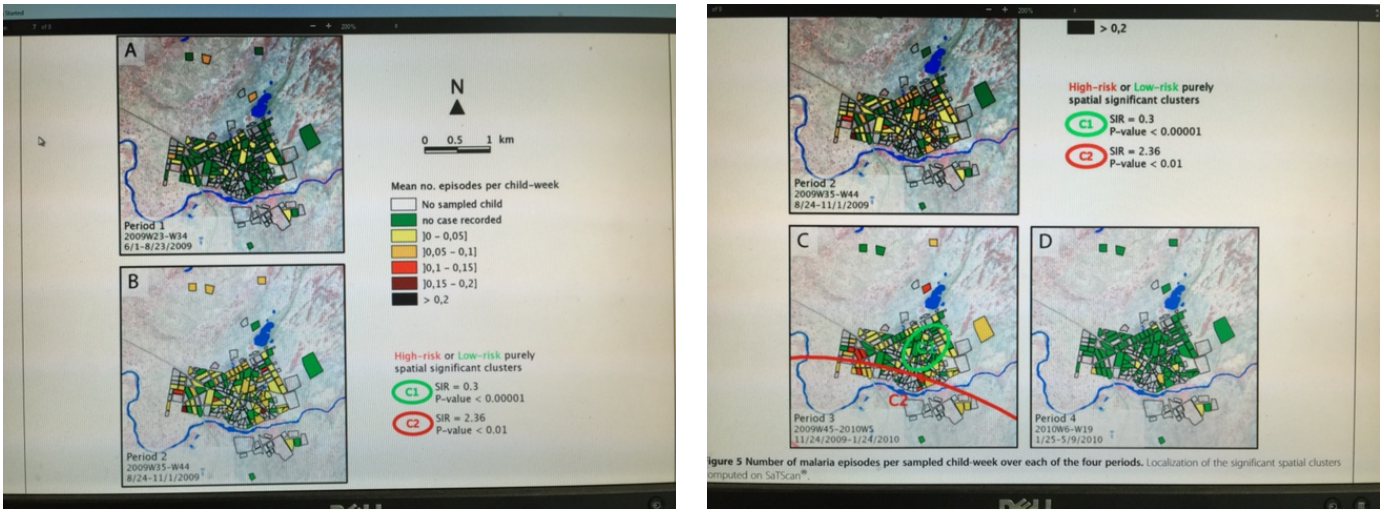



Figure 12: Photo copy of Figure 5 from [5] show the area in which infection spreads in Figs. A–C.

not result in infection; furthermore, pathogens may be unable to find a within-host substrate before degradation. Therefore, the onset of infection requires the direct or indirect (vector-driven) uptake of a minimum dose of infectious agents (virus, bacteria, protozoa, or helminths). Once the host has been colonized above the minimum infectious dose, the pathogen's virulence and the host-pathogen interaction determine pathogen environmental shedding, which in turn determines the dynamics of epidemics. Therefore, we can now take into account within-host dynamics to determine environmental pathogen population density, and the dynamics of interaction between pathogen, vectors and hosts. Such formulation results in very complicated segmentations of the spatial domain for each species considered, and it is in this context that bivariate spline solutions can produce informative numerical results.

Diffusive models in ecology are usually limited to the study of population dispersal occurring when the underlying process is a random walk with a step-length size following a normal distribution. But not all dispersal of organisms follow this pattern. However, it is possible to create patterns of dispersal that are orders of magnitude different on the same domain by carefully controlling the function $D(p, \mathbf{x})$ in 1.2. This direction is other natural extension of the present work.

We have shown specific examples of regular and irregular domains for which bivariate spline calculations can easily be obtained to study populations with Allee effect dynamics. Note that the properties of the discrete weak solution were determined with the generic form

$$p_t = \text{ div } (D(p, \mathbf{x})\nabla p(\mathbf{x}, t))) + F(p(\mathbf{x}, t)),$$

that is to say, with a generic $F(p) = pF_1(p)$ for some continuous function $F_1$. Spline solutions can be used to numerically solve such diffusive PDE in general form. Nevertheless, we focused on the dynamics of the Allee effect because of its intrinsic relationship with the ecological problems we are interested in. Hence the results presented here are of ample applicability and can be extended beyond the problem of Allee dynamics, and outside the realm of ecology.

