ESAIM: M2AN 51 (2017) 1223–1243
DOI: 10.1051/m2an/2016058

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

A FINITE VOLUME METHOD FOR NONLOCAL COMPETITION-MUTATION EQUATIONS WITH A GRADIENT FLOW STRUCTURE

Wenli Cai[1] and Hailiang Liu[2]

Abstract. In this paper, we design, analyze and numerically validate energy dissipating finite volume schemes for a competition-mutation equation with a gradient flow structure. The model describes the evolution of a population structured with respect to a continuous trait. Both semi-discrete and fully discrete schemes are demonstrated to satisfy the two desired properties: positivity of numerical solutions and energy dissipation. These ensure that the positive steady state is asymptotically stable. Moreover, the discrete steady state is proven to be the same as the minimizer of a discrete energy function. As a comparison, the positive steady state can also be produced by a nonlinear programming solver. Finally, a series of numerical tests is provided to demonstrate both accuracy and the energy dissipation property of the numerical schemes. The numerical solutions of the model with small mutation are shown to be close to those of the corresponding model with linear competition.

Mathematics Subject Classification. 35B40, 65M08, 92D15.

Received March 8, 2016. Accepted September 1, 2016.

1. Introduction

In this work, we are concerned with the problem governed by

$$\partial_t f(t,x) = \Delta f(t,x) + \frac{1}{2} f(t,x)\left(a(x) - \int_X b(x,y) f^2(t,y)\mathrm{d}y\right),\ t>0,\ x\in X, \tag{1.1a}$$

$$f(0,x) = f_0(x) \geq 0, \quad x \in X, \tag{1.1b}$$

$$\frac{\partial f}{\partial \nu} = 0, \quad x \in \partial X, \tag{1.1c}$$

where X is a subdomain of $\mathbb{R}^d$, and ν is the unit outward normal at a point x on the boundary ∂X. This is an intergro-differential equation that describes the evolution of a population of density $f(t,x)$ structured with respect to a continuous trait x. In this model, the diffusion term plays certain role of mutations in the population dynamics. In the model, coefficient $a(x)$ is the intrinsic growth rate of individuals with trait x, and $b(x,y) > 0$ represents the competitive interaction between individuals. The trait dependent competition as such appears in many population balance models of Lotka-Volterra type, see *e.g.*, [4, 12, 15, 16]. In particular, the nonlinear

Keywords and phrases. Selection-mutation dynamics, evolutionary stable distribution, energy dissipation.

[1] Department of Mathematics, China University of Mining and Technology, Beijing 100083, P.R. China. cwl15@cumtb.edu.cn

[2] Department of Mathematics, Iowa State University, Ames, IA 50011, USA. hliu@iastate.edu

competition effect does appear in the model for fish species population introduced in [33] in the study of the effect of exploitation on these species. Their model when the number of fish species tend to infinity formally leads to a continuous model of the form

$$\partial_t f(t,x) = \frac{1}{2} f(t,x)\left(a(x) - \int_X b(x,y)(f(t,y) - d(x,y))^2 \mathrm{d}y\right).$$

This equation with $d = 0$ when augmented with a mutation term Δf is exactly (1.1a).

For rare mutation, the diffusion term may be dropped from model (1.1a), the resulting equation under the transformation $u = f^2$ reduces to the model with linear competition,

$$\partial_t u(t,x) = u(t,x)\left(a(x) - \int_X b(x,y)u(t,y)\mathrm{d}y\right).$$

Such a simplified model with the usual mutation has been derived from random stochastic models of finite populations (see [7,8]). This competition model or its variation arises not only in evolution theory but also in ecology for non-local resources (and x denotes the location there, see *e.g.* [5,13,18]).

The model without mutation is interesting from the point of view of asymptotic behavior; one expects that the population density concentrates at large times, see, *e.g.*, [2,6,12,20,32]. The singular steady-state solutions of the competition model correspond to highly concentrated population densities of the form of well separated Dirac masses, which have been shown to happen only asymptotically in models with mutation [3,9,26–28,30,31].

The main attractive feature of model (1.1a) is its gradient flow structure in the sense that (1.1a) can be written as

$$\partial_t f = -\frac{1}{2}\frac{\delta F}{\delta f}, \tag{1.2}$$

where the corresponding energy functional is

$$F[f] = \frac{1}{4}\int\int b(x,y)f^2(t,x)f^2(t,y)\mathrm{d}x\mathrm{d}y - \frac{1}{2}\int a(x)f^2(t,x)\mathrm{d}x + \int |\nabla_x f(t,x)|^2\mathrm{d}x, \tag{1.3}$$

so that the energy dissipation law $\frac{\mathrm{d}}{\mathrm{d}t}F[f] = -2\int|\partial_t f|^2\mathrm{d}x \le 0$ holds for all $t > 0$, at least for classical solutions. There is a large literature on the subject, in terms of modeling and analysis, around the model with linear competition and mutation of the form

$$\partial_t u(t,x) = \Delta u + u(t,x)\left(a(x) - \int_X b(x,y)u(t,y)\mathrm{d}y\right). \tag{1.4}$$

However, the large time solution behavior of this model remains a challenging issue due to lack of a gradient flow structure. Nevertheless, for $b(x,y) = \eta(y) > 0$, the asymptotic solution behavior when mutation tends to vanish has been well studied, see *e.g.*, [3,28,31]. There are also other similar models, for example, model (1.4) with $b(x,y) = a(y)$ is studied in [1] through explicit solutions due to its special structure.

The aim of the present investigation is to design finite volume schemes to produce numerical solutions with satisfying long-time dynamics. This is achieved through a proper discretization, so that numerical solution

$$f_j^n \sim \frac{1}{h}\int_{I_j} f(n\Delta t, x)\mathrm{d}x,$$

approximates $f(n\Delta t, x)$ over the cell I_j indexed by $1 \le j \le N$ (in one dimensional case), with $\cup I_j = X$, where Δt is the time step and h the spatial mesh size; and the discrete energy

$$F(f) = \left[\frac{h}{4}\sum_{j,i=1}^{N}\bar{b}_{ji}f_j^2 f_i^2 - \frac{1}{2}\sum_{j=1}^{N}\bar{a}_j(f_j)^2 + \sum_{j=1}^{N-1}\left(\frac{f_{j+1}-f_j}{h}\right)^2\right]h, \tag{1.5}$$

satisfies the energy dissipation inequality

$$F(f^{n+1}) - F(f^n) \le -h \sum_{j=1}^{N} \frac{(f_j^{n+1} - f_j^n)^2}{\Delta t}$$

for Δt relative to the spatial mesh size is suitably small.

Under reasonable assumptions we are able to show that the problem of finding the discrete positive steady state is equivalent to solving a nonlinear optimization problem:

$$\begin{aligned} &\min_{f \in \mathbb{R}^N} F(f), \\ &\text{subject to} \quad f \in \{f | \ f \ge 0\}. \end{aligned}$$

The discrete energy (1.5), as a powerful tool, is further employed to prove the asymptotic stability of the steady state for both the semi-discrete and fully discrete numerical schemes as time tends to infinity. Note that in this work we derive the semi-discrete scheme by directly taking the cell average of the PDE. It may also be seen as a finite difference scheme and can be derived from the discrete version of the energy defined in (1.5) *via*,

$$\frac{\mathrm{d}}{\mathrm{d}t} f_j = -\frac{1}{2h} \partial_{f_j} F(f). \tag{1.6}$$

For derivation of structure preserving finite difference schemes from some discrete energy for a set of PDEs other than (1.1) we refer to the book [14] by Furihata and Matsuo.

Numerical experiments are performed to test both accuracy and the energy dissipation property of the numerical schemes. Numerical solutions at different times are given to show time-asymptotic convergence toward the positive steady state. Moreover, for small mutation coefficients, we illustrate how the amount of mutation affects the solution behavior.

The organization of this paper is as follows. In Section 2, we first propose the semi-discrete finite volume scheme, and then prove the existence and uniqueness of positive steady state through the equivalence between the problem of finding the positive steady state and the associated nonlinear optimization problem. Thus, the positive steady state can also be independently obtained by a well-established nonlinear programming solver. In Section 2.2, we prove that the semi-discrete scheme satisfies both positivity and energy dissipation property under some conditions on the discrete coefficients, which guarantees that the positive steady state for the semi-discrete scheme is asymptotically stable. Section 3 is devoted to a fully discrete scheme, which is semi-implicit and easy to implement. The fully discrete scheme is shown to be positivity preserving and energy decreasing under an appropriate restriction on the time step. In addition, the positive steady state for the fully discrete scheme is also shown to be asymptotically stable. In Section 4 we discuss how to design structure-preserving high order schemes. Section 5 is devoted to some numerical tests using the proposed schemes and the computation of the positive steady state through a nonlinear optimization solver. The numerical solutions of the scheme with small mutation are also compared with the solutions to the direct competition model. Finally, some concluding remarks are given in Section 6.

