

# Numerical Scheme for the Generalised Serre-Green-Naghdi Model

Jordan Pitt – `jordan.pitt@anu.edu.au`, Christopher Zoppou – `christopher.zoppou@anu.edu.au`, Stephen Roberts – `stephen.roberts@anu.edu.au`

---

---

## 1. Abstract

We describe a numerical scheme for solving the recently derived generalised Serre-Green-Naghdi equations. This numerical scheme is an extension to the numerical scheme for the classical Serre-Green-Naghdi equations which solved an auxiliary elliptic equation and then solved the remaining equations in conservation law form using the finite volume method. We then describe a second-order numerical method as an example implementation of this numerical scheme which is validated using known analytic solutions; the travelling solitary wave and the dam-break problem. The numerical method is further validated for general members of this family of equations using forced solutions. The described numerical method is conservative, robust and second-order accurate.

## 2. Introduction

The generalised Serre-Green-Naghdi Equations (gSGNE) were recently derived by Clamond and Dutykh [1]. They are a family of equations that generalise the classical Serre-Green-Naghdi Equations (SGNE) first derived by Serre [2] with the addition of two parameters  $\beta_1$  and  $\beta_2$ . The equations describe the behaviour of water waves in shallow water where the typical water depth  $h_0$  is much smaller than the wavelength  $\lambda$  so that the shallowness parameter  $\sigma = h_0/\lambda \ll 1$ . The gSGNE are of particular interest to the wave modelling community as their dispersion relationship well approximates the dispersion relationship of the linear wave theory [3]. The choice of  $\beta$  values allows for the dispersion relationship of the gSGNE to be accurate up to  $\mathcal{O}(\sigma^2)$  terms,  $\mathcal{O}(\sigma^4)$  terms and even  $\mathcal{O}(\sigma^6)$  terms [1, 4]. Therefore, these equations provide a way to study the behaviour of water waves as higher powers of  $\sigma$  are retained in the wave model.

In previous work we have developed and validated numerical methods for a conservative reformulation of the SGNE [5, 6, 7, 8] given by  $\beta_1 = 2/3$  and  $\beta_2 = 0$  in the gSGNE. These numerical methods have all been based on the following approach; firstly solve an auxiliary elliptic equation containing only spatial derivatives to obtain  $u$  thereby obtaining all the primitive variables required by the conservation equation. Then evolve the conservation equation using the finite volume method. The benefit of the approach was the maintenance of the fundamental conservation properties of the SGNE [8] and the robustness of the method in the presence of steep gradients [9]. The approach has been shown to produce the theoretical accuracy of the underlying methods up to third-order [7, 8] with the desired conservation and linear dispersion properties [8]. The approach has also been successfully extended to include the effects of varying bathymetry and allow for the presence of dry beds [8].

It was demonstrated by Clamond and Dutykh [1] that the gSGNE also possesses a conservation reformulation and thus the technique described above for the SGNE can be extended to the gSGNE. The success of this technique for the SGNE makes this an attractive option resulting in an explicit, robust and conservative method for solving the gSGNE. Hence, the purpose of this paper is to describe the extension of the numerical scheme of Zoppou et al. [7] for the recently developed gSGNE [1].

This paper begins by introducing the gSGNE and highlighting their important properties with regards to the developed numerical scheme. The second-order implementation of the numerical scheme of Zoppou et al. [7] is then described.

The numerical method is then validated against analytic solutions of the SGNE and the Shallow Water Wave Equations (SWWE), demonstrating its convergence rate and conservation properties. Forced solutions are also used to validate that all terms in the gSGNE are being approximated to the correct order of accuracy. Forced solutions are necessary to validate the numerical method for other members of this family of equations,

because no analytic solutions are currently known. Together these validations demonstrate the capability of the numerical scheme to produce robust and accurate numerical methods.

### 3. Generalised Serre-Green-Naghdi Equations

The gSGNE were derived by Clamond and Dutykh [1] using a Lagrangian field theory approach. These equations generalise the SGNE that describe a depth averaged approximation to the Euler equations, where  $h(x, t)$  is the water depth,  $u(x, t)$  is the depth averaged horizontal velocity and  $g$  is the acceleration due to gravity. The gSGNE generalise the SGNE by introducing two free parameters  $\beta_1$  and  $\beta_2$ , with each pair of  $\beta$  values corresponding to one member of this family of equations. In particular, the family of equations captured by gSGNE contains the SWWE, the SGNE, a family of regularised SWWE studied by Clamond and Dutykh [1] and a family of improved dispersion SGNE equations studied by Clamond et al. [4].

The gSGNE describe the conservation of mass ( $h$ ), momentum ( $uh$ ) and energy ( $\mathcal{E}$ ) for water waves subject to gravitational forces like so

$$\frac{\partial h}{\partial t} + \frac{\partial(hu)}{\partial x} = 0 \quad (1a)$$

$$\frac{\partial(hu)}{\partial t} + \frac{\partial}{\partial x} \left( hu^2 + \frac{1}{2}gh^2 + \frac{1}{2}h^2\Gamma \right) = 0 \quad (1b)$$

$$\frac{\partial(\mathcal{E})}{\partial t} + \frac{\partial}{\partial x} \left[ hu \left( \frac{1}{2}u^2 + \frac{1}{4}\beta_1 h^2 \frac{\partial u}{\partial x} \frac{\partial u}{\partial x} + gh \left( 1 + \frac{1}{4}\beta_2 \frac{\partial h}{\partial x} \frac{\partial h}{\partial x} \right) + \frac{1}{3}h\Gamma \right) + \frac{1}{2}\beta_2 gh^3 \frac{\partial h}{\partial x} \frac{\partial u}{\partial x} \right] = 0 \quad (1c)$$

where

$$\Gamma = \beta_1 h \left[ \frac{\partial u}{\partial x} \frac{\partial u}{\partial x} - \frac{\partial^2 u}{\partial x \partial t} - u \frac{\partial^2 u}{\partial x^2} \right] - \beta_2 g \left[ h \frac{\partial^2 h}{\partial x^2} + \frac{1}{2} \frac{\partial h}{\partial x} \frac{\partial h}{\partial x} \right] \quad (1d)$$

$$\mathcal{E} = \frac{1}{2}hu^2 + \frac{1}{4}\beta_1 h^3 \frac{\partial u}{\partial x} \frac{\partial u}{\partial x} + \frac{1}{2}gh^2 \left( 1 + \frac{1}{2}\beta_2 \frac{\partial h}{\partial x} \frac{\partial h}{\partial x} \right). \quad (1e)$$

When  $\beta_1 = \beta_2 = 0$  these equations reduce to the SWWE and the SGNE are recovered when  $\beta_1 = 2/3$  and  $\beta_2 = 0$ .

Equation (1) contains three partial differential equations with only two unknowns,  $h$  and  $u$ . Since there are only two unknowns it is sufficient to solve only two of the three equations, typically either the conservation of mass (1a) and the conservation of momentum equations (1b) or the conservation of mass (1a) and the conservation of energy equations (1c). If  $h$  and  $u$  are sufficiently smooth, then the solution of any pair of these equations satisfies all equations (1a), (1b) and (1c) simultaneously. However, for particular  $\beta$  values, for example those corresponding to the SWWE, it is possible to obtain non-smooth solutions of  $h$  and  $u$  for any pair of these equations that no longer satisfies all three equations simultaneously [10]. Since the mass (1a) and momentum (1b) equations are commonly solved this results in dissipation of energy around discontinuities in the solution. Conversely, when the mass (1a) and energy equations (1c) are solved this results in dissipation of momentum and an incorrect propagation speed for discontinuities in the solution.

When solutions are sufficiently smooth all equations hold simultaneously and the total amounts of all these quantities remain constant in time if the system is closed. This can be seen by integrating (1) over the domain, and observing that the temporal derivative of the spatial integrals of mass ( $h$ ), momentum ( $uh$ ) and energy ( $\mathcal{E}$ ) is zero when there is no flux across the domain boundaries. This conservation property of the gSGNE will be used to validate the numerical method and its solutions.

#### 3.1. Dispersion Relation of the Linearised gSGNE

The linear dispersion properties of water wave equations have been of particular interest [4, 11, 12], as the scope of wave modelling expands to include dispersive effects. Indeed, the gSGNE are especially relevant because they provide better dispersion properties when compared to the linear theory for water waves [3] than the SWWE for which  $\beta_1 = \beta_2 = 0$ .

To obtain the dispersion relationship of the linearised gSGNE, (1) is first linearised by considering small waves superimposed on water with a mean depth  $h_0$  and a mean velocity  $u_0$ . The dispersion relationship of the linearised gSGNE is then obtained by seeking travelling wave solutions of the form  $\exp(i(kx - \omega t))$  to obtain

$$\omega_{\text{gSGNE}}^{\pm} = u_0 k \pm k \sqrt{gh_0} \sqrt{\frac{\beta_2 h_0^2 k^2 + 2}{\beta_1 h_0^2 k^2 + 2}} \quad (2)$$

which provides the angular frequency  $\omega$  of travelling wave solutions of the linearised gSGNE equations for waves with wavenumber  $k$ . The dispersion relation has a positive and negative branch corresponding to the direction of these waves. This dispersion relation (2) is equivalent to the dispersion relation derived by Clamond and Dutykh [1] for the gSGNE when  $u_0 = 0$ .

The dispersion relation of the gSGNE approximates the dispersion relationship of linear wave theory, as can be seen by comparing their power series approximations provided below

$$\begin{aligned} \omega_{\text{linear}}^{\pm} &= u_0 k \pm \sqrt{gk \tanh(kh_0)} \\ &= \left(u_0 \pm \sqrt{gh_0}\right) k \pm \frac{-1}{6} \sqrt{gh_0} h_0^2 k^3 \pm \frac{19}{360} \sqrt{gh_0} h_0^4 k^5 + \mathcal{O}(k^7) \end{aligned} \quad (3a)$$

$$\omega_{\text{gSGNE}}^{\pm} = \left(u_0 \pm \sqrt{gh_0}\right) k \pm \frac{(\beta_2 - \beta_1)}{4} \sqrt{gh_0} h_0^2 k^3 \pm \frac{(3\beta_1^2 - 2\beta_2\beta_1 - \beta_2^2)}{32} \sqrt{gh_0} h_0^4 k^5 + \mathcal{O}(k^7). \quad (3b)$$

Which is  $k^2$  accurate for all  $\beta$  values, accurate to the  $k^4$  term when  $\beta_1 = \beta_2 + 2/3$  and accurate up to the  $k^6$  term when  $\beta_1 = \beta_2 + 2/3$  and  $\beta_2 = 2/15$  [4]. Since  $k = 2\pi/\lambda$ , these power of  $k$  terms have a corresponding shallowness parameter,  $\sigma$  term. The gSGNE equations have a dispersion relationship that can be accurate up to  $\mathcal{O}(\sigma^2)$  terms,  $\mathcal{O}(\sigma^4)$  terms and  $\mathcal{O}(\sigma^6)$  terms. Since the  $\beta$  values allow us to alter the accuracy of the dispersion relationship, the gSGNE allow us to consistently compare the effect of the dispersion relationship on numerical solutions and thus the physical phenomena.

