High Resolution Schemes for a Hierarchical Size-Structured Model

Jun Shen¹, Chi-Wang Shu² and Mengping Zhang³

Abstract

In this paper we discuss two explicit finite difference schemes, namely a first order upwind scheme and a second order high resolution scheme, for solving a hierarchical size-structured population model with nonlinear growth, mortality and reproduction rates. We prove stability and convergence for both schemes and provide numerical examples to demonstrate their capability in solving smooth and discontinuous solutions.

Key Words: hierarchical size-structured population model, upwind scheme, high resolution scheme, stability, convergence

AMS(MOS) subject classification: 65M06, 65M12, 92-08

¹Department of Mathematics, University of Science and Technology of China, Hefei, Anhui 230026, P.R. China. E-mail: jshen3@mail.ustc.edu.cn

²Division of Applied Mathematics, Brown University, Providence, RI 02912, USA. E-mail: shu@dam.brown.edu. Research partially supported by the Chinese Academy of Sciences while the author was in residence at the University of Science and Technology of China (grant 2004-I-8) and at the Institute of Computational Mathematics and Scientific / Engineering Computing. Additional support is provided by ARO grant W911NF-04-1-0291, NSF grant DMS-0510345 and AFOSR grant FA9550-05-1-0123.

³Department of Mathematics, University of Science and Technology of China, Hefei, Anhui 230026, P.R. China. E-mail: mpzhang@ustc.edu.cn. Research partially supported by the Chinese Academy of Sciences grant 2004-I-8.
1 Introduction

In this paper we develop stable and convergent finite difference schemes for a hierarchical size-structured population model given by the following equations

\[ u_t + (g(x, Q(x,t)) u)_x + m(x, Q(x,t)) u = 0, \quad (x, t) \in (0, L) \times (0, T) \]
\[ g(0, Q(0,t)) u(0,t) = C(t) + \int_0^L \beta(x, Q(x,t)) u(x,t) dx, \quad t \in (0, T) \quad (1.1) \]
\[ u(x, 0) = u^0(x), \quad x \in [0, L] \]

where \( u(x,t) \) is the density of individuals having size \( x \) at time \( t \), and the non-local term \( Q(x,t) \) is defined by

\[ Q(x,t) = \alpha \int_0^x w(\xi) u(\xi,t) d\xi + \int_x^L w(\xi) u(\xi,t) d\xi, \quad 0 \leq \alpha < 1 \quad (1.2) \]

for some given function \( w \). \( Q(x,t) \) depends on the density \( u \) in a global way and is usually referred to as the environment.

A special feature of equation (1.1) is the boundary condition at size \( x = 0 \), which involves the function \( g \) representing the growth rate of an individual, and a global dependency on the density \( u(x,t) \) for all \( x \in (0, L] \). The function \( m \) in (1.1) represents the mortality rate of an individual. The function \( \beta \) in the boundary condition of (1.1) represents the reproduction rate of an individual, and the function \( C \) represents the inflow rate of zero-size individual from an external source. We assume that the functions \( g, m \) and \( \beta \) are functions of both the size \( x \) and the environment \( Q \), which in turn depends globally on the density \( u \), hence the problem is highly nonlinear.

Hierarchical structured population models have been studied in the literature in, e.g. [2, 3, 5, 6, 10, 12, 18], usually with more restrictive assumptions on the functions \( g, \beta \) and \( m \). For example, in [3] the model (1.1) was considered for the special situation \( g = g(Q), \beta = \beta(Q), m = m(Q) \) and \( C = 0 \). In [2], the model (1.1) was studied with the functions \( g \) and \( \beta \) depending linearly on the size \( x \), \( m \) independent of \( x \), and \( C = 0 \). In [12], (1.1) was investigated with \( \alpha = 0 \). The model (1.1) with the complete generality as stated above
was studied in [1], in which an implicit first order finite difference scheme was analyzed and its stability and convergence, as well as the existence, uniqueness and well-posedness (in $L^1$ norm) of bound variation weak solutions for (1.1) were proved. However, the scheme in [1] is not very practical for actual numerical simulation, because it is implicit and only first order accurate.

In this paper we develop and analyze two explicit finite difference schemes, namely a first order upwind scheme and a second order high resolution scheme, for solving (1.1). We prove stability and convergence for both schemes. Many aspects of our proof are based on the techniques in [1, 4, 7, 14], but it is not a routine generalization, because of the complication due to the explicit time marching, second order accuracy, and global constraints in the equation. We also provide numerical examples to demonstrate the capability of these schemes in solving smooth and discontinuous solutions.

As in [1], we make the following assumptions on the model functions:

- (H1) $g(x, Q)$ is twice continuously differentiable with respect to $x$ and $Q$, $g(x, Q) > 0$ for $x \in [0, L)$, $g(L, Q) = 0$, and $g_Q(x, Q) \leq 0$.
- (H2) $m(x, Q)$ is nonnegative continuously differentiable with respect to $x$ and $Q$.
- (H3) $\beta(x, Q)$ is nonnegative continuously differentiable with respect to $x$ and $Q$. Furthermore, there is a constant $\omega_1 > 0$ such that $\sup_{(x, Q) \in [0, L] \times [0, \infty)} \beta(x, Q) \leq \omega_1$.
- (H4) $w(x)$ is nonnegative continuously differentiable.
- (H5) $C(t)$ is nonnegative continuously differentiable.
- (H6) $u^0 \in BV[0, L]$ and $u^0(x) \geq 0$.

In section 2, we present an explicit, first order upwind scheme for solving (1.1) and prove its stability and convergence. In section 3, we present an explicit, second order high resolution scheme for solving (1.1) and prove its stability and convergence. Section 4 contains numerical
examples demonstrating the capability of these two numerical schemes. Concluding remarks are given in section 5.

2 A first order upwind finite difference scheme

First, we briefly describe the standard notations to be used in this paper. We assume the spatial domain $[0, L]$ is divided into $N$ cells with cell boundary points denoted by $x_j$, for $0 \leq j \leq N$, $x_0 = 0$ and $x_N = L$. For simplicity of presentation we will assume that the mesh is uniform of size $\Delta x$, namely $x_j = j \Delta x$. This assumption is not essential for the analysis or the numerical computation, more general meshes can be easily considered. We also denote the time step by $\Delta t$. In fact, this time step $\Delta t = \Delta t^n = t^{n+1} - t^n$ could change from one step to the next step, based on stability conditions, but we use the same notation $\Delta t$ without the superscript $n$ since we will only consider one-step time discretization (forward Euler or Runge-Kutta), hence there will be no confusion. We shall denote by $u^n_j$ and $Q^n_j$ the finite difference approximations of $u(x_j, t^n)$ and $Q(x_j, t^n)$, respectively. We also denote

$$g^n_j = g(x_j, Q^n_j), \quad \beta^n_j = \beta(x_j, Q^n_j), \quad m^n_j = m(x_j, Q^n_j), \quad w_j = w(x_j), \quad C^n = C(t^n).$$

We define the standard finite difference operators

$$D^-(u^n_j) = \frac{u^n_j - u^n_{j-1}}{\Delta x}, \quad \Delta^+(u^n_j) = u^n_{j+1} - u^n_j, \quad \Delta^-(u^n_j) = u^n_j - u^n_{j-1},$$

and we define the standard discrete $L^1$ and $L^\infty$ norms and $TV$ seminorm of the grid function $u^n$ by

$$\|u^n\|_1 = \sum_{j=1}^N |u^n_j| \Delta x, \quad \|u^n\|_\infty = \max_{0 \leq j \leq N} |u^n_j|, \quad TV(u^n) = \sum_{j=0}^{N-1} |u^n_{j+1} - u^n_j|.$$

The explicit, first order upwind finite difference scheme for (1.1) that we consider in this section is defined by

$$\frac{u^n_{j+1} - u^n_j}{\Delta t} + \frac{g^n_j u^n_j - g^n_{j-1} u^n_{j-1}}{\Delta x} + m^n_j u^n_j = 0, \quad 1 \leq j \leq N \quad (2.1)$$
with the left boundary condition implemented by
\[ g_0^n u_0^n = C^n + \sum_{j=1}^{N} \beta_j^n u_j^n \Delta x, \tag{2.2} \]
the environment is computed by
\[ Q_j^n = \alpha \sum_{i=1}^{j} w_i u_i^n \Delta x + \sum_{i=1}^{N} w_i u_i^n \Delta x \tag{2.3} \]
and the initial condition is taken as
\[ u_j^0 = u^0(x_j), \quad j = 1, 2, \ldots, N. \]

We denote \( \lambda = \frac{\Delta t}{\Delta x} \), and rewrite the scheme (2.1) as
\[ u_j^{n+1} = u_j^n - \lambda (g_j^n u_j^n - g_{j-1}^n u_{j-1}^n) - \Delta t m_j^n u_j^n = (1 - \lambda g_j^n - \Delta t m_j^n) u_j^n + \lambda g_{j-1}^n u_{j-1}^n, \quad j \geq 1. \tag{2.4} \]
Since we consider only one step explicit schemes, the right side of (2.4) contains only terms at time level \( t^n \). Hence sometimes we will omit the superscript \( n \) when it does not cause confusion.

We first prove the \( L^1 \) boundedness of the numerical solution \( u^n \) for \( t^n \leq T \), under the assumption that \( u_j^n \geq 0 \). We will prove the validity of this assumption later.