2. Scheme formulation and discrete properties

In order to design an energy dissipating method for capturing the large time behavior of selection-mutation dynamics (1.1), we make the following basic assumptions:

$$a \in L^\infty(X), \quad |\{x \mid \quad a(x) > 0\}| \ne 0; \tag{2.1a}$$

$$b \in L^\infty(X \times X), \quad \operatorname{essinf}_{x,x' \in X} b(x,x') > 0, \quad b(x,y) = b(y,x); \tag{2.1b}$$

$$\forall g \in L^1(X) \backslash \{0\}, \quad \int\int b(x,y) g(x) g(y) \mathrm{d}x \mathrm{d}y > 0. \tag{2.1c}$$

These assumptions are sufficient for the existence of solutions of the problem under consideration. For simplicity of presentation, we restrict ourselves to one dimensional setting with $X = [-1, 1]$. Partition X into subcells $I_j = (x_{j-1/2}, x_{j+1/2})(j = 1, \dots, N)$ of uniform mesh $h = 2/N$ so that $x_{j-1/2} = x_{1/2} + (j-1)h$ with $x_{1/2} = -1$, $x_{N+1/2} = 1$. We consider the following semi-discrete scheme

$$\frac{\mathrm{d}}{\mathrm{d}t} f_j = \frac{f_{j+1} - 2f_j + f_{j-1}}{h^2} + \frac{1}{2} f_j \left(\bar{a}_j - h \sum_{i=1}^{N} \bar{b}_{ji} f_i^2 \right), \quad 1 \le j \le N, \tag{2.2}$$

where

$$f_0 = f_1, \quad f_{N+1} = f_N,$$

$$\bar{a}_j = \frac{1}{h} \int_{I_j} a(x) \mathrm{d}x, \quad \bar{b}_{ji} = \frac{1}{h^2} \int_{I_i} \int_{I_j} b(x, y) \mathrm{d}x \mathrm{d}y, \tag{2.3}$$

and the numerical solution $f_j(t)$ approximates the cell average of the exact solution f,

$$\bar{f}_j(t) = \frac{1}{h} \int_{I_j} f\left(t, x\right) \mathrm{d}x.$$

2.1. Existence and uniqueness of the discrete positive steady state

We denote $f = (f_1, f_2, \dots, f_N)^{\mathrm{T}}$, and define the feasible set

$$S = \{ f \in \mathbb{R}^N \mid f \ge 0 \}.$$

As a nonlinear dynamical system, the large time behavior of solutions to (2.2) is closely related to the steady states $\tilde{f} \in S$ satisfying

$$\frac{-\tilde{f}_1 + \tilde{f}_2}{h^2} + \frac{1}{2} \tilde{f}_1 \left(\bar{a}_1 - h \sum_{i=1}^{N} \bar{b}_{1i} \tilde{f}_i^2 \right) = 0, \tag{2.4a}$$

$$\frac{\tilde{f}_{j-1} - 2\tilde{f}_j + \tilde{f}_{j+1}}{h^2} + \frac{1}{2} \tilde{f}_j \left(\bar{a}_j - h \sum_{i=1}^{N} \bar{b}_{ji} \tilde{f}_i^2 \right) = 0, \quad j = 2, \dots, N-1, \tag{2.4b}$$

$$\frac{\tilde{f}_{N-1} - \tilde{f}_N}{h^2} + \frac{1}{2} \tilde{f}_N \left(\bar{a}_N - h \sum_{i=1}^{N} \bar{b}_{Ni} \tilde{f}_i^2 \right) = 0. \tag{2.4c}$$

From assumptions (2.1) it follows at the discrete level some similar assumptions:

$$|\bar{a}_j| \le \|a\|_{L^\infty}, \quad \{1 \le j \le N, \bar{a}_j > 0\} \ne \emptyset; \tag{2.5a}$$

$$0 < b_f \le \bar{b}_{ji} \le \|b\|_{L^\infty} \text{ and } \bar{b}_{ji} = \bar{b}_{ij}, \text{ for } 1 \le i, j \le N; \tag{2.5b}$$

$$\sum_{j=1}^{N} \sum_{i=1}^{N} \bar{b}_{ji} g_i g_j > 0 \text{ for any } g_j \text{ such that } \sum_{j=1}^{N} |g_j|^2 \ne 0. \tag{2.5c}$$

(2.5c) implies that $B = (\bar{b}_{ij})_{N \times N}$ is a positive definite matrix. These assumptions ensure the existence and uniqueness of positive steady states.

We first show the uniqueness.

Theorem 2.1. *Let $\tilde{f}$ be a non-negative solution to (2.4), then*

(i) *either $\tilde{f} > 0$ or $\tilde{f} = 0$, and*

(ii) *positive steady state $\tilde f$ is unique.*

Proof.

(i) Suppose $\tilde f_{j_0} = 0$ for some $1 \le j_0 \le N$, then the j_0th equation of (2.4) can be written as

$$\frac{\tilde f_{j_0-1} + \tilde f_{j_0+1}}{h^2} = 0,$$

and note that $\tilde f_j \ge 0$ for all $1 \le j \le N$, we must have $\tilde f_{j_0\pm 1} = 0$. Repeating this procedure, we can conclude that $\tilde f_j = 0$ for all $1 \le j \le N$.

(ii) Suppose $\tilde f$ and $\tilde g$ are two positive steady solutions to (2.4) so that

$$I := h^2 \sum_{j,i=1}^{N} \bar b_{ji}(\tilde f_j^2 - \tilde g_j^2)(\tilde f_i^2 - \tilde g_i^2) \ge 0. \tag{2.6}$$

On the other hand,

$$\begin{aligned} I &= h^2 \sum_{j,i=1}^{N} \bar b_{ji}(\tilde f_j^2 - \tilde g_j^2)\tilde f_i^2 - h^2 \sum_{j,i=1}^{N} \bar b_{ji}(\tilde f_j^2 - \tilde g_j^2)\tilde g_i^2 \\ &= h\sum_{j=1}^{N}(\tilde f_j - \tilde g_j^2/\tilde f_j)\tilde f_j h \sum_{i=1}^{N} \bar b_{ji}\tilde f_i^2 + h\sum_{j=1}^{N}(\tilde g_j - \tilde f_j^2/\tilde g_j)\tilde g_j h\sum_{i=1}^{N}\bar b_{ji}\tilde g_i^2. \end{aligned}$$

Then, using equation (2.4), we proceed

$$\begin{aligned} I &= h\sum_{j=1}^{N}\Bigg[(\tilde f_j - \tilde g_j^2/\tilde f_j)\left(\frac{2(\tilde f_{j-1} - 2\tilde f_j + \tilde f_{j+1})}{h^2} + \bar a_j \tilde f_j\right) \\ &\quad + (\tilde g_j - \tilde f_j^2/\tilde g_j)\left(\frac{2(\tilde g_{j-1} - 2\tilde g_j + \tilde g_{j+1})}{h^2} + \bar a_j \tilde g_j\right)\Bigg] \\ &= \frac{2}{h}\sum_{j=1}^{N}\left(\tilde f_{j-1}\tilde f_j + \tilde f_j\tilde f_{j+1} - \frac{\tilde f_{j-1}\tilde g_j^2}{\tilde f_j} - \frac{\tilde f_{j+1}\tilde g_j^2}{\tilde f_j} + \tilde g_{j-1}\tilde g_j + \tilde g_j\tilde g_{j+1} - \frac{\tilde f_j^2\tilde g_{j-1}}{\tilde g_j} - \frac{\tilde f_j^2 \tilde g_{j+1}}{\tilde g_j}\right), \end{aligned}$$

when using the notation $\tilde f_0 = \tilde f_1$, $\tilde f_{N+1} = \tilde f_N$, $\tilde g_0 = \tilde g_1$ and $\tilde g_{N+1} = \tilde g_N$. We have

$$\begin{aligned} I &= \frac{2}{h}\sum_{j=1}^{N-1}\left(2\tilde f_j\tilde f_{j+1} - \frac{\tilde f_j\tilde g_{j+1}^2}{\tilde f_{j+1}} - \frac{\tilde f_{j+1}\tilde g_j^2}{\tilde f_j} + 2\tilde g_j\tilde g_{j+1} - \frac{\tilde f_{j+1}^2\tilde g_j}{\tilde g_{j+1}} - \frac{\tilde f_j^2\tilde g_{j+1}}{\tilde g_j}\right) \\ &= -\frac{2}{h}\sum_{j=1}^{N-1}\left[\frac{(\tilde f_j\tilde g_{j+1} - \tilde f_{j+1}\tilde g_j)^2}{\tilde g_j\tilde g_{j+1}} + \frac{(\tilde f_j\tilde g_{j+1} - \tilde f_{j+1}\tilde g_j)^2}{\tilde f_j\tilde f_{j+1}}\right] \le 0, \end{aligned}$$

which means that I is both nonnegative and nonpositive according to (2.6). Therefore $\tilde f = \tilde g$. ∎

For the existence of $\tilde f$, we define the discrete Lyapunov functional as follows

$$F(f) = \left(\frac{1}{4}\sum_{j,i=1}^{N}\bar b_{ji} f_j^2 f_i^2 h - \frac{1}{2}\sum_{j=1}^{N}\bar a_j f_j^2 + \sum_{j=1}^{N-1}\left(\frac{f_{j+1} - f_j}{h}\right)^2\right) h. \tag{2.7}$$

In fact, the positive steady state can be expressed as a solution to the following optimization problem

$$\min_{f\in\mathbb{R}^N} F, \tag{2.8a}$$

$$\text{subject to} \quad f \in S = \{f \in \mathbb{R}^N | f \geq 0\}. \tag{2.8b}$$

Lemma 2.2. *If* (2.5) *holds, then the minimizer of nonlinear optimization problem* (2.8) *is the same as the positive solution of* (2.4).

Proof. Let g be a solution to problem (2.8), we now show $g = \tilde{f}$, which is the unique positive steady state. A simple calculation gives

$$\frac{\partial F}{\partial f_j}\Big|_{f=g} = -2h\left\{\frac{g_{j-1} - 2g_j + g_{j+1}}{h^2} + \frac{1}{2}g_j\left(\bar{a}_j - h\sum_{i=1}^N \bar{b}_{ji}g_i^2\right)\right\}, \quad 1 \leq j \leq N$$

with notations $g_0 = g_1$ and $g_{N+1} = g_N$. If one component of $g \in S$ is zero, we must have

$$\frac{\partial F}{\partial f_j}\Big|_{f=g} \geq 0,$$

which when combined with $g \in S$ ensures that $g \equiv 0$. Otherwise, if g is strictly positive, then we have

$$\frac{\partial F}{\partial f_j}\Big|_{f=g} = 0, \quad 1 \leq j \leq N,$$

which is the same relation as (2.4). By the uniqueness of the positive solution to (2.4), we much have $g = \tilde{f}$.

