Energy conservation and physical optics for discrete long wave equations

G. Pedersen

Department of Mathematics, University of Oslo, PO box 1053, 0316 Oslo, Norway

Abstract

In a preceding paper, [7], physical optics was developed for a selection of discrete long wave solutions by application a perturbation expansion. Herein, we approach discrete optics from analysis of discrete energy, in the sense of discrete expressions fulfilling a local energy conservation law. The expressions for energy density and flux are found from finite difference and element discretizations of linear and non-linear shallow water equations as well as linear Boussinesq equations with constant depth.

The discrete energies are generally not perfectly positive definite and ambiguity in their definition is encountered and discussed. Still, the discrete energies appear as well behaved, which is confirmed through calculation of averaged energy quantities for harmonics, that reproduces the physical optics from [7] under the assumption of negligible diffraction. The optics and the derived expressions for the energies are verified through direct solution of discrete long wave equations.

Discrete energies are also discussed in relation to Hamilton's principle. The discrete momentum equation can be obtained from Hamilton's principle inserted energy like quadratures. However, those energy forms are not consistent with the energies that are conserved.

1   Introduction

Among researchers in wave theory a key question concerning numerical solution of wave equations is to what extent the numerical procedure defines a virtual medium with properties that are similar to those of the physical medium. Hence, the performance of numerical methods is often discussed in terms like numerical dispersion, numerical diffusion, spurious reflection etc., rather than more or less abstract error norms. Generally, this is a useful approach, but there are pitfalls as well. Frequently, physical terms are used boldly for numerical models while their role and modified behaviour are not sufficiently investigated.

The present work is a continuation of the preceding article, [7], subsequently referred to as (I), on amplification in shoaling water described by an optical theory derived from a discrete generalization of the WKBJ expansion. One of the key results was a simple numerical counterpart to the well known Greens law. In the absence of a mean current, physical optics can also be approached through energy conservation. The transport equation for the amplitude may be derived employing the assumptions that diffraction is negligible and that the wave locally is a single harmonic mode fulfilling the same dispersion relation and possessing the same averaged energy density and flux, expressed in terms of amplitude and wave numbers, as in homogeneous medium. A crucial question then becomes: Do the transport equations in (I)

possess a corresponding link to some properly defined discrete energy ? However, even though the energy concepts provide an excellent and simple basis for the understanding of physical wave phenomena, their role in connection with numerical solutions is less clear.

There is a tradition in numerical analysis for employing energy-like positive definite forms in stability analysis. In, for instance, [1, 2, 6], a global energy measure is established for a mixed system difference/differential equations in the sense that time is not discretized. Hence, the concept and context is different from the one herein, where we couple local energy densities to amplification and refraction in non-uniform media. In a study of an atmospheric convection model [4] reported a derivation of local discrete energy-conservation laws with similarities to those for wave equations presented herein.

We derive expressions for the discrete energy density and flux from the discretized equations of motion. This is a slightly cumbersome task due to the non-common arithmetics involved and difficulties that arise from ambiguity. Averaged densities are then discussed and related to the discrete optics in (I). We follow the structure of the preceding paper by presenting the most elaborate and general results for a very basic difference method for the shallow water equations, while a wider spectrum of methods is investigated for uniform grid and unidirectional wave propagation. Inevitably, this will involve quite some details on the numerics, even though derivations are put in appendices. Readers that are not particularly interested in the numerics may skip the sections 2.4 and 3.3 through 5 and still receive the general idea. For all cases from (I) we are able find energy laws and reproduce the optical results, even though some caution is needed concerning ambiguity for constant depth. For completeness also energy densities for some nonlinear methods are reported and the discrete energy concept is investigated through Hamilton's principle. It turns out that in the discrete case the energy does not play its usual role in the variational principle.

2 Basic theory

2.1 Scaling and equations

Marking dimensional quantities by a star we introduce a curvilinear, orthogonal coordinate system with horizontal axes $ox^\star$, $oy^\star$ in the undisturbed water level and $oz^\star$ pointing vertically upwards. Moreover, we assume a bottom at $z^\star = -h^\star$ and denote the surface elevation and averaged horizontal particle velocity by and $\eta^\star$ and $\vec{v}^\star$ respectively. Applying the maximum depth, h_0, and a characteristic wavelength, L, as "vertical" and "horizontal" length scales we are then led to the following definition of non-dimensional variables

$$\left.\begin{array}{ll} x^\star = L^\star x, & y^\star = L^\star y, \quad t^\star = L^\star (g h_0^\star)^{-\frac{1}{2}} t, \\ \eta^\star = \alpha h_0^\star \eta, & z^\star = h_0^\star z, \quad \vec{v}^\star = \alpha (g h_0^\star)^{\frac{1}{2}} \vec{v}, \end{array}\right\} \tag{1}$$

where g is the acceleration of gravity and α is an amplitude measure. Moreover, we denote the map factors by m_x and m_y. This means that a line segment $\mathrm{d}x$ on the x-axis corresponds to a physical arc-length $\mathrm{d}x/m_x$ etc. Provided that α and $\beta \equiv (h_0/L)^2$ are sufficiently small, the flow is governed by the shallow water equations

$$\frac{1}{m_x m_y}\frac{\partial \eta}{\partial t} = -\frac{\partial}{\partial x}\left(\frac{hu}{m_y}\right) - \frac{\partial}{\partial y}\left(\frac{hv}{m_x}\right), \tag{2}$$

$$\frac{\partial u}{\partial t} = -m_x\frac{\partial \eta}{\partial x}, \quad \frac{\partial v}{\partial t} = -m_y\frac{\partial \eta}{\partial y}. \tag{3}$$

We recognize the former as the depth integrated continuum equation, whereas the latter two prescribe momentum conservation. In addition to the above equations we will also employ nonlinear and dispersive equations for plane waves (one horizontal direction, only) and unitary map factors. Taking into account terms of order β we obtain a set of Boussinesq equations (see, for instance, [13] and [9])

$$\frac{\partial \eta}{\partial t} = -\frac{\partial}{\partial x} \left((h + \alpha\eta)u \right) \tag{4}$$

$$\frac{\partial u}{\partial t} + \frac{1}{2}\alpha\frac{\partial u^2}{\partial x} = -\frac{\partial \eta}{\partial x} + \frac{1}{2}\beta h\frac{\partial^2}{\partial x^2}\left(h\frac{\partial u}{\partial t} \right) - \frac{1}{6}\beta h^2 \frac{\partial^3 u}{\partial^2 x \partial t}, \tag{5}$$

Multiplying the two equations in (3) by $hu/m_x m_y$ and $hv/m_x m_y$, respectively, adding the results and and invoking the continuum equation we derive the energy equation

$$\frac{\partial E}{\partial t} + \nabla \cdot \vec{F} = 0, \tag{6}$$

where

$$E \equiv \frac{1}{2m_x m_y}\left(h(u^2 + v^2) + \eta^2 \right), \quad \vec{F} \equiv \eta\left(\frac{hu}{m_y}\vec{i} + \frac{hv}{m_x}\vec{j} \right), \tag{7}$$

and $\vec{v} \equiv u\vec{i} + v\vec{j}$. Naturally, the terms in E are kinetic and potential energy, respectively, while it is easily realized that $\vec{F}$ is the effect of the pressure work and the advection of potential energy. This process of identifying each of the terms and assure that they inherit every physical transport mechanism is crucial when conservation laws are derived from a set of governing equations. In fact, for any twice differentiable vector field $\vec{G}$, (6) implies a new conservation law with the modified density $E_m = E + \nabla \cdot \vec{G}$ and the modified flux $\vec{F}_m = \vec{F} - \partial\vec{G}/\partial t$. Hence, to name (6) as an energy equation we must justify that E and $\vec{F}$ correspond exactly to the energy density and flux. This is generally rather straightforward. On the other hand, when seeking discrete conservation laws we cannot rely on exact physical interpretations. We will face this problem subsequently.

We note that the energy density in (7) is not the energy per area in the "physical", horizontal plane, but the energy per area in the curvilinear coordinate system. Naturally, the flux density, $\vec{F}$, is defined accordingly. In the subsequent sections we will define also wave numbers and group velocity with respect to the curvilinear system rather than the physical plane. Still, it is easily realized that the quantities maintain their usual meaning and interrelations. Naturally, this is also implicit in the discrete results given subsequently, that will reproduce the continuous case when the grid increments approach zero.