**Proposition 2.1.** If \( u_j^n \geq 0 \), then \( \|u^n\|_1 \) is bounded when \( t^n \leq T \).

**Proof.** Since \( u_j^n \geq 0 \), \( m_j^n \geq 0 \) (assumption H2 in section 1) and \( g_0^n = 0 \) (assumption H1 in section 1), we have
\[
\begin{align*}
\frac{\|u^{n+1}\|_1 - \|u^n\|_1}{\Delta t} &= \sum_{j=1}^{N} \frac{u_j^{n+1} - u_j^n}{\Delta t} \Delta x \\
&= - \sum_{j=1}^{N} (g_j^n u_j^n - g_{j-1}^n u_{j-1}^n) - \sum_{j=1}^{N} m_j^n u_j^n \Delta x \\
&\leq - \sum_{j=1}^{N} (g_j^n u_j^n - g_{j-1}^n u_{j-1}^n) \\
&= g_0^n u_0^n \\
&= C^n + \sum_{j=1}^{N} \beta_j^n u_j^n \Delta x \\
&\leq C + \omega_1 \|u^n\|_1
\end{align*}
\]
where $C$ denotes the upperbound of $C(t)$ for $t \in [0, T]$ and $\omega_1$ is the upperbound of $\beta(x, Q)$ given in assumption H3 in section 1. If $\Delta t$ is a constant, we immediately have

$$\|u^n\|_1 \leq (1 + \omega_1 \Delta t) \|u^{n-1}\|_1 + C \Delta t$$

$$\leq (1 + \omega_1 \Delta t)^n \|u^0\|_1 + \sum_{j=0}^{n-1} (1 + \omega_1 \Delta t)^j C \Delta t$$

$$\leq e^{\omega_1 T} \|u^0\|_1 + \frac{C e^{\omega_1 T}}{\omega_1} \equiv \omega_2,$$

where the constant $\omega_2$, as well as a sequence of such constants $\omega_k$ to be defined later, depend only on the given functions $g, m, C, \beta, w$, the final time $T$ and the initial condition $u^0$. This proof is clearly also valid, with a minor modification, for the situation when $\Delta t$ is not a constant. ■

With this $L^1$ bound on the density $u^j_i$, we can easily obtain the following upper bound for the environment $Q$

$$|Q^n_j| = \left| \alpha \sum_{i=1}^{j} w_i u^n_i \Delta x + \sum_{i=j+1}^{N} w_i u^n_i \Delta x \right|$$

$$\leq \|w\|_{\infty} \max_{n} \|u^n\|_1 \leq \omega_2 \|w\|_{\infty} \equiv Q_{\text{max}}.$$

We now have a bounded closed domain $D = \{(x, Q) \in [0, L] \times [0, Q_{\text{max}}]\}$ that $x$ and $Q$ reside in, hence by the smoothness assumptions of $g, m, \beta$ and $w$, we have a fixed constant $\omega_3$ such that

$$\sup_{D} |f(x, Q)| \leq \omega_3, \quad \sup_{0 \leq x \leq L} |h(x)| \leq \omega_3$$

for

$$f(x, Q) = g(x, Q), g_x(x, Q), g_Q(x, Q), g_{xx}(x, Q), g_{xQ}(x, Q), g_{QQ}(x, Q),$$

$$m(x, Q), m_x(x, Q), m_Q(x, Q), \beta(x, Q),$$

$$h(x) = w(x), w'(x).$$

Thus when $\Delta t \leq \Delta t_0 \equiv \frac{1}{2\omega_3}$ and $\lambda \leq \lambda_0 \equiv \frac{1}{2\omega_3}$, we have

$$1 - \lambda g^n_j - \Delta tm^n_j \geq 0, \quad 1 \leq j \leq N. \quad (2.5)$$
This clearly implies \( u^n_j \geq 0 \) by (2.4). Notice that we can either choose \( \lambda = \lambda_0 \) as a constant or a variable depending on the time level \( t^n \). We have thus verified the assumption made in Proposition 2.1 about the non-negativity of \( u^n_j \).

Next we will prove the \( L^\infty \) boundedness of the numerical solution.

**Proposition 2.2.** \( \|u^n\|_\infty \) is bounded for \( t^n \leq T \).

**Proof.** First we have

\[
g_0^n u^n_0 = C^n + \sum_{i=1}^{N} \beta^n_i u^n_i \Delta x \leq C + \omega_1 \|u^n\|_1 \leq C + \omega_1 \omega_2.
\]

Since \( g \) is continuous and \( g(0, Q) > 0 \) by assumption H1 in section 1, we can take

\[
\mu = \min_{Q \in [0, Q_{max}]} g(0, Q) > 0. \quad (2.6)
\]

Therefore

\[
|u^n_0| \leq \frac{C + \omega_1 \omega_2}{\mu}. \quad (2.7)
\]

As for \( j \geq 1 \), we use (2.4), (2.5) and the non-negativity of \( m \) to obtain

\[
|u^n_j| \leq (1 - \lambda g^n_{j-1} - \Delta tm^n_{j-1}) \|u^{n-1}\|_\infty + \lambda g^n_{j-1} \|u^{n-1}\|_\infty
\]

\[
\leq \|u^{n-1}\|_\infty - \lambda (g^n_{j-1} - g^n_{j-1}) \|u^{n-1}\|_\infty.
\]

We have

\[
g^n_{j-1} - g^n_{j-1} = g(x_j, Q^n_{j-1}) - g(x_{j-1}, Q^n_{j-1}) + g(x_{j-1}, Q^n_{j-1}) - g(x_{j-1}, Q^n_{j-1})
\]

\[
= \Delta x (\dot{x}_j, Q^n_{j-1} \Delta x + g_Q(x_{j-1}, Q^n_{j-1}) (Q^n_{j-1} - Q^n_{j-1})
\]

\[
= \Delta x (\dot{x}_j, Q^n_{j-1} \Delta x + g_Q(x_{j-1}, Q^n_{j-1}) (\alpha - 1) w_j u^{n-1}_j \Delta x.
\]

Here and below \( \dot{z}_j \) denotes a value between \( z_{j-1} \) and \( z_j \), for \( z = x \) or \( z = Q \). By assumption, \( \alpha < 1, g_Q(x, Q) \leq 0 \), we clearly have

\[-\lambda g_Q(x_{j-1}, Q_j) (\alpha - 1) w_j u^{n-1}_j \Delta x \leq 0.
\]
Hence we obtain immediately, for $j \geq 1$,

\[
|u_j^n| \leq \|u^{n-1}\|_\infty + \sup_P \left| g_x(x, Q) \right| \|u^{n-1}\|_\infty \Delta t \
\leq (1 + \omega_3 \Delta t) \|u^{n-1}\|_\infty
\]

This, together with (2.7), clearly implies

\[
\|u^n\|_\infty \leq \max \{ e^{\omega_3 T} \|u^0\|_\infty, \frac{1}{\mu} (C + \omega_1 \omega_2) \} \equiv \omega_4.
\]

Before proving the total variation stability of the scheme, we would need to prove the following results.

**Lemma 2.3.** There exist positive constants $\omega_5$, $\omega_6$ and $\omega_7$ such that

\[
\max_{1 \leq j \leq N} |Q_j^n - Q_{j-1}^n| \leq \omega_5 \Delta x, \quad \max_{1 \leq j \leq N} |g_j^n - g_{j-1}^n| \leq \omega_5 \Delta x, \quad \max_{1 \leq j \leq N} |m_j^n - m_{j-1}^n| \leq \omega_5 \Delta x;
\]

for $1 \leq j \leq N$,

\[
|g_{j+1}^n - 2g_j^n + g_{j-1}^n| \leq \omega_6 \left( \Delta x^2 + \Delta x |u_{j+1}^n - u_j^n| \right),
\]

for $1 \leq j \leq N - 1$, and

\[
\max_{1 \leq j \leq N} |Q_j^{n+1} - Q_j^n| \leq \omega_7 \Delta t \left( 1 + TV(u^n) \right), \quad \max_{1 \leq j \leq N} |g_j^{n+1} - g_j^n| \leq \omega_7 \Delta t \left( 1 + TV(u^n) \right),
\]

\[
\max_{1 \leq j \leq N} |\beta_j^{n+1} - \beta_j^n| \leq \omega_7 \Delta t \left( 1 + TV(u^n) \right),
\]

for $0 \leq j \leq N$.