We next prove $F(\tilde{f}) < F(0)$, which ensures that g cannot be zero. A direct calculation using summation by parts gives

$$\sum_{j=1}^{N-1} \frac{(\tilde{f}_{j+1} - \tilde{f}_j)^2}{h} = -\sum_{j=1}^N \frac{\tilde{f}_j(\tilde{f}_{j+1} - 2\tilde{f}_j - \tilde{f}_{j-1})}{h}. \tag{2.9}$$

Multiplying the jth equation in (2.4) by $\tilde{f}_j h$, and summing from $j = 1$ to N, we have

$$-\sum_{j=1}^N \frac{\tilde{f}_j(\tilde{f}_{j+1} - 2\tilde{f}_j - \tilde{f}_{j-1})}{h} = \frac{h}{2}\sum_{j=1}^N \bar{a}_j \tilde{f}_j^2 - \frac{h^2}{2}\sum_{j,i=1}^N \bar{b}_{ji}\tilde{f}_j^2\tilde{f}_i^2. \tag{2.10}$$

With (2.9) and (2.10), the functional $F(\tilde{f})$ can be reduced to

$$F(\tilde{f}) = -\frac{h^2}{4}\sum_{j,i=1}^N \bar{b}_{ji}\tilde{f}_j^2\tilde{f}_i^2 < 0 = F(0). \tag{2.11}$$

Finally, let $\tilde{f}$ be the positive steady state, then $\tilde{f}$ must be the unique critical point of F in the interior of S. Assume that a minimizer g can be achieved on the boundary of S, then the above argument shows that $g = 0$. This together with the fact that $F(\tilde{f}) < F(0)$ and $F(\infty) = \infty$ suggests that $\tilde{f}$ is also the solution to (2.8). □

Armed with this result, we are left to show $F(\cdot)$ admits a unique minimizer among elements in S.

Lemma 2.3. *For fixed $N > 0$, if $\bar{a}_i > 0$ for some i, $\bar{b}_{ij} \geq 0$ for all i, j and $(\bar{b}_{ij})$ is positive definite, then there exists only one nontrivial vector $g \in S$ such that*

$$F(g) = \min_{f \in S} F(f).$$

Proof. Note that S is nonempty and closed. For positive definite matrix B, $F(f)$ is bounded from below, *i.e.*,

$$\frac{-|a|_\infty^2 N}{4\lambda_{\min}} \leq \frac{h^2}{4}\lambda_{\min}\sum_{j=1}^N f_j^4 - \frac{h}{2}\|a\|_{L^\infty}\sum_{j=1}^N f_j^2 \leq F(f), \tag{2.12}$$

where $\lambda_{\min} > 0$ is the smallest eigenvalue of B. Set $m := \inf_{f\in S} F(f)$, then m is finite. Select a minimizing sequence $\{f^{(n)}\}_{n=1}^\infty$, so that

$$F(f^{(n)}) \to m.$$

Thus, (2.12) implies that $\{f^{(n)}\}_{n=1}^\infty$ is bounded in $S \subset \mathbb{R}^N$. By the Bolzano–Weierstrass theorem, there exists a subsequence $\{f^{(n_k)}\}_{k=1}^\infty$ that converges to g. We assert that the limit $g \in S$ for S is closed. By continuity of $F(f)$,

$$m = \lim_{k\to\infty} F(f^{(n_k)}) = F(g).$$

Hence, $F(g) = m = \min_{f\in S} F(f)$, and $g = \tilde{f} > 0$ is unique by the claim of Lemma 2.2. $\square$

Combining Lemma 2.2 with Lemma 2.3, we have the following.

Theorem 2.4. *If (2.5) holds, then there exists a unique positive steady state as defined in (2.4).*

Hence, the positive steady state can be obtained by solving nonlinear optimization problem (2.8).

2.2. Properties of the semi-discrete scheme

As noted above, when applicable, the vector notation $f(t) = (f_1(t), \ldots, f_N(t))^{\mathrm{T}}$ is used.

Theorem 2.5. *Assume (2.5) holds, and let $f(t)$ be the numerical solution to the semi-discrete scheme (2.2). Then,*

(i) *if $f_j(0) \geq 0$ for every $1 \leq j \leq N$, then*

$$f_j(t) \geq 0 \ \textit{ for any } \ 1 \leq j \leq N \ \textit{ and any } \ t > 0.$$

(ii) *F is non-increasing in time. Moreover,*

$$\frac{\mathrm{d}F}{\mathrm{d}t} = -2h\sum_{j=1}^N \left(\frac{\mathrm{d}f_j}{\mathrm{d}t}\right)^2 \leq 0. \tag{2.13}$$

Proof.

(i) Define

$$\Omega = \{f \in \mathbb{R}^N \mid \ f \geq 0\}.$$

The claimed positivity preserving property follows if Ω is an invariant region. Note that $\mathbf{0}$ is an equilibrium solution to (2.2), it suffices to show

$$\frac{\mathrm{d}f}{\mathrm{d}t}\cdot\nu := \sum_{j=1}^N \nu_j\frac{\mathrm{d}f_j}{\mathrm{d}t} < 0, \quad \forall f \in \partial\Omega\backslash\{\mathbf{0}\},$$

where ν is the outward normal vector at $\partial\Omega$, which can be defined by

$$\nu_j = \begin{cases} -1, & \text{if } j \in S_0, \\ 0, & \text{elsewhere.} \end{cases}$$

Here we use the index set

$$S_0 = \{1 \le j \le N \mid f_j = 0, f \in \partial\Omega\}.$$

It follows that

$$\frac{\mathrm{d}f}{\mathrm{d}t} \cdot \nu = -\sum_{j \in S_0} \frac{f_{j-1} + f_{j+1}}{h^2} < 0,$$

for one of the following two cases must appear: there exists $j_0 \in S_0$, then

$$f_{j_0} = 0, \quad f_{j_0+1} > 0 \quad \text{or} \quad f_{j_0} = 0, \quad f_{j_0-1} > 0.$$

This is due to the fact that S is neither empty nor $\{1, \ldots, N\}$ for $f \in \partial\Omega\backslash\{\mathbf{0}\}$. The desired positivity property is thus proved.

(ii) By a direct calculation,

$$\begin{aligned} \frac{\mathrm{d}F}{\mathrm{d}t} &= h\sum_{j=1}^{N}\left(\frac{1}{2}f_j\frac{\mathrm{d}f_j}{\mathrm{d}t}h\sum_{i=1}^{N}\bar{b}_{ji}f_i^2 + \frac{1}{2}f_j^2 h\sum_{i=1}^{N}\bar{b}_{ji}f_i\frac{\mathrm{d}f_i}{\mathrm{d}t} - \bar{a}_j f_j\frac{\mathrm{d}f_j}{\mathrm{d}t}\right) \\ &\quad + h\sum_{j=1}^{N-1}\left[\frac{2(f_{j+1}-f_j)}{h^2}\left(\frac{\mathrm{d}f_{j+1}}{\mathrm{d}t} - \frac{\mathrm{d}f_j}{\mathrm{d}t}\right)\right] \\ &=: -h\sum_{j=1}^{N}\frac{\mathrm{d}f_j}{\mathrm{d}t}f_j\left(\bar{a}_j - h\sum_{i=1}^{N}\bar{b}_{ji}f_i^2\right) + 2T, \end{aligned} \tag{2.14}$$

where $\bar{b}_{ij} = \bar{b}_{ji}$ has been used. Next, we estimate T. Using the notation $f_0 = f_1$ and $f_{N+1} = f_N$, we have

$$\begin{aligned} T &= \frac{1}{h}\left[\sum_{j=1}^{N-1}\frac{\mathrm{d}f_{j+1}}{\mathrm{d}t}(f_{j+1}-f_j) - \sum_{j=1}^{N-1}\frac{\mathrm{d}f_j}{\mathrm{d}t}(f_{j+1}-f_j)\right] \\ &= -h\sum_{j=1}^{N}\frac{\mathrm{d}f_j}{\mathrm{d}t}\frac{f_{j-1}-2f_j+f_{j+1}}{h^2}. \end{aligned}$$

Substitution of this into (2.14), together with (2.2) leads to the desired property (2.13). □

We may also examine the large time behavior of $f(t)$.

Theorem 2.6. *For fixed $N \ge 0$, assume (2.5) holds, and $\tilde{f}$ is the positive steady state. Let the initial data $f(0)$ be positive and $F(f(0)) < 0$, and $f(t)$ be the numerical solution generated from semi-discrete scheme (2.2), then*

$$\lim_{t\to\infty} f(t) = \tilde{f}. \tag{2.15}$$

Proof. Define a set

$$\Sigma = \{f = (f_1, f_2, \ldots, f_N) \in \mathbb{R}^N | f > 0,\ F(f) < 0\}.$$

Then both $\tilde{f}$ and initial data $f(0)$ are in Σ.

On this set we define a functional $V(f)$ by

$$V(f) = F(f) - F(\tilde{f}),$$

which implies that $\frac{\mathrm{d}}{\mathrm{d}t}V = \frac{\mathrm{d}}{\mathrm{d}t}F$. We see that $V(\tilde{f}) = \frac{\mathrm{d}}{\mathrm{d}t}V(\tilde{f}) = 0$. Furthermore, V satisfies the following properties:

(1) $V(f) > 0$ for any $f \in \Sigma \backslash \tilde{f}$(positive definite);
(2) $\frac{\mathrm{d}}{\mathrm{d}t}V \leq 0$ for all $f \in \Sigma$(non-increasing in time);
(3) $V(f) \to \infty$, if $|f| \to \infty$.

Note also that the set $\{\dot{V}(f) = 0\} \bigcap \Sigma$ contains no trajectory of system (2.2) except the trivial trajectory $f(t) = \tilde{f}$ for $t \geq 0$. We can therefore apply the local Krasovskii–LaSalle principle [21] to conclude that

$$\lim_{t\to\infty} f(t) = \tilde{f},$$

which leads to the desired result.