The analysis of energy conservation in relation to Boussinesq equations is far from straightforward. While the energy is conserved to $O(\beta^2)$ most formulations, as (45), are not exactly conservative. The linearized version, on the other hand, is conservative, but the occurrence of a variable depth in the dispersion terms leads to extremely complex arithmetics for the discrete energy. Hence, we confine the analysis to constant depth and $\alpha = 0$ and find the energy densities

$$E = \frac{1}{2}hu^2 + \frac{1}{2}\eta^2 + \frac{\beta}{6}h\left(\frac{\partial \eta}{\partial t} \right)^2, \quad U = \eta hu - \frac{\beta}{3}h^3 u\frac{\partial^2 u}{\partial x \partial t}, \tag{8}$$

where the $O(\beta)$ terms in E and U are the contributions from the vertical velocity component and non-hydrostatic part of the pressure, respectively. On the other hand, retaining non-linear

terms and variable depth, while deleting the terms of order β, we find

$$E = \frac{1}{2}\left((h+\alpha\eta)\,u^2 + \eta^2\right), \quad U = (h+\alpha\eta)u(\eta + \frac{1}{2}\alpha u^2). \tag{9}$$

We now observe that the $O(\alpha)$ terms stem from correction of the height of fluid columns and advection of kinetic energy.

2.2 Discrete formalism

In the present paper we perform a lot of arithmetics on discrete quantities. We employ the same simple formalism as in (I) that is also similar to that of [4]. For completeness we describe this formalism, with a few extensions, also herein. The approximation to a quantity f at a grid-point with coordinates $(\beta\Delta x, \gamma\Delta y, \kappa\Delta t)$ where Δx, Δy and Δt are the grid increments, is denoted by $f_{\beta,\gamma}^{(\kappa)}$. To make the difference equations more compact and legible we introduce the symmetric difference and average operators, δ_x and $^{-x}$ by

$$\delta_x f_{\beta,\gamma}^{(\kappa)} = \frac{1}{\Delta x}(f_{\beta+\frac{1}{2},\gamma}^{(\kappa)} - f_{\beta-\frac{1}{2},\gamma}^{(\kappa)}), \quad (\overline{f}^x)_{\beta,\gamma}^{(\kappa)} = \frac{1}{2}(f_{\beta+\frac{1}{2},\gamma}^{(\kappa)} + f_{\beta-\frac{1}{2},\gamma}^{(\kappa)}). \tag{10}$$

We note that the differences and averages are defined at intermediate grid locations as compared to f. Difference and average operators with respect to the other coordinates y and t are defined correspondingly. It is easily shown that these operators are commutative in all combinations. To abbreviate the expressions further we also group terms of identical indices inside square brackets, leaving the super- and subscripts outside the right bracket. A "leap-frog" difference can be expressed as the combination of the average and difference in (10), but it is sometimes convenient to have particular notation for the "leap-frog" average

$$[\overline{f}^{L_x}]_{\beta,\gamma}^{(\kappa)} = \frac{1}{2}\left(f_{\beta+1,\gamma}^{(\kappa)} + f_{\beta-1,\gamma}^{(\kappa)}\right). \tag{11}$$

In nonlinear terms and the expressions for discrete energy it is convenient to introduce a special notation for the squared temporal geometrical mean

$$[f^{(*2)}]^{(n)} \equiv f^{(n+\frac{1}{2})}f^{(n-\frac{1}{2})}. \tag{12}$$

Using the above definitions we derive a series of relations that are useful for calculation of the discrete energy quantities. Some of the relations are given in the appendices together with a few other details.

2.3 Difference equations; standard method

For the set (2) and (3) we employ the same standard discretization as in [7]. With $\eta_{i,j}^{(n)}$, $u_{i+\frac{1}{2},j}^{(n+\frac{1}{2})}$ and $v_{i,j+\frac{1}{2}}^{(n+\frac{1}{2})}$ as primary unknowns the difference equations read

$$[M\delta_t\eta = -\delta_x(h^{(u)}\gamma_y u) - \delta_y(h^{(v)}\gamma_x v)]_{i,j}^{(n+\frac{1}{2})}, \tag{13}$$

$$[\delta_t u = -m_x\delta_x\eta]_{i+\frac{1}{2},j}^{(n)}, \quad [\delta_t v = -m_y\delta_y\eta]_{i,j+\frac{1}{2}}^{(n)}, \tag{14}$$
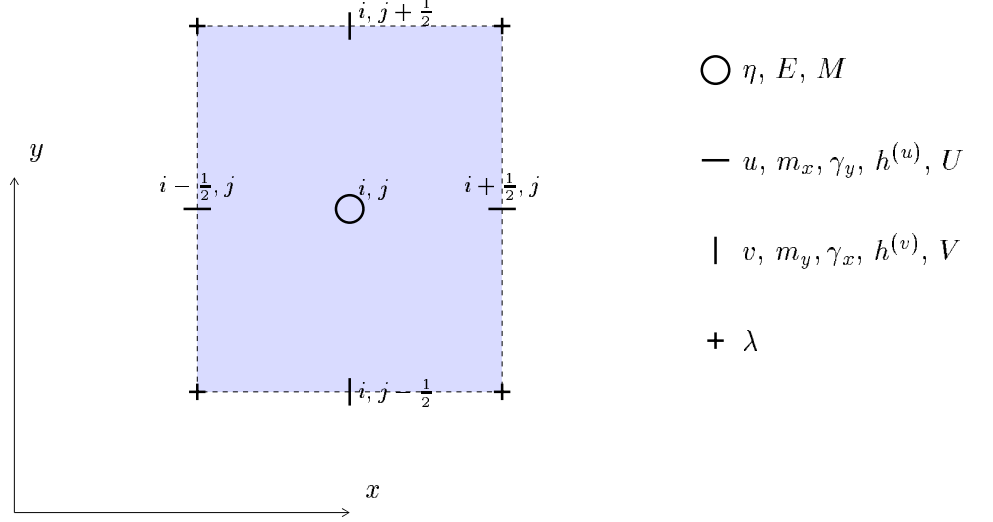
4

Figure 1: Control cell, in xy plane, for mass and energy with generic indices. Location of discrete quantities are shown to the right.

where M, γ_x and γ_y are some discrete representations of $1/(m_x m_y)$, $1/m_x$ and $1/m_y$. In (I) they needed to be of second order. We will retain this requirement even though it is not equally crucial for all purposes herein. Correspondingly, $h^{(u)}$ and $h^{(v)}$ are approximations to h at grid sites for u and v respectively. A grid cell with enumeration and location of variables, including the λ that will be defined in appendix B.2, is shown in figure 1.

For constant h and Cartesian coordinates, that correspond to constant map factors, there exist harmonic solutions with amplitudes A, $\hat{u}$ and $\hat{v}$ for η, u and v respectively. Denoting the wavenumber by $\vec{k} = k\vec{i} + \ell\vec{j}$ we then obtain the discrete dispersion and amplitude relations

$$\tilde{\omega}^2 = h\left(m_x^2\tilde{k}^2 + m_y^2\tilde{\ell}^2\right), \quad \hat{u} = \frac{m_x\tilde{k}}{\tilde{\omega}}A, \quad \hat{v} = \frac{m_y\tilde{\ell}}{\tilde{\omega}}A, \tag{15}$$

where the involved quantities are defined according to

$$\tilde{k} \equiv \frac{2}{\Delta x}\sin\left(\frac{k\Delta x}{2}\right), \quad \tilde{\ell} \equiv \frac{2}{\Delta y}\sin\left(\frac{\ell\Delta y}{2}\right). \tag{16}$$

We note that $\tilde{k} = k + O(\Delta x^2)$ etc. and that the dispersion relation of the differential equations, (2) and (3), is reproduced to second order in the grid increments. The stability criterion, that readily follows from (15), reads

$$h\Delta t^2 \leq \frac{\Delta x^2 \Delta y^2}{m_y^2\Delta x^2 + m_x^2\Delta y^2}. \tag{17}$$

From (15) we may also obtain the formal group velocity

$$\vec{c}_g \equiv \frac{\partial\tilde{\omega}}{\partial k}\vec{i} + \frac{\partial\tilde{\omega}}{\partial l}\vec{j} = \frac{h}{C_t\tilde{\omega}}\left(m_x^2 C_x\tilde{k}\vec{i} + m_y^2 C_y\tilde{\ell}\vec{j}\right), \tag{18}$$

where the quantities C_x, C_y and C_t are defined according to

$$C_t \equiv \cos(\frac{\omega \Delta t}{2}), \quad C_x \equiv \cos(\frac{k \Delta x}{2}), \quad C_y \equiv \cos(\frac{\ell \Delta y}{2}). \tag{19}$$

2.4 Other methods investigated

For the set (45), or simplifications therefrom, we employ the same selection of methods as (I), except that we now include some non-linearity for completeness. Like [3] and [8] we discretize the advection term by a geometric mean. Using mid point representation for the remaining terms we generalize the scheme in (13,14) according to

$$[\delta_t \eta = -\delta_x(h + \alpha \overline{\eta}^{xt})u)]_i^{(n+\frac{1}{2})}, \tag{20}$$

$$\left[\delta_t u + \frac{1}{2}\alpha \delta_x \overline{u^{(*2)}}^x = -\delta_x \eta + \frac{1}{3}\beta h \delta_x^2(h \delta_t u) - \frac{1}{6}\beta h^2 \delta_x^2 \delta_t u\right]_{i+\frac{1}{2}}^{(n)}, \tag{21}$$

where we have replaced $h^{(u)}$ by h. The purpose of the geometric mean is to avoid nonlinear implicit terms in the time stepping.