**Proof.** Clearly, for $1 \leq j \leq N$,

\[
|Q_j^n - Q_{j-1}^n| = \left| \frac{1}{\alpha} \sum_{i=1}^{j} w_i u_i^n + \sum_{i=j+1}^{N} w_i u_i^n - \frac{1}{\alpha} \sum_{i=1}^{j-1} w_i u_i^n - \sum_{i=j}^{N} w_i u_i^n \right| \Delta x \\
= \left| (\alpha - 1) w_j u_j^n \Delta x \leq \|u\|_\infty \|u^n\|_\infty \Delta x \leq \omega_3 \omega_4 \Delta x; \right.
\]
\[ |g^n_j - g^{n-1}_j| = |g(x_j, Q^n_j) - g(x_{j-1}, Q^n_{j-1}) + g(x_{j-1}, Q^n_j) - g(x_{j-1}, Q^n_{j-1})| \]
\[ \leq |g_x(\hat{x}_j, Q^n_j)|\Delta x + |g_Q(x_{j-1}, \hat{Q}_j^n)| |Q^n_j - Q^n_{j-1}| \]
\[ \leq \omega_3 \Delta x + \omega_3(\omega_3 \omega_4 \Delta x) = \omega_3(1 + \omega_3 \omega_4) \Delta x; \]

\[ |m^n_j - m^{n-1}_j| = |m(x_j, Q^n_j) - m(x_{j-1}, Q^n_{j-1})| \]
\[ \leq |m_x(\hat{x}_j, Q^n_j)|\Delta x + |m_Q(x_{j-1}, \hat{Q}_j^n)| |Q^n_j - Q^n_{j-1}| \]
\[ \leq \omega_3 \Delta x + \omega_3(\omega_3 \omega_4 \Delta x) = \omega_3(1 + \omega_3 \omega_4) \Delta x. \]

We have thus proved (2.8) with

\[ \omega_5 = \max(\omega_3 \omega_4, \omega_3(1 + \omega_3 \omega_4)). \]

As to (2.9), we have, for \( 1 \leq j \leq N - 1, \)

\[ |g^n_{j+1} - 2g^n_j + g^n_{j-1}| = |(g^n_{j+1} - g^n_j) - (g^n_j - g^n_{j-1})| \]
\[ = \Delta_+ \left( g_x(\hat{x}_j, Q^n_j)\Delta x + g_Q(x_{j-1}, \hat{Q}_j^n)(Q^n_j - Q^n_{j-1}) \right) \]
\[ \leq |g_x(\hat{x}_{j+1}, Q^n_{j+1}) - g_x(\hat{x}_j, Q^n_j)|\Delta x \]
\[ + |g_Q(x_j, \hat{Q}_j^n)w_{j+1}u^n_{j+1} - g_Q(x_{j-1}, \hat{Q}_j^n)w_ju^n_j|(1 - \alpha)\Delta x \]
\[ = I + II \]

where

\[ I = |g_x(\hat{x}_{j+1}, Q^n_{j+1}) - g_x(\hat{x}_j, Q^n_j)|\Delta x \]
\[ = |g_{xx}(\hat{x}_{j+1}, Q^n_{j+1})(\hat{x}_{j+1} - \hat{x}_j) + g_{xQ}(\hat{x}_j, \hat{Q}_{j+1}^n)(Q^n_{j+1} - Q^n_j)|\Delta x \]
\[ \leq 2\omega_3 \Delta x^2 + \omega_3 \omega_5 \Delta x^2; \]
\[ II = \left| g_Q(x_j, \hat{Q}_{j+1}^n)w_{j+1}u_{j+1}^n - g_Q(x_{j-1}, \hat{Q}_j^n)w_ju_j^n \right| (1 - \alpha) \Delta x \]

\[
= \left| (g_Q(x_j, \hat{Q}_{j+1}^n) - g_Q(x_{j-1}, \hat{Q}_j^n))w_{j+1}u_{j+1}^n + g_Q(x_{j-1}, \hat{Q}_j^n)u_{j+1}^n (w_{j+1} - w_j) \\
+ g_Q(x_{j-1}, \hat{Q}_j^n)w_j(u_{j+1}^n - u_j^n) \right| (1 - \alpha) \Delta x \\
\leq \left| g_{Qx}(\hat{x}_j, \hat{Q}_{j+1}^n)\Delta x + g_{QQ}(x_{j-1}, \hat{Q}_{j+1}^n)(\hat{Q}_{j+1}^n - \hat{Q}_j^n) \right| ||w||_\infty ||u^n||_\infty \Delta x \\
+ \omega_3 ||u^n||_\infty ||w_x||_\infty \Delta x^2 + \omega_3 ||w||_\infty |u_{j+1}^n - u_j^n| \Delta x \\
\leq \omega_2^2 \omega_4 \Delta x^2 + \omega_2^2 \omega_4 \Delta x (2\omega_5 \Delta x) + \omega_3^2 \omega_4 \Delta x^2 + \omega_3^2 |u_{j+1}^n - u_j^n| \Delta x \\
\leq 2\omega_3^2 \omega_4 (1 + \omega_5) \Delta x^2 + \omega_3^2 \Delta x |u_{j+1}^n - u_j^n|.
\]

Hence we have proved (2.9) with

\[ \omega_6 = \max \left( \omega_3(2 + \omega_5 + 2\omega_3\omega_4 + 2\omega_3\omega_4\omega_5), \; \omega_3^2 \right). \]

For (2.10), we have, for \(0 \leq j \leq N\),

\[ |Q_{j+1}^n - Q_j^n| = \left| \alpha \sum_{i=1}^{j} w_iu_i^{n+1} + \sum_{i=j+1}^{N} w_iu_i^n - \alpha \sum_{i=1}^{j} w_iu_i^n - \sum_{i=j+1}^{N} w_iu_i^n \right| \Delta x \\
\leq \sum_{i=1}^{N} |u_i^{n+1} - u_i^n| w_i \Delta x \leq \omega_3 \sum_{i=1}^{N} |u_i^{n+1} - u_i^n| \Delta x. \]

Since, for \(i \geq 1\),

\[ |u_i^{n+1} - u_i^n| = |\lambda (g_i^n u_i^n - g_{i-1}^n u_{i-1}^n) + m_i^n u_i^n \Delta t| = \lambda |g_i^n u_i^n - u_i^n| + \lambda |g_i^n - g_{i-1}^n| u_i^{n-1} + ||m||_\infty ||u^n||_\infty \Delta t \]

\[ \leq \lambda \omega_3 |u_i^n - u_i^{n-1}| + \omega_4 \omega_5 \Delta t + \omega_3 \omega_4 \Delta t, \]

we have, therefore,

\[ |Q_{j+1}^n - Q_j^n| \leq \omega_3 \sum_{i=1}^{N} (\lambda \omega_3 |u_i^n - u_i^{n-1}| + \omega_4 \omega_5 \Delta t + \omega_3 \omega_4 \Delta t) \Delta x = \omega_3^2 TV(u^n) \Delta t + \omega_3 \omega_4 (\omega_3 + \omega_5) L \Delta t. \]
We now also have

\[
|g_j^{n+1} - g_j^n| = |g(x_j, Q_j^{n+1}) - g(x_j, Q_j^n)| \\
= |g_{\bar{Q}}(x_j, \bar{Q}_j)| ||Q_j^{n+1} - Q_j^n| \leq \omega_3 |Q_j^{n+1} - Q_j^n| \\
\leq \omega_3^2 TV(u^n) \Delta t + \omega_3^2 \omega_1 (\omega_3 + \omega_5) L \Delta t,
\]

\[
|\beta_j^{n+1} - \beta_j^n| = |\beta(x_j, Q_j^{n+1}) - \beta(x_j, Q_j^n)| \\
= |\beta_{\bar{Q}}(x_j, \bar{Q}_j)| ||Q_j^{n+1} - Q_j^n| \leq \omega_3 |Q_j^{n+1} - Q_j^n| \\
\leq \omega_3^2 TV(u^n) \Delta t + \omega_3^2 \omega_1 (\omega_3 + \omega_5) L \Delta t.
\]

Therefore we have proved (2.10) with

\[
\omega_7 = \max(\omega_3^2, \omega_3^3, \omega_3 \omega_1 (\omega_3 + \omega_5) L, \omega_3^2 \omega_1 (\omega_3 + \omega_5) L).
\]

\[\blacksquare\]

We are now ready to prove the total variation stability of the scheme.

**Proposition 2.4.** TV \((u^n)\) is bounded for \(t^n \leq T\).

**Proof.** First, we rewrite the scheme (2.1) as

\[
u_j^{n+1} = u_j^n - \lambda g_j^n (u_j^n - u_{j-1}^n) - \lambda (g_j^n - g_{j-1}^n) u_{j-1}^n - \Delta t m_j^n u_j^n.
\]

We then have, for \(1 \leq j \leq N - 1\),

\[
u_j^{n+1} - u_j^n = ((u_j^{n+1} - u_j^n) - \lambda g_{j+1}^n (u_{j+1}^n - u_j^n) + \lambda g_j^n (u_j^n - u_{j-1}^n)) \\
+ (-\lambda u_j^n (g_{j+1}^n - g_j^n) + \lambda u_{j-1}^n (g_j^n - g_{j-1}^n)) + (-\Delta t (m_{j+1}^n u_{j+1}^n - m_j^n u_j^n))
\]

\[
= A_j^n + B_j^n + C_j^n.
\]

Hence

\[
TV(u^{n+1}) = \sum_{j=0}^{N-1} |u_j^{n+1} - u_j^n| \leq \sum_{j=1}^{N-1} |A_j^n| + \sum_{j=1}^{N-1} |B_j^n| + \sum_{j=1}^{N-1} |C_j^n| + |u_1^{n+1} - u_0^n|.
\]