From the dispersion relation (2), the phase speed  $v_p$  and the group speed  $v_g$  can be derived as follows

$$v_p^{\pm} = \frac{\omega_{\text{gSGNE}}^{\pm}}{k} = u_0 \pm \sqrt{gh_0} \sqrt{\frac{\beta_2 h_0^2 k^2 + 2}{\beta_1 h_0^2 k^2 + 2}}, \quad (4a)$$

$$v_g^{\pm} = \frac{\partial \omega_{\text{gSGNE}}^{\pm}}{\partial k} = u_0 \pm \sqrt{gh_0} \sqrt{\frac{\beta_2 h_0^2 k^2 + 2}{\beta_1 h_0^2 k^2 + 2}} \left[ 1 + \frac{\beta_2 - \beta_1}{\left(\frac{1}{2}\beta_2 h_0^2 k^2 + 1\right) \left((\beta_1 - \frac{1}{3}) h_0^2 k^2 + 1\right)} \right]. \quad (4b)$$

### 3.1.1. Phase Speed Bounds

Using the phase speed bounds of the SGNE Le Métayer et al. [13] and Zoppou et al. [7] applied approximate Riemann solvers to solve the SGNE in conservation law form. Thus, if the gSGNE can also be shown to have bounded phase speeds then these same solvers can be applied to solve the gSGNE in conservation law form.

To demonstrate that the phase speeds are bounded, observe that when  $\beta_1 \geq 0$ ,  $\beta_2 \geq 0$  and  $h_0 k \geq 0$  then

$$f(h_0 k) = \frac{\beta_2 (h_0 k)^2 + 2}{\beta_1 (h_0 k)^2 + 2},$$

is a monotone function over  $h_0 k$ . This can be seen by reformulating and taking the derivative with respect to  $h_0 k$ , to obtain that

$$\frac{\partial(f(h_0 k))}{\partial(h_0 k)} = [\beta_2 - \beta_1] \frac{4(h_0 k)}{(\beta_1 (h_0 k)^2 + 2)^2}.$$

The derivative is greater than zero and thus  $f(h_0 k)$  is monotone non-decreasing if  $\beta_1 \leq \beta_2$ . Whilst the derivative is greater than zero and thus  $f(h_0 k)$  is monotone non-increasing if  $\beta_1 \geq \beta_2$ . Since  $v_p^+ = u_0 + \sqrt{gh_0 f(h_0 k)}$  and  $v_p^- = u_0 - \sqrt{gh_0 f(h_0 k)}$ , given the above properties of  $f(h_0 k)$  under the initial assumptions  $v_p^+$  is monotone non-decreasing and  $v_p^-$  is monotone non-increasing when  $\beta_1 \leq \beta_2$ . Whereas when  $\beta_1 \geq \beta_2$  then  $v_p^+$  is monotone non-increasing and  $v_p^-$  is monotone non-decreasing.

In addition to the monotonicity of  $v_p^\pm$ , when  $k \rightarrow 0$  then  $v_p^\pm \rightarrow u_0 \pm \sqrt{gh_0}$ , whilst as  $k \rightarrow \infty$  then  $v_p^\pm \rightarrow u_0 \pm \sqrt{gh_0} \sqrt{\beta_2/\beta_1}$ . Therefore,  $v_p^\pm$  is monotonic and bounded at the limits of the domain, and thus bounded for all  $\beta$  values provided that  $\beta_1 = 0$  when  $\beta_2 = 0$ , otherwise in the  $k \rightarrow \infty$  limit the phase speed is no longer bounded. Consequently, the methods of Le Métayer et al. [13] and Zoppou et al. [7] for the SGNE can be extended to the gSGNE because the phase speeds are bounded.

In addition, to the phase speed bounds we also have the following chain of inequalities when  $\beta_1 \geq \beta_2$

$$u_0 - \sqrt{gh_0} \leq v_p^- \leq u_0 - \sqrt{gh_0} \sqrt{\frac{\beta_2}{\beta_1}} \leq u_0 \leq u_0 + \sqrt{gh_0} \sqrt{\frac{\beta_2}{\beta_1}} \leq v_p^+ \leq u_0 + \sqrt{gh_0}. \quad (5)$$

We designate this region of  $\beta$  values, as ‘Region 1’. In this region the minimum and maximum phase speed values are  $u_0 - \sqrt{gh_0}$  and  $u_0 + \sqrt{gh_0}$  respectively. Region 1 is then characterised by trailing dispersive waves when  $\beta_2 < \beta_1$  with an upper boundary where waves are not dispersive when  $\beta_2 = \beta_1$ . Region 1 includes the SWWE and the SGNE and is consistent with the behaviour of the dispersive waves given by the linear theory for water waves [3].

When  $\beta_2 > \beta_1$  the inequality chain becomes

$$u_0 - \sqrt{gh_0} \sqrt{\frac{\beta_2}{\beta_1}} \leq v_p^- \leq u_0 - \sqrt{gh_0} \leq u_0 \leq u_0 + \sqrt{gh_0} \leq v_p^+ \leq u_0 + \sqrt{gh_0} \sqrt{\frac{\beta_2}{\beta_1}} \quad (6)$$

This will be denoted as ‘Region 2’ and it is characterised by advancing dispersive waves. Since we are only interested in water waves with trailing dispersive waves, this paper will be restricted to studying Region 1, although the numerical scheme is valid for all regions.

The regions and location of important members of the gSGNE in terms of  $\beta$  values is summarised in Figure 1. All equations in the gSGNE family are accurate up to  $\mathcal{O}(k^2)$  in the dispersion relationship (3). Furthermore, there is a family of equations that obtain  $\mathcal{O}(k^4)$  accuracy in the dispersion relationship with one particular member of this family, the improved SGNE, obtaining a the dispersion relationship which is  $\mathcal{O}(k^6)$  accurate. Additionally, there are also illustrative numerical solutions that demonstrates the effect of the  $\beta$  values and hence the different dispersion relationships on solutions to the dam-break problem described in Section 5.2. The dam-break problem consists of initially still water with a left water depth, in this case of  $2m$  and a right water depth, in this case  $1m$  with a discontinuous jump between the two. The numerical solutions have been well resolved using 25600 cells and the simulated water depth  $h$  at  $t = 35s$  is plotted. The same numerical solver has been used for each with the only change being the  $\beta$  parameters.

The numerical solutions for the SWWE compare well to analytic solutions for the SWWE in Section 5.2, and the numerical solutions of the SGNE compare well to the study of the dam-break problems performed by Pitt et al. [9] for the SGNE. By comparing the numerical solution to its location in the  $\beta$  plot we can see that the behaviour of solutions is described well by the chain of inequalities (5) when  $\beta_1 \geq \beta_2$  and (6) when  $\beta_1 < \beta_2$ . In particular, for the SWWE and the regularised SWWE  $\beta_1 = \beta_2$  which results in a phase speed which is independent of  $k$  and thus there is no dispersion. Consequently, we observe no wave train between the shock and the rarefaction fan for these non-dispersive wave models. For the dispersive wave models such as the classical SGNE and the improved SGNE with  $\mathcal{O}(k^4)$  and  $\mathcal{O}(k^6)$  accuracy in the dispersion relationship respectively, we observe a dispersive wave train between the shock and the rarefaction fan. In Region 2 of the  $\beta$  plot we observe advancing dispersive wave trains behind the shock and ahead of the shock due to the change of the chain of inequalities when  $\beta_1 < \beta_2$  in (6).

### 3.2. Alternative Conservative Form of the gSGNE

Clamond and Dutykh [1] provided a reformulation of (1b) for the gSGNE, in an analogous to the way the SGNE [7, 13, 14] were reformulated. The purpose of this reformulation is to remove the mixed spatial-temporal derivative in the flux term of the momentum equation (1b), which is difficult to treat numerically. This reformulation is obtained by introducing a new conserved quantity

$$G = hu - \frac{\beta_1}{2} \frac{\partial}{\partial x} \left( h^3 \frac{\partial u}{\partial x} \right)$$