While (20,21) extends (13,14), the Crank-Nicholson scheme inherits a different time discretization

$$[\delta_t \eta = -\delta_x Q]_j^{(n+\frac{1}{2})}, \quad [Q \equiv (h + \alpha \overline{\eta}^{xt})\overline{u}^t]_{j+\frac{1}{2}}^{(n+\frac{1}{2})}, \tag{22}$$

$$[\delta_t u + \frac{1}{2}\alpha \delta_x \overline{u^2}^{xt} = -\delta_x \overline{\eta}^t]_{j+\frac{1}{2}}^{(n+\frac{1}{2})}. \tag{23}$$

We note that this method is staggered in space but not in time and that we have omitted the dispersion terms for simplicity.

Discrete version of Green law were established for two different finite element schemes in (I). For both we omit nonlinear and dispersive terms. The first method is of a standard type that mimic a staggered grid by use of different trial functions for the surface elevation and the velocities. A derivation of the method is sketched in appendix B.1.5. Expressed in terms of finite differences it reads

$$[\delta_t \eta = -\delta_x(hu)]_j^{(n+\frac{1}{2})} \text{ (a)}, \quad [(1 + \frac{1}{6}\Delta x^2 \delta_x^2)\delta_t u = -\delta_x \eta]_{j+\frac{1}{2}}^{(n)} \text{ (b)}. \tag{24}$$

The other finite element technique is based on a Bernoulli equation expressed with the velocity potential as unknown, instead of a momentum equation for the velocity. Details are given in [3] and (I). From nodal values of the velocity potential, $\phi_j^{(n+\frac{1}{2})}$, velocities are defined by $u \equiv \delta_x \phi$ and we are led to

$$[(1 + \frac{1}{6}\Delta x^2 \delta_x^2)\delta_t \eta = -\delta_x(hu)]_j^{(n+\frac{1}{2})} \text{ (a)}, \quad [\delta_t u = -\delta_x \eta]_j^{(n)} \text{ (b)}. \tag{25}$$

For the Crank-Nicholson method we obtain relations inherent in (15), with $m_x = 1$ and $\hat{v} = \tilde{\ell} = 0$, save that $\tilde{\omega}$ must be replaced by $\frac{2}{\Delta t}\tan(\frac{1}{2}\omega \Delta t)$. The group velocity then becomes $c_g = C_t^2 C_x h^{\frac{1}{2}}$.

6

Dispersion relations for the linearized version of the Boussinesq equations (20,21), and the element methods, (24) and (25), may all be stated in the form

$$\tilde{\omega} = \frac{h^{\frac{1}{2}}\tilde{k}}{\left(1 + \mu\tilde{k}^2\right)^{\frac{1}{2}}},$$ (26)

where $\mu = \frac{1}{3}\beta h^2$ for (20,21) and $\mu = -\frac{1}{6}\Delta x^2$ for both the element methods. The amplitude relation may be expressed $\hat{u} = \tilde{\omega}A/(h\tilde{k})$ for the linearized Boussinesq equations and the second element formulation, (25), while the first element method, (24), yields $\hat{u} = \tilde{\omega}A/((h + \frac{1}{6}\Delta x^2)\tilde{k})$. The group velocity belonging to a dispersion relation $\tilde{\omega} = F(\tilde{k})$, like (26), is readily found as $c_g = C_x F'(\tilde{k})/C_t$.

2.5   Optics

Geometrical and physical optics are standard methods for waves propagating in a slowly varying medium. Rather complete treatises are given in, for instance, [10] and [5]. In (I) a discrete optics was developed by means of perturbation techniques applied to the finite difference equations (13,14) as well as the linear versions of the methods described in section 2.4. We will briefly refer some results that are relevant in the present context.

The analysis of (I) was confined to strict periodicity in time, though the methods employed could have been extended to include a slow temporal modulation. For propagation in shoaling water there exists a stopping depth, h_c, where full reflection takes place. This minimum depth depends on the frequency, ω, and the grid increments. The WKBJ-like expansion gave the transport equation

$$\nabla \cdot \left(\frac{A^2 h}{m_x m_y}\left(m_x^2 C_x \tilde{k}\vec{i} + m_y^2 C_y \tilde{\ell}\vec{j}\right)\right) = 0,$$ (27)

where A is the local amplitude. The analytic counterpart, obtained by letting $\Delta x, \Delta y, \Delta t \to 0$, has the interpretation that the energy flux along a beam of rays is constant. We will subsequently investigate if a similar statement holds for (27). For normal incidence and uniform grids discrete Green's laws were obtained according to

Method	$\frac{A}{B}(h - h_c)^{\frac{1}{4}}$	h_c
Standard case $(13, 14)$	1	$\frac{\tilde{\omega}^2 \Delta x^2}{4}$
Lin. Boussinesq $(20, 21)$	$\left(1 - \frac{\beta}{3}h\tilde{\omega}^2\right)^{-\frac{1}{2}}\left(1 - \frac{\beta}{3}(h + h_c)\tilde{\omega}^2\right)^{-\frac{1}{4}}$	$\frac{\Delta x^2 \tilde{\omega}^2}{2\left(1 + \left(1 - \frac{\beta}{3}\Delta x^2 \tilde{\omega}^4\right)^{\frac{1}{2}}\right)}$
Lin. Crank-Nic. $(22, 23)$	1	$\frac{\Delta x^2 \tan^2\left(\frac{1}{2}\omega\Delta t\right)}{\Delta t^2}$
Element method (24)	$\frac{h}{h + \frac{1}{6}\Delta x^2 \tilde{\omega}^2}$	$\frac{\tilde{\omega}^2 \Delta x^2}{12}$
Element method (25)	1	$\frac{\tilde{\omega}^2 \Delta x^2}{12}$

(28)

where B is a constant. The analytic Green's law is consistent with claiming that the temporal mean of the energy flux is constant. One of the main objectives of the present paper is to investigate a corresponding claim for the discrete case.

7

For the hydrostatic case (28) implies that the method (25) reproduces the analytic result most closely. However, this method is inferior to the standard method concerning numeric dispersion as well as by inheriting an unfavourable stability criterion. Anyhow, in the present paper we will not be primarily concerned with the overall performance of the different methods, but focus on the existence and properties of discrete energy.

3 Discrete energy account

3.1 Conservation of discrete energy

We seek a discrete counterpart to (6) where the involved quantities, as far as possible, maintain their interpretations as energy density and flux. First the locations of the nodes for the discrete density and fluxes must be decided. Certainly this is a crucial step in the procedure. It may be argued that no definition of such nodes are needed and that the energy can be given as a sum over the entire domain. However, we intend to discuss local conservation of discrete energy and it is then convenient to establish controle volums with an energy density node at the centre and flux nodes at the surfaces. Guided by the structure of the C-grid and (7) we denote the x and y components of the energy flux by U and V respectively and seek the discrete quantities

$$E_{i,j}^{(n)}, \quad U_{i+\frac{1}{2},j}^{(n+\frac{1}{2})}, \quad V_{i,j+\frac{1}{2}}^{(n+\frac{1}{2})}.$$

The numerical counterpart to (6) then reads

$$[\delta_t E + \delta_x U + \delta_y V = 0]_{i,j}^{(n+\frac{1}{2})}. \tag{29}$$

According to the discussion at the end of section 2.1 we must expect problems connected with ambiguity in this equation. When $\Delta x, \Delta y, \Delta t \to 0$ the discrete fluxes and densities must approach those in (7). Beyond this requirement the physical identification of energy cannot lead us to the "correct" choice for E, U and V (see section 3.3).