11
We now estimate each term separately. First we have
\[
\sum_{j=1}^{N-1} |A^n_j| \leq \sum_{j=1}^{N-1} \left( (1 - \lambda g^n_{j+1}) |u^n_{j+1} - u^n_j| + \lambda g^n_j |u^n_j - u^n_{j-1}| \right)
\]
\[
= \sum_{j=1}^{N-1} |u^n_{j+1} - u^n_j| + \lambda g^n_j |u^n_j - u^n_0| - \lambda g^n_N |u^n_N - u^n_{N-1}|
\]
\[
= \sum_{j=1}^{N-1} |u^n_{j+1} - u^n_j| + \lambda g^n_j |u^n_j - u^n_0|
\]
where in the first inequality we have used (2.5), and in the last equality we have used the fact \(g(x_N, Q) = 0\). The term \(B^n_j\) can be rearranged as
\[
B^n_j = -\lambda u^n_j (g^n_{j+1} - g^n_j) + \lambda u^n_{j-1} (g^n_j - g^n_{j-1})
\]
\[
= \lambda \left( (g^n_{j+1} - g^n_j)(u^n_j - u^n_{j-1}) - (g^n_{j+1} - 2g^n_j + g^n_{j-1})u^n_j \right).
\]
We can now use Lemma 2.3 to obtain
\[
\sum_{j=1}^{N-1} |B^n_j| \leq \sum_{j=1}^{N-1} \lambda |g^n_{j-1} - g^n_j| |u^n_j - u^n_{j-1}| + \sum_{j=1}^{N-1} \lambda |g^n_{j+1} - 2g^n_j + g^n_{j-1}| |u^n_j|
\]
\[
\leq \omega_5 \Delta t \sum_{j=1}^{N-1} |u^n_j - u^n_{j-1}| + \omega_4 \omega_6 \Delta t \left( \sum_{j=1}^{N-1} \Delta x + \sum_{j=1}^{N-1} |u^n_j - u^n_{j-1}| \right)
\]
\[
\leq \omega_4 \omega_6 L \Delta t + (\omega_5 + \omega_4 \omega_6) \Delta t TV(u^n).
\]

The term \(C^n_j\) can be estimated as
\[
|C^n_j| = \Delta t |m^n_{j+1} u^n_{j+1} - m^n_j u^n_j + m^n_j u^n_{j+1} - m^n_{j+1} u^n_j|
\]
\[
\leq \Delta t \omega_3 \Delta x \omega_4 + \omega_3 |u^n_{j+1} - u^n_j| \Delta t
\]
Hence we have
\[
\sum_{j=1}^{N-1} |C^n_j| \leq \omega_3 \omega_4 L \Delta t + \omega_3 \Delta t TV(u^n).
\]

Let
\[
\omega_8 = \max \left( (\omega_3 + \omega_6) \omega_4 L, \omega_3 + \omega_5 + \omega_4 \omega_6 \right),
\]
we have
\[
TV(u^{n+1}) \leq \omega_8 \Delta t + \omega_8 \Delta t TV(u^n) + \sum_{j=1}^{N-1} |u^n_j - u^n_{j-1}| + \lambda g^n_j |u^n_j - u^n_0| + |u^{n+1}_1 - u^{n+1}_0|.
\]
Next we discuss $|u_1^{n+1} - u_0^{n+1}|$. Using (2.1), (2.2) and Lemma 2.3, we have

$$|u_1^{n+1} - u_0^{n+1}| = |(1 - \lambda g_1^n - \Delta t m^n_1)u_1^n + \lambda g_0^n u_0^n - u_0^{n+1}|$$
$$= |(1 - \lambda g_1^n)(u_1^n - u_0^n) - m^n_1 u_1^n \Delta t - \lambda(g_1^n - g_0^n)u_0^n - (u_0^{n+1} - u_0^n)|$$
$$\leq (1 - \lambda g_1^n)|u_1^n - u_0^n| + \omega_3 \omega_4 \Delta t + \omega_5 \omega_4 \Delta t + |u_0^{n+1} - u_0^n|.$$ 

We then have, with $\omega_9 = \omega_8 + \omega_4(\omega_3 + \omega_5),$

$$TV(u^{n+1}) \leq \omega_9 \Delta t + \omega_9 \Delta t TV(u^n) + TV(u^n) + |u_0^{n+1} - u_0^n|.$$ 

Finally, we must estimate $|u_0^{n+1} - u_0^n|$. From (2.2), we have

$$g_0^{n+1} u_0^{n+1} - g_0^n u_0^n = g_0^{n+1}(u_0^{n+1} - u_0^n) + (g_0^{n+1} - g_0^n) u_0^n$$
$$= C_{n+1} - C_n + \sum_{j=1}^{N}(\beta_{j}^{n+1} u_{j+1}^{n+1} - \beta_{j}^{n} u_{j}^{n}) \Delta x$$
$$= C_{n+1} - C_n + \sum_{j=1}^{N}(\beta_{j}^{n+1} (u_{j+1}^{n+1} - u_j^n) + (\beta_{j}^{n+1} - \beta_{j}^{n}) u_j^n) \Delta x.$$ 

Noticing that $g_0^{n+1} \geq \mu > 0$ by (2.6), we have, by Lemma 2.3 and (2.11),

$$|u_0^{n+1} - u_0^n| \leq \frac{1}{\mu} |g_0^{n+1} - g_0^n| u_0^n + \frac{1}{\mu} |C_{n+1} - C_n| + \frac{\omega_3}{\mu} \sum_{j=1}^{N} |u_{j+1}^{n+1} - u_j^n| \Delta x + \frac{\omega_4}{\mu} \sum_{j=1}^{n} |\beta_{j}^{n+1} - \beta_{j}^{n}| \Delta x$$
$$\leq \frac{\omega_4}{\mu} (\omega_7 TV(u^n) \Delta t + \omega_7 \Delta t) + \frac{\omega_3}{\mu} \Delta t + \frac{\omega_3}{\mu} \sum_{j=1}^{N} (\lambda \omega_3 |u_{j}^{n+1} - u_{j-1}^{n+1}| + \omega_4 (\omega_3 + \omega_5) \Delta t) \Delta x$$
$$+ \frac{\omega_4}{\mu} L (\omega_7 TV(u^n) \Delta t + \omega_7 \Delta t)$$
$$\leq \omega_1 \omega_1 TV(u^n) \Delta t + \omega_1 \Delta t,$$

with

$$\omega_1 = \max \left( \frac{\omega_4 \omega_7}{\mu} (1 + L) + \frac{\omega_3}{\mu} (1 + L) + \frac{\omega_3 \omega_4}{\mu} (\omega_3 + \omega_5) L \right).$$

Now, with $\omega_{11} = \omega_9 + \omega_1$, we have

$$TV(u^{n+1}) \leq (1 + \omega_{11} \Delta t) TV(u^n) + \omega_{11} \Delta t$$

which implies the boundedness of $TV(u^n)$ for $t^n \leq T$. ■

13
Next, we show the Lipschitz stability in $t$.

**Proposition 2.5.** There exists a positive constant $M$ such that for any $q > p$, we have

$$\sum_{j=1}^{N} \left| \frac{u_j^q - u_j^p}{\Delta t} \right| \Delta x \leq M(q - p).$$

**Proof.** Using (2.1) and Lemma 2.3, we have

$$\sum_{j=1}^{N} \left| \frac{u_j^{n+1} - u_j^n}{\Delta t} \right| \Delta x = \sum_{j=1}^{N} \left| D^{-}(g_j^n u_j^n) + m_j^n u_j^n \right| \Delta x$$

$$= \sum_{j=1}^{N} \left| \left( \frac{g_j^n - g_j^{n-1}}{\Delta x} + m_j^n \right) u_j^n + g_j^{n-1} D^{-}(u_j^n) \right| \Delta x$$

$$\leq \omega_4 \omega_5 L + \omega_3 \omega_4 L + \omega_3 TV(u^n) \leq M.$$ 

Thus,

$$\sum_{j=1}^{N} \left| \frac{u_j^q - u_j^p}{\Delta t} \right| \Delta x \leq \sum_{n=p}^{\infty} \sum_{j=1}^{N} \left| \frac{u_j^{n+1} - u_j^n}{\Delta t} \right| \Delta x \leq M(q - p).$$

\[\blacksquare\]

Following [17] we can define a family of functions $\{U_{\Delta x, \Delta t}\}$ by

$$U_{\Delta x, \Delta t}(x, t) = u_j^n$$

for $x \in [x_{j-1}, x_j)$, $t \in [t^{n-1}, t^n)$, $j = 1, \cdots, N$ and $n = 1, \cdots, l$. Then, the set of functions $\{U_{\Delta x, \Delta t}\}$ is compact in the topology of $L^1((0, L) \times (0, T))$, and we have the following result of convergence.

**Proposition 2.6.** There exists a subsequence $\{U_{\Delta x_{i_1}, \Delta t_{i_1}}\} \subset \{U_{\Delta x, \Delta t}\}$ which converges to a $BV([0, L] \times [0, T])$ function $u(x, t)$ in the sense that

$$\int_0^L |U_{\Delta x_{i_1}, \Delta t_{i_1}}(x, 0) - u^0(x)| dx \to 0$$

and

$$\int_0^T \int_0^L |U_{\Delta x_{i_1}, \Delta t_{i_1}}(x, t) - u(x, t)| dx dt \to 0.$$
as \( i \to \infty \). Furthermore, the function \( u \), satisfying

\[
\| u \|_{BV([0,L] \times [0,T])} \leq E(\| u^0 \|_{BV([0,L])}, \| C \|_{C^1([0,T])}),
\]

is the unique \( BV([0,L] \times [0,T]) \) solution \( u(x,t) \) for (1.1), and the numerical solution \( \{U_{\Delta x, \Delta t}\} \) converges to it when \( \Delta x \to 0 \).