The last two properties are easy to verify by (2.13) and the form of F. We next only verify property (1). For arbitrary $\tau \geq 0$ and $\forall g$ such that $f = \tilde{f} + \tau g \in \Sigma$, we let $s(\tau) = F(f) - F(\tilde{f})$, then a direct calculation upon regrouping leads to

$$s'(\tau) = -2h \sum_{j=1}^{N} g_j \left[\frac{1}{2} f_j \left(\bar{a}_j - h \sum_{i=1}^{N} \bar{b}_{ji} f_i^2 \right) + \frac{f_{j-1} - 2f_j + f_{j+1}}{h^2} \right]. \tag{2.16}$$

In order that $s'(\tau) = 0$ for arbitrary $f \in \Sigma$ and arbitrary g, the following is necessarily satisfied

$$\frac{1}{2} f_j \left(\bar{a}_j - h \sum_{i=1}^{N} \bar{b}_{ji} f_i^2 \right) + \frac{f_{j-1} - 2f_j + f_{j+1}}{h^2} = 0.$$

Hence we have $f = \tilde{f}$ by the uniqueness of the positive steady solution, *i.e.*, $\tau = 0$. This says that $\tilde{f}$ is the only critical point of $V(f)$, and $V(\tilde{f}) = 0$, which when combined with (3) ensures (1). $\square$

3. Time discretization

We propose the following fully discrete scheme

$$\frac{f_j^{n+1} - f_j^n}{\Delta t} = \frac{f_{j-1}^{n+1} - 2f_j^{n+1} + f_{j+1}^{n+1}}{h^2} + \frac{1}{2} f_j^{n+1} \left(\bar{a}_j - h \sum_{i=1}^{N} \bar{b}_{ji} (f_i^n)^2 \right), \tag{3.1}$$

with

$$f_0^n = f_1^n, \quad f_{N+1}^n = f_N^n.$$

This scheme is semi-implicit, linear in f^{n+1}, hence easy to implement. We next show that such a scheme also preserves both positivity and energy dissipation property under certain restriction on the time step.

To proceed, we set the discrete energy as

$$F^n = F(f^n) = \left[\frac{1}{4} \sum_{j,i=1}^{N} \bar{b}_{ji} (f_j^n)^2 (f_i^n)^2 h - \frac{1}{2} \sum_{j=1}^{N} \bar{a}_j (f_j^n)^2 + \sum_{j=1}^{N-1} \left(\frac{f_{j+1}^n - f_j^n}{h} \right)^2 \right] h. \tag{3.2}$$

For later use, we present a uniform L^1-bound and L^2-bound of the numerical solution when $b \geq b_f > 0$. Let $\gamma = \|a\|_{L^\infty}/b_f$ which will be used to quantify the uniform bound.

Lemma 3.1. *Assume* (2.5) *holds. Let* f^n *be the numerical solution generated from scheme* (3.1) *with nonnegative initial data* $f^0 \geq 0$, *and* $\|f^0\|_2 < \infty$. *Then, for any* $n \in \mathbb{N}$,

$$\|f^n\|_2 \leq \max\left\{\|f^0\|_2, \sqrt{\gamma}\right\}, \quad \|f^n\|_1 \leq \max\left\{\sqrt{2}\|f^0\|_2, \sqrt{2\gamma}\right\}, \tag{3.3}$$

provided

$$\Delta t \leq \frac{1}{\|a\|_{L^\infty}}. \tag{3.4}$$

Proof. Let $Q^n = h\sum_{j=1}^{N}(f_j^n)^2 = \|f^n\|_2^2$. Multiplying $f_j^{n+1}h$ on both sides of scheme (3.1) and then summing over j from 1 to N, we have

$$h\sum_{j=1}^{N}\frac{(f_j^{n+1})^2 - f_j^{n+1}f_j^n}{\Delta t} = \sum_{j=1}^{N}\frac{f_{j-1}^{n+1}f_j^{n+1} - 2(f_j^{n+1})^2 + f_j^{n+1}f_{j+1}^{n+1}}{h} + \frac{h}{2}\sum_{j=1}^{N}(f_j^{n+1})^2\left(\bar{a}_j - h\sum_{i=1}^{N}\bar{b}_{ji}(f_i^n)^2\right). \tag{3.5}$$

The term on the left hand side can be written as

$$h\sum_{j=1}^{N}\left[\frac{(f_j^{n+1})^2 - (f_j^n)^2}{2\Delta t} + \frac{(f_j^{n+1} - f_j^n)^2}{2\Delta t}\right],$$

and the second term on the right hand side of (3.5) can be estimated as

$$\begin{aligned}&\leq \frac{1}{2}\max_{1\leq j\leq N}|\bar{a}_j - h\sum_{i=1}^{N}\bar{b}_{ji}(f_i^n)^2|Q^{n+1}\\ &\leq \frac{1}{2}\left(\|a\|_{L^\infty} - b_f Q^n\right)Q^{n+1}.\end{aligned}$$

The first term on the right hand side of (3.5) is non-negative since

$$\begin{aligned}&\frac{1}{h}\left[2\sum_{j=2}^{N}f_j^{n+1}f_{j-1}^{n+1} - 2\sum_{j=1}^{N}(f_j^{n+1})^2 + (f_1^{n+1})^2 + (f_N^{n+1})^2\right] \leq\\ &\frac{1}{h}\left[\sum_{j=2}^{N}(f_j^{n+1})^2 + \sum_{j=2}^{N}(f_{j-1}^{n+1})^2 - 2\sum_{j=1}^{N}(f_j^{n+1})^2 + (f_1^{n+1})^2 + (f_N^{n+1})^2\right] = 0,\end{aligned}$$

where we have used the notation $f_0 = f_1$ and $f_{N+1} = f_N$, and the Cauchy–Schwarz inequality.

Combining the above estimates, we obtain

$$Q^{n+1} \leq Q^n + \Delta t\left(\|a\|_{L^\infty} - b_f Q^n\right)Q^{n+1}.$$

There are two cases to distinguish:

(i) if $Q^n \geq \gamma$, then $Q^{n+1} \leq Q^n$;

(ii) if $Q^n < \gamma$, we rewrite

$$Q^{n+1} - \gamma \leq (Q^n - \gamma)(1 - \Delta t Q^{n+1} b_f).$$

According to (3.4), we have

$$Q^{n+1} - \gamma \leq (Q^n - \gamma) \left(1 - \frac{Q^{n+1}}{\gamma}\right),$$

which leads to $Q^{n+1} \leq \gamma$. Hence,

$$Q^{n+1} \leq \max\{Q^n, \gamma\} \leq \ldots \leq \max\{Q^0, \gamma\}.$$

Furthermore, using the Cauchy–Schwarz inequality, we have $\|f^n\|_1 \leq \sqrt{2}\|f^n\|_2$. These ensure the claimed estimate (3.3). ☐

Theorem 3.2. *Assume (2.5) is satisfied, and let f^n be the numerical solution to the fully-discrete scheme (3.1) with time step satisfying*

$$\Delta t \leq \min\left\{\frac{1}{\|a\|_{L^\infty}},\ \frac{h}{\frac{h}{2}\left(\|a\|_{L^\infty} + C\|b\|_{L^\infty}\right) + 2C\|b\|_{L^\infty}}\right\}, \tag{3.6}$$

where $C = \max\{\|f^0\|_2^2, \|a\|_{L^\infty}/b_f\}$. Then,

(i) *if $f_j^n > 0$ for every $1 \leq j \leq N$, then $f_j^{n+1} > 0$ for any $1 \leq j \leq N$ and any $n \in \mathbb{N}$;*
(ii) *F^n is a decreasing sequence in n. Moreover,*

$$F^{n+1} - F^n \leq -h \sum_{j=1}^{N} \frac{(f_j^{n+1} - f_j^n)^2}{\Delta t}. \tag{3.7}$$

Proof.

(i) Rearranging (3.1), we have

$$-\frac{\Delta t}{h^2} f_{j-1}^{n+1} + \left\{1 + \frac{2\Delta t}{h^2} - \frac{\Delta t}{2}\left[\bar{a}_j - h\sum_{i=1}^{N} \bar{b}_{ji}(f_i^n)^2\right]\right\} f_j^{n+1} - \frac{\Delta t}{h^2} f_{j+1}^{n+1} = f_j^n.$$

Let $f_{j_0}^{n+1} = \min_{1\leq j\leq N} f_j^{n+1}$, then the j_0th equation gives

$$\begin{aligned} 0 \leq f_{j_0}^n &= -\frac{\Delta t}{h^2} f_{j_0-1}^{n+1} + \left\{1 + \frac{2\Delta t}{h^2} - \frac{\Delta t}{2}\left[\bar{a}_{j_0} - h\sum_{i=1}^{N} \bar{b}_{j_0 i}(f_i^n)^2\right]\right\} f_{j_0}^{n+1} - \frac{\Delta t}{h^2} f_{j_0+1}^{n+1} \\ &\leq \left\{1 - \frac{\Delta t}{2}\left[\bar{a}_{j_0} - h\sum_{i=1}^{N} \bar{b}_{j_0 i}(f_i^n)^2\right]\right\} f_{j_0}^{n+1}. \end{aligned}$$

From $\Delta t \leq 1/\|a\|_{L^\infty}$ as implied by (3.6), we see that the coefficient of $f_{j_0}^{n+1}$ is strictly positive, therefore $f_{j_0}^{n+1} \geq 0$, which means $f_j^{n+1} \geq 0$ for all $1 \leq j \leq N$. Specially, if $f_j^n > 0$ for every $1 \leq j \leq N$, then $f_j^{n+1} > 0$ for all $1 \leq j \leq N$.