Different procedures can be employed for the determination of E, U and V. Herein we present the most illustrative one, which is also closest to the standard approach for deriving energy equations from differential equations for momentum and mass conservation. Starting with multiplying the discrete components of the momentum equation with the volume flux components, we then use (13), (14) and the identities of appendix A to change the terms into total differences. The procedure is a little bit tricky, but performing our calculations as straightforward as possible we may then hope to obtain a useful result. In view of the problem with non-uniqueness and the somewhat uncommon arithmetics involved, we sketch the calculations briefly in appendix B.1. We finally arrive at a conservation law of type (29) with

$$E = \overline{\left(\frac{\gamma_y h^{(u)}}{2m_x}u^{(*2)}\right)}^x + \overline{\left(\frac{\gamma_x h^{(v)}}{2m_y}v^{(*2)}\right)}^y + \frac{M}{2}\eta^2, \tag{30}$$

$$U = \gamma_y h^{(u)}\overline{\eta}^{xt}u, \quad V = \gamma_x h^{(v)}\overline{\eta}^{yt}v \tag{31}$$

The physical energy density as defined in (7) is obviously positive definite, which implies that the presence of a wave always leads to an increased energy relative to the equilibrium state. On the other hand, the discrete E given in (30) is *not* positive definite due to the geometrical

8

averaging in the kinetic energy. This is most easily demonstrated for plane waves propagating parallel to the x-axis. For a given energy node the right hand side of (30) then involves the corresponding η node and the four neighbouring u nodes in the x-t plane. An inspection of the numerical scheme reveals that all of those may be regarded as independent in the sense that no internal constraints on the five node values are imposed by the difference equations (13) and (14). This, as well as the consequence for positive definiteness, is easily demonstrated through an example. For simplicity we assume that h is constant and that all map factors equal unity. We specify initial conditions for u and η at respectively $t = -\frac{1}{2}\Delta t$ and $t = 0$, say. We are then free to choose, for instance, $u_{I-\frac{1}{2}}^{(-\frac{1}{2})} = u_{I+\frac{1}{2}}^{(-\frac{1}{2})} = 1$, $\eta_{I-1}^{(0)} = -\frac{2\Delta x}{\Delta t}$, $\eta_I^{(0)} = 0$ and $\eta_{I+1}^{(0)} = \frac{2\Delta x}{\Delta t}$. From (14) we then find $u_{I-\frac{1}{2}}^{(\frac{1}{2})} = u_{I+\frac{1}{2}}^{(\frac{1}{2})} = -1$ and $E_I^{(0)} = -\frac{1}{2}h$. It should be noted that in actual numerical simulations, as reported in section 5, the negative energies occur sparsely and with small absolute values.

3.2   Averaged energy densities and physical optics

For a uniform medium the difference equations have simple harmonic solutions as described in section 2.3. Even though the discrete energy density in (30) is not positive definite in general, it is easily shown that the energy density of a stable single harmonic mode is non-negative everywhere.

Since a period generally does not correspond to an integral number of Δt, we define an average of a discrete variable by

$$\langle F \rangle = \lim_{N \to \infty} \frac{1}{N+1} \sum_{n=n_0}^{n_0+N} F^{(n)}. \tag{32}$$

Naturally, a spatial average may be defined accordingly, but the two averages will be equivalent wherever used subsequently. For products of harmonics it is readily shown that the result is independent of n_0 and the spatial location and that the square of a cosine yields $\frac{1}{2}$ for the average etc. (see appendix A). Substitution of a single stable harmonic mode, that obeys the relations (15) and (16), into (30) and (31) and averaging give

$$\langle E \rangle = \frac{1}{2m_x m_y}C_t^2 A^2, \quad \langle U \rangle = \frac{m_x C_t C_x \tilde{k}}{2m_y \tilde{\omega}}hA^2, \quad \langle V \rangle = \frac{m_y C_t C_y \tilde{\ell}}{2m_x \tilde{\omega}}hA^2. \tag{33}$$

We immediately observe that the discrete transport equation, (27), can be expressed as $\nabla \cdot (\langle U \rangle \vec{i} + \langle V \rangle \vec{j}) = 0$, which reflect energy conservation. A trivial substitution confirms that the identities

$$\langle U \rangle = c_{gx}\langle E \rangle, \quad \langle V \rangle = c_{gy}\langle E \rangle, \tag{34}$$

where $\vec{c}_g$ is as given in (18), hold also for discrete harmonics. Accordingly, the transport equation (27) can be rephrased as $\nabla \cdot (\vec{c}_g \langle E \rangle) = 0$.

When a discrete wave mode propagates in a slowly varying bathymetry we may now assume that (33) is locally valid. For plane waves in a two dimensional bathymetry conservation of discrete energy leads to

$$h^{\frac{1}{2}}A^2 C_x = 2\frac{U m_y}{C_t} = \text{const.} \tag{35}$$

9

Employing the discrete dispersion relation (15) we then reproduce the discrete counterpart of Green's law (28). Naturally, the stopping depth, h_c, is the depth at which the group velocity becomes zero, with total reflection as the expected result.

The discrete energy conservation law (29), with involved quantities given by (30) and (31), is valid regardless of the stability criterion (17). When the criterion is violated there exist modes with complex ω. It turns out that for these modes E changes sign periodically in space. This is a necessity for reconciling the growth of instable modes with global conservation of discrete energy. This is most clearly demonstrated for eigen oscillations in a closed basin with no-flux conditions applied at the side walls. For modes with $\text{Im}\,\omega \neq 0$ the total (summed) discrete energy then equals zero.

3.3 Alternative conservation laws

It is an extremely difficult task to detect and classify all possible replacements of (30) and (31). However, a few classes of alternative densities are easily spotted.

The most obvious ambiguity may be linked to any discrete field $G^{(n)}_{i+\frac{1}{2}}$, say, that is of order $\Delta x^2, \Delta t^2$ and that is a function of the field variables and their differences. Naturally, this is a discrete counterpart to the continuous field, also named G, discussed in section 2.1. Alternative energy quantities can then be defined according to $E_m = E + \delta_x G$ and $U_m = U - \delta_t G$. However, when a harmonic is inserted G becomes periodic and it can be shown that the additional terms vanish in the averaging process.

When a set of energy quantities, like (30,31), is found, other sets may readily be found by applying one of the average operators to the density and the fluxes. The new set also fulfills an equation like (29), though at a different grid location. It is obvious that such a procedure corresponds to a redefinition of the control volume. Moreover, in the modified energy conservation law we may rewrite and transfer terms between densities and fluxes, as explained in the preceding paragraph, to obtain compact formulas rather similar to (30,31). Such results may also be obtained directly by modifying the calculations in appendix B.1. Anyhow, any number of average operators applied to E, say, will not change the temporal average (3.2). Hence, we will obtain the same optics from all energy fluxes of this kind.

We may also modify the calculations of appendix B.1 by applying an additional number of, N, say, time averages both to the difference equations (13), (14) and to the volume flux components by which we multiply the latter. Obviously, we end up with replacing η, u and v in (30) and (31) by their N'th order temporal averages. Correspondingly, $\langle E \rangle$, $\langle U \rangle$ and $\langle V \rangle$ will all be multiplied by C_t^{2N} and we still reproduce both (34) and the physical optics. Naturally, also this procedure leads to a shift in the location of the energy nodes for odd N. Correspondingly, when we assume that the depth and the map factors are constant we may also apply the spatial average operators before calculating the energy densities and obtain expressions inherent in (30,31), save that u and η are replaced by spatial averages. Whereas (34) still holds, we will *not* reproduce the results of physical optics. For instance, for the case of normal incidence in a plane bathymetry, employment of a single average with respect to x yields an $\langle U \rangle$ that is C_x^2 times the result in (33). As long as the wave propagates in constant depth the factor is constant, but this is not the case for variable depth. Consequently, $\langle U \rangle$=const. is no longer consistent with the discrete Green's law. Thus, even though the energy flux expressions in constant depth suffice to find transport equations for the amplitude, fluxes and densities that obey a conservation law in constant depth, only, cannot be used. A less obvious example is encountered in section 4 and

appendix B.1.5 for the element formulation (24).

Another kind of ambiguity is inherent already in the differential equations that may possess more than one energy-like conservation law. The shallow water equations may describe problems of various physical origin. One well known example is waves on an elastic membrane. The one-dimensional counterpart, the elastic string, is discussed in appendix C. It yields energy quantities that are very different, even though they do produce the same transport equation for the amplitude. Finally, we note that all linear combination of densities and fluxes that obey (29) will fulfill this relation themselves. A less straightforward calculation than the one in appendix B.1 will easily lead to such a combination. In fact, an approach that was attempted in the early stages of the present work lead to a combination of (33) and the densities of the elastic string.

3.4   The variational approach

Often differential equations may be equivalent to variational principles, where a functional, usually defined as the integral of some Lagrangian density, is required to attain an extreme or display zero variation, at least. Profound discussions of calculus of variations, in the context of wave propagation, are found in [12] and [10].

Variational principles have been widely used as starting point for numerical methods such as global Ritz and finite element techniques, as described in standard textbooks like, for instance, [11]. However, they are seldom used to derive finite difference techniques. In the present paper we regard the topic from a slightly different angle; we employ Hamilton's principle in an attempt to shed more light on the discrete energy concept. A crucial question may read: is a discretized Hamilton's principle, based on the discrete densities in (30), consistent with the the finite difference equations (13,14) ? In that case we have established a sound criterion for proper definition of discrete energy. Unfortunately, as the subsequent derivations will show, this is not the case.