**Proof.** The convergence of a subsequence to a \( BV \) function \( u(x,t) \) and the fact that \( u(x,t) \) is a \( BV \) weak solution of (1.1) follow from Propositions 2.1, 2.2, 2.4, 2.5 and [17]. The uniqueness of bounded variation weak solutions of (1.1) is proved in [1]. Using this uniqueness we easily deduce the convergence of the numerical solution \( \{U_{\Delta x, \Delta t}\} \) towards \( u(x,t) \) when \( \Delta x \to 0 \). \( \blacksquare \)

### 3 A second order high resolution finite difference scheme

The first order scheme defined in the previous section is very diffusive and would need many grid points to achieve acceptable resolution. In this section we develop and analyze a second order high resolution finite difference scheme for (1.1), following the minmod based MUSCL schemes [7, 13]. We remark, however, that the analysis is significantly more complicated because of the global constraints in (1.1). We note that our scheme can be easily generalized to the more accurate generalized MUSCL type scheme similar to the one in [14] and the total variation bounded modified minmod based scheme in [15] without affecting the analysis. The second order high resolution finite difference scheme that we consider in this section is defined by

\[
\frac{u_j^{n+1} - u_j^n}{\Delta t} + \frac{\hat{f}^n_{j+1/2} - \hat{f}^n_{j-1/2}}{\Delta x} + m^n_j u_j^n = 0, \quad 1 \leq j \leq N
\]

(3.1)

where the numerical flux \( \hat{f}^n_{j+1/2} \) is defined by

\[
\hat{f}^n_{j+1/2} = \left\{ \begin{array}{ll}
g^+_j u_j^n & \text{if } j = 2, \ldots, N - 2 \\
g^-_j u_j^n & \text{if } j = 0, 1, N - 1, N
\end{array} \right.
\]

where the minmod function \( m \) is defined by [7]

\[
m(a, b) = \frac{\text{sign}(a) + \text{sign}(b)}{2} \min(|a|, |b|).
\]

(3.2)
Clearly, this scheme is second order accurate except at the boundary, where it is first order accurate. This guarantees second order accuracy in the global $L^1$ norm. The global boundary condition at the left is implemented by a second order composite trapezoid rule

\[ g_0^n u_0^n = C^n + \sum_{j=0}^{N} \beta_j^n u_j^n \Delta x, \quad (3.3) \]

where the special summation notation is defined by

\[ \sum_{j=j_1}^{j_2} \alpha_j = \frac{1}{2} a_{j_1} + \frac{1}{2} a_{j_2} + \sum_{j=j_1+1}^{j_2-1} a_j \]

if $j_2 - j_1 \geq 1$, and of course

\[ \sum_{j=j_1}^{j_2} \alpha_j = 0 \quad \text{if} \quad j_2 \leq j_1. \]

The environment is computed also by a second order composite trapezoid rule, except for the integral over the first interval which is computed by the right-ended rectangular rule to avoid using $u_0^n$. That is,

\[ Q_0^n = \omega_1 u_1^n \Delta x + \sum_{i=1}^{N} \alpha_i u_i^n \Delta x, \quad Q_1^n = \alpha_\omega_1 u_1^n \Delta x + \sum_{i=1}^{N} \alpha_i u_i^n \Delta x, \]

\[ Q_j^n = \alpha \omega_1 u_1^n \Delta x + \alpha \sum_{i=1}^{j} \alpha_i u_i^n \Delta x + \sum_{i=j}^{N} \alpha_i u_i^n \Delta x, \quad 2 \leq j \leq N. \quad (3.4) \]

Notice that this approximation to $Q_j^n$ is second order accurate. The initial condition is still taken as

\[ u_j^0 = u^0(x_j), \quad j = 1, 2, \ldots, N. \]

Still using the notation $\lambda = \frac{\Delta t}{\Delta x}$, we can write the scheme (3.1) as

\[ u_{j+1}^n = u_j^n - \lambda (\hat{f}_{j+1/2}^n - \hat{f}_{j-1/2}^n) - \Delta \tau u_j^n, \quad j \geq 1. \quad (3.5) \]

We denote

\[ A_j^n = \begin{cases} \frac{1}{2} \left( g_{j+1}^n + g_j^n + g_j^n \frac{m(\Delta u_j^n, \Delta u_{j+1}^n)}{\Delta u_j^n} - g_{j-1}^n \frac{m(\Delta u_j^n, \Delta u_{j-1}^n)}{\Delta u_j^n} \right) & : j = 3, \ldots, N - 2 \\ \frac{1}{2} \left( g_{j+1}^n + g_j^n + g_j^n \frac{m(\Delta u_j^n, \Delta u_{j+1}^n)}{\Delta u_j^n} \right) & : j = 2 \\ \frac{1}{2} \left( 2g_j^n - g_{j-1}^n \frac{m(\Delta u_j^n, \Delta u_{j-1}^n)}{\Delta u_j^n} \right) & : j = N - 1 \\ g_j^n & : j = 1, N \end{cases} \]
\begin{equation*}
B_j^n = \begin{cases}
\frac{1}{2}(\Delta_+ g_j^n + \Delta_0 g_j^n) & : j = 3, \ldots, N - 2 \\
\frac{1}{2}\Delta_+ g_j^n & : j = 2 \\
\frac{1}{2}\Delta_0 g_j^n & : j = N - 1 \\
\Delta_0 g_j^n & : j = 1, N
\end{cases}
\end{equation*}

and rewrite the scheme (3.5) as

\begin{equation*}
u_j^{n+1} = (1 - \lambda A_j^n - m_j^n \Delta t)u_j^n + \lambda (A_j^n - B_j^n)u_{j-1}^n, \quad j \geq 1. \tag{3.6}
\end{equation*}

As before, we first prove the \(L^1\) boundedness of the numerical solution \(u^n\) for \(t^n \leq T\), under the assumption that \(u^n_j \geq 0\). We will prove the validity of this assumption later.

**Proposition 3.1.** If \(u^n_j \geq 0\), then \(\|u^n\|_1\) is bounded when \(t^n \leq T\).

**Proof.** As before, since \(u^n_j \geq 0\), \(m_j^n \geq 0\) and \(g_N^n = 0\), we have

\[
\frac{\|u^{n+1}\|_1 - \|u^n\|_1}{\Delta t} = \sum_{j=1}^{N} \frac{u_{j}^{n+1} - u_{j}^{n}}{\Delta t} \Delta x \\
= - \sum_{j=1}^{N} (\hat{j}_{j+1/2}^n - \hat{j}_{j-1/2}^n) - \sum_{j=1}^{N} m_j^n u_j^n \Delta x \\
\leq - \sum_{j=1}^{N} (\hat{j}_{j+1/2}^n - \hat{j}_{j-1/2}^n) \\
= g_0^n u_0^n \\
= C^n + \frac{1}{2} \beta_0^n u_0^n \Delta x + \frac{1}{2} \beta_N^n u_N^n \Delta x + \sum_{j=1}^{N-1} \beta_j^n u_j^n \Delta x \leq C^n + \frac{1}{2} \beta_0^n u_0^n \Delta x + \omega_1 \|u^n\|_1.
\]

We now assume, for the time being, that \(u_k^0 \leq \theta\) where \(\theta\) is a constant. This assumption will be justified later. If \(\Delta x \leq 2C/\omega_1 \theta\), where again \(C\) denotes the upperbound of \(C(t)\) for \(t \in \[0, T]\) and \(\omega_1\) is the upperbound of \(\beta(x, Q)\), then we have \(\frac{1}{2} \beta_0^n u_0^n \Delta x \leq C\). For constant \(\Delta t\), we then immediately have

\[
\|u^n\|_1 \leq (1 + \omega_1 \Delta t)\|u^{n-1}\|_1 + 2C \Delta t \\
\leq (1 + \omega_1 \Delta t)^n\|u^0\|_1 + \sum_{j=0}^{n-1} (1 + \omega_1 \Delta t)^j 2C \Delta t \\
\leq e^{\omega_1 T}\|u^0\|_1 + \frac{2C e^{\omega_1 T}}{\omega_1} \equiv M_2
\]
where the constant $M_2$, as well as a sequence of such constants $M_k$ to be defined later, depend only on the given functions $g$, $m$, $C$, $\beta$, $w$, the final time $T$ and the initial condition $u^0$. This proof is clearly also valid, with a minor modification, for the situation when $\Delta t$ is not a constant.