(ii) We proceed to estimate $F^{n+1} - F^n$ as follows:

$$\begin{aligned} F^{n+1} - F^n = & \frac{h^2}{4} \sum_{i,j=1}^{N} \bar{b}_{ji} \left[(f_j^{n+1})^2 (f_i^{n+1})^2 - (f_j^n)^2 (f_i^n)^2 \right] \\ & - \frac{h}{2} \sum_{j=1}^{N} \bar{a}_j (f_j^{n+1} - f_j^n)(f_j^{n+1} + f_j^n) \\ & + \frac{1}{h} \sum_{j=1}^{N-1} \left[(f_{j+1}^{n+1} - f_j^{n+1})^2 - (f_{j+1}^n - f_j^n)^2 \right]. \end{aligned} \tag{3.8}$$

Due to symmetry of $\bar{b}_{ij}$, the first term on the right can be written as

$$\frac{h^2}{4} \sum_{i,j=1}^{N} \bar{b}_{ji} \left[(f_j^{n+1})^2 - (f_j^n)^2 \right] \left[(f_i^{n+1})^2 + (f_i^n)^2 \right].$$

Summation by parts when applied to the third term on the right of (3.8) gives

$$\begin{aligned} I_3 := & \frac{1}{h} \sum_{j=1}^{N-1} \left[f_{j+1}^{n+1} - f_{j+1}^n - (f_j^{n+1} - f_j^n) \right] \left[f_{j+1}^{n+1} + f_{j+1}^n - (f_j^{n+1} + f_j^n) \right] \\ = & - \frac{1}{h} \sum_{j=1}^{N} (f_j^{n+1} - f_j^n)(\omega_j^{n+1} + \omega_j^n), \end{aligned}$$

where $\omega_j^{n+1} = f_{j-1}^{n+1} - 2f_j^{n+1} + f_{j+1}^{n+1}$.

Substitution of these into (3.8) gives

$$F^{n+1} - F^n = - h \sum_{j=1}^{N} \left(f_j^{n+1} - f_j^n \right) A_j$$

where, using the scheme (3.1), we have

$$\begin{aligned} A_j = & \frac{\omega_j^{n+1} + \omega_j^n}{h^2} + \frac{1}{2}(f_j^{n+1} + f_j^n) \left[\bar{a}_j - h \sum_{i=1}^{N} \bar{b}_{ji} (f_i^n)^2 - \frac{h}{2} \sum_{i=1}^{N} \bar{b}_{ji} \left((f_i^{n+1})^2 - (f_i^n)^2 \right) \right] \\ = & 2 \frac{f_j^{n+1} - f_j^n}{\Delta t} - \frac{\omega_j^{n+1} - \omega_j^n}{h^2} - \frac{f_j^{n+1} - f_j^n}{2} \left(\bar{a}_j - h \sum_{i=1}^{N} \bar{b}_{ji} (f_i^n)^2 \right) \\ & - \frac{h}{4} (f_j^{n+1} + f_j^n) \sum_{i=1}^{N} \bar{b}_{ji} \left((f_i^{n+1})^2 - (f_i^n)^2 \right). \end{aligned}$$

Hence

$$
\begin{aligned}
F^{n+1}-F^{n}=&-2h\sum_{j=1}^{N}\frac{\left(f_j^{n+1}-f_j^{n}\right)^2}{\Delta t}+\frac{1}{h}\sum_{j=1}^{N}\left(f_j^{n+1}-f_j^{n}\right)\left(\omega_j^{n+1}-\omega_j^{n}\right)\\
&+\frac{h}{2}\sum_{j=1}^{N}\left(f_j^{n+1}-f_j^{n}\right)^2\left[\bar{a}_j-h\sum_{i=1}^{N}\bar{b}_{ji}(f_i^{n})^2\right]\\
&+\frac{h^2}{4}\sum_{j=1}^{N}\left(f_j^{n+1}-f_j^{n}\right)\left(f_j^{n+1}+f_j^{n}\right)\sum_{i=1}^{N}\bar{b}_{ji}\left(f_i^{n+1}-f_i^{n}\right)\left(f_i^{n+1}+f_i^{n}\right)\\
=:&-2T_1+T_2+T_3+T_4.
\end{aligned}
$$

Since $T_1>0$, we only need to prove $T_2,\ T_3,\ T_4\le C\Delta t T_1$ for suitably small Δt. Note that for $p_j=f_j^{n+1}-f_j^{n}$ we have

$$
\begin{aligned}
T_2&=\frac{1}{h}\sum_{j=1}^{N}p_j(p_{j+1}-2p_j+p_{j-1})\\
&=-\frac{1}{h}\sum_{j=1}^{N-1}(p_{j+1}-p_j)^2\le 0.
\end{aligned}
$$

In virtue of Lemma 3.1, we have

$$
\begin{aligned}
T_3&=\frac{h}{2}(\|a\|_{L^\infty}+\|b\|_{L^\infty}\|f^n\|_2^2)\sum_{j=1}^{N}\left(f_j^{n+1}-f_j^{n}\right)^2\\
&\le c_1\Delta t T_1,
\end{aligned}
$$

where $c_1=\frac{1}{2}\left(\|a\|_{L^\infty}+\|b\|_{L^\infty}\max\{\|f^0\|_2^2,\|a\|_{L^\infty}/b_f\}\right)$.

Using the fact that B is symmetric, we obtain

$$
\begin{aligned}
T_4&\le\frac{h^2}{4}\sum_{j=1}^{N}\left(f_j^{n+1}-f_j^{n}\right)^2\left(f_j^{n+1}+f_j^{n}\right)\sum_{i=1}^{N}\bar{b}_{ji}\left(f_i^{n+1}+f_i^{n}\right)\\
&\le\frac{1}{4}\|b\|_{L^\infty}\max_{1\le j\le N}h\left(|f_j^{n+1}|+|f_j^{n}|\right)\sum_{i=1}^{N}\left(|f_i^{n+1}|h+|f_i^{n}|h\right)\sum_{j=1}^{N}\left(f_j^{n+1}-f_j^{n}\right)^2\\
&\le\|b\|_{L^\infty}\max_{n}\|f^n\|_1^2\sum_{j=1}^{N}\left(f_j^{n+1}-f_j^{n}\right)^2\le c_2\Delta t T_1,
\end{aligned}
$$

where $c_2=2\|b\|_{L^\infty}\max\{\|f^0\|_2^2,\|a\|_{L^\infty}/b_f\}/h$.

For Δt satisfying (3.6), putting together the above estimates, we have

$$
F^{n+1}-F^{n}\le-\left[2-(c_1+c_2)\,\Delta t\right]T_1\le -h\sum_{j=1}^{N}\frac{(f_j^{n+1}-f_j^{n})^2}{\Delta t},
$$

which is as desired. $\square$

The established energy dissipation inequality (3.7) ensures the following time-asymptotic result.

Theorem 3.3. *For fixed $N \geq 0$, assume (2.5) is satisfied and $\tilde{f}$ is the positive steady state. Let the initial data f^0 be positive and $F(f^0) < 0$, and f^n be the numerical solution generated from fully-discrete scheme (3.1). If Δt is small so that (3.6) holds, then*

$$\lim_{n\to\infty} f^n = \tilde{f}. \tag{3.9}$$

Proof. Define the set

$$\Sigma = \{f = (f_1, f_2, \ldots, f_N) \in \mathbb{R}^N | f > 0,\ F(f) < 0\}.$$

Then we have $\tilde{f} \in \Sigma$ and $f^0 \in \Sigma$. Denote

$$\varphi^n(f^0) = f^n,$$

and $\omega(f^0)$ be a set of limit points of convergent subsequences of the numerical solution $\varphi^n(f^0)$. The boundedness of $\|f^n\|_2$, as stated in Lemma 3.1, ensures that there always exists a convergent subsequence of $\varphi^n(f^0)$, hence $\omega(f^0)$ is nonempty.

From inequality (3.7), we see that F^n is a decreasing sequence in n, and also bounded from below by

$$F^n \geq -\frac{1}{2}\|a\|_{L^\infty}\|f^n\|_2^2 \geq -\frac{1}{2}\|a\|_{L^\infty} \max\left\{\|f^0\|_2^2, \gamma\right\}.$$

Hence the limit of F^n exists when n tends to ∞; that is

$$\lim_{n\to\infty} F^n = \lim_{n\to\infty} F(f^n) = F^* \leq F^0 < 0.$$

For any $g \in \omega(f^0)$, there exists a subsequence $\{f^{(n_l)}\}$ that converges to g, so that

$$F(g) = \lim_{l\to\infty} F(f^{(n_l)}) = F^* < 0,$$

therefore $g \geq 0$ and $g \neq 0$, because $F(0) = 0$.

Note that

$$F^* \leq F(\varphi^{n+n_l}(f^0)) \leq F(\varphi^{n_l}(f^0))$$

for any $n \in \mathbb{N}$. Let l tend to ∞, by the continuity of F, we obtain

$$F^* \leq F(\varphi^{(n)}(g)) \leq F(g) = F^*,$$

that is, $F(\varphi^{n+1}(g)) = F(\varphi^n(g)) = F^*$ for $n = 0, 1, \ldots$ From inequality (3.7) we see that

$$\varphi_j^{n+1}(g) = \varphi_j^n(g) \geq 0$$

for every $1 \leq j \leq N$ and any $n \in \mathbb{N} \cup \{0\}$. Since $g \geq 0$ and $g \neq 0$, scheme (3.1) leads to $\varphi^n(g) \geq 0$ and $\varphi^n(g) \neq 0$, and further $\varphi^n(g) = \tilde{f}$ for any $n \in \mathbb{N} \cup \{0\}$. Hence, $g = \tilde{f}$, that is $\omega(f^0) = \{\tilde{f}\}$. The desired limit (3.9) is thus proved. □

4. Spatially high order scheme

In this section we discuss a way of designing spatially high-order energy dissipative schemes in the context of discontinuous Galerkin (DG) discretization. We denote by v^+ and v^- the right and left limits of function v crossing cell interfaces, and

$$[v] = v^+ - v^-, \quad \{v\} = \frac{v^+ + v^-}{2}.$$

Define an k-degree discontinuous finite element space

$$V_h = \left\{v \in L^2(\Omega), \quad v|_{I_j} \in P^k(I_j), j = 1, \ldots, N\right\},$$

where $P^k(I_j)$ denotes the set of all polynomials of degree at most k on I_j.