Hamilton's principle for a system reads:

$$\mathcal{D} \int \mathcal{L} \mathrm{d}t = 0, \quad \mathcal{L} \equiv T - \Phi, \tag{36}$$

where T and Φ are the kinetic and potential energy of the system, respectively, and the variations are assumed to vanish at the limits of the integral. For a body of fluid the energies are given as volume integrals of densities and the continuum equation is employed as a constraint. Hamilton's principle will then imply momentum conservation. Application of Hamilton's principle according to (7) and (2) is rather standard, but a brief description is still given in appendix B.2. In the discrete analogue the energies are given as sums and a discrete continuum equation, in our case (13), must be used as constraint. The crucial step is selecting the expressions for the energies, that may be regarded as sums of local densities. Again the discrete calculations become a bit lengthy, as well as rich in details, and are left for appendix B.2. Surprisingly it turns out that $\mathcal{L}$ defined from the false energy density

$$\hat{E} = \overline{\frac{\gamma_y h^{(u)}}{2m_x} u^2}^{xt} + \overline{\frac{\gamma_x h^{(v)}}{2m_y} v^2}^{yt} + \frac{M}{2} \eta^2, \tag{37}$$

will result in the momentum equations (13). However, (37) is not conserved for the numerical scheme in question and is positive definite even for unstable modes. On the other hand, defining

$\mathcal{L}$ from (30) we are led to momentum equations of the form

$$\delta_t \overline{u}^{L_t} = -m_x \delta_x \eta, \tag{38}$$

that corresponds to a method that may be formally stable, but is otherwise ridiculous. It must be concluded that the energy that is locally conserved for a numerical method and the quadratic energy form needed to derive the method from Hamilton's principle generally are different. Hamilton's principle does not help us to establish a firm and unique discrete energy concept.

4   Extensions

As in (I) we have at first confined our analysis to a particular finite difference scheme applied to the linear hydrostatic equations before we proceed to present a selection of extensions. These are mostly the same as in the preceding paper, but also some non-linearity is included. For simplicity all cases are restricted to uniform grids and one dimensional wave propagation and may be listed as follows:

1. *The nonlinear version of the standard scheme; (20,21) with $\beta = 0$.*

$$
\begin{aligned}
E &= \tfrac{1}{2}\overline{(h + \alpha\overline{\eta}^x)u^{(*2)}}^x + \tfrac{1}{2}\eta^2, \\
U &= Q\overline{\eta}^{xt} + \alpha\tfrac{1}{2}\overline{\left(\overline{\overline{Q}^x u^{(*2)}}^{xt} - \tfrac{\Delta x^2}{4}\delta_x(Q\delta_x\overline{u^{(*2)}}^t)\right)}^x,
\end{aligned} \tag{39}
$$

   where $Q = (h + \alpha\overline{\eta}^{xt})u$.

2. *The nonlinear Cranck-Nicholsen method; (22,23).*

$$
\begin{aligned}
&[E = \overline{\left(\tfrac{1}{2}(h + \alpha\overline{\eta}^x)u^2\right)}^x + \tfrac{1}{2}\eta^2]_j^{(n)}, \\
&[U = Q\overline{\eta}^{xt} + \tfrac{1}{2}\alpha\overline{\left(\overline{\overline{Q}^x \overline{u^2}}^{xt} - \tfrac{\Delta x^2}{4}\delta_x(Q\delta_x\overline{u^2})\right)}^x]_{j+\frac{1}{2}}^{(n+\frac{1}{2})},
\end{aligned} \tag{40}
$$

   where $Q = (h + \alpha\overline{\eta}^{xt})\overline{u}^t$. For the linearized version the average density of a harmonic mode becomes $\langle E \rangle = \tfrac{1}{2}A^2$, which is identical to the result for the differential equations.

3. *The linearized Boussinesq equations for constant depth. (20,21) with $\alpha = 0$.*

$$
E = \frac{1}{2}h\overline{u^{(*2)}}^x + \frac{1}{2}\eta^2 + \frac{\beta h}{6}(\delta_t\eta)^{(*2)}, \quad U = hu\overline{\eta}^{xt} - \frac{\beta h^3}{3}u\delta_x\delta_t\overline{u}^{xt}. \tag{41}
$$

   As for the hydrostatic case we find $\langle E \rangle = \tfrac{1}{2}C_t^2 A^2$ for stable modes.

4. *The finite element method, based on velocities; (24).*

$$
\begin{aligned}
E &= \tfrac{1}{2}\overline{hu^{(*2)}}^x + \tfrac{1}{2}\eta^2 + \tfrac{\Delta x^2}{12}\overline{(\delta_t u)^2}^x - \tfrac{\Delta x^4}{72}\overline{(\delta_x\delta_t u)^2}^{xx}, \\
U &= hu\overline{\eta}^{xt} + \tfrac{\Delta x^2}{6}\left(hu\delta_x\delta_t\overline{u}^{xt} + \delta_t\overline{\eta}^x\delta_t\overline{u}^t\right) + \tfrac{\Delta x^4}{36}\overline{\delta_t\overline{u}^{xt}\delta_t^2\delta_x u}^x.
\end{aligned} \tag{42}
$$

   Averaging on stable harmonics yields $\langle E \rangle = \tfrac{1}{2}C_t^2(1 + \tfrac{\Delta x^2\tilde{\omega}^2}{6h})A^2$.

12

5. *The finite element method, based on velocity potential; (25).*

$$E = \frac{1}{2} h \overline{u^{(*2)}}^x + \frac{1}{2}\eta^2 - \frac{\Delta x^2}{12}\overline{(\delta_x\eta)^2}^x, \quad U = h u \overline{\eta}^{xt} + \frac{\Delta x^2}{6}\overline{\eta}^{xt}\delta_x\delta_t\eta. \tag{43}$$

For stable modes we now find $\langle E \rangle = \frac{1}{2}C_t^2(\frac{1}{3} + \frac{2}{3}C_x^2)A^2$.

Details on the derivations are given in appendix B.1. To obtain compact formulas for $\langle E \rangle$ we have employed the numerical dispersion relations as well as amplitude relations like those in (15). For all sets of energy quantities, as given above, direct calculation proves that $\langle U \rangle = c_g\langle E \rangle$ and that $\langle U \rangle =$const. reprodues the WKBJ solutions (27,28) from (I). Energy densities for constant depth is all that is needed to produce the transport equations leading to Green's law. Still, as stated in section 3.3, we cannot employ every set of densities that fulfill the energy equation on constant depth. This was experienced for the element formulation (24) that attains the same form as the Boussinesq model for constant depth (se appendix B.1.5). Generally, we should employ only energy laws that are extendable to variable depths, even though we had noe problems with the linearized Boussinesq equation.

For the linear Crank-Nicholson method we obtain an E that is positive definite, in contrast to the result, (30), for the temporal staggered grid. Naturally, this reflects that the Crank-Nicholson method is unconditionally stable for the linear equations.

5   Verification, tests

For the standard difference scheme (I) verified the discrete optics through comparisons with direct solutions of the finite difference equations. Herein, we will verify the discrete Green's law for the other methods. It is particularly important to check the Boussinesq solution and the element formulation (24), due to the false densities found in constant depth.

As in (I) we confine the study to plane waves and uniform grids, assume temporal periodicity, separate variables, and eliminate one unknown to obtain a second order difference equation. Generally it is most convenient to retain the velocity, rather than the surface elevation as in (I). Writing $u_{j+\frac{1}{2}}^{(n+\frac{1}{2})} = w_{j+\frac{1}{2}}e^{-i\omega(n+\frac{1}{2})\Delta t}$ and $\eta_j^{(n)} = \zeta_j e^{-i\omega n\Delta t}$ we find

$$[(1 + \gamma\tilde{\omega}^2)\delta_x^2(hw) + \nu\tilde{\omega}^2\delta_x^2 w + \tilde{\omega}^2 w = 0]_{j+\frac{1}{2}}, \quad [\zeta = -\frac{i}{\tilde{\omega}}\delta_x\left((h + \mu\tilde{\omega}^2)w\right)]_j, \tag{44}$$

where $\gamma = \nu = 0$ for the standard method, $\gamma = -\frac{1}{2}\beta h$, $\nu = \frac{1}{6}\beta h^2$ for the linearized Boussinesq model and $\gamma = 0$, $\nu = \frac{1}{6}\Delta x^2$ for both element methods. The coefficient μ is $\frac{1}{6}\Delta x^2$ for (25) and zero otherwise. We employ the same test problem as (I) : an incident wave encountering a smooth slope leading up to a shelf (figure 2). Boundary conditions and incident waves are specified as in (I). As seen in figure 3 the discrete Green's law is very accurate for all methods when the slope is moderately steep.