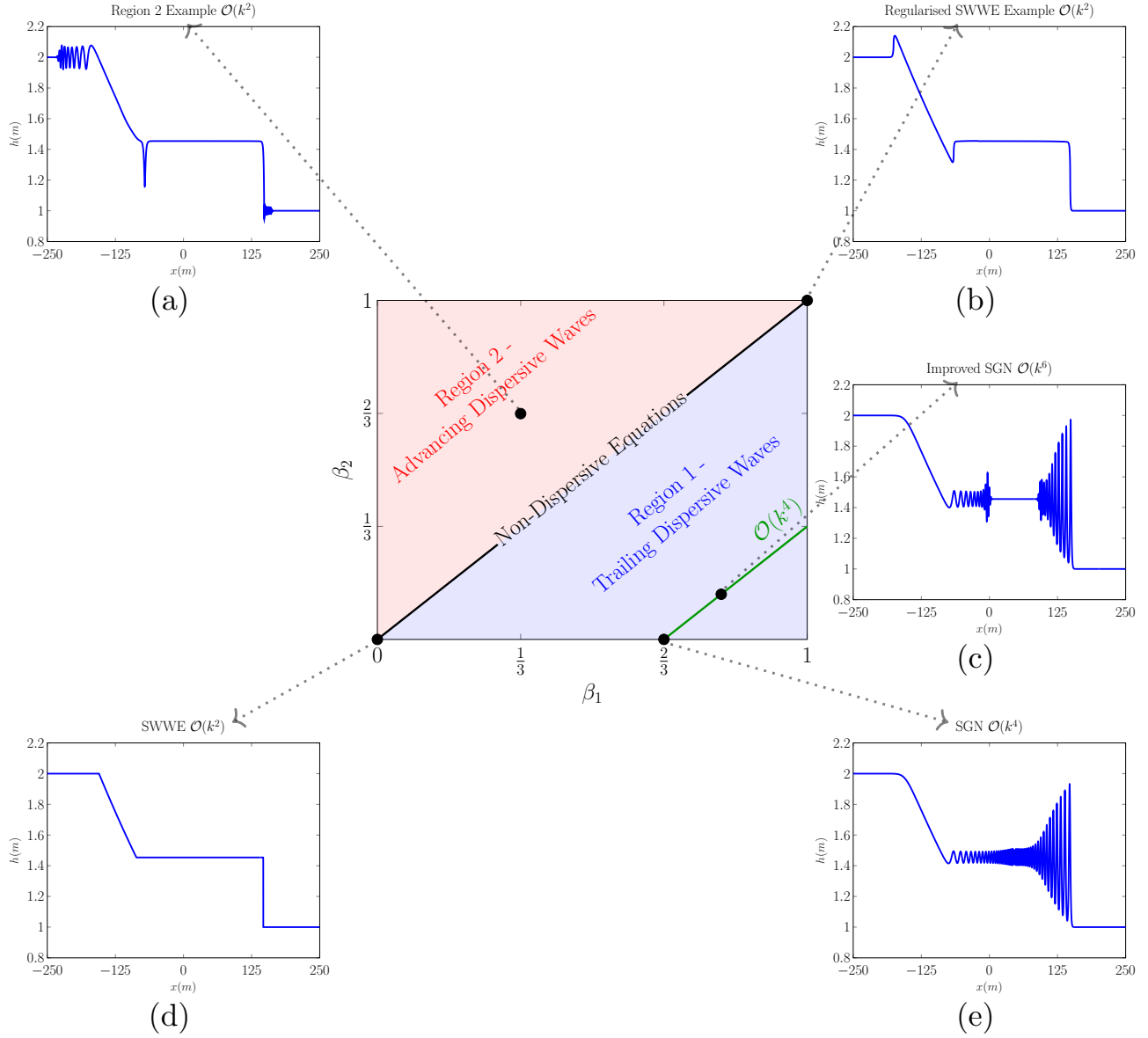


Figure 1: Phase speed regions of gSGNE in terms of  $(\beta_1, \beta_2)$  showing important equations, their associated dispersion relationship accuracy (3) and an example numerical solution for the dam-break problem (25) solved using  $(1/3, 2/3)$  in (a),  $(1, 1)$  in (b),  $(2/3 + 2/15, 2/15)$  in (c),  $(0, 0)$  in (d) and  $(2/3, 0)$  in (e).

and thus (1b) can be written in conservation law form for  $G$  as

$$\frac{\partial G}{\partial t} + \frac{\partial}{\partial x} \left( uG + \frac{gh^2}{2} - \beta_1 h^3 \frac{\partial u}{\partial x} \frac{\partial u}{\partial x} - \frac{1}{2} \beta_2 g h^2 \left[ h \frac{\partial^2 h}{\partial x^2} + \frac{1}{2} \frac{\partial h}{\partial x} \frac{\partial h}{\partial x} \right] \right) = 0$$

When  $\beta_1 = 2/3$  then  $G$  is the same conserved quantity introduced for the SGNE [7, 13, 14].

This reformulation, provides the gSGNE in conservation law form for both  $h$  and the new conserved quantity  $G$ . Thus the conservative gSGNE are

$$\frac{\partial h}{\partial t} + \frac{\partial(uh)}{\partial x} = 0 \quad (7a)$$

$$\frac{\partial G}{\partial t} + \frac{\partial}{\partial x} \left( uG + \frac{gh^2}{2} - \beta_1 h^3 \frac{\partial u}{\partial x} \frac{\partial u}{\partial x} - \frac{1}{2} \beta_2 g h^2 \left[ h \frac{\partial^2 h}{\partial x^2} + \frac{1}{2} \frac{\partial h}{\partial x} \frac{\partial h}{\partial x} \right] \right) = 0. \quad (7b)$$

with

$$G = uh - \frac{\beta_1}{2} \frac{\partial}{\partial x} \left( h^3 \frac{\partial u}{\partial x} \right). \quad (7c)$$

Now that the gSGNE are written in conservation law form and have a bound on the phase speeds, they can be solved numerically using the hybrid finite volume technique described by Zoppou et al. [7] for the SGNE.

#### 4. Numerical Scheme

The proposed numerical scheme for the gSGNE extends those previously published for the SGNE [7, 13] to allow any  $\beta$  values and the associated additional terms. We begin the description of the numerical scheme by giving a brief outline of the finite volume method and the discretisation therein. We then provide an overview of the numerical technique and then describe the simplest second-order implementation of this scheme in detail. With previous studies [7, 8] showing the sufficiency and necessity of second-order methods for the SGNE.

##### 4.1. Finite Volume Method and Discretisation

The core of these hybrid finite volume methods for the gSGNE is the finite volume method used to solve the equations in conservation law form. The finite volume method solves equations of the form

$$\frac{\partial q}{\partial t} + \frac{\partial f(q)}{\partial x} = 0 \quad (8)$$

where  $q$  is a generic conserved quantity. This is the form of the gSGNE after the reformulation (7). In the finite volume method (8) is integrated over cells of fixed width  $\Delta x$  in space and over time steps of fixed length  $\Delta t$ . The midpoint of the  $j^{th}$  cell is given by  $x_j = x_0 + j\Delta x$  while the cell edges of the  $j^{th}$  cell are given by  $x_{j-1/2} = x_0 + (j - \frac{1}{2})\Delta x$  and  $x_{j+1/2} = x_0 + (j + \frac{1}{2})\Delta x$ . Likewise the  $n^{th}$  time step is given by  $t^n = t^0 + n\Delta t$ .

Integrating (8) over both space and time results in

$$\bar{q}_j^{n+1} = \bar{q}_j^n + \frac{\Delta t}{\Delta x} \left[ F_{j+1/2}^n - F_{j-1/2}^n \right] \quad (9)$$

where

$$\bar{q}_j^n = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} q(x, t^n) dx$$

is the average of  $q$  over the  $j^{th}$  cell at time  $t^n$  and

$$F_{j\pm 1/2}^n = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} f(q(x_{j\pm 1/2}, t)) dt$$

is the average flux of  $q$  across the cell edge from time  $t^n$  to  $t^{n+1}$ . Therefore, if  $F_{j\pm 1/2}^n$  can be approximated with the appropriate order of accuracy, then we have an explicit method for updating the cell average of the conserved quantities through time for equations in conservation law form (8). In this paper we only consider approximations to  $F_{j\pm 1/2}^n$  which are first-order accurate in time.

#### 4.2. Overview

We will now describe the numerical scheme, which uses the above finite volume formulation to solve (7a) and (7b). When describing this broad overview it is also useful to consider collections of the point-wise values  $q$  or cell averaged values  $\bar{q}$  described above over the whole domain at a particular time. Thus for  $q$ , we define  $\mathbf{q}^n$  to be the vector of  $q_j^n$  values and  $\bar{\mathbf{q}}^n$  to be the vector of  $\bar{q}_j^n$  values for all cells in the domain at time  $t^n$ .

The numerical scheme for the gSGNE based on the finite volume method (9) then proceeds as follows. Begin at the  $n^{th}$  time step where the vectors of cell averages of the conserved quantities  $\bar{\mathbf{h}}^n$  and  $\bar{\mathbf{G}}^n$  are either known or provided as initial conditions.

1. Solve (7c) using  $\bar{\mathbf{h}}^n$  and  $\bar{\mathbf{G}}^n$  to obtain an approximation to  $\bar{\mathbf{u}}^n$ , which can be expressed as

$$\mathcal{A}(\bar{\mathbf{h}}^n, \bar{\mathbf{G}}^n) \rightarrow \bar{\mathbf{u}}^n.$$

2. Solve (7a) and (7b), using the finite volume method (9) with an approximate Riemann solver such as the one described by Kurganov et al. [15], to obtain  $\bar{h}_j^{n+1}$  and  $\bar{G}_j^{n+1}$  at the next time step, like so

$$\mathcal{F}(\bar{\mathbf{h}}^n, \bar{\mathbf{G}}^n, \bar{\mathbf{u}}^n) \rightarrow \bar{\mathbf{h}}^{n+1}, \bar{\mathbf{G}}^{n+1}.$$

3. Since  $\mathcal{F}$  is only first-order accurate in time, Steps 1 and 2 are repeated and combined using a convex combination to obtain a Strong Stability Preserving (SSP) Runge Kutta time stepping method [16] to approximate  $\bar{\mathbf{h}}^{n+1}$  and  $\bar{\mathbf{G}}^{n+1}$  with the correct accuracy in time.

This numerical scheme produces the numerical methods of Le Métayer et al. [13], Zoppou et al. [7] and Pitt [8] when  $\beta_1 = 2/3$  and  $\beta_2 = 0$ . Additionally, when  $\beta_1 = \beta_2 = 0$  the gSGNE reduces to the SWWE, and this numerical scheme reduces to a traditional finite volume method [17].

#### 4.3. Example Implementation

To demonstrate the utility of this numerical scheme we present a simple second-order implementation, and then validate it in the following section. The description of this numerical method will be broken up into the steps described in Section 4.2 for simplicity and also to highlight the interchangeability of the different parts of the scheme.

##### 4.3.1. Step 1 - Solution of the Elliptic Equation

To solve (7c) given  $\bar{\mathbf{h}}^n$  and  $\bar{\mathbf{G}}^n$  to obtain  $\bar{\mathbf{u}}^n$  we use the observation that the cell average and the cell nodal value are equal up to second-order accuracy, so that  $\bar{\mathbf{q}}^n = \mathbf{q}^n + \mathcal{O}(\Delta x^2)$  for all the quantities of interest. Assuming that  $u$  is sufficiently smooth, then a second-order central finite difference approximation can be used to accurately solve (7c).