In one dimensional setting, the DG scheme for (1.1) is to find $f_h \in V_h$ such that for all $v \in V_h$ and $j = 1, \ldots, N$,

$$\begin{aligned}\int_{I_j} \partial_t f_h v \mathrm{d}x = & -\int_{I_j} \partial_x f_h \partial_x v \mathrm{d}x + \left(\widehat{\partial_x f_h} v + \partial_x v(f_h - \{f_h\})\right)|_{\partial I_j} \\ & + \frac{1}{2}\int_{I_j} v f_h(t,x)\left(a(x) - \int_\Omega b(x,y) f_h^2(t,y)\mathrm{d}y\right)\mathrm{d}x.\end{aligned} \tag{4.1}$$

Here

$$v|_{\partial I_j} = v(x^-_{j+1/2}) - v(x^+_{j-1/2}),$$

and $\widehat{\partial_x f_h}$ is the numerical flux taken as

$$\widehat{\partial_x f_h} = \beta_0 \frac{[f_h]}{h} + \{\partial_x f_h\},\ j = 1, \ldots N-1, \tag{4.2}$$

where $h = \Delta x$ for uniform meshes at $x_{j+1/2}$ ($h = (\Delta x_j + \Delta x_{j+1})/2$ for non-uniform meshes). The terms evaluated at two ends are zero due to the natural boundary condition. The initial data $f_h(0,x) \in V_h$ is obtained by the piecewise L^2 projection of $f_0(x)$.

Taking $v = f_h$ and summing (4.1) over j we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} F_h = -2\int_\Omega |\partial_t f_h|^2 \mathrm{d}x,$$

where

$$F_h = \frac{1}{4}\int\int b(x,y) f_h^2(x) f_h^2(y)\mathrm{d}x\mathrm{d}y - \frac{1}{2}\int a(x) f_h^2(x)\mathrm{d}x + A(f_h, f_h)$$

with

$$A(f_h, f_h) = \sum_{j=1}^{N}\int_{I_j} |\partial_x f_h|^2 \mathrm{d}x + \sum_{j=1}^{N-1}(\widehat{\partial_x f_h} + \{\partial_x f_h\})[f_h]\Big|_{j+1/2}$$

is a high order approximation of F. It is known from [24,25] that $A(u,v)$ is symmetric and there exists $\gamma > 0$ such that

$$A(f_h, f_h) \geq \gamma \|f_h\|_E^2, \quad \|f_h\|_E^2 = \sum_{j=1}^{N}\int_{I_j} |\partial_x f_h|^2 \mathrm{d}x + \sum_{j=1}^{N-1} h^{-1}[f_h]^2\Big|_{j+1/2}$$

provided $\beta_0 > k^2$. As for the time variable, we may apply a similar explicit-implicit discretization, coupled with the more accurate Gauss quadrature in the integral terms in (4.1). The positivity of numerical solutions is known to be difficult to achieve when using higher order schemes. These issues and further numerical validation are hence left for future investigation.

5. Numerical implementation and examples

In this section we present several numerical tests to illustrate both accuracy and capability of scheme (3.1).

In order to see the numerical performance at large times, we present an alternative way to compute the steady state: this is based on the established result that the positive steady state is the same as the solution to a nonlinear optimization problem. For general nonlinear optimization problem, a variety of algorithms has been proposed in the literature, including trust-region-reflective algorithm, active-set algorithm, interior-point algorithm and sequence quadratic programming algorithm(sqp for short) (see [10,11,17,19,29,34]). We shall use the Matlab code *fmincon.m* to implement the sqp algorithm.

TABLE 1. Errors and the convergence orders of the numerical solution on uniform meshes of N cells.

	$f_0(x) = 0.5(\sin(100x) + 2)$			
N	L^∞ error	order	L^1 error	order
10	0.1583	–	0.2456	–
20	0.0568	1.4787	0.0821	1.5809
40	0.0139	2.0308	0.0138	2.5727
80	0.0034	2.0315	0.0014	3.3012
160	0.0016	1.0874	7.1341e-04	0.9726
320	6.7877e-04	1.2371	3.5278e-04	1.0159
640	2.2410e-04	1.5988	1.7611e-04	1.0023

5.1. One-dimensional tests to scheme (3.1)

We numerically test several selected examples.

Recall that the positivity of b when $b(x, y) = K(x - y)$ is equivalent to the positivity of the Fourier transform of K (see [20], p. 502). In addition to the Gaussian kernel, we also use $K = \frac{1}{1+x^2}$. In fact, a simple calculation using the Cauchy integral formula in complex plane, one obtains

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-ix\xi}}{1+x^2} \mathrm{d}x = \sqrt{\frac{\pi}{2}} \mathrm{e}^{-|\xi|}.$$

Therefore, the b's used in (5.1) also satisfies the positivity condition as required.

Example 1 (Accuracy and energy test)**.** Following the setting used in [22, 23], we consider

$$a(x) = 10(x-1)^2(x-0.1005)^2(x+1)^2, \quad b(x, y) = \frac{1}{1 + (40(x-y))^2}, \tag{5.1}$$

which when using the 3-point Gaussian quadrature rule gives the needed discrete data, $\bar{a}_j$ and $\bar{b}_{ji}$. For initial data given by

$$f_0(x) = 0.5(\sin(100x) + 2), \tag{5.2}$$

the initialization is given by

$$f_j^0 = \frac{1}{h} \int_{I_j} f_0(x) \mathrm{d}x, \quad j = 1, \ldots, N.$$

This evaluation is carried out by the 3-point Gaussian quadrature rule. Let f_j^n denote the numerical solution in the jth cell of N cells, and $\tilde{f}_i^n$ the reference solution in the ith cell of mN cells. The L^∞ error and the L^1 error are defined as

$$\max_{1 \le j \le N} \max_{1 \le l \le m} |\, f_j^n - \tilde{f}_{m(j-1)+l}^n \,|, \quad \sum_{j=1}^{N} \sum_{l=1}^{m} |\, f_j^n - \tilde{f}_{m(j-1)+l}^n \,| \frac{h}{m},$$

respectively. In our simulation, the numerical solution of 1280 cells is taken as the reference solution. Let the final time $T = n\Delta t$, the accuracy of numerical scheme (3.1) at $T = 1.0$ with time step $\Delta t = 0.1$ is given in Table 1, which confirms the first-order accuracy. Here the choice of Δt may be determined according to the bound in Theorem 3.2. Actually, Δt can be taken slightly larger as long as time-asymptotic convergence is obtained. Table 2 gives the temporal change of the discrete energy (3.2).

TABLE 2. The change of the discrete energy (3.2) with $N = 80$ and $\Delta t = 0.1$.

t_n	0	1	3	5	8	10	12
$F(f^n)$	826.6761	–1.5902	–3.7552	–4.5460	–4.6092	–4.6096	–4.6096

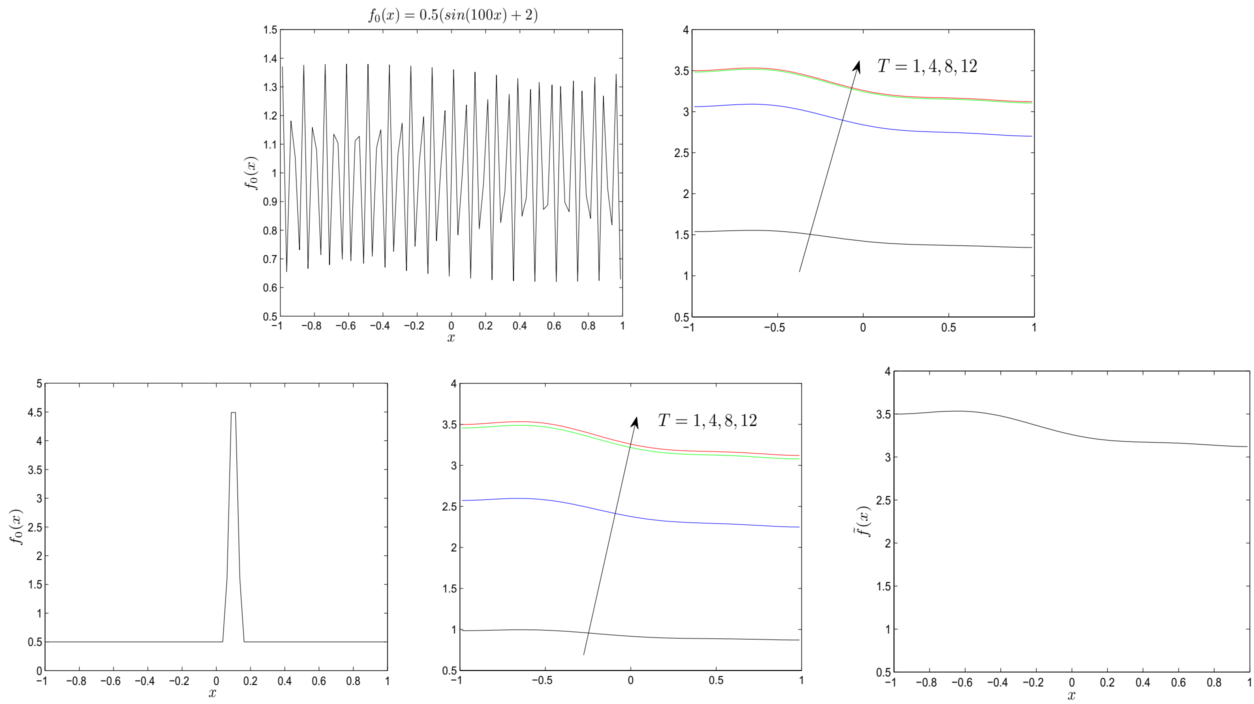

FIGURE 1. The first row: initial data (5.2) (*left*); corresponding numerical solutions (*right*). The second row: initial data (5.3) (*left*); corresponding numerical solutions (*middle*); positive steady state (*right*). $N = 80$ and $\Delta t = 0.1$.

Example 2 (Large time behavior and positive steady state)**.** In addition to initial data (5.2), we also test with another positive initial data of the form

$$f_0(x) = \begin{cases} 2(\cos(2\pi(x - 0.1)) + 1) + 0.5, & |x - 0.1| \leq 0.03, \\ 0.5, & \text{else}. \end{cases} \tag{5.3}$$

Numerical solutions to scheme (3.1) with coefficients (5.1) are convergent to the positive steady state, which is shown in Figure 1.