## Acknowledgments

This research has been partially supported by the International Centers for Excellence in Malaria Research - Center for non-Amazonian regions of Latin America (CLAIM), NIH's NIAID cooperative

agreement U19AI089702. We would like to thank two anonymous reviewers who helped improve this paper substantially.

The second author Ming-Jun Lai would like to thank the Simons Foundation for the support through a Simon Collaboration grant #280646 from 2013–2018.

## References

[1] Warder Clyde Allee. *The social life of animals*. Beacon Press, Boston, MA, 1958.

[2] Gerard Awanou, Ming-Jun Lai, and Paul Wenston. The multivariate spline method for scattered data fitting and numerical solutions of partial differential equations. *Wavelets and splines: Athens*, pages 24–74, 2005.

[3] Robert Stephen Cantrell and Chris Cosner. *Spatial ecology via reaction-diffusion equations*. John Wiley & Sons, 2004.

[4] C Cosner, JC Beier, RS Cantrell, D Impoinvil, L Kapitanski, MD Potts, A Troyo, and S Ruan. The effects of human movement on the persistence of vector-borne diseases. *Journal of theoretical biology*, 258(4):550–560, 2009.

[5] Drissa Coulibaly, Stanislas Rebaudet, Mark Travassos, Youssouf Tolo, Matthew Laurens, Abdoulaye Kone, Karim Traore, Ando Guindo, Issa Diarra, Amadou Niangaly, Modibo Daou, Ahmadou Dembele, Mody Sissoko, Bourema Kouriba, Nadine Dessay, Jean Gaudart, Renaud Piarroux, Mahamadou Thera, Christopher Plowe, and Ogobara Doumbo. Spatio-temporal analysis of malaria within a transmission season in Bandiagara, Mali. *Malaria Journal*, 12(1):82, 2013.

[6] Alassane Dicko, Carsten Mantel, Boureima Kouriba, Issaka Sagara, Mahamadou A. Thera, Seydou Doumbia, Mouctar Diallo, Belco Poudiougou, Mahamadou Diakite, and Ogobara K. Doumbo. Season, fever prevalence and pyrogenic threshold for malaria disease definition in an endemic area of mali. *Tropical Medicine & International Health*, 10(6):550–556, 2005.

[7] Gerald Farin. Triangular bernstein-bézier patches. *Computer Aided Geometric Design*, 3(2):83–127, 1986.

[8] Ronald Aylmer Fisher. The wave of advance of advantageous genes. *Annals of Eugenics*, 7(4):355–369, 1937.

[9] GF Gause. Experimental demonstration of volterra's periodic oscillations in the numbers of animals. *Journal of Experimental Biology*, 12(1):44–48, 1935.

[10] Xian-Liang Hu, Dan-Fu Han, and Ming-Jun Lai. Bivariate splines of various degrees for numerical solution of partial differential equations. *SIAM Journal on Scientific Computing*, 29(3):1338–1354, 2007.

[11] CB Huffaker. Biological control of weeds with insects. *Annual review of entomology*, 4(1):251–276, 1959.

[12] Derek M Johnson, Andrew M Liebhold, Patrick C Tobin, and Ottar N Bjørnstad. Allee effects and pulsed invasion by the gypsy moth. *Nature*, 444(7117):361–363, 2006.

[13] Gerry F Killeen, Aklilu Seyoum, Chadwick Sikaala, Amri S Zomboko, John E Gimnig, Nicodem J Govella, and Michael T White. Eliminating malaria vectors. *Parasit Vectors*, 6:172, 2013.

[14] I Petrovsky A Kolomogorov and N Piscouno. Etude de l'equation de la diffusion avec croissance de la quantite de la matiere et son application a un problem biologique. *Moscow University Bull. Math*, 1:1–25, 1937.

[15] Mark Kot, Mark A Lewis, and Pauline van den Driessche. Dispersal data and the spread of invading organisms. *Ecology*, 77(7):2027–2042, 1996.

[16] Ming-Jun Lai, Chun Liu, and Paul Wenston. Numerical simulations on two nonlinear biharmonic evolution equations. *Applicable Analysis*, 83(6):563–577, 2004.

[17] Ming-Jun Lai, Chun Liu, and Paul Wenston. On two nonlinear biharmonic evolution equations: existence, uniqueness and stability. *Applicable Analysis*, 83(6):541–562, 2004.

[18] Ming-Jun Lai and Christof Meile , Scattered data interpolation with nonnegative preservation using bivariate splines, Computer Aided Geometric Design, vol. 34 (2015) pp. 37–49.