The second-order central finite difference approximation to (7c) can be written as

$$\mathbf{u}^n = \mathbf{A}^{-1} \mathbf{G}^n \tag{10}$$

where  $\mathbf{A}$  is a tri-diagonal matrix with the sub-diagonal, diagonal and super-diagonal given by

$$A_{j,j-1} = -\frac{\beta_1}{2} \left[ \frac{(h_j^n)^3}{\Delta x^2} - \frac{3(h_j^n)^2}{2\Delta x} \frac{h_{j+1}^n - h_{j-1}^n}{2\Delta x} \right] \tag{11a}$$

$$A_{j,j} = h_j^n + \beta_1 \frac{(h_j^n)^3}{\Delta x^2} \tag{11b}$$

$$A_{j,j+1} = -\frac{\beta_1}{2} \left[ \frac{(h_j^n)^3}{\Delta x^2} + \frac{3(h_j^n)^2}{2\Delta x} \frac{h_{j+1}^n - h_{j-1}^n}{2\Delta x} \right] \tag{11c}$$

respectively.

The finite difference approximation (10) can be solved with any matrix solver, for this method we use the explicit Thomas Algorithm [18]. This is an efficient method provided  $\mathbf{A}$  is non-singular which is the case as long as  $h_j^n > 0$ , which is true for all the problems of interest in this paper.

Thus as desired we have

$$\mathcal{A}(\bar{\mathbf{h}}^n, \bar{\mathbf{G}}^n) = \mathbf{A}^{-1} \mathbf{G}^n = \mathbf{u}^n \rightarrow \bar{\mathbf{u}}^n. \tag{12}$$

#### 4.3.2. Step 2 - Finite Volume Method - Reconstruction

The finite volume method of Kurganov et al. [15] relies on approximations of all the terms in the spatial derivative of the flux at the cell edges  $x_{j\pm 1/2}$ . Thus the following quantities require second-order approximations at the cell edges:  $u$ ,  $h$ ,  $G$ ,  $\partial u/\partial x$ ,  $\partial h/\partial x$ ,  $\partial^2 h/\partial x^2$ . In this paper we are interested in reproducing solutions to smooth analytic and forced solutions for all members of the gSGNE. We are also interested in reproducing the discontinuous dam-break solution for only the SWWE where  $\beta_1 = \beta_2 = 0$  and thus only  $h$ ,  $u$  and  $G = uh$  need to be approximated. For this reason our gSGNE solver uses approximations to  $u$ ,  $h$ ,  $G$  that allow for discontinuities and thus use limiting. Whereas the approximations to the derivatives  $\partial u/\partial x$ ,  $\partial h/\partial x$ ,  $\partial^2 h/\partial x^2$  will assume that these quantities are smooth and thus do not require limiting.

Figure 1 demonstrates the sufficiency of this reconstruction approach even in situations where the derivatives are present in the equations but are not smooth. A main reason for this is that the elliptic equation (7c) helps control the smoothness of  $u$ . However, this does not explain why we get reasonable results when assuming the derivatives of  $h$  are smooth when they are not smooth in the initial conditions. This will be the focus of further research as we compare different derivative approximations in the presence of discontinuities for the gSGNE. For the validations performed in this paper, the assumption on the smoothness of the derivatives  $\partial u/\partial x$ ,  $\partial h/\partial x$ ,  $\partial^2 h/\partial x^2$  is adequate.

The reconstructions for  $u$ ,  $h$  and  $G$  can be summarised for a general quantity  $q$  like so

$$q_{j-1/2}^+ = \bar{q}_j - \frac{\Delta x}{2} s_j \quad \text{and} \quad q_{j+1/2}^- = \bar{q}_j + \frac{\Delta x}{2} s_j \quad (13)$$

where

$$s_j = \text{minmod} \left( \theta \frac{\bar{q}_j - \bar{q}_{j-1}}{\Delta x}, \frac{\bar{q}_{j+1} - \bar{q}_{j-1}}{2\Delta x}, \theta \frac{\bar{q}_{j+1} - \bar{q}_j}{\Delta x} \right). \quad (14)$$

The minmod function is defined as follows

$$\text{minmod}(a, b, c) = \begin{cases} \min(a, b, c) & \text{when } a < 0, b < 0, c < 0 \\ \max(a, b, c) & \text{when } a > 0, b > 0, c > 0 \\ 0 & \text{otherwise} \end{cases}. \quad (15)$$

This gives a reconstruction at each cell edge, so cell  $j$  gives rise to  $q_{j-1/2}^+$  and  $q_{j+1/2}^-$  while cell  $j+1$  gives rise to  $q_{j+1/2}^+$  and  $q_{j+3/2}^-$ , all of which are required in the flux approximation.

For  $\partial u/\partial x$ ,  $\partial h/\partial x$  and  $\partial^2 h/\partial x^2$  we assume that the quantities are smooth, and so using the appropriate order finite difference approximation at the cell edges is sufficient. We demonstrate these approximations for  $x_{j+1/2}$  and a generic quantity  $q$

$$\begin{aligned} \left[ \frac{\partial q}{\partial x} \right]_{j+1/2} &= \frac{q_{j+1} - q_j}{\Delta x}, \\ \left[ \frac{\partial^2 q}{\partial x^2} \right]_{j+1/2} &= \frac{q_{j+2} - q_{j+1} - q_j + q_{j-1}}{2\Delta x^2}. \end{aligned}$$

#### 4.3.3. Step 2 - Finite Volume Method - Flux Approximation

To solve (7a) and (7b) which are both in conservation law form (8) we use (9). Since we begin with  $\bar{\mathbf{h}}^n$  and  $\bar{\mathbf{G}}^n$ , we only require an approximation to  $F_{j\pm 1/2}^n$  to obtain  $\bar{\mathbf{h}}^{n+1}$  and  $\bar{\mathbf{G}}^{n+1}$ .

In this numerical scheme the approximate Riemann solver described by Kurganov et al. [15] is used to calculate  $F_{j\pm 1/2}^n$ . The major advantage of this scheme is that it only requires bounds on the phase speeds. Only the calculation of the flux term  $F_{j+1/2}^n$  is demonstrated as the process to calculate the flux term  $F_{j-1/2}^n$  is identical but with different cells. For  $q$  the flux is approximated by

$$F_{j+\frac{1}{2}}^n = \frac{a_{j+\frac{1}{2}}^+ f(q_{j+\frac{1}{2}}^-) - a_{j+\frac{1}{2}}^- f(q_{j+\frac{1}{2}}^+)}{a_{j+\frac{1}{2}}^+ - a_{j+\frac{1}{2}}^-} + \frac{a_{j+\frac{1}{2}}^+ a_{j+\frac{1}{2}}^-}{a_{j+\frac{1}{2}}^+ - a_{j+\frac{1}{2}}^-} [q_{j+\frac{1}{2}}^+ - q_{j+\frac{1}{2}}^-] \quad (16)$$



where  $a_{j+\frac{1}{2}}^+$  and  $a_{j+\frac{1}{2}}^-$  are given by the phase speed bounds and all quantities on the right hand side are computed at time  $t^n$ . Applying the phase speed bounds (5) and (6) we obtain

$$a_{j+\frac{1}{2}}^- = \min \left\{ 0, u_{j+1/2}^- - \max \left( 1, \sqrt{\frac{\beta_2}{\beta_1}} \right) \sqrt{gh_{j+1/2}^-}, u_{j+1/2}^+ - \max \left( 1, \sqrt{\frac{\beta_2}{\beta_1}} \right) \sqrt{gh_{j+1/2}^+} \right\}, \quad (17)$$

$$a_{j+\frac{1}{2}}^+ = \max \left\{ 0, u_{j+1/2}^- + \max \left( 1, \sqrt{\frac{\beta_2}{\beta_1}} \right) \sqrt{gh_{j+1/2}^-}, u_{j+1/2}^+ + \max \left( 1, \sqrt{\frac{\beta_2}{\beta_1}} \right) \sqrt{gh_{j+1/2}^+} \right\} \quad (18)$$

where the  $\max \left( 1, \sqrt{\beta_2/\beta_1} \right)$  accounts for the different phase speed bounds (5) and (6), which depend on the choice of  $\beta$  values.

The flux functions  $f(q_{j+\frac{1}{2}}^-)$  and  $f(q_{j+\frac{1}{2}}^+)$  are evaluated using the reconstructed values of the  $j^{th}$  and  $(j+1)^{th}$  cell respectively. From the continuity equation (7a) we have

$$f \left( h_{j+\frac{1}{2}}^\pm \right) = u_{j+1/2}^\pm h_{j+1/2}^\pm.$$

For the evolution of  $G$  equation (7b) we have

$$\begin{aligned} f \left( G_{j+\frac{1}{2}}^\pm \right) &= u_{j+1/2}^\pm G_{j+1/2}^\pm + \frac{g}{2} \left( h_{j+1/2}^\pm \right)^2 - \beta_1 \left( h_{j+1/2}^\pm \right)^3 \left( \left[ \frac{\partial u}{\partial x} \right]_{j+1/2} \right)^2 \\ &\quad - \frac{1}{2} \beta_2 g \left( h_{j+1/2}^\pm \right)^2 \left[ h_{j+1/2}^\pm \left[ \frac{\partial^2 h}{\partial x^2} \right]_{j+1/2} + \frac{1}{2} \left( \left[ \frac{\partial h}{\partial x} \right]_{j+1/2} \right)^2 \right] \end{aligned} \quad (19)$$

where the derivatives are assumed to be smooth across the cell edge, and thus do not require different super-scripts.

Since the reconstructions were given above for all these quantities, we can approximate  $F_{j\pm 1/2}^n$  using (16) and thus employ (9) to obtain  $\bar{h}^{n+1}$  and  $\bar{G}^{n+1}$  resulting in

$$\mathcal{F} \left( \bar{h}^n, \bar{G}^n, \bar{u}^n \right) = \bar{q}_j^n + \frac{\Delta t}{\Delta x} \left[ F_{j+1/2}^n - F_{j-1/2}^n \right] \rightarrow \bar{h}^{n+1}, \bar{G}^{n+1} \quad (20)$$

as desired.