We also consider Gaussian coefficients given by:

$$a(x) = G(x, \sigma_1), \quad b(x, y) = G(x - y, \sigma_2), \tag{5.4}$$

where

$$G(x, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \mathrm{e}^{-\frac{x^2}{2\sigma}}.$$

For random initial data, we test the time-asymptotic convergence of numerical solutions toward the positive steady state. The results are given in Figure 2.

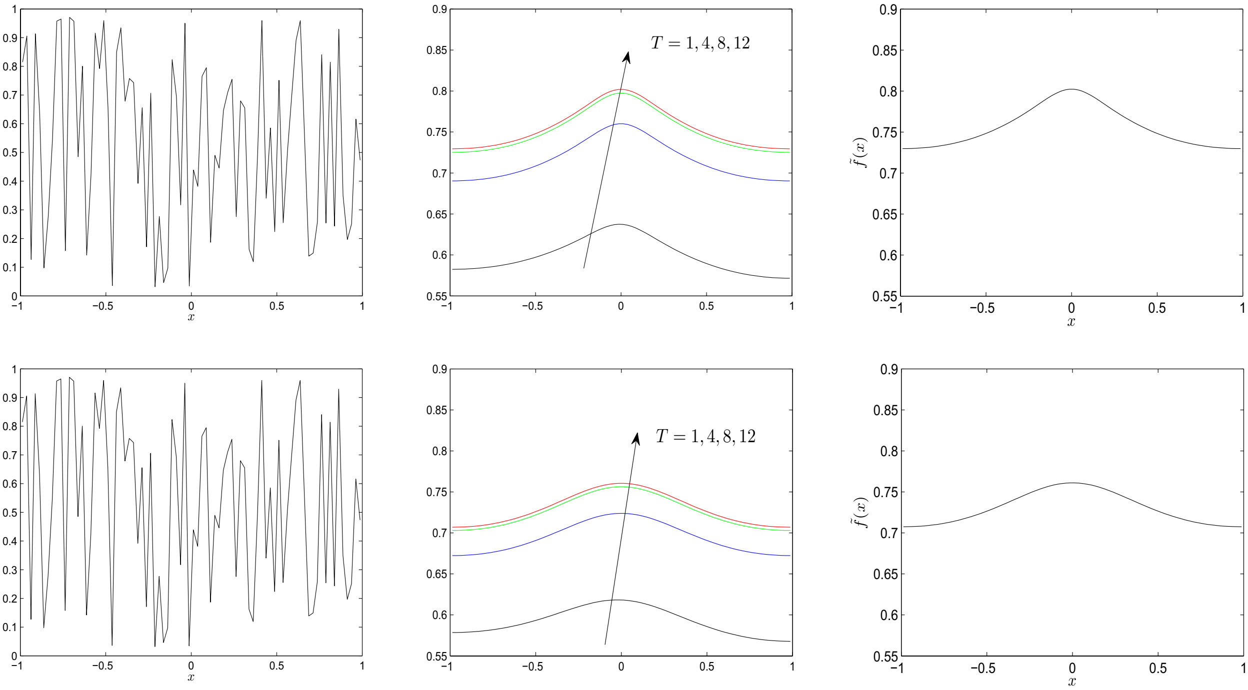

FIGURE 2. The first row: random initial data (*left*); numerical solutions for $\sigma_1 = 0.01 < \sigma_2 = 0.05$ (*middle*); corresponding steady state (*right*). The second row: random initial data (*left*); numerical solutions for $\sigma_1 = 0.05 > \sigma_2 = 0.01$ (*middle*); corresponding steady state (*right*). Final time $T = 1, 4, 8, 12$, $N = 80$ and $\Delta t = 0.1$.

5.2. Numerical tests with small mutation

The previous numerical tests show that numerical solutions are almost uniformly continuous, instead of concentration. That is because the diffusion, of size $O(1)$, is strong in spreading the density profile. Note that the diffusion in the model represents certain mutation mechanism, however, mutation is rare in practical applications. We hence investigate how the amount of diffusion could affect the solution behavior over time. Consider the following model

$$\partial_t f(t,x) = \epsilon f_{xx}(t,x) + \frac{1}{2} f(t,x)\left(a(x) - \int_X b(x,y) f^2(t,y)\mathrm{d}y\right),\ t > 0,\ x \in X = [-1,1], \tag{5.5}$$

and the corresponding fully discrete scheme

$$\frac{f_j^{n+1} - f_j^n}{\Delta t} = \epsilon \frac{f_{j-1}^{n+1} - 2f_j^{n+1} + f_{j+1}^{n+1}}{h^2} + \frac{1}{2} f_j^{n+1}\left(\bar{a}_j - h\sum_{i=1}^{N} \bar{b}_{ji}(f_i^n)^2\right),\ 1 \le j \le N, \tag{5.6}$$

with

$$f_0^n = f_1^n, \quad f_{N+1}^n = f_N^n.$$

Model (5.5), when the mutation rate $\epsilon \to 0$, formally leads to

$$\frac{\partial}{\partial t} f(t,x) = \frac{1}{2} f(t,x)\left(a(x) - \int_X b(x,y) f^2(t,y)\mathrm{d}y\right), \tag{5.7}$$

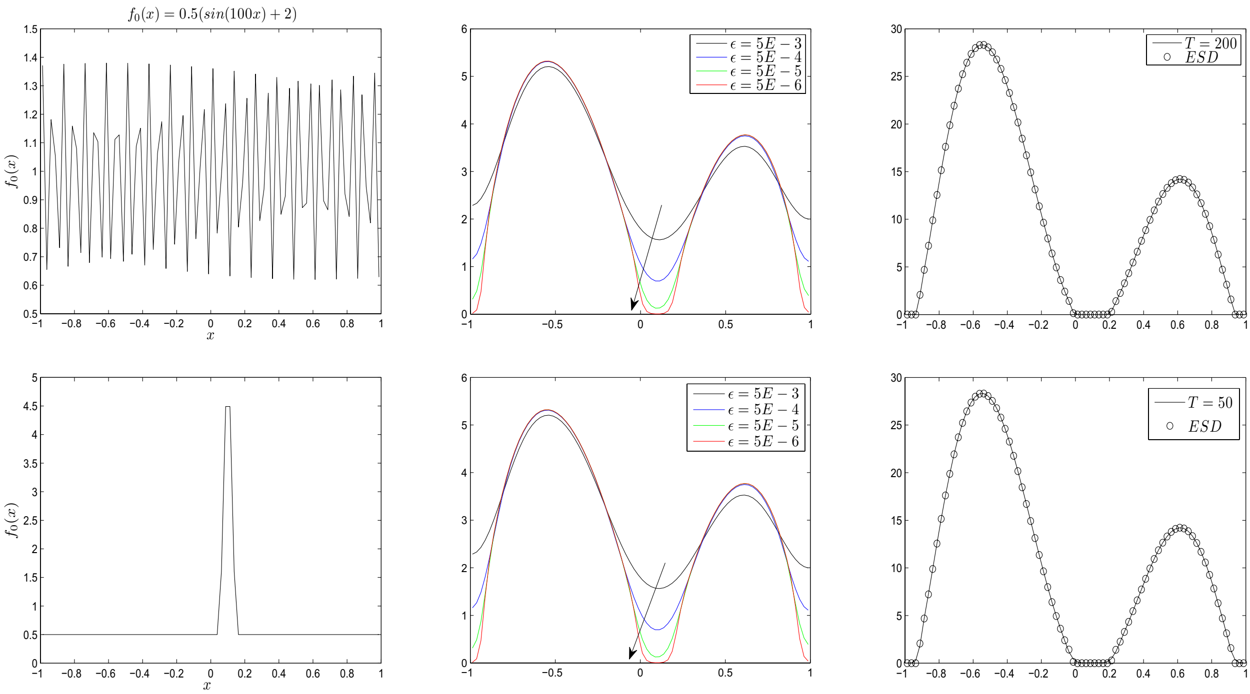

FIGURE 3. Initial data (5.2) and (5.3) (*left*); numerical solutions to scheme (5.6) for $\epsilon = 5E-3, 5E-4, 5E-5, 5E-6$ at $T = 200$ and $\Delta t = 0.9$ (*middle*); with $\Delta t = 0.01$, numerical solutions of model (5.8) at $T = 200$ and $T = 50$, respectively, and the ESD generated by a nonlinear programming solver in [22] (*right*).

or the direct competition model for $u = f^2$, of the form

$$\frac{\partial}{\partial t}u(t,x) = u(t,x)\left(a(x) - \int_X b(x,y)u(t,y)\mathrm{d}y\right), \tag{5.8}$$

which is numerically investigated in [22].

Rearranging scheme (5.6), we have

$$-\frac{\epsilon\Delta t}{h^2}f_{j-1}^{n+1} + \left[1 + \frac{2\epsilon\Delta t}{h^2} - \frac{\Delta t}{2}\left(\bar{a}_j - h\sum_{i=1}^{N}\bar{b}_{ji}(f_i^n)^2\right)\right]f_j^{n+1} - \frac{\epsilon\Delta t}{h^2}f_{j+1}^{n+1} = f_j^n. \tag{5.9}$$

This scheme for $\Delta t < \frac{2}{\|a\|_{L^\infty}}$ is shown to be positivity preserving. We test numerical solutions to scheme (5.6) with data (5.1) and (5.4). In the numerical simulation, we compare numerical solutions of scheme (5.6) for a series of sufficiently small ϵ with that of the direct competitive model as obtained in [22].

Example 3 (Numerical solutions with data (5.1))**.** For fixed sufficiently large final time and an appropriate time step, we display the changes of numerical solutions for two kinds of initial data with a series of decreasing ϵ in Figure 3. The results indicate that the numerical solution for smaller ϵ is getting closer to the corresponding discrete Evolutionarily Stable Distribution (ESD for short) of model (5.8), in which the ESD is also obtained using a nonlinear programming solver as obtained in [22].