When time dependence is restored in a solution of (44), the real parts of η and u will form a solution to the original difference equations. As an extra check on the arithmetics leading to E and U, we insert the solution into the appropriate expression from section 4 and compute the residual $R = \delta_t E + \delta_x U$. Apart from errors due to finite precision arithmetics R should be zero. Employing 32 bits precision we find residuals ranging from order 10^{-7} for 75 gridpoints (case in figure 3) to 10^{-5} for 1500 gridpoints. Changing to 64 bits precision the residuals are decreased
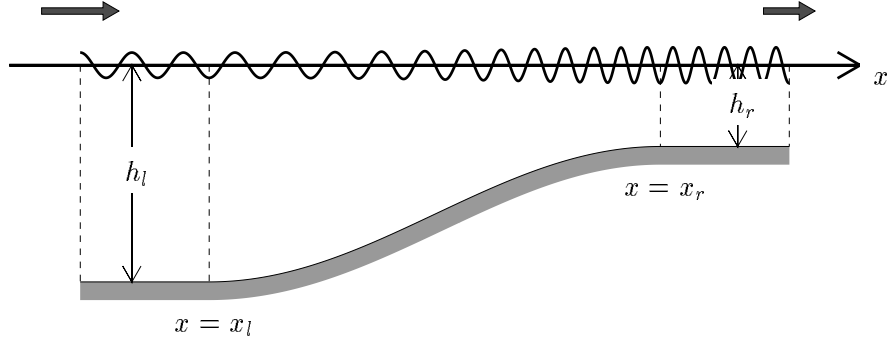
13

Figure 2: Definition sketch of test problem.

by a factor 10^{-8}, say. Thus, it is apparent that a nonzero R is due to round off errors only. For the dispersive case, where we have no exact conservation law in variable depth, we have instead employed a simple harmonic, equal to the incident wave, on constant depth. In the nonlinear case, (20,21) with $\beta = 0$, we cannot employ a single harmonic in time as test case. Instead we specify a pulse by the initial conditions

$$\eta_j^{(0)} = 2A_0 e^{-\frac{(j\Delta x)^2}{\lambda_e^2}}, \quad [\overline{u}^t]_{j+\frac{1}{2}}^{(0)} = 0, \tag{45}$$

that produces an incident waves of approximate amplitude A_0 provided a no-flux condition is invoked at $x = 0$. Again the discrete energy conservation law, this time given by (39), is fulfilled save for finite arithmetics effects. For illustration we have shown a rather extreme example in figure 4, where the wave actually has broken. Since no diffusion or shock capturing terms are included in the computations, discrete energy still is conserved and breaking manifest itself as a disintegration of the wavefront into short oscillations. In view of the strong nonlinearity and the short fluctuations it is noteworthy how close E follows the square of η.

6   Discussion

Following the key steps of the derivation of discrete energy density and flux from the differential equations we were able to find corresponding discrete quantities for all numerical schemes that were investigated. Employment of a set of discrete identities were essential for this procedure. However, it should be noted that all investigated methods are built from purely symmetric differences and averages which imply that all discretization errors are of even order. Hence, they will not likely introduce numerical diffusion that would have made any sort of energy conservation unlikely. Moreover, only a few of the numerous possible representations of nonlinear terms have been studied and the Boussinesq equations were linearized and restricted to constant depth only.

The discrete energy quantities are well behaved in the sense that they approach the corresponding analytic (from differential equations) quantities when the grid increments approach zero. On the other hand, unlike the analytic ones the discrete energy densities are not always positive when the equilibrium state is disturbed. Still, for a single harmonic the local energy density is always non-negative, provided that the stability criterion is fulfilled. Hence, the total energy of, for instance, a closed rectangular basin is always positive when waves are present and
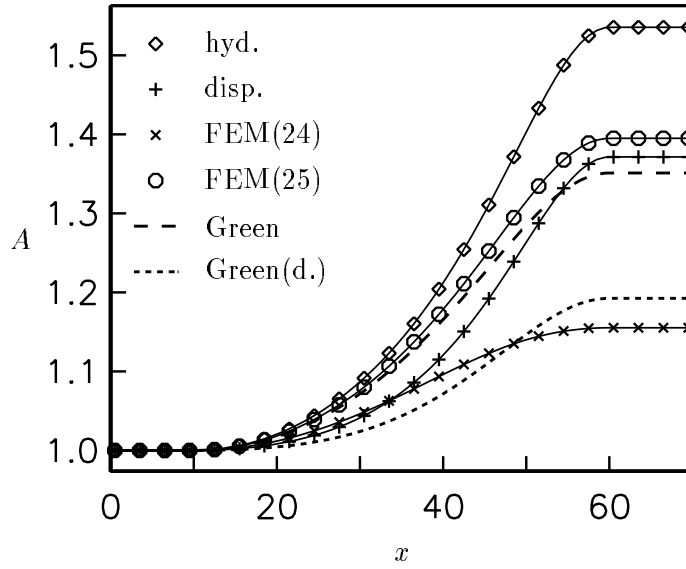
14

Figure 3: Amplitude variation at slope. Exact discrete solutions are marked by symbols for every third node. Optical approximations are depicted with lines, the hydrostatic Green's law, $A = Bh^{-1/4}$, with long dashes and the dispersive Green's law with short dashes. Parameters are $x_l = 10$, $x_r = 60$, $h_l = 1$, $h_r = 0.3$, $\Delta x = \Delta t = 1$, $L^* = h_0^*$ and $\omega = \pi/4$ that yield a wavelength equal to 8 in deep water for the standard method.
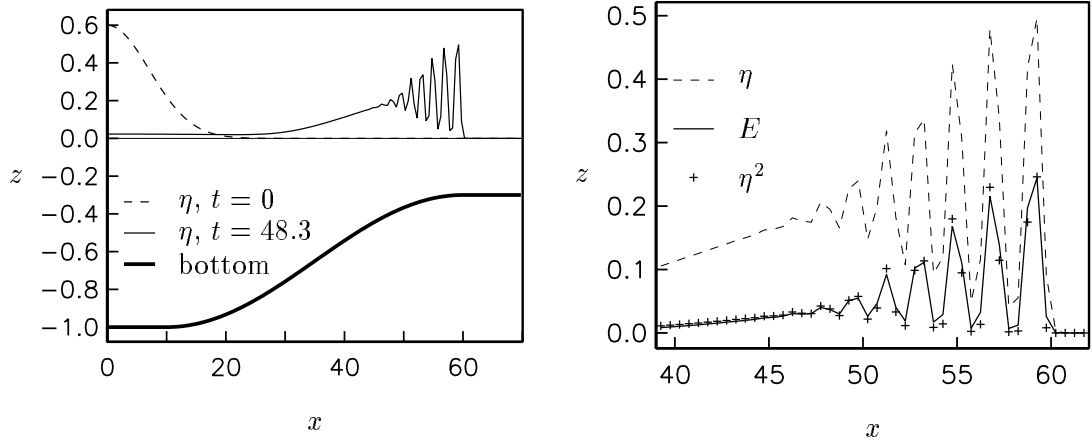

Figure 4: Nonlinear simulation of pulse defined according to $A_0 = 0.3$ and $\lambda_e = 10$ and with additional parameters $\Delta x = 0.5$, $\Delta t = 0.3$, $x_l = 10$, $x_r = 60$ and $h_r = 0.3$. Left panel: initial condition and solution after $t = 48.3$. Right panel: η and E at $t = 48.3$.

we may denote the discrete energy measures as weakly positive definite. For unstable modes the average discrete energy, or total energy in a closed basin, is zero. With the exception of the linear Crank-Nicholson method, that yielded a strictly positive definite energy, the conservation of a discrete energy was no guarantee of stability.

In the linear case the transport equations for the amplitude, as reported by (I), were invariably reproduced by claiming conservation of average discrete energies. But, it must be emphasized that the definition of discrete energies is ambiguous and that energy flux densities from derivations valid for constant depth only, may lead to false results. For variable depth there are still ambiguity, but it is very unlikely that any energy density that is weakly positive definite, as explained above, will lead to erraneous transport equations. Experience with ambiguities that have been encountered supports this presumption. Direct calculations confirmed that the averaged energy flux density equaled the average energy density times the group velocity for all linear cases, while the averages of discrete potential and kinematic energies are generally different. We may now present a quasi physical explanation for the over-amplification according to the discrete Green's law reported by (I). Claiming that $c_g \langle E \rangle$ is constant we now observe that $\langle E \rangle$ is reduced relative to the analytical value by a uniform factor (see (33)), while the corresponding ratio for the group velocities decreases in shoaling water due to the decreasing wavelength. Thus, to maintain a constant flux the discrete amplification must be larger than the analytical one.

An attempt to tie in the discrete energies with a discrete Hamiltons principle did not give the anticipated result. Discrete momentum equations were derived from Hamiltons principle, but they were not consistent with conservation of the energy forms used to construct the Lagrangian.