#### 4.3.4. Step 3 - Runge-Kutta Time Stepping

Combining both Steps 1 and Steps 2, provides a spatially second-order scheme with first-order time stepping. To arrive at a fully second-order method, we repeat Steps 1 and Steps 2 according to the second-order SSP Runge Kutta method [16]. Doing this we obtain the following scheme making use of the above implementation of  $\mathcal{A}$ , (12) and  $\mathcal{F}$ , (20)

$$\begin{aligned} \bar{h}^{(1)}, \bar{G}^{(1)} &= \mathcal{F} \left( \bar{h}^n, \bar{G}^n, \mathcal{A} \left( \bar{h}^n, \bar{G}^n \right) \right) \\ \bar{h}^{(2)}, \bar{G}^{(2)} &= \mathcal{F} \left( \bar{h}^{(1)}, \bar{G}^{(1)}, \mathcal{A} \left( \bar{h}^{(1)}, \bar{G}^{(1)} \right) \right) \\ \bar{h}^{n+1}, \bar{G}^{n+1} &= \frac{1}{2} \left( \bar{h}^n + \bar{h}^{(2)} \right), \frac{1}{2} \left( \bar{G}^n + \bar{G}^{(2)} \right). \end{aligned}$$

obtaining a stable fully second-order method for solving (7).

#### 4.4. Boundary Conditions

For the purposes of the validation below and for simplicity, we have applied Dirichlet boundary conditions using ghost cells over which the values of  $h$ ,  $G$  and  $u$  are known. Thus in addition to the  $m$  cells inside the boundary we have an additional  $l$  cells either side of it so that we have the left ghost cells  $-l, -l+1, \dots, -1$

and the right ghost cells  $m, m + 1, \dots, m + l - 1$ . Since the numerical method has a maximum stencil that extends 2 cells beyond the target cell, then  $l$  must be greater than 1.

This boundary condition is applied to Step 1 by extending the matrix equation (10) to vectors containing the ghost cells  $\hat{\mathbf{u}}^n$  and  $\hat{\mathbf{G}}^n$  like so

$$\begin{aligned}\hat{\mathbf{u}}^n &= [u_{-l}^n \dots u_{-1}^n u_0^n \dots u_{m-1}^n u_m^n \dots u_{m+l-1}^n]^T \\ \hat{\mathbf{G}}^n &= [G_{-l}^n \dots G_{-1}^n G_0^n \dots G_{m-1}^n G_m^n \dots G_{m+l-1}^n]^T.\end{aligned}$$

Likewise the associated extended matrix  $\hat{\mathbf{A}}$  in (10) is the same as  $\mathbf{A}$  when  $0 \leq j \leq m - 1$  (11) with the additional associated ghost cell values given by

$$\hat{A}_{j,j-1} = 0 \quad \hat{A}_{j,j} = 1 \quad \hat{A}_{j,j+1} = 0$$

when  $j < 0$  or  $j > m - 1$ .

For Step 2 there is no need to reconstruct the quantities inside the ghost cells because  $h$ ,  $G$  and  $u$  are given.

#### 4.5. Courant-Frederichs-Lewy Condition

To ensure the stability of the finite volume method (9) the Courant-Friedrichs-Lewy (CFL) condition [19] is used. The CFL condition is necessary for stability and ensures that time steps are small enough so that information is only transferred between neighbouring cells. For the gSGNE the CFL condition is

$$\Delta t \leq \frac{Cr}{\max_j \{a_{j+1/2}^\pm\}} \Delta x \quad (21)$$

where  $a_{j+1/2}^\pm$  are the phase speed bounds used in the flux approximation (18) and  $0 \leq Cr \leq 1$  is the Courant number.

## 5. Validation

The numerical method described above is validated using analytic solutions for particular  $\beta$  values that correspond to the SGNE and the SWWE as well as a forced solution. Together these tests demonstrate the ability of the method to reproduce analytic solutions to important members of the gSGNE family, as well as assessing the accuracy of the numerical method's approximation of all terms in the gSGNE.

### 5.1. Convergence and Conservation Measures

To validate the numerical solutions we make use of measures of convergence and conservation. The measure of convergence will be the relative deviation of the numerical solution from the equivalent analytic or forced solution using the  $L_2$  norm. While the conservation properties of the numerical method will be measured by numerically approximating the energy in the initial conditions and the numerical solution and comparing them using a measure called  $C_1$ .

For a quantity  $q$  with a vector of its analytic or forced solution at the cell midpoints  $\mathbf{q}$  and the numerical solution at the cell midpoints  $\mathbf{q}^*$ , the discrete non-dimensional  $L_2$  norm is

$$L_2(\mathbf{q}, \mathbf{q}^*) = \left( \frac{\sum_j (q_j - q_j^*)^2}{\sum_j q_j^2} \right)^{\frac{1}{2}} \quad (22)$$

where the time-step superscripts were suppressed for simplicity.

For a quantity  $q$  with a vector of its values at the  $n^{\text{th}}$  time step  $\mathbf{q}^n$ , the total amount of the quantity is approximated by  $C(\mathbf{q}^n)$ . The method for this is the same as the method described by Zoppou et al. [7], which

has a higher order of accuracy than the numerical method used to solve the gSGNE. Using the numerical approximation to the total amounts, the conservation error is obtained using

$$C_1(\mathbf{q}^0, \mathbf{q}^n) = \begin{cases} \frac{|C(\mathbf{q}^0) - C(\mathbf{q}^n)|}{|C(\mathbf{q}^0)|} & , \quad |C(\mathbf{q}^0)| > 0 \\ |C(\mathbf{q}^0) - C(\mathbf{q}^n)| & , \quad |C(\mathbf{q}^0)| = 0 \end{cases} . \quad (23)$$

## 5.2. Analytic Solutions

The analytic solutions used to validate the numerical method, are the solitary travelling wave solution of the SGNE and the dam-break solution of the SWWE. The solitary travelling wave solution is a smooth travelling wave solution, where the non-linear and dispersive terms are in balance in the gSGNE. Whereas the dam-break solution of the SWWE demonstrates the robustness of the method in the presence of steep gradients.

### 5.2.1. SGNE - Solitary Travelling Wave Solution

When  $\beta_1 = 2/3$  and  $\beta_2 = 0$  the gSGNE are equivalent to the SGNE which admit the following travelling wave solution [20]

$$h(x, t) = a_0 + a_1 \text{sech}^2(\kappa(x - ct)), \quad (24a)$$

$$u(x, t) = c \left( 1 - \frac{a_0}{h(x, t)} \right), \quad (24b)$$

where

$$\kappa = \frac{\sqrt{3a_1}}{2a_0\sqrt{a_0 + a_1}} \quad (24c)$$

and

$$c = \sqrt{g(a_0 + a_1)} \quad (24d)$$

is the speed of the wave.

This travelling wave solution is maintained due to a balance between the dispersive terms and the non-linear terms in the momentum equation (1b). This balance results in a solitary wave that is advected with a constant speed without a change in shape. Validating the numerical solutions for the gSGNE solver using this solution tests the balance between these terms in (1b), and allows us to verify the method's conservation of energy as the solution is smooth. To enable a comparison between the numerical method and the SGNE solver of Zoppou et al. [7] and Pitt [8], the chosen solitary travelling wave parameters were  $a_0 = 1m$  and  $a_1 = 0.7m$  and the acceleration due to gravity at the Earth's surface,  $g = 9.81m^2/s$  was used.

The numerical solution was solved over the domain  $[-200m, 200m]$  from  $t = 0s$  until  $t = 30s$ . The spatial resolution was varied like so  $\Delta x = 400/(100 \times 2^l)$ , where  $l$  was increased from 0 to 12. To satisfy the CFL condition (21) the time step  $\Delta t = \Delta x/(2\sqrt{g(a_0 + a_1)})$  was chosen. The limiting parameter  $\theta = 1.2$ , was chosen to match previous numerical experiments [7, 8].

Example numerical solutions for  $h$ ,  $u$  and  $G$  with  $\Delta x = 400/(100 \times 2^6) \approx 0.06m$  are plotted in Figure 2. These examples demonstrate that the numerical solutions can reproduce the analytic solutions well.

The convergence properties of the method as  $\Delta x$  varies are given in Figure 3 where Figure 3a shows the  $L_2$  norm calculated over the whole domain and Figure 3b shows the  $L_2$  norm at the peak only. The  $L_2$  norm over the whole domain demonstrates that all quantities are approximated with second-order accuracy, and thus the method is second-order accurate. The  $L_2$  norm at the peak was measured by taking the relative difference between the numerical solutions value at the peak of the wave which are given by  $\max(\mathbf{h}^n)$ ,  $\max(\mathbf{u}^n)$  to the analytic values which are  $a_0 + a_1$  and  $ca_0/(a_0 + a_1)$  respectively. The  $L_2$  norm at the peak demonstrates that even with the slope limiting and the diffusion introduced by the flux approximation [9], the numerical method introduces little diffusion and is second-order accurate.

The conservation error  $C_1$  (23) of the numerical solutions as  $\Delta x$  varies is given in Figure 4. This conservation measure demonstrates that the finite volume method conserves  $h$  and  $G$  up to round-off error, which accumulates as  $\Delta x$  increases due to the additional number of calculations. Since  $uh$  and  $\mathcal{E}$  are not conserved quantities in our finite volume method which solves (7a) and (7b), they are not conserved up to round-off error. However,

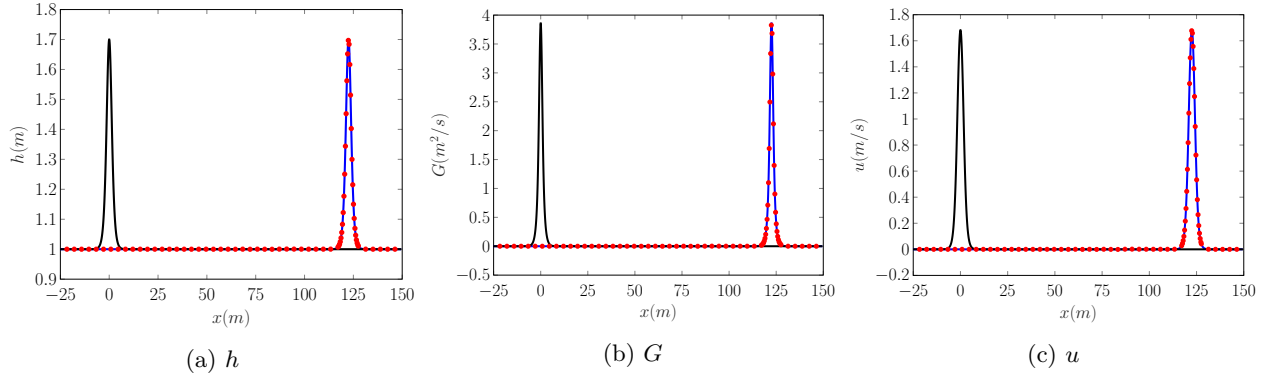


Figure 2: Plot of comparing initial (—), analytic solution (—), and numerical solution with  $\Delta x \approx 0.06m$  (•) at  $t = 30s$ .