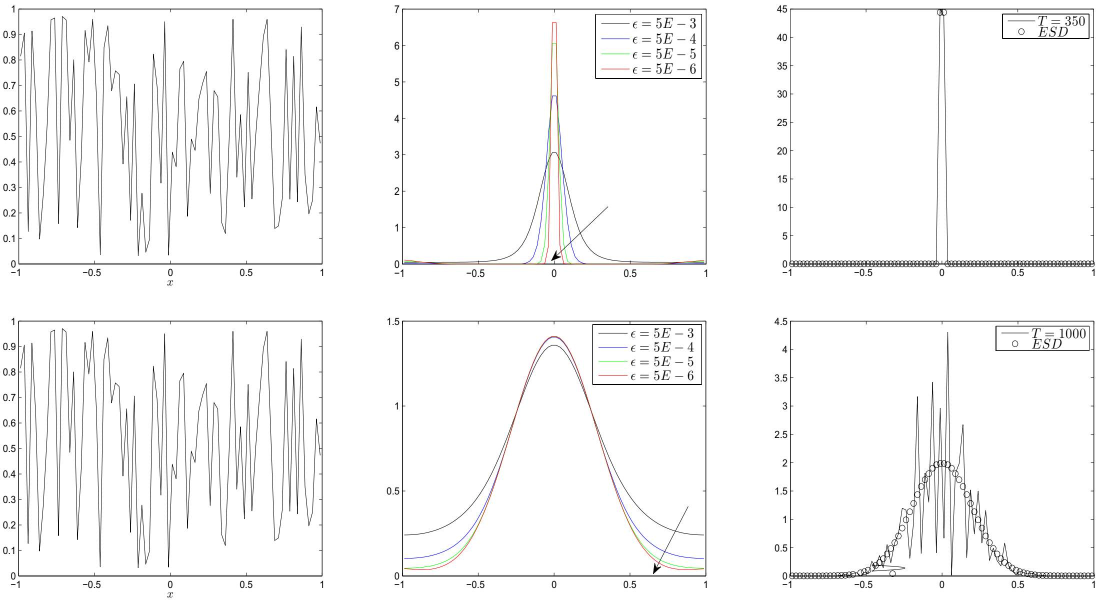

FIGURE 4. Random initial data (*left*); numerical solutions to scheme (5.6) for $\epsilon = 5E-3, 5E-4, 5E-5, 5E-6$ at $T = 350$ and $\Delta t = 0.5$, and at $T = 1000$ and $\Delta t = 1.1$, respectively (*middle*); with $\Delta t = 0.01$, numerical solutions of model (5.8) at $T = 350$ and $T = 1000$, respectively, and the ESD generated by a nonlinear programing solver in [22] (*right*).

Example 4 (Numerical solutions with Gaussian data (5.4))**.** We test numerical solutions to scheme (5.6) with random initial data for different ϵ. It is shown in Fig.4 that the numerical solution will tend to a Dirac mass concentrating on 0 for $\sigma_1 = 0.01 < \sigma_2 = 0.05$, but to a Gaussian function for $\sigma_1 = 0.01 > \sigma_2 = 0.05$, as time becomes large. In other words, the numerical solutions to scheme (5.6) for smaller ϵ are getting closer to the ESD of the direct competition model (5.8).

6. Summary

In this work, we have developed finite volume schemes for solving the competition-mutation model with a gradient flow structure. The schemes are easy to compute and are shown to preserve both positivity and the energy dissipation property. We have also established the equivalence between the existence of a unique discrete positive steady state and the existence of a unique minimizer of the nonlinear optimization problem. The proposed energy functional is further explored to prove that numerical solutions of both semi-discrete and fully discrete schemes converge asymptotically toward the positive steady state as time becomes large. Numerical examples have confirmed both accuracy and entropy dissipation properties of the numerical schemes, and have shown the efficiency to capture the large time behavior of numerical solutions. Furthermore, numerical solutions of the model with small mutation are shown to be close to those of the direct competition model.

Acknowledgements. The authors thank Professor P.E. Jabin for suggesting the model and stimulating discussions. Cai was supported by the China Scholarship Council and by the Fundamental Research Funds for the Central Universities under Grant 00-800015JH TianYuan Special Funds of National Natural Science Foundation of China (No. 11626228) and National Natural Science Foundation of China (No. 11601514); and Liu was supported by the National Science Foundation under Grant DMS13-12636 and by NSF Grant RNMS (Ki-Net) 1107291.

References

[1] A. Alfaro and R. Carles, Explicit solutions for replicator-mutator equation: extinction vs. acceleration. *SIAM J. Appl. Math.* **74** (2014) 1919–1934.

[2] G. Barabás and G. Meszéna, When the exception becomes the rule: the disappearance of limiting similarity in the Lotka-Volterra model. *J. Math. Biol.* **258** (2009) 89–94.

[3] G. Barles, S. Mirrahimi and B. Perthame, Concentration in Lotka-Volterra parabolic or integral equations: a general convergence result. *Methods Appl. Anal.* **16** (2009) 321–340.

[4] H. Berestycki, G. Nadin, B. Perthame and L. Ryzhik, The non-local Fisher–KPP equation: traveling waves and steady states. *Nonlinearity* **22** (2009) 2813–2844.

[5] B. Bolker and W. Pacala, Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theor. Popul. Biol.* **52** (1997) 179–197.

[6] R. Bürger, The mathematical theory of selection, recombination and mutation. Wiley, New York (2000).

[7] N. Champagnat, R. Ferrière and S. Méléard, Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. *Theor. Popul. Biol.* **69** (2006) 297–231.

[8] N. Champagnat, R. Ferrière and S. Méléard, From individual stochastic processes to macroscopic models in adaptive evolution. *Stoch. Models* **24** (2008) 2–44.

[9] Y. Cohen and G. Galiano, Evolutionary distributions and competition by way of reaction-diffusion and by way of convolution. *Bull. Math. Biol.* **75** (2013) 2305–2323.

[10] T.F. Coleman and Y. Li, On the convergence of reflective newton methods for large-scale nonlinear minimization subject to bounds. *Math. Program.* **67** (1994) 189–224.

[11] T.F. Coleman and Y. Li, An interior trust region approach for nonlinear minimization subject to bounds. *SIAM J. Optim.* **6** (1996) 418–445.

[12] L. Desvillettes, P.E. Jabin, S. Mischler and G. Raoul, On selection dynamics for continuous structured populations. *Commun. Math. Sci.* **6** (2008) 729–747.

[13] N. Fournier and S. Méléard, A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab.* **14** (2004) 1880–1919.

[14] D. Furihata and T. Matsuo, Discrete variational derivative method: a structure-preserving numerical method for partial differential equations. Chapman and Hall/CRC numerical analysis and scientific computation series. ISBN 978-1-4200-9445-9 (2011).

[15] S. Génieys and B. Perthame, Dynamics of Nonlocal Fisher concentration points: a nonlinear analysis of Turing patterns. *Math. Model. Nat. Phenomenon* **2** (2007) 135–151.

[16] S. Génieys, V. Volpert and P. Auger, Adaptive dynamics: modeling Darwin's divergence principle. *C. R. Biol.* **329** (2006) 876–879.

[17] P.E. Gill, W. Murray and M.H. Wright, Practical Optimization. London, Academic Press (1981).

[18] S.A. Gourley, Traveling front of a nonlocal Fisher equation. *J. Math. Biol.* **41** (2000) 272–284.

[19] S.P. Han, A globally convergent method for nonlinear programming. *J. Optimiz. Theory Appl.* **22** (1977) 297–309.

[20] P.E. Jabin and G. Raoul, On selection dynamics for competitive interactions. *J. Math. Biol.* **63** (2011) 493–517.

[21] J. LaSalle, Some extensions of Liapunov's second method. *IRE Trans. Circuit Theory* **7** (1960) 520–527.

[22] H.L. Liu, W.L. Cai and N. Su, Entropy satisfying schemes for computing selection dynamics in competitive interactions. *SIAM J. Numer. Anal.* **53** (2015) 1393–1417.

[23] H.L. Liu and H. Yu, Entropy/energy stable schemes for evolutionary dispersal models. *J. Comput. Phys.* **256** (2014) 656–677.

[24] H.L. Liu, Optimal error estimates of the direct discontinuous Galerkin method for convection-diffusion equations. *Math. Comput.* **84** (2015) 2263–2295.

[25] H.L. Liu HL and J. Yan, The direct discontinuous Galerkin (DDG) method for diffusion with interface corrections. *Commun. Comput. Phys.* **8** (2010) 541–564.

[26] A. Lorz, S. Mirrahimi and B. Perthame, Dirac mass dynamics in multidimensional nonlocal parabolic equations. *Commun. Partial Differ. Eq.* **36** (2011) 1071–1098.

[27] S. Mirrahimi, B. Perthame, E. Bouin and P. Millien, Population formulation of adaptative meso-evolution: theory and numerics. *Mathematics and Biosciences in Interaction* (2011) 159–174.

[28] S. Mirrahimi, B. Perthame and J.Y. Wakano, Direct competition results from strong competition for limited resource. *J. Math. Biol.* **68** (2014) 931–949.

[29] J. Nocedal and S.J. Wright, Numerical Optimization, 2nd edn. *Springer Series in Operations Research.* Springer Verlag (2006).

[30] B. Perthame and G. Barles, Dirac concentrations in Lotka-Volterra parabolic PDEs. *Indiana Univ. Math. J.* **57** (2008) 3275–3301.

[31] G. Raoul, Long time evolution of populations under selection and vanishing mutations. *Acta Appl. Math.* **114** (2011) 1–14.

[32] G. Raoul, Local stability of evolutionary attractors for continuous structured populations. *Monatsh. Math.* **165** (2012) 117–144.

[33] K. Shirakihara and S. Tanaka, Two fish species competition model with nonlinear interactions and equilibrium catches. *Res. Popul. Ecol.* **20** (1978) 123–140.

[34] R.A. Waltz, J.L. Morales, J. Nocedal and D. Orban, An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Math. Program.* **107** (2006) 391–408.