[19] Ming-Jun Lai and Larry L Schumaker. On the approximation power of bivariate splines. *Advances in Computational Mathematics*, 9(3-4):251–279, 1998.

[20] Ming-Jun Lai and Larry L Schumaker. *Spline functions on triangulations*. Number 110. Cambridge University Press, 2007.

[21] Ming-Jun Lai and Paul Wenston. Bivariate splines for fluid flows. *Computers & fluids*, 33(8):1047–1073, 2004.

[22] MA Lewis and P Kareiva. Allee dynamics and the spread of invading organisms. *Theoretical Population Biology*, 43(2):141–158, 1993.

[23] Wen-Xiu Ma and B Fuchssteiner. Explicit and exact solutions to a kolmogorov-petrovskii-piskunov equation. *International Journal of Non-Linear Mechanics*, 31(3):329–338, 1996.

[24] Thomas Robert Malthus. *An Essay on the Principle of Population Or a View of Its Past and Present Effects on Human Happiness, an Inquiry Into Our Prospects Respecting the Future Removal Or Mitigation of the Evils which it Occasions by Rev. TR Malthus*. Reeves and Turner, 1872.

[25] James D Murray. Mathematical biology i: An introduction, vol. 17 of interdisciplinary applied mathematics, 2002.

[26] Otto Richter, Sylvia Moenickes, and Frank Suhling. Modelling the effect of temperature on the range expansion of species by reaction–diffusion equations. *Mathematical biosciences*, 235(2):171–181, 2012.

[27] Lionel Roques, Alain Roques, Henri Berestycki, and André Kretzschmar. A population facing climate change: joint influences of allee effects and environmental boundary geometry. *Population Ecology*, 50(2):215–225, 2008.

[28] John Gordon Skellam. Random dispersal in theoretical populations. *Biometrika*, pages 196–218, 1951.

[29] Steven T Stoddard, Amy C Morrison, Gonzalo M Vazquez-Prokopec, Valerie Paz Soldan, Tadeusz J Kochel, Uriel Kitron, John P Elder, and Thomas W Scott. The role of human movement in the transmission of vector-borne pathogens. *PLoS Negl Trop Dis*, 3(7):e481, 2009.

[30] MF Suarez, ML Quiñones, JD Palacios, A Carrillo, et al. First record of ddt resistance in anopheles darlingi. *J Am Mosq Control Assoc*, 6(1):72–74, 1990.

[31] Pierre-François Verhulst. Notice sur la loi que la population poursuit dans son accroissement correspondance mathématique et physique 10: 113-121. Technical report, Retrieved 09/08, 2009.

[32] Feng Wang, Haiyan Wang, and Kuai Xu. Diffusive logistic model towards predicting information diffusion in online social networks. In *Distributed Computing Systems Workshops (ICDCSW), 2012 32nd International Conference on*, pages 133–139. IEEE, 2012.

[33] World Health Organization Expert Committee on Malaria: World malaria report. Geneva: WHO; 2011.

[34] World Health Organization Expert Committee on Malaria: 18th Report. World Health Organ Tech Rep; 1986:735.

[35] World Health Organization Expert Committee on Malaria: 20th Report. World Health Organ Tech Rep; 2000:892.

## 8. Appendix A: Preliminary on Bivariate Splines

In this section, we explain bivariate spline functions of any degree $d$ and smoothness $r \geq 1$ over arbitrary triangulation $\triangle$. Most of the following discussion can be found in [20]. We outline these functions here just for convenience. Let $\Omega$ be a polygonal domain in $\mathbb{R}^2$ and $\triangle$ a triangulation of $\Omega$. That is, $\triangle$ is a finite collection of triangles $T \subset \Omega$ such that $\cup_{T\in\triangle} T = \Omega$ and the intersection of any two triangles is either the empty set, a common edge, or a common vertex. For each $T \in \triangle$, let $|T|$ denote the length of the longest edge of $T$, and let $\rho_T$ be the radius of the inscribed circle of $T$. The longest edge length in the triangulation $\triangle$ is denoted by $|\triangle|$ and is referred to as the size of the triangulation. For any triangulation $\triangle$ we define its shape parameter by

$$\kappa_\triangle := \frac{|\triangle|}{\rho_\triangle}, \tag{8.1}$$

where $\rho_\triangle$ is the minimum of the radii of the in-circles of the triangles of $\triangle$. The shape parameter for a single triangle, $\kappa_T$, satisfies

$$\kappa_T := \frac{|T|}{\rho_T} \leq \frac{2}{\tan(\theta_T/2)} \leq \frac{2}{\sin(\theta_T/2)}, \tag{8.2}$$

where $\theta_T$ is the smallest angle in the triangle $T$. The shape of a given triangulation affects how well we can approximate a function over the triangulation. Hence we have the following definition of a $\beta$-quasi-uniform triangulation.