It must be concluded that a single discrete energy concept, reproducing all important features of physical energy, does not exist. Still, a discrete energy concept may be useful in the contexts of amplification, focusing, artificial reflection, performance of boundaries etc., but must be exercised only with great caution.

A Useful discrete identities

Using the definitions (10) and (12) we develop some relations that simplify the discrete calculations. Some of these are employed also in the references, though sometimes in a different notation. Omitting the dummy specifications of grid sites we may first state the product rules

$$\delta_x(fg) = \overline{g}^x \delta_x f + \overline{f}^x \delta_x g, \quad \overline{fg}^x = \overline{f}^x \overline{g}^x + \frac{\Delta x^2}{4}(\delta_x f)(\delta_x g). \tag{46}$$

It is easily realized that two successive applications of a difference operator produce the standard three point approximation to the second derivative. For the second order average operator we then find

$$\overline{f}^{xx} = f + \frac{\Delta x^2}{4} \delta_x^2 f. \tag{47}$$

Combining (46) and (47) we obtain the useful identity

$$\overline{f \delta_x g}^x = \delta_x(f \overline{g}^x) - g \delta_x f. \tag{48}$$

Another relation, that comes in handy, is obtained in similar manner

$$\overline{g\overline{f}^x}^x = \overline{g}^x f + \frac{\Delta x^2}{4}\delta_x(g\,\delta_x f) \tag{49}$$

We may relate the geometric mean to the arithmetic mean through relations like

$$f^{(*2)} = (\overline{f}^t)^2 - \frac{\Delta t^2}{4}(\delta_t f)^2, \quad (\overline{u}^x)^{(*2)} + \frac{\Delta x^2}{4}(\delta_x u)^{(*2)} = \overline{u^{(*2)}}^x. \tag{50}$$

By means of (46) a product like $\overline{f}^x \delta_x f$ is easily rewritten as the total difference $\delta_x(\frac{1}{2}f^2)$. Invoking also (47) and (50) we derive the related formula

$$f\delta_t(\overline{f}^t) = \delta_t(\frac{1}{2}f^{(*2)}), \tag{51}$$

that will prove itself very helpful.

For a trigonometric function $f_\alpha = A\exp(\mathrm{i}k\alpha\Delta x)$, defined for any real α, application of the difference and average operators yields the simple results

$$\delta_x f = \mathrm{i}\tilde{k}f, \quad \overline{f}^x = C_x f, \quad f^{(*2)} = f^2 \tag{52}$$

where $\tilde{k} \equiv \frac{2}{\Delta x}\sin(\frac{1}{2}k\Delta x)$ and $C_x = \cos(\frac{1}{2}k\Delta x)$ as in section (2.3). When $f_\alpha^{(n)} = A\cos(k\alpha\Delta x - \omega n\Delta t + \theta_0)$ we obtain relations like

$$\langle f^{(*2)}\rangle = (C_t^2 - \frac{1}{2})A^2, \quad \langle \overline{f^2}^x\rangle = \frac{1}{2}A^2, \quad \text{etc.} \tag{53}$$

These relations will be useful for calculating discrete dispersion relations and averaged energy quantities.

For the discrete version of the variational calculus we also need some relations involving sums. From (46) we readily obtain a "summation by part" formula

$$\sum_{n=n_0}^{n_1}[\overline{F}^t\delta_t G]^{(n)} = \frac{1}{\Delta t}\left([FG]^{(n_1+\frac{1}{2})} - [FG]^{(n_0-\frac{1}{2})}\right) - \sum_{n=n_0}^{n_1}[\overline{G}^t\delta_t F]^{(n)}. \tag{54}$$

If F and G are defined at alternating grid positions we may rewrite a product sum

$$\sum_{n=n_0}^{n_1}[F\overline{G}^t]^{(n+\frac{1}{2})} = \frac{1}{2}G^{(n_0)}F^{(n_0+\frac{1}{2})} + \frac{1}{2}G^{(n_1+1)}F^{(n_1+\frac{1}{2})} + \sum_{n=n_0+1}^{n_1}[\overline{F}^t G]^{(n)}. \tag{55}$$

The variation of a sum of geometric means my be transformed according to

$$\begin{aligned}
\mathcal{D}\sum_{n=n_0}^{n_1}[F^{(*2)}]^{(n+\frac{1}{2})} &= \sum_{n=n_0}^{n_1}(F^{(n+1)}\mathcal{D}F^{(n)} + F^{(n)}\mathcal{D}F^{(n+1)}) \\
&= F^{(n_0+1)}\mathcal{D}F^{(n_0)} + F^{(n_1)}\mathcal{D}F^{(n_1+1)} + \sum_{n=n_0+1}^{n_1}\left[\overline{F}^{L_x}\mathcal{D}F\right]^{(n)}.
\end{aligned} \tag{56}$$

17

B Discrete calculations

B.1 Energies

B.1.1 The standard case

We start the derivation of (30) by averaging the x-component of the momentum equation in (14) with respect to t. Then we multiply the resulting equation with $1/m_x$ and the x-component of the flux as represented in (13), namely $\gamma_y h^{(u)} u$. This leads to

$$\frac{\gamma_y h^{(u)}}{m_x} u \delta_t \overline{u}^t = -\gamma_y h^{(u)} u \delta_x \overline{\eta}^t. \tag{57}$$

By means of (51) we recognize the left hand side as a total time difference. Averaging the equation with respect to x we then employ (48) to rewrite the right hand side according to

$$\delta_t \overline{\left(\frac{\gamma_y h^{(u)}}{2m_x} u^{(*2)}\right)}^x = -\overline{\gamma_y h^{(u)} u \delta_x \overline{\eta}^t}^x = -\delta_x(\gamma_y h^{(u)} \overline{\eta}^{xt} u) + \overline{\eta}^t \delta_x(\gamma_y h^{(u)} u). \tag{58}$$

We observe that last term on the right hand side contains a flux difference identical to one term in the discrete continuum equation (29). Applying the corresponding operations to the y-component of (14), adding the results and invoking the discrete continuity equation (13) we recognize the sum of the rightmost term of (58) and its counterpart from the y-component as the time difference of a discrete potential energy. The resulting discrete conservation law is then identified as (29) inserted (30). We emphasize that the discrete derivation follows the most straightforward derivation of (6) closely, apart from the averaging and the slightly increased complexity due to the calculus of differences. Other derivations, that were less straightforward, gave slightly different results.

B.1.2 Nonlinear equations

Deleting the $O(\beta)$ terms in (20,21) we then follow the procedure of appendix B.1.1, multiplying the time average of the momentum equation by the flux, given as $Q \equiv (h + \alpha \overline{\eta}^{xt}) u$ this time, and transforming first the left hand side. The local acceleration term may be rewritten by means of (51) and (46)

$$Q \delta_t \overline{u}^t = \frac{1}{2}(h + \alpha \overline{\eta}^{xt}) \delta_t u^{(*2)} = \delta_t \left(\frac{1}{2}(h + \alpha \overline{\eta}^x) u^{(*2)}\right) - \frac{\alpha}{2} \overline{u^{(*2)}}^t \delta_t \overline{\eta}^x. \tag{59}$$

Applying (48), after recasting the sequence of terms, we may transform the contribution from the advection term

$$Q \overline{\delta_x \frac{1}{2} u^{(*2)}}^{xt} = \delta_x \left(\frac{1}{2} \overline{Q}^x \overline{u^{(*2)}}^{xt}\right) - \frac{1}{2} \overline{\overline{u^{(*2)}}^{xt} \delta_x Q}^x. \tag{60}$$

Adding (59) and (60) we combine the last term of each to a total difference by (20) followed by (49) with $\delta_x Q$ as g and $\overline{u^{(*2)}}^t$ as f. After being averaged with respect to x the right hand side of the momentum equation can be transformed as before, again by means of (48) and (20)

$$\overline{Q \delta_x \overline{\eta}^t}^x = \delta_x \left(Q \overline{\eta}^{xt}\right) - \overline{\eta}^t \delta_x Q = \delta_x \left(Q \overline{\eta}^{xt}\right) + \delta_t(\frac{1}{2} \eta^2). \tag{61}$$

Assembling all terms we then find the expressions in (39).

18

B.1.3 Boussinesq equations

Assuming constant depth in (20,21) and deleting the $O(\alpha)$ terms, we observe that a single dispersion term in the momentum equations has appeared as compared to the case in B.1.1. To calculate the contributions from the this term, an extra application of (48) and (51) is all that we need to find (41)

$$\overline{u\delta_x^2\delta_t\overline{u}^t}^x = \delta_x\left(u\delta_x\delta_t\overline{u}^{xt}\right) - \delta_x\delta_t\overline{u}^t\delta_x u = \delta_x\left(u\delta_x\delta_t\overline{u}^{xt}\right) - \delta_t\left(\frac{1}{2}(\delta_x u)^{(*2)}\right), \tag{62}$$

where the constant factor $\frac{1}{3}\beta h^3$ has been omitted in every term.