We now look at the bound of $Q^n_j$. For $0 \leq j \leq N$, we have, by the definition of $Q^n_j$ in (3.4), that

$$Q^n_j \leq \omega_1 u^n_1 \Delta x + \sum_{i=1}^N w_i u^n_i \Delta x \leq \frac{3}{2} \omega_3 M_2 \equiv Q_{\max},$$

and therefore

$$g^n_0 \geq \min_{0 \leq Q \leq Q_{\max}} g(0, Q) \equiv \mu > 0. \quad (3.7)$$

Thus if $\Delta x \leq \mu / \omega_1$, we have $g^n_0 - \frac{1}{2} \omega_0 \Delta x \geq \mu / 2$, hence from (3.3) we deduce

$$u^n_0 \leq \frac{2}{\mu} (\omega_1 M_2 + C). \quad (3.8)$$

The constants on the right hand side of the inequality above does not depend on $\theta$, hence the assumption on the boundedness of $u^n_0$ is justified. ■

As before, we now have a bounded closed domain $\mathcal{D} = \{(x, Q) \in [0, L] \times [0, Q_{\max}]\}$ that $x$ and $Q$ reside in, hence by the smoothness assumptions of $g$, $m$, $\beta$ and $w$, we have a fixed constant $M_3$ such that

$$\sup_{\mathcal{D}} |f(x, Q)| \leq M_3, \quad \sup_{0 \leq x \leq L} |h(x)| \leq M_3, \quad \sup_{0 \leq t \leq T} |\eta(t)| \leq M_3$$

for

$$f(x, Q) = g(x, Q), g_x(x, Q), g_Q(x, Q), g_{xx}(x, Q), g_{xQ}(x, Q), g_{QQ}(x, Q),$$

$$m(x, Q), m_x(x, Q), m_Q(x, Q), \beta(x, Q),$$

$$h(x) = w(x), w'(x), \quad \eta(t) = C(t), C'(t).$$

It can be easily shown that

$$|A^n_j| \leq \frac{3}{2} \max_{\mathcal{D}} |g(x, Q)| \leq \frac{3}{2} M_3,$$
thus when $\Delta t \leq \Delta t_0 \equiv \frac{1}{2M_3}$ and $\lambda \leq \lambda_0 \equiv \frac{1}{3M_3}$, we have

$$1 - \lambda A^n_j - \Delta tm^n_j \geq 0, \quad 1 \leq j \leq N. \quad (3.9)$$

Noticing that

$$2(A^n_j - B^n_j) =
\begin{cases}
g^n_j \left( 1 + \frac{m(\Delta_+ u^n_{j+1}, \Delta_- u^n_j)}{\Delta_- u^n_j} \right) + g^n_{j-1} \left( 1 - \frac{m(\Delta_+ u^n_{j-1}, \Delta_- u^n_j)}{\Delta_- u^n_j} \right) & : \ j = 3, \cdots, N - 2, \\
g^n_j \left( 2 + \frac{m(\Delta_+ u^n_{j+1}, \Delta_- u^n_j)}{\Delta_- u^n_j} \right) & : \ j = 2, \\
g^n_j + g^n_{j-1} \left( 1 - \frac{m(\Delta_+ u^n_{j-1}, \Delta_- u^n_j)}{\Delta_- u^n_j} \right) & : \ j = N - 1, \\
2g^n_{j-1} & : \ j = 1, N.
\end{cases}$$

which implies, by the definition of the minmod function (3.2), that

$$A^n_j - B^n_j \geq 0, \quad 1 \leq j \leq N. \quad (3.10)$$

This, together with (3.9), clearly implies $u^n_j \geq 0$ by (3.6). Notice that we can either choose $\lambda = \lambda_0$ as a constant or a variable depending on the time level $t^n$. We have thus verified the assumption made in Proposition 3.1 about the non-negativity of $u^n_j$.

Next we will prove the $L^\infty$ boundedness of the numerical solution.

**Proposition 3.2.** $||u^n||_\infty$ is bounded for $t^n \leq T$.

**Proof.** First, we have already shown the boundedness of $u^n_0$ in (3.8). As for $j \geq 1$, we use (3.6), (3.9), (3.10) and the non-negativity of $m$ to obtain

$$|u^n_j| \leq (1 - \lambda A^n_{j-1} - \Delta tm^n_{j-1}) ||u^{n-1}||_\infty + \lambda (A^n_{j-1} - B^n_{j-1}) ||u^{n-1}||_\infty$$

$$\leq ||u^{n-1}||_\infty - \lambda B^n_{j-1} ||u^{n-1}||_\infty.$$ 

We can easily verify that, for $2 \leq j \leq N$,

$$Q_j - Q_{j-1} = \frac{1}{2}(\alpha - 1)(w_j u_j + w_{j-1} u_{j-1}) \Delta x. \quad (3.11)$$

For $j = 1$, we have a similar formula

$$Q_1 - Q_0 = (\alpha - 1) \omega_1 u_1. \quad (3.12)$$
Therefore, we have
\[
g_j^{n-1} - g_j^{n-1} = g(x_j, Q_j^{n-1}) - g(x_{j-1}, Q_j^{n-1}) + g(x_{j-1}, Q_{j-1}^{n-1}) - g(x_{j-1}, Q_{j-1}^{n-1})
\]
\[
= g_x(\dot{x}_j, Q_j^{n-1}) \Delta x + g_Q(x_{j-1}, Q_j^{n-1})(Q_j^{n-1} - Q_{j-1}^{n-1})
\]

By assumption, \(\alpha < 1\), \(g_Q(x, Q) \leq 0\). We clearly have, for \(2 \leq j \leq N\),
\[
-g_Q(x_{j-1}, \hat{Q}^{n-1}_j)(Q_j^{n-1} - Q_{j-1}^{n-1}) = -\frac{1}{2}g_Q(x_{j-1}, \hat{Q}^{n-1}_j)(\alpha - 1)(w_j u_j^{n-1} + w_{j-1} u_{j-1}^{n-1}) \Delta x \leq 0,
\]
also
\[
-g_Q(x_0, \hat{Q}^{n-1}_1)(Q_1^{n-1} - Q_0^{n-1}) = -g_Q(x_0, \hat{Q}^{n-1}_1)(\alpha - 1)\omega_1 u_1^{n-1} \Delta x \leq 0.
\]

Now, noticing that \(B_j \leq \max_i (g_i - g_i)\) for \(j \geq 1\), we obtain immediately, for \(j \geq 1\),
\[
|u_j^n| \leq ||u^{n-1}||_\infty + \sup_{\gamma} |g_x(x, Q)| ||u^{n-1}||_\infty \Delta t \leq (1 + M_3 \Delta t) ||u^{n-1}||_\infty.
\]

This, together with (3.8), clearly implies
\[
||u^n||_\infty \leq \max \{e^{M_3 T} ||u^0||_\infty, \frac{2}{\mu} (\omega_1 M_2 + C)\} \equiv M_4.
\]

\[
\]

Before proving the total variation stability of the scheme, we would need to prove the following results.

**Lemma 3.3.** There exist positive constants \(M_5\), \(M_6\) and \(M_7\) such that
\[
\max_{1 \leq j \leq N} |Q_j^n - Q_{j-1}^n| \leq M_5 \Delta x, \quad \max_{1 \leq j \leq N} |g_j^n - g_{j-1}^n| \leq M_5 \Delta x, \quad \max_{1 \leq j \leq N} |m_j^n - m_{j-1}^n| \leq M_5 \Delta x,
\]
(3.13)
for \(1 \leq j \leq N\);
\[
|g_{j+1}^n - 2g_j^n + g_{j-1}^n| \leq M_6 \Delta x \left(\Delta x + |u_j^n - u_{j-1}^n| + |u_{j+1}^n - u_j^n|\right), \quad 1 \leq j \leq N - 1 (3.14)
\]
\[
|B_j^n - B_{j-1}^n| \leq M_6 \Delta x \left(\Delta x + |u_j^n - u_{j-1}^n| + |u_{j+1}^n - u_j^n|\right), \quad 4 \leq j \leq N - 2;
\]
and
\[
|Q_j^{n+1} - Q_j^n| \leq M_7 TV(u^n) \Delta t + M_7 \Delta t, \quad |g_j^{n+1} - g_j^n| \leq M_7 TV(u^n) \Delta t + M_7 \Delta t, \quad (3.15)
\]
\[ |\beta_j^{n+1} - \beta_j^n| \leq M_f TV(u^n) \Delta t + M_f \Delta t, \]

for \(0 \leq j \leq N\).

**Proof.** By (3.11), we have, for \(2 \leq j \leq N\),

\[ |Q_j^n - Q_{j-1}^n| = \left| \frac{1}{2}(\alpha - 1)(w_j u_j^n + w_{j-1} u_{j-1}^n) \right| \Delta x \leq \|w\|_\infty \|u^n\|_\infty \Delta x \leq M_3 M_4 \Delta x, \]

which is clearly also valid for \(j = 0\) by (3.12). Therefore,

\[
\begin{align*}
|g_j^n - g_{j-1}^n| &= |g(x_j, Q_j^n) - g(x_{j-1}, Q_{j-1}^n)| \\
&\leq \|g\|_\infty \Delta x + \|g\|_\infty \Delta x
\end{align*}
\]

\[
\leq M_3 \Delta x + M_3 (M_3 M_4 \Delta x) = M_3 (1 + M_3 M_4) \Delta x
\]

\[
|m_j^n - m_{j-1}^n| = |m(x_j, Q_j^n) - m(x_{j-1}, Q_{j-1}^n)| \\
\leq \|m\|_\infty \Delta x + \|m\|_\infty \Delta x = M_3 \Delta x + M_3 (M_3 M_4 \Delta x) = M_3 (1 + M_3 M_4) \Delta x
\]

We have thus proved (3.13) with

\[ M_5 = \max (M_3 M_4, M_3 (1 + M_3 M_4)). \]