**Definition 8.1** ($\beta$-Quasi-Uniform Triangulation)**.** *Let $0 < \beta < \infty$. A triangulation $\triangle$ is a $\beta$-quasi-uniform triangulation provided that*

$$\frac{|\triangle|}{\rho_\triangle} \leq \beta.$$

Once we have a triangulation, we define the spline space of degree $d$ and smoothness $r$ over that triangulation as follows:

**Definition 8.2** (Spline Space)**.** *Let $\triangle$ be a given triangulation of a domain $\Omega$. Then we define the spline space of smoothness $r$ and degree $d$ over $\triangle$ by,*

$$S_d^r(\triangle) = \{s \in C^r(\Omega) \mid s|_T \in \mathcal{P}_d, \ \forall \ T \in \triangle\},$$

*where $\mathcal{P}_d$ is the space of polynomials of degree at most $d$.*

We next explain how to represent a spline function in $S_d^r(\triangle)$. Let $T = \langle (x_1, y_1), (x_2, y_2), (x_3, y_3) \rangle$. For any point $(x, y)$, let $b_1, b_2, b_3$ be the solution of

$$\begin{array}{rcl} x & = & b_1 x_1 + b_2 x_2 + b_3 x_3 \\ y & = & b_1 y_1 + b_2 y_2 + b_3 y_3 \\ 1 & = & b_1 + b_2 + b_3. \end{array}$$

$(b_1, b_2, b_3)$ is the so-called barycentric coordinates of $(x, y)$ with respect to $T$. Note that $b_i$ is a linear polynomial of $(x, y)$ for $i = 1, 2, 3$. Fix a degree $d > 0$. For $i + j + k = d$, let

$$B_{ijk}^T(x, y) = \frac{d!}{i!j!k!} b_1^i b_2^j b_3^k$$

which is called Bernstein-Bézier polynomial. Let

$$S|_T = \sum_{i+j+k=d} c_{ijk}^T B_{ijk}^T(x, y).$$

We use $\mathbf{s} = (c_{ijk}^T, i + j + k = d, T \in \triangle)$ to represent the coefficient vector for spline function $S \in S_d^{-1}(\triangle)$. In order to make $S \in S_d^0(\triangle)$, we have to construct a smoothness matrix $H$ such that $H\mathbf{s} = 0$ ensure that $S$ is a continuous function. Such a smoothness matrix is known and in fact it is known for any smoothness $r \geq 0$ (cf. [7]).

Note that Bernstein-Bézier representation of spline functions is very convenient for basic evaluation, derivatives and integration. We use the de Casteljau algorithm to evaluate a Bernstein-Bézier polynomial at any point inside the triangle. It is a simple and stable computation. See [20]. Let $T = \langle \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \rangle$ and $S|_T = \sum_{i+j+k=d} c_{ijk} B_{ijk}(x, y)$. Then the directional derivative $D_{\mathbf{v}_2 - \mathbf{v}_1} S|_T$ is

$$D_{\mathbf{v}_2 - \mathbf{v}_1} S|_T = d \sum_{i+j+k=d-1} (c_{i,j+1,k} - c_{i+1,j,k}) B_{ijk}(x, y).$$

Similar for $D_{\mathbf{v}_3 - \mathbf{v}_1} S|_T$. $D_x$ and $D_y$ are linearly combinations of these two directional derivatives. Let $s$ be a spline with $s|_T = \sum_{i+j+k=d} c_{ijk}^T B_i jk(x, y), T \in \triangle$ in $S_d^r(\triangle)$. Then

$$\int_\Omega s(x, y) dx dy = \sum_{T \in \triangle} \frac{A_T}{\binom{d+2}{2}} \sum_{i+j+k=d} c_{ijk}^T.$$