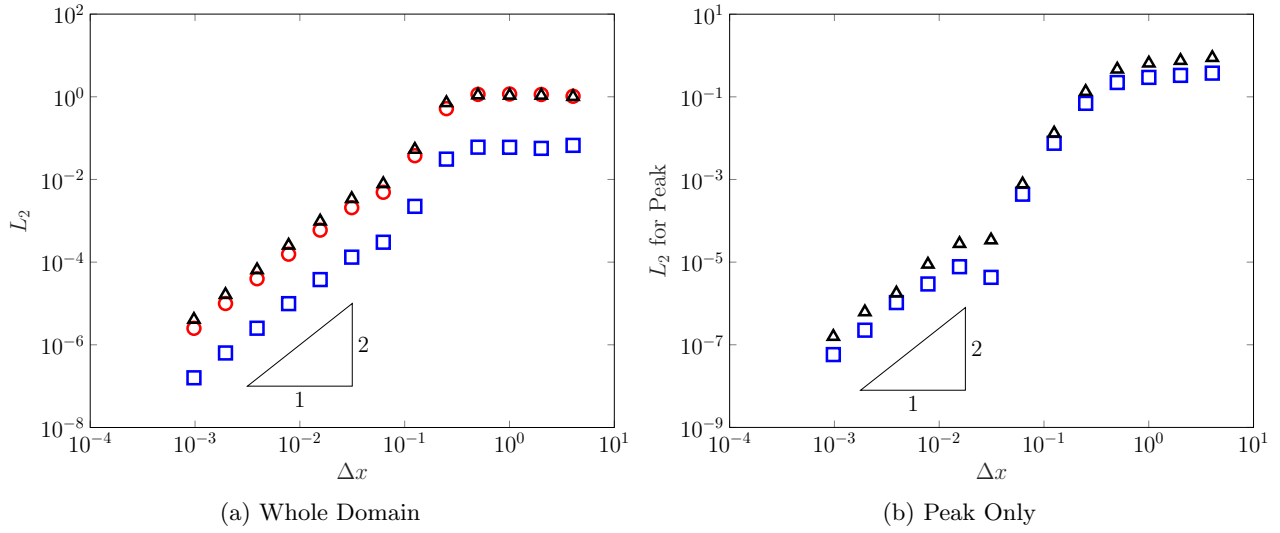


Figure 3: Convergence plots for  $h$  (□),  $G$  (○) and  $u$  (△) as  $\Delta x$  varies.

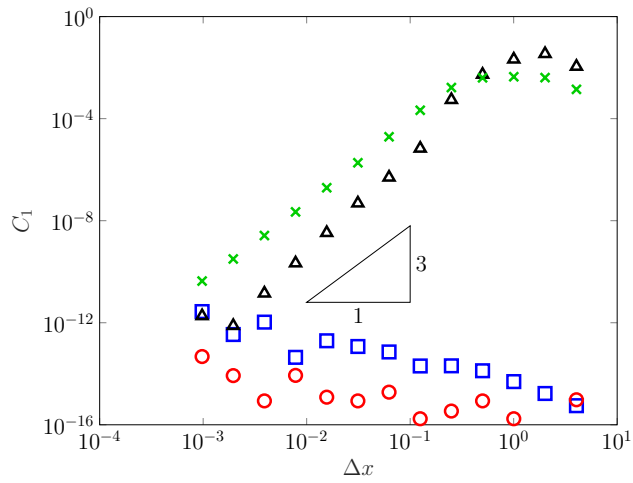


Figure 4: Conservation plot for  $h$  (□),  $G$  (○),  $uh$  (△) and  $\mathcal{E}$  (×) as  $\Delta x$  varies.

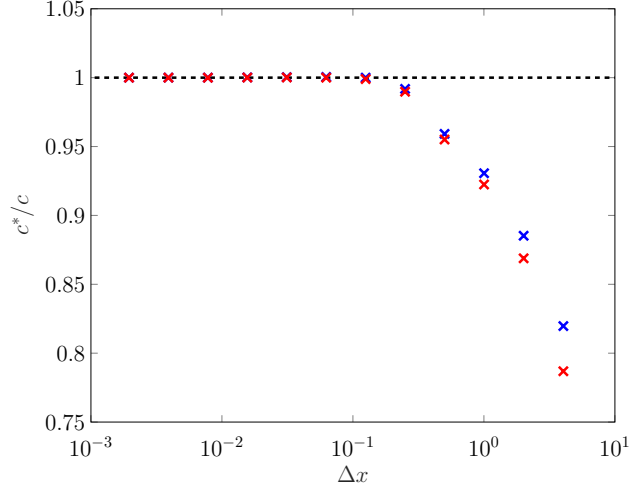


Figure 5: Plot of  $c^*_+/c$  (X),  $c^*_-/c$  (X) and the analytic value (— —) as  $\Delta x$  varies.

the conservation properties of  $uh$  and  $\mathcal{E}$  are good, with both exhibiting conservation errors that have better than second-order convergence in  $\Delta x$ .

By locating the cell on which  $h^n$  achieves its maximum, we can provide an upper and lower bound for the speed of the numerical solution to the travelling wave problem where the lower bound is  $c^*_- = (x_{\text{peak cell}} - 0.5\Delta x)/t$  and the upper bound is  $c^*_+ = (x_{\text{peak cell}} + 0.5\Delta x)/t$ . These upper and lower bounds are compared to the analytic value  $c$  given by (24d) in Figure 5. This figure demonstrates the diffusion of the numerical scheme when  $\Delta x$  is large because both the upper and lower bounds on the wave speed are below the analytic value. However, this diffusion becomes negligible when  $\Delta x$  is small, and for the lowest  $\Delta x$  value we observe that  $c^*_- < c < c^*_+$ , indicating that the peak is travelling at the correct speed up to cell width accuracy.

These results agree well with the numerical solutions of Pitt [8], who compared various numerical methods for the SGNE. This demonstrates that the numerical method solving the gSGNE when  $\beta_1 = 2/3$  and  $\beta_2 = 0$  accurately reproduces the analytic solution of the SGNE.

### 5.2.2. SWWE - Dam-break Solution

When  $\beta_1 = \beta_2 = 0$  the gSGNE equations reduce to the SWWE. Consequently, (7b) and (1b) become identical with  $G = uh$ . The SWWE possess an analytic solution to the dam-break problem given by the initial conditions

$$h(x, 0) = \begin{cases} h_0 & x < 0 \\ h_1 & x \geq 0 \end{cases} \quad (25a)$$

$$u(x, 0) = 0 \quad (25b)$$

$$G(x, 0) = 0. \quad (25c)$$

The solution to the dam-break problem using the conservation of mass (1a) and momentum (1b) equations is given by

$$h(x, t) = \begin{cases} h_0 & , & x \leq -t\sqrt{gh_0} \\ \frac{4}{9g} (\sqrt{gh_0} - \frac{x}{2t})^2 & , & -t\sqrt{gh_0} < x \leq t(u_2 - \sqrt{gh_2}) \\ h_2 & , & t(u_2 - \sqrt{gh_2}) < x \leq tS \\ h_1 & , & tS \leq x \end{cases} \quad (26a)$$

$$u(x, t) = \begin{cases} 0 & , & x \leq -t\sqrt{gh_0} \\ \frac{2}{3} (\sqrt{gh_0} + \frac{x}{t}) & , & -t\sqrt{gh_0} < x \leq t(u_2 - \sqrt{gh_2}) \\ u_2 & , & t(u_2 - \sqrt{gh_2}) < x \leq tS \\ 0 & , & tS \leq x \end{cases} \quad (26b)$$

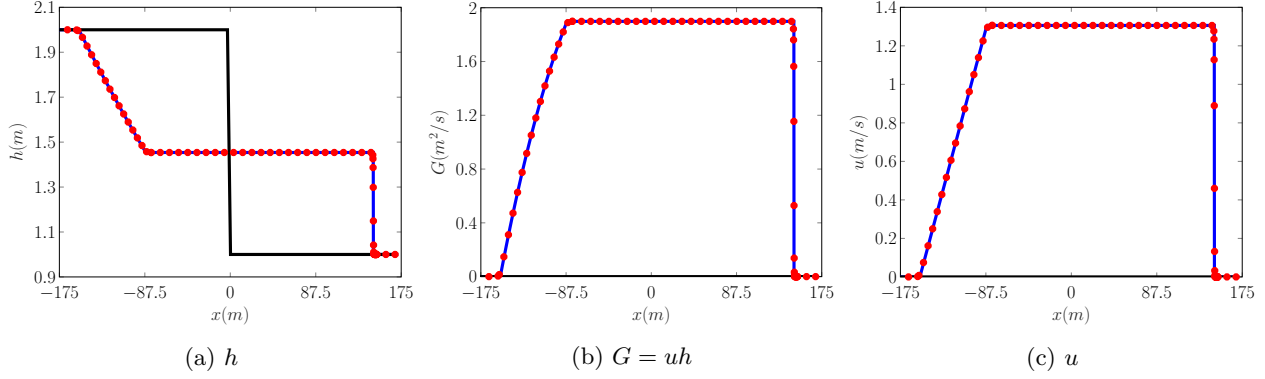


Figure 6: Comparison of initial (—), analytic solution (—), and numerical solution with  $\Delta x \approx 0.15m$  (•) at  $t = 35s$ .

The constant state values  $h_2$  and  $u_2$  and the shock speed  $S$  can be calculated for any initial conditions by solving

$$h_2 = \frac{h_0}{2} \left( \sqrt{1 + 8 \left( \frac{2h_2}{h_2 - h_0} \left( \frac{\sqrt{gh_1} - \sqrt{gh_2}}{\sqrt{gh_0}} \right)^2 - 1 \right)} - 1 \right) \quad (27a)$$

$$u_2 = 2 \left( \sqrt{gh_1} - \sqrt{gh_2} \right), \quad (27b)$$

$$S = \frac{2h_2}{h_2 - h_1} \left( \sqrt{gh_0} - \sqrt{gh_2} \right). \quad (27c)$$

The initial conditions as well as the analytic solution are discontinuous. Due to the discontinuities the solutions to the initial conditions are not unique, as solving the momentum equation (1b) or the energy equation (1c) results in solutions with similar structures but different shock speeds. The solution presented above is solution of the mass (1a) and momentum (1b) equations, as these equations are the basis of the numerical method.