As to (3.14), for \(4 \leq j \leq N - 2\), we can easily verify

\[
|B_j^n - B_{j-1}^n| = \left| \frac{1}{2}(g_{j+1}^n - 2g_j^n + g_{j-1}^n) + \frac{1}{2}(g_j^n - 2g_{j-1}^n + g_{j-2}^n) \right| \\
\leq \max_i |g_{i+1}^n - 2g_i^n + g_{i-1}^n|
\]

hence we only need to prove the first inequality in (3.14), the proof of which is similar to that for (2.9) in Lemma 2.3, using (3.11), with

\[ M_6 = \max \left( 2M_3^2 M_4 (1 + M_5), \frac{1}{2} M_3^2 \right). \]

For (3.15), \(0 \leq j \leq N\), we have, by the definition of \(Q_j\) in (3.4), that

\[ |Q_j^{n+1} - Q_j^n| \leq \frac{3}{2} \sum_{i=1}^N |u_i^{n+1} - u_i^n| |w_i \Delta x \leq \frac{3}{2} M_3 \sum_{i=1}^N |u_i^{n+1} - u_i^n| \Delta x. \]

21
From (3.6) and the definition of $A^n_i$ and $B^n_i$, we have, for $1 \leq i \leq N$,

\[
|u^n_{i+1} - u^n_i| = |-\lambda A^n_i u^n_i + \lambda(A^n_i - B^n_i)u^n_{i-1} - m^n_i u^n_i \Delta t|
\leq \lambda |A^n_i| |u^n_i - u^n_{i-1}| + \lambda |B^n_i| |u^n_{i-1}| + ||m^n||_\infty \|u^n\|_\infty \Delta t
\leq 2\lambda \sup_P \{ g(x, Q) \} |u^n_i - u^n_{i-1}| + \lambda M_4 \max_k |g^n_k - g^n_{k-1}| + M_3 M_4 \Delta t
\leq 2\lambda M_3 |u^n_i - u^n_{i-1}| + M_4 M_5 \Delta t + M_3 M_4 \Delta t.
\]

Thus we have

\[
|Q^{n+1}_j - Q^n_j| \leq \frac{3}{2} M_3 \sum_{i=1}^N (2\lambda M_3 |u^n_i - u^n_{i-1}| + M_4 M_5 \Delta t + M_3 M_4 \Delta t) \Delta x
= 3 M_3^2 TV(u^n) \Delta t + \frac{3}{2} M_3 M_4 (M_3 + M_5) L \Delta t,
\]

which implies

\[
|g^{n+1}_j - g^n_j| = |g(x_j, Q^{n+1}_j) - g(x_j, Q^n_j)|
= |g_Q(x_j, \bar{Q}_j)| |Q^{n+1}_j - Q^n_j| \leq M_3 |Q^{n+1}_j - Q^n_j|
\leq 3 M_3^2 TV(u^n) \Delta t + \frac{3}{2} M_3^2 M_4 (M_3 + M_5) L \Delta t,
\]

\[
|\beta^{n+1}_j - \beta^n_j| = |\beta(x_j, Q^{n+1}_j) - \beta(x_j, Q^n_j)|
= |\beta_Q(x_j, \bar{Q}_j)| |Q^{n+1}_j - Q^n_j| \leq M_3 |Q^{n+1}_j - Q^n_j|
\leq 3 M_3^2 TV(u^n) \Delta t + \frac{3}{2} M_3^2 M_4 (M_3 + M_5) L \Delta t.
\]

We have thus proved (3.15) with

\[
M_7 = \max \left(3 M_3^2, 3 M_3^3, \frac{3}{2} M_3 M_4 (M_3 + M_5) L, \frac{3}{2} M_3^2 M_4 (M_3 + M_5) L \right).
\]

We are now ready to prove the total variation stability of the scheme.

**Proposition 3.4.** $TV(u^n)$ is bounded for $t^n \leq T$. 

22
Proof. First, we rewrite the scheme (3.6) as
\[
u_{j+1}^{n+1} = \nu_j^n - \lambda A_j^n (\nu_j^n - \nu_{j-1}^n) - \lambda B_j^n \nu_{j-1}^n - \Delta t m_j^n \nu_j^n, \quad j \geq 1.
\]
We then have
\[
u_{j+1}^{n+1} - \nu_j^{n+1} = \left[ (1 - \lambda A_{j+1}^n) (\nu_{j+1}^n - \nu_j^n) + \lambda (A_j^n - B_{j}^n) (\nu_j^n - \nu_{j-1}^n) \right]
+ [-\lambda \nu_j^n (B_{j+1}^n - B_j^n)] + [-\Delta t (m_{j+1}^n \nu_{j+1}^n - m_j^n \nu_j^n)]
= D_j^n + E_j^n + F_j^n, \quad j = 1, 2, \ldots, N - 1.
\]
Hence
\[
TV(u^{n+1}) = \sum_{j=0}^{N-1} |u_{j+1}^{n+1} - u_j^{n+1}| \leq \sum_{j=1}^{N-1} |D_j^n| + \sum_{j=1}^{N-1} |E_j^n| + \sum_{j=1}^{N-1} |F_j^n| + |u_1^{n+1} - u_0^{n+1}|.
\]
We now estimate each term separately. First we have
\[
\sum_{j=1}^{N-1} |D_j^n| \leq \sum_{j=1}^{N-1} (1 - \lambda A_{j+1}^n) |u_{j+1}^n - u_j^n| + \lambda (A_j^n - B_j^n) |u_j^n - u_{j-1}^n|
= \sum_{j=1}^{N-1} |u_{j+1}^n - u_j^n| - \lambda \sum_{j=1}^{N-1} B_j^n |u_j^n - u_{j-1}^n| + \lambda g_1^n |u_1^n - u_0^n| - \lambda g_N^n |u_N^n - u_{N-1}^n|
\leq \sum_{j=1}^{N-1} |u_{j+1}^n - u_j^n| + M_5 \Delta t TV(u^n) + \lambda g_1^n |u_1^n - u_0^n|
\]
where in the first inequality we have used (3.9) and (3.10), and in the last inequality we have used Lemma 3.3, the fact that
\[
|B_j| \leq \max_i |g_i - g_{i-1}|, \quad 1 \leq j \leq N, \quad (3.17)
\]
and the fact that \(g(x_N, Q) = 0\). Using again Lemma 3.3 and (3.17), we have
\[
\sum_{j=1}^{N-1} |E_j^n| \leq \sum_{j=1}^{N-2} \lambda |B_{j+1}^n - B_j^n| |u_j^n| + \sum_{j=1,2,3,N-1} \lambda |B_{j+1}^n - B_j^n| |u_j^n|
\leq M_1 M_6 \Delta t \left( \sum_{j=4}^{N-2} \Delta x + \sum_{j=4}^{N-2} (|u_{j+2}^n - u_{j+1}^n| + |u_{j+1}^n - u_j^n|) \right) + 8 \lambda M_4 \|B^n\|_\infty
\leq M_1 M_6 L \Delta t + 2 M_1 M_6 \Delta t TV(u^n) + 8 M_4 M_5 \Delta t.
\]
The term $F_j^n$ can be estimated as
\[
|F_j^n| = \Delta t|m_{j+1}^n u_{j+1}^n - m_j^n u_j^{n+1} + m_j^n u_j^{n+1} - m_j^n u_j^n| \\
\leq M_3 M_4 \Delta t \Delta x + M_3 |u_{j+1}^n - u_j^n| \Delta t
\]
Hence we have
\[
\sum_{j=1}^{N-1} |F_j^n| \leq M_3 M_4 L \Delta t + M_3 \Delta t TV(u^n).
\]
Let
\[
M_8 = \max \{ M_4 ((M_3 + M_6) L + 8 M_5), M_3 + M_5 + 2 M_4 M_6 \},
\]
we have
\[
TV(u^{n+1}) \leq M_8 \Delta t + M_8 \Delta t TV(u^n) + \sum_{j=1}^{N-1} |u_j^n - u_{j-1}^n| + \lambda g_j^n |u_1^n - u_0^n| + |u_1^{n+1} - u_0^{n+1}|.
\]
Next we discuss $|u_1^{n+1} - u_0^{n+1}|$. This boundary term has the same form as in the first order case
\[
|u_1^{n+1} - u_0^{n+1}| = |(1 - \lambda g_1^n - \Delta t m_1^n)u_1^n + \lambda g_0^n u_0^n - u_0^{n+1}| \\
= |(1 - \lambda g_1^n)(u_1^n - u_0^n) - m_1^n u_1^n \Delta t - \lambda (g_1^n - g_0^n)u_0^n - (u_0^{n+1} - u_0^n)| \\
\leq (1 - \lambda g_1^n)|u_1^n - u_0^n| + M_3 M_4 \Delta t + M_5 M_4 \Delta t + |u_0^{n+1} - u_0^n|.
\]
We then have
\[
TV(u^{n+1}) \leq M_9 \Delta t + M_9 \Delta t TV(u^n) + TV(u^n) + |u_0^{n+1} - u_0^n|
\]
where
\[
M_9 = M_8 + M_4 (M_3 + M_5).
\]
Finally, we must estimate $|u_0^{n+1} - u_0^n|$. From (3.3), we have
\[
g_0^{n+1} u_0^{n+1} - g_0^n u_0^n = g_0^{n+1} (u_0^{n+1} - u_0^n) + (g_0^{n+1} - g_0^n) u_0^n \\
= C^{n+1} - C^n + \sum_{j=0}^{N} \left( \beta_j^{n+1} u_j^{n+1} - \beta_j^n u_j^n \right) \Delta x \\
= C^{n+1} - C^n + \sum_{j=0}^{N} \left( \beta_j^{n+1} (u_j^{n+1} - u_j^n) + (\beta_j^{n+1} - \beta_j^n) u_j^n \right) \Delta x.
\]
24
Rearranging terms and using (3.16) and the results of Lemma 3.2, we obtain