B.1.4 The Crank-Nicholsen method

This time we do not average the momentum equation in time, but otherwise the derivation of discrete energies follows the one in appendix B.1.2 closely. Instead of (59) we now find

$$Q\delta_t u = \delta_t\left(\frac{1}{2}(h + \alpha\overline{\eta}^x)u^2\right) - \frac{\alpha}{2}\overline{u^2}^t\delta_t\overline{\eta}^x, \tag{63}$$

while (60) is replaced by

$$Q\delta_x\overline{\frac{1}{2}u^2}^{xt} = \delta_x\left(\frac{1}{2}\overline{Q}^x\overline{u^2}^{xt}\right) - \overline{\frac{1}{2}\overline{u^2}^{xt}\delta_x Q}^x. \tag{64}$$

We observe that $u^{(*2)}$ is replaced by u^2 and that the temporal locations are changed as compared to the staggered scheme in appendix B.1.2. The last terms of the above expressions will again combine to a total difference. After averaging with respect to x the left hand side of the momentum equation transforms as in (61) and we arrive at (40).

B.1.5 Two finite element schemes

For both methods investigated the element j is defined as the interval $(j-\frac{1}{2})\Delta x \leq x \leq (j+\frac{1}{2})\Delta x$. In the first formulation the variables are approximated according to

$$u \approx \tilde{u} = \sum_j u_{j+\frac{1}{2}}(t)N_{j+\frac{1}{2}}(x), \quad \eta \approx \tilde{\eta} = \sum_j \eta_j(t)K_j(x), \tag{65}$$

where $N_{j+\frac{1}{2}}$ is the continuous and piece-wise linear "hat" function, that is 1 at node $j+\frac{1}{2}$ and zero at all other nodes, while K_j is one at element j and zero elsewhere. Choosing different shape functions for different unknowns, in such a manner, is a standard procedure to mimic the use of staggered grids in finite difference methods. Using $N_{j+\frac{1}{2}}$ and K_j as weight functions for the momentum and continuum equation, respectively, we integrate to arrive at the residual formulation

$$\int \frac{\partial \tilde{u}}{\partial t}N_{j+\frac{1}{2}}\mathrm{d}x = \int \tilde{\eta}\frac{\mathrm{d}N_{j+\frac{1}{2}}}{\mathrm{d}x}\mathrm{d}x + .., \quad \int \frac{\partial \tilde{\eta}}{\partial t}K_j\mathrm{d}x = -\int \frac{\tilde{\partial}}{\partial x}(h\tilde{u})K_j\mathrm{d}x, \tag{66}$$

where the momentum equation has been integrated by parts, since $\tilde{\eta}$ is only integrable, and the resulting boundary terms have been omitted. With the same time discretization as the standard

method, (13,14), the difference equivalent of (66) is readily worked out as (24) where the exact definition of $h_{j+\frac{1}{2}}$ depends on the integration rule for the flux term. We note that a mass matrix term, involving δ_x^2 have appeared on the left hand side of the momentum equation. For constant depth (24) becomes equal to linearized Boussinesq equation, provided β is replaced by $-\frac{1}{2}\Delta x^2/h^2$. The densities found for those equations should then apply. However, even though we then have established a discrete energy conservation law, with the correct behaviour in the limit $\Delta x, \Delta t \to 0$, it does *not* reproduce the discrete Greens law. The presumed reason is that this energy conservation cannot be generalized to variable depth, even though we have not proven that such an extension is impossible. Even if the constant depth version provides the quantitative information for physical optics, it must be an instance of a more general conservation law, involving variable depth, for the energy argument to apply in the first place. This objection can be raised also against the treatment of the linearized Boussinesq equations, where we may have been saved by a stroke of luck. Another energy conservation law, that is also valid for variable depth, has been found. Unfortunately, the calculations are a bit lengthy and the result somewhat messy. Again using h for $h^{(u)}$ we follow the standard method in appendix B.1.1 for all terms, apart from the extra term in the momentum equation, due to consistent mass. The treatment of this term may be referred as the employment of a series of identities, where the rightmost term is repeatedly rewritten until it becomes a total temporal or spatial difference. Before each identity we list the specific formulas that are used.

$$
\begin{aligned}
48, 24(a) \quad &: \quad \overline{hu\delta_x^2\delta_t\overline{u}^t}^x = \delta_x\left(\delta_x\delta_t\overline{u}^{xt}hu\right) + \delta_x\delta_t\overline{u}^t\delta_t\eta \\
48 \quad &: \quad \delta_x\delta_t\overline{u}^t\delta_t\eta = \delta_x\left(\delta_t\overline{\eta}^x\delta_t\overline{u}^t\right) - \overline{\delta_t\overline{u}^t\delta_x\delta_t\eta}^x \\
24(b), 46 \quad &: \quad \delta_t\overline{u}^t\delta_x\delta_t\eta = -\tfrac{1}{2}\delta_t\left((\delta_t u)^2\right) - \tfrac{\Delta x^2}{6}\delta_t\overline{u}^t\delta_t^2\delta_x^2 u \\
48, 46 \quad &: \quad \delta_t\overline{u}^t\delta_t^2\delta_x^2 u = \delta_x\left(\delta_t\overline{u}^{xt}\delta_t^2\delta_x u\right) - \tfrac{1}{2}\delta_t\overline{(\delta_t\delta_x u)^2}^x
\end{aligned}
\tag{67}
$$

Backward substitution in the above sequence then produces E and U as defined in (42). There might exist simpler energy laws, but they have not been found by a reasonable effort. For constant depth we now find averaged densities that are $1 + \tilde{\omega}^2\Delta x^2/(6h)$ times the ones obtained by adapting the result from the Boussinesq equations as explained above.

The second method corresponds to (25). This time an extra term has appeared in the continuum equation, as compared to the standard difference scheme. We follow the procedure of the standard method until the rightmost term in (58) is transformed by means of the continuum equation. We then obtain an extra term that is rewritten in a few steps

$$
\begin{aligned}
25(a) \quad &: \quad -\overline{\eta}^t\delta_x(hu) = \delta_t(\tfrac{1}{2}\eta^2) + \tfrac{\Delta x^2}{6}\overline{\eta}^t\delta_x^2\delta_t\eta \\
48, 46 \quad &: \quad \overline{\eta}^t\delta_x^2\delta_t\eta = \delta_x(\overline{\eta}^{xt}\delta_x\delta_t\eta) - \delta_t\overline{\tfrac{1}{2}(\delta_x\eta)^2}^x.
\end{aligned}
\tag{68}
$$

The result is stated in (43). It might be objected that the element methods now have been treated in a "finite difference" fashion, while a more appropriate approach is to define energies through integrals of energy like products of the trial functions, $\tilde{u}$ and $\tilde{\eta}$. However, such a procedure will eventually lead to a quadratic form in nodal values, just as the above procedure, while specific non-trivial choices may be necessary for integration rules etc.

B.2 Calculus of variation

For shallow water waves we may write, according to (7)

$$\mathcal{L} = \int \int \frac{1}{2m_x m_y} \left(h(u^2 + v^2) - \eta^2 \right) \mathrm{d}x \mathrm{d}y.$$

To obtain the momentum equations we must apply Hamilton's principle with the continuity equation, (2) as constraint. This is invoked by the introduction of a Lagrangian multiplier λ. Since we are going to mimic the procedure for the discrete case, we refer the important steps briefly. (36) yields

$$\int \int \int \frac{1}{m_x m_y} \left(h(u\mathcal{D}u + v\mathcal{D}v) - \eta\mathcal{D}\eta \right) \mathrm{d}x \mathrm{d}y \mathrm{d}t = 0, \tag{69}$$

where the variations of the different field variables are dependent. The constraint, (2), is varied, multiplied with λ, integrated over the fluid domain as well as in time and integrated by parts to give

$$\int \int \int \frac{1}{m_x m_y} \left(\frac{\partial \lambda}{\partial t}\mathcal{D}\eta + \frac{\partial \lambda}{\partial x}\frac{h}{m_x}\mathcal{D}u + \frac{\partial \lambda}{\partial y}\frac{h}{m_y}\mathcal{D}v \right) \mathrm{d}x \mathrm{d}y \mathrm{d}t = 0, \tag{70}$$

where we have assumed, for simplicity, that the variations vanish at fluid boundaries. Subtracting the above equations from (69), choosing λ to make the term containing $\mathcal{D}\eta$, say, to vanish, the independence of $\mathcal{D}u$ and $\mathcal{D}v$ then result in

$$\frac{u}{m_x} = \frac{\partial \lambda}{\partial x}, \quad \frac{v}{m_y} = \frac{\partial \lambda}{\partial y}, \quad \eta = -\frac{\partial \lambda}{\partial t}. \tag{71}$$

Elimination of