A number of numerical experiments were run for the dam-break problem with  $h_0 = 2m$  and  $h_1 = 1m$ . The domain of the solution was  $[-250m, 250m]$  with a final time of  $t = 35s$ . The spatial resolution was varied like so  $\Delta x = 500/(100 \times 2^l)$  where  $l$  was increased from 0 to 12. To satisfy the CFL condition (21) the time step length  $\Delta t = \Delta x / (2\sqrt{gh_0})$  was used. The limiting parameter  $\theta = 1.0$  and the acceleration due to gravity  $g = 9.81m^2/s$  were used.

Example numerical solutions with the spatial resolution  $\Delta x = 500/(100 \times 2^5) \approx 0.15m$  and the analytic solutions for  $h$ ,  $G$  and  $u$  at the final time are plotted in Figure 6. These figures demonstrate that the method is robust in the presence of steep gradients and accurately reproduces the analytic solution.

The presence of discontinuities in the analytic solution makes accurately assessing convergence of the numerical solutions difficult. To circumvent these issues, the convergence measure has been restricted to comparing the numerical and analytic solutions for the constant region between the rarefaction fan and the shock. This modified convergence measure as  $\Delta x$  is varied is plotted in Figure 7a. This figure demonstrates that the scheme retains its second-order accuracy away from discontinuities, as desired.

Since the analytic solution contains discontinuities all three conservation laws for  $h$ ,  $G$  and  $\mathcal{E}$  are not all satisfied simultaneously [10]. Since we solved equations for  $h$  and  $G$ , these quantities are conserved in the analytic solutions however,  $\mathcal{E}$  is no longer conserved and energy is lost as the shock propagates. Therefore, we introduce a new measure  $\mathcal{E}^*$  which measures the conservation error of the energy by comparing the total energy in the numerical solution with the total energy in the analytic solution at the final time. The conservation error for mass, momentum and energy as calculated normally and the new corrected energy are all compared in Figure 7b. This figure demonstrates that due to the use of the finite volume method even in the presence of discontinuities the conserved quantities  $h$  and  $G$  are conserved up to round-off error, leading to their conservation error increasing as  $\Delta x$  decreases. Energy is not conserved by the analytic solution, and this can be seen as the conservation error of  $\mathcal{E}$  does not improve as  $\Delta x$  decreases. However, when accounting for the lack of energy conservation by using  $\mathcal{E}^*$  we are able to recover first-order accuracy. Given that this solution contains a

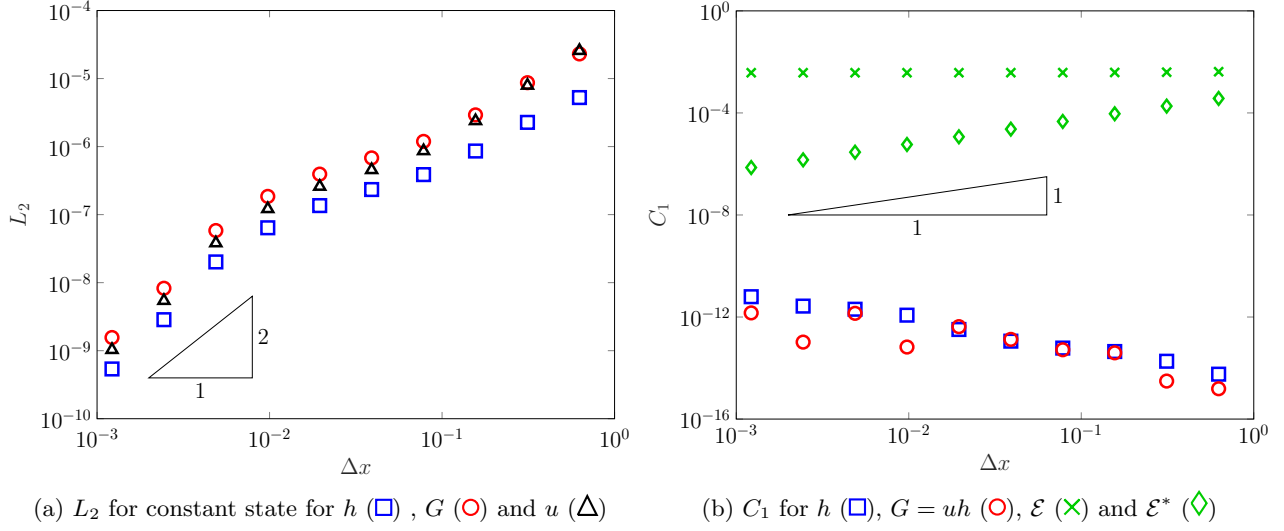


Figure 7: Convergence and conservation plots.

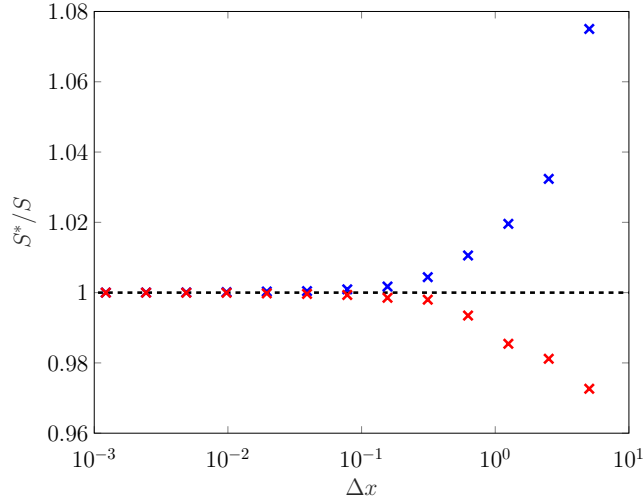


Figure 8: Plot of  $S_+^*/S$  ( $\times$ ),  $S_-^*/S$  ( $\times$ ) and analytic value ( $-$ ) as  $\Delta x$  varies.

discontinuity these are good conservation results. [reference...]

To further justify the ability of the method to resolve the discontinuous analytic solution, I have found lower  $x_{\text{lower}}$  and upper  $x_{\text{upper}}$  bounds for the location of the shock in the numerical solution and used this to bound the shock speed in the numerical solution. For  $x_{\text{lower}}$  this was accomplished by finding the first cell with  $\bar{h}_j \leq h_1 + \frac{9}{10}(h_2 - h_1)$  and for  $x_{\text{upper}}$  this was achieved by finding the first cell with  $\bar{h}_j \leq h_1 + \frac{1}{10}(h_2 - h_1)$ . Consequently, the lower bound for the shock speed  $S_-^* = x_{\text{lower}}/t$  and the upper bound for the shock speed  $S_+^* = x_{\text{upper}}/t$  were calculated and compared to the analytic value  $S$  given by (27c) in Figure 8. This figure demonstrates that as  $\Delta x$  decreases the numerical solutions are better resolving the shock. Furthermore it demonstrates that  $x_{\text{lower}}$  and  $x_{\text{upper}}$  provide a bound for the true location of the shock.

These results demonstrate that the analytic solution of the SWWE has been accurately reproduced by the numerical method which solves the gSGNE when  $\beta_1 = \beta_2 = 0$ .

### 5.3. Forced Solutions

There are currently no known analytic solutions to the gSGNE for other  $\beta$  values. Hence, to demonstrate the validity and versatility of the method to solve the gSGNE for other  $\beta$  values, forced solutions are necessary.

This is vital for the gSGNE equations in particular because for the  $\beta$  values tested above,  $\beta_2$  is always zero, and thus the known analytic solutions do not assess the numerical methods accuracy for the  $\beta_2$  term.

To generate a forced solution the forced gSGNE are considered

$$\frac{\partial h}{\partial t} + \frac{\partial(uh)}{\partial x} = \frac{\partial h^*}{\partial t} + \frac{\partial(u^*h^*)}{\partial x} \quad (28a)$$

$$\begin{aligned} \frac{\partial G}{\partial t} + \frac{\partial}{\partial x} \left( uG + \frac{gh^2}{2} - \beta_1 h^3 \frac{\partial u}{\partial x} \frac{\partial u}{\partial x} - \frac{1}{2} \beta_2 g h^2 \left[ h \frac{\partial^2 h}{\partial x^2} + \frac{1}{2} \frac{\partial h}{\partial x} \frac{\partial h}{\partial x} \right] \right) = \\ \frac{\partial G^*}{\partial t} + \frac{\partial}{\partial x} \left( u^* G^* + \frac{g(h^*)^2}{2} - \beta_1 (h^*)^3 \frac{\partial u^*}{\partial x} \frac{\partial u^*}{\partial x} - \frac{1}{2} \beta_2 g (h^*)^2 \left[ h^* \frac{\partial^2 h^*}{\partial x^2} + \frac{1}{2} \frac{\partial h^*}{\partial x} \frac{\partial h^*}{\partial x} \right] \right). \end{aligned} \quad (28b)$$

The forced gSGNE admit the solutions  $h^*$ ,  $u^*$  and  $G^*$  assuming  $G^*$  satisfies (7c). Since these equations are satisfied for any chosen functions for  $h^*$ ,  $u^*$  and  $G^*$  and any  $\beta$  values, forced solutions can be used to verify the method for a larger class of problems than permitted by currently known analytic solutions. Since the left hand-side of these modified equations are approximated by the numerical method, combining the numerical method with the analytic expressions for the right hand-side, produces a method that approximates the forced gSGNE (28) with the same convergence properties as the underlying numerical method for the gSGNE.

The following forced solution

$$h^*(x, t) = a_0 + a_1 \exp \left( \frac{(x - a_2 t)^2}{2a_3} \right) \quad (29a)$$

$$u^*(x, t) = a_4 \exp \left( \frac{(x - a_2 t)^2}{2a_3} \right) \quad (29b)$$

where  $G^*$  is given by (7c), were used. These forced solutions describe Gaussian bumps in  $h$  and  $u$  that travel at a constant speed  $a_2$ . This forced solution was chosen because it is smooth and the terms in (28) are not constant over the whole domain. Smoothness is necessary for the current description of the forced solutions, since it requires the derivatives to be defined in the classical strong sense. This ensures that the forced solutions include all the terms in the gSGNE when assessing the numerical method.