\[
\left| \left( g_0^{n+1} - \frac{1}{2} \beta_0^{n+1} \Delta x \right) (u_0^{n+1} - u_0^n) \right|
\leq |C^{n+1} - C^n| + |g_0^{n+1} - g_0^n| u_0^n + \frac{1}{2} u_0^n \Delta x |\beta_0^{n+1} - \beta_0^n| + \sum_{j=1}^N (|\beta_j^{n+1} - \beta_j^n| u_j^n + |\beta_j^{n+1} - \beta_j^n| u_j^n) \Delta x
\leq M_3 \Delta t + M_4 M_7 \Delta t (1 + TV(u^n)) + \frac{1}{2} M_1 M_7 \Delta t (1 + TV(u^n))
\]

\[
+ M_3 \sum_{j=1}^N \left( 2 \lambda M_3 |u_j^n - u_{j-1}^n| + M_4 (M_3 + M_5) \Delta t \right) \Delta x + M_4 M_7 L \Delta t (1 + TV(u^n))
\leq \left( M_4 M_7 (2 + L) + 2 M_3^2 \right) \Delta t TV(u^n) + (M_3 + M_4 M_7 (2 + L) + M_3 M_4 (M_3 + M_5) L) \Delta t
\]

where in the last inequality we have assumed \( \Delta x \leq 2 \). Noticing that, by (3.7), \( g_0^{n+1} \geq \mu > 0 \).

Hence if \( \Delta x \leq \frac{\mu}{M_0} \), we have \( g_0^{n+1} - \frac{1}{2} \beta_0^{n+1} \Delta x \geq \frac{\mu}{2} > 0 \). Hence

\[
|u_0^{n+1} - u_0^n| \leq M_{10} TV(u^n) \Delta t + M_{10} \Delta t
\]

with

\[
M_{10} = \frac{1}{\mu} \max \left( M_4 M_7 (2 + L) + 2 M_3^2, M_3 + M_4 M_7 (2 + L) + M_3 M_4 (M_3 + M_5) L \right).
\]

Now, with \( M_{11} = M_9 + M_{10} \), we have

\[
TV(u^{n+1}) \leq (1 + M_{11} \Delta t) TV(u^n) + M_{11} \Delta t
\]

which implies the boundedness of \( TV(u^n) \).  

Next, we show the Lipschitz stability in \( t \).

**Proposition 3.5.** There exists a positive constant \( M \) such that for any \( q > p \), we have

\[
\sum_{j=1}^N \left| \frac{u_j^q - u_j^p}{\Delta t} \right| \Delta x \leq M(q - p).
\]

**Proof.** Using (3.6), (3.17) and the definition of \( A_j^n \) and \( B_j^n \), we obtain

\[
\sum_{j=1}^N \left| \frac{u_j^{n+1} - u_j^n}{\Delta t} \right| \Delta x = \sum_{j=1}^N \left| \left( B_j^n \Delta x + m_j^n \right) u_j^n + (A_j^n - B_j^n) D^-(u_j^n) \right| \Delta x
\leq \sum_{j=1}^N \max_i |g_i^n - g_i^{n-1}| u_j^n + M_3 \sum_{j=1}^N u_j^n \Delta x + 3 M_3 \sum_{j=1}^N |u_j^n - u_{j-1}^n|
\leq M_4 M_5 L + M_3 M_4 L + 3 M_3 TV(u^n) \leq M.
\]
Thus,
\[
\sum_{j=1}^{N} \left| \frac{u_j^q - u_j^p}{\Delta t} \right| \Delta x \leq \sum_{n=p}^{q} \sum_{j=1}^{N} \left| \frac{u_j^{n+1} - u_j^n}{\Delta t} \right| \Delta x \leq M(q-p).
\]

If we again define a family of functions \( \{ U_{\Delta x, \Delta t} \} \) by
\[
U_{\Delta x, \Delta t}(x, t) = u_j^n
\]
for \( x \in [x_{j-1}, x_j) \), \( t \in [t^{n-1}, t^n) \), \( j = 1, \ldots, N \) and \( n = 1, \ldots, l \), then we have the following proposition. The proof is the same as that for Proposition 2.6.

**Proposition 3.6.** The numerical solution \( \{ U_{\Delta x, \Delta t} \} \) converges to the unique \( BV([0, L] \times [0, T]) \) solution \( u(x, t) \) for (1.1) when \( \Delta x \to 0 \).

Finally, we remark that the scheme (3.1) is second order in space but only first order in time. We should use the following second order TVD Runge-Kutta time discretization [16]
\[
u^{(1)} = u^n + \Delta t L(u^n); \quad u^{n+1} = \frac{1}{2} (u^n + u^{(1)} + \Delta t L(u^{(1)}))
\]
where \( L \) is the spatial operator. This will yield a second order (in space and time) scheme which shares the same stability and convergence properties as the scheme (3.1). See also [8, 9].

### 4 Numerical examples

In this section we perform numerical experiments to demonstrate the properties of the schemes developed in previous sections. We take the initial condition as \( u^0(x) = -x^2 + x + 1 \), with the parameters and functions in (1.1) and (1.2) taken as \( L = 1 \), \( \alpha = 0.5 \), \( w(x) = 1 \), \( g(x, Q) = (1 - x)(3 - x + x^2 / 2 - Q) \), \( m(x, Q) = 4 + 2Q + (1 - x)^2 / 2 \), \( \beta(x, Q) = (1 + x)(2 - Q) \).

For the second order scheme, based on a local truncation error analysis, it is more accurate to adjust the mesh size for the second interval \( x_2 - x_1 \) from \( \Delta x \) to \( \frac{3}{2} \Delta x \), and the mesh size for the second last interval \( x_{N-1} - x_{N-2} \) from \( \Delta x \) to \( \frac{1}{2} \Delta x \) (not the actual mesh sizes in the
physical space, just that used in the scheme), hence we have made this adjustment in the computation. This apparently does not affect the stability and convergence analysis as the analysis does not require uniform meshes.

First we demonstrate that the schemes are non-oscillatory in the presence of solution discontinuities. For this purpose we take $C(t) = 3$, which causes an incompatibility of the boundary data and the initial condition at the origin. The solution then has a discontinuity emitted from the left boundary and traveling to the right, until it moves outside the right boundary. See Fig. 4.1, left, for the evolution of the solution until $t = 2$. When $t = 0.5$, the solution still contains a discontinuity. The numerical solutions using $N = 100$ uniformly spaced grid points for both the first order scheme and the second order scheme are plotted in Fig. 4.1, right, against a reference solution which is obtained by the second order scheme with $N = 2,000$ grid points. We can see clearly that both schemes can resolve the discontinuity without oscillation, and the second order scheme resolves the discontinuity much better without introducing spurious numerical oscillations. This verifies the high resolution property of the second order scheme.

Figure 4.1: Left: the evolution of the solution to $t = 2$. Right: numerical solutions using $N = 100$ uniform grid points using the first order scheme (triangles) and using the second order scheme (circles), versus the reference solution (solid line) obtained by the second order scheme using $N = 2,000$ grid points.
Next, we demonstrate that the schemes can achieve their designed accuracy for smooth solutions. For this purpose we take \( C(t) = \frac{38}{2t} + t \), which ensures the compatibility of the boundary data and the initial condition at the origin. The solution then is continuous but has a discontinuous derivative (a kink) emitted from the left boundary and traveling to the right, until it moves outside the right boundary. When \( t = 2 \), the kink has already moved out of the right boundary and solution becomes smooth. Since we do not know the exact solution, we use the second order scheme with \( N = 10, 240 \) grid points to produce a reference solution and then compute the \( L^1 \) errors of the first and second order schemes using coarser meshes, see Table 4.1. We can see that the designed orders of accuracy are obtained by the first and second order schemes for this smooth solution.

Table 4.1: \( L^1 \) errors and numerical order of accuracy of the first and second order schemes using \( N \) uniformly spaced mesh points.

<table>
<thead>
<tr>
<th>( N )</th>
<th>First order scheme</th>
<th>Second order scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>( L^1 ) error</td>
<td>order</td>
</tr>
<tr>
<td>10</td>
<td>9.49E-02</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>4.61E-02</td>
<td>1.04</td>
</tr>
<tr>
<td>40</td>
<td>2.25E-02</td>
<td>1.03</td>
</tr>
<tr>
<td>80</td>
<td>1.10E-02</td>
<td>1.03</td>
</tr>
<tr>
<td>160</td>
<td>5.44E-03</td>
<td>1.02</td>
</tr>
</tbody>
</table>

5 Concluding remarks

We have developed a first order explicit upwind scheme and a second order explicit high resolution scheme for solving a hierarchical size-structured population model with nonlinear growth, mortality and reproduction rates, which contains global terms both for the boundary condition and for the coefficients in the equations. Stability and convergence are proved for both schemes for solutions with bounded total variation, which include discontinuous solutions. Numerical results are provided to demonstrate the capability of these schemes in resolving smooth as well as discontinuous solutions. Future work will include the design

References