If $p = \sum_{i+j+k=d} a_{ijk} B_{ijk}(x,y)$ and $q = \sum_{i+j+k=d} b_{ijk} B_{ijk}(x,y)$ over a triangle $T$, then

$$\int_T p(x,y)q(x,y)dxdy = \mathbf{a}^\top M_d \mathbf{b},$$

where $\mathbf{a} = (a_{ijk}, i+j+k=d)^\top$, $\mathbf{b} = (b_{ijk}, i+j+k=d)^\top$, $M_d$ is a symmetric matrix with known entries (a formula for these entries is known (cf. [20]). These elementary operations have been implemented in MATLAB. See [2]. Many different linear and nonlinear partial differential equations have been solved by using these bivariate spline functions. See [21], [2], [10].

When $d \geq 3r+2$ the spline space $S_d^r(\triangle)$ possesses an optimal approximation order which is achieved by the use of a quasi-interpolation operator. Let $\|f\|_{L_p(\Omega)}$ denote the usual $L_p$ norm of $f$ over $\Omega$, $|f|_{m,p,\Omega}$ denotes the $L_p$ norm of the $m^{th}$ derivatives of $f$ over $\Omega$, and $W_p^{m+1}(\Omega)$ stands for the usual Sobolev space over $\Omega$.

To define the quasi-interpolation operator, we need linear functionals $\{\lambda_{ijk,T}\}_{i+j+k=d}, T \in \triangle$ which are based on values of $f$ at the set of domain points over triangles in $\triangle$, that is

$$\lambda_{ijk,T}(f) = \sum_{|\nu|=d} a_\nu^{ijk} f(\xi_\nu^T), \tag{8.3}$$

where $\xi_\nu^T = (i\mathbf{v}_1^T + j\mathbf{v}_2^T + k\mathbf{v}_3^T)/d$ for $\nu = (i,j,k)$ with $i+j+k=d$ and $\mathbf{v}_i, i = 1,2,3$ are vertexes of triangle $T$.

A quasi-interpolation operator of $f$ is defined by

$$Qf := \sum_{T\in\triangle} \sum_{i+j+k=d} \lambda_{ijk,T}(f) B_{ijk}^T. \tag{8.4}$$

Now, we are ready to state a theorem on optimal approximation order (cf. [19] and [20]).

**Theorem 8.1** (Optimal Approximation Order)**.** *Assume $d \geq 3r+2$ and let $\triangle$ be a triangulation of $\Omega$. Then there exists a quasi-interpolatory operator $Qf \in S_d^r(\triangle)$ mapping $f \in L_1(\Omega)$ into $S_d^r(\triangle)$ such that $Qf$ achieves the optimal approximation order: if $f \in W_p^{m+1}(\Omega)$,*

$$\|D_x^\alpha D_y^\beta (Qf - f)\|_{L_p(\Omega)} \leq C|\triangle|^{m+1-\alpha-\beta} |f|_{m+1,p,\Omega} \tag{8.5}$$

*for all $\alpha+\beta \leq m+1$ with $0 \leq m \leq d$, where $D_x$ and $D_y$ denote the derivatives with respect to the first and second variables and the constant $C$ depends only on the degree $d$ and the smallest angle $\theta_\triangle$ and may be dependent on the Lipschitz condition on the boundary of $\Omega$.*

We sometimes need to use the so-called Markov inequality to compare the size of the derivative of a polynomial with the size of the polynomial itself on a given triangle $t$. As a spline function is a piecewise polynomial function, this inequality can be also applied to any spline function. See [20] for a proof.

**Theorem 8.2.** *Let $t := \langle v_1, v_2, v_3 \rangle$ be a triangle, and fix $1 \leq q \leq \infty$. Then there exists a constant $K$ depending only on $d$ such that for every polynomial $p \in \mathcal{P}_d$, and any nonnegative integers $\alpha$ and $\beta$ with $0 \leq \alpha+\beta \leq d$,*

$$\|D_1^\alpha D_2^\beta p\|_{q,t} \leq \frac{K}{\rho_t^{\alpha+\beta}} \|p\|_{q,t}, \quad 0 \leq \alpha+\beta \leq d, \tag{8.6}$$

*where $\rho_t$ denotes the radius of the largest circle inscribed in $t$.*

More detail on the theory of bivariate splines can be found in [20] and their computational schemes in [2]. As our PDE (1.2) is nonlinear, we have to extend the MATLAB codes used in [2] to handle this nonlinear PDE discussed in this paper.