The particular parameter values  $a_0 = 1$ ,  $a_1 = 0.5$ ,  $a_3 = 20$  and  $a_4 = 0.3$  were chosen in this investigation, while multiple  $\beta$  values were tested we will be focusing on  $\beta_1 = 2/15 + 2/3$  and  $\beta_2 = 2/15$  below. This choice of  $\beta$  values corresponds to the improved SGNE which have a  $\mathcal{O}(k^6)$  accurate dispersion relationship. There is an example numerical solution for these equations in Figure 1 (c).

The numerical solutions were produced over the domain  $[-100m, 100m]$  with a final time of  $t = 10s$ . The spatial resolution was varied like so  $\Delta x = 200/(100 \times 2^l)$  with  $l$  varied from 0 to 11. To satisfy the CFL condition the time step size  $\Delta t = \Delta x / \left( 2 \left[ a_4 + a_2 + \sqrt{g(a_0 + a_1)} \right] \right)$  was chosen. The acceleration due to gravity  $g = 9.81m^2/s$  was used. For the forced solutions the limiting on the reconstruction on  $h$ ,  $u$  and  $G$  was removed. Because these forced solutions are smooth, such reconstruction is not necessary.

Figure 9 compares an example numerical solution at the final time for  $h$ ,  $G$  and  $u$  with  $\Delta x = 200/(100 \times 2^5) \approx 0.06m$  with the analytic solution. These example solutions demonstrate that the numerical method is able to reproduce the forced solution well, validating the numerical methods approximation to all terms in (7).

Figure 10 demonstrates the convergence of the numerical scheme as  $\Delta x$  decreases. All quantities of interest are converging at the expected second-order. Since the right hand-sides of (28) are evaluated analytically, the observed error is caused by the numerical method alone. Therefore, these results demonstrate that the scheme is second-order for all terms in the gSGNE.

## 6. Conclusion

A modified version of the numerical scheme for the SGNE outlined by Zoppou et al. [7] was used to solve the gSGNE [1, 4]. This numerical scheme for the gSGNE was validated by describing a fully second-order



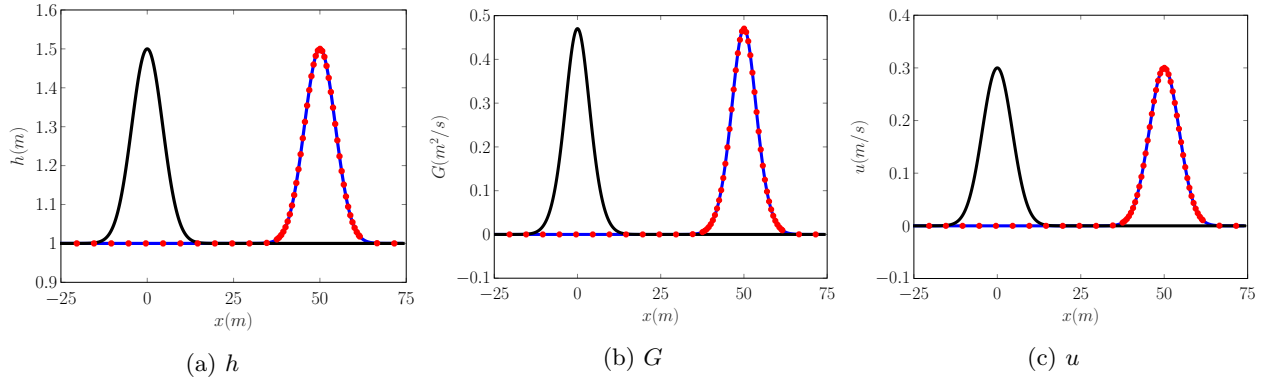


Figure 9: Example plots of initial conditions (—), analytic solution (—), and numerical solution with  $\Delta x \approx 0.06m$  (•).

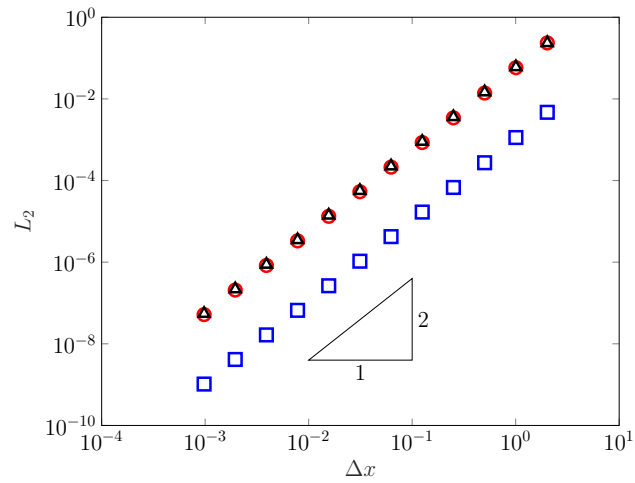


Figure 10: Convergence plot of  $h$  (□),  $G$  (○),  $u$  (△) for the forced solutions for various  $\Delta x$  values.

implementation of the numerical method. This numerical method was validated against analytic solutions of the SGNE and SWWE and forced solutions. The analytic solutions demonstrate that the gSGNE solver accurately reproduces important members of the family of equations described by the gSGNE whilst conserving the quantities of interest. The forced solutions demonstrate that the method remains second-order for all values of the free parameters,  $\beta_1$  and  $\beta_2$ . The gSGNE method described above is the first validated numerical method for the gSGNE. Additionally, the described numerical scheme can be straightforwardly extended to higher-order methods as done by Zoppou et al. [7] for the SGNE. However, an open question remains; the necessity and implementation of limiting in the reconstruction of the derivatives.

## 7. References

- [1] D. Clamond and D. Dutykh. Non-dispersive conservative regularisation of nonlinear shallow water (and isentropic euler equations). *Communications in Nonlinear Science and Numerical Simulation*, 55:237–247, 2018.
- [2] F. Serre. Contribution à l’étude des écoulements permanents et variables dans les canaux. *La Houille Blanche*, 6:830–872, 1953.
- [3] G. B. Whitham. Non-linear dispersion of water waves. *Journal of Fluid Mechanics*, 27(2):399–412, 1967.
- [4] D. Clamond, D. Dutykh, and D. Mitsotakis. Conservative modified Serre-Green-Naghdi equations with improved dispersion characteristics. *Communications in Nonlinear Science and Numerical Simulation*, 45: 245–257, 2017.
- [5] C. Zoppou. *Numerical Solution of the One-dimensional and Cylindrical Serre Equations for Rapidly Varying Free Surface Flows*. PhD thesis, Australian National University, Mathematical Sciences Institute, College of Physical and Mathematical Sciences, Australian National University, Canberra, ACT 2600, Australia, 2014.
- [6] C. Zoppou, S.G Roberts, and J. Pitt. A solution of the conservation law form of the Serre equations. *The Australia and New Zealand Industrial and Applied Mathematics Journal*, 57(4):385–394, 2016.
- [7] C. Zoppou, J. Pitt, and S. Roberts. Numerical solution of the fully non-linear weakly dispersive Serre equations for steep gradient flows. *Applied Mathematical Modelling*, 48:70–95, 2017.
- [8] J.P.A. Pitt. *Simulation of Rapidly Varying and Dry Bed Flow using the Serre equations solved by a Finite Element Volume Method*. PhD thesis, Australian National University, Mathematical Sciences Institute, College of Physical and Mathematical Sciences, Australian National University, Canberra, ACT 2600, Australia, 2019.
- [9] J.P.A. Pitt, C. Zoppou, and S.G. Roberts. Behaviour of the Serre equations in the presence of steep gradients revisited. *Wave Motion*, 76(1):61–77, 2018.
- [10] Y. Pu, R.L. Pego, D. Dutykh, and D. Clamond. Weakly singular shock profiles for a non-dispersive regularization of shallow-water equations. *Communications in Mathematical Sciences*, 16:1361–1378, 2018.
- [11] A. G. Filippini, M. Kazolea, and M. Ricchiuto. A flexible genuinely nonlinear approach for nonlinear wave propagation, breaking and run-up. *Journal of Computational Physics*, 310:381–417, 2016.
- [12] J.S.A. do Carmo, J.A. Ferreira, and L. Pinto. On the accurate simulation of nearshore and dam break problems involving dispersive breaking waves. *Wave Motion*, 85:125 – 143, 2019.
- [13] O. Le Métayer, S. Gavriluk, and S. Hank. A numerical scheme for the Green-Naghdi model. *Journal of Computational Physics*, 229(6):2034–2045, 2010.
- [14] M. Li, P. Guyenne, F. Li, and L. Xu. High order well-balanced CDG-FE methods for shallow water waves by a Green-Naghdi model. *Journal of Computational Physics*, 257(1):169–192, 2014.

- [15] A. Kurganov, S. Noelle, and G. Petrova. Semidiscrete central-upwind schemes for hyperbolic conservation laws and Hamilton-Jacobi equations. *Journal of Scientific Computing, Society for Industrial and Applied Mathematics*, 23(3):707–740, 2002.
- [16] S. Gottlieb, C. Shu, and E. Tadmor. Strong stability-preserving high-order time discretization methods. *Review, Society for Industrial and Applied Mathematics*, 43(1):89–112, 2001.
- [17] S. Roberts and C. Zoppou. Explicit schemes for dam-break simulations. *Journal of Hydraulic Engineering*, 129:11–34, 2003.
- [18] S.D. Conte and C. De Boor. *Elementary numerical analysis: An algorithmic approach*. International Series in Pure and Applied Mathematics. McGraw-Hill Inc., New York, 3rd edition, 1980.
- [19] R. Courant, K. Friedrichs, and H. Lewy. On the partial difference equations of mathematical physics. *IBM Journal of Research and Development*, 11(2):215–234, 1967.
- [20] G.A. El, R.H.J. Grimshaw, and N.F. Smyth. Unsteady undular bores in fully nonlinear shallow-water theory. *Physics of Fluids*, 18(2):027104, 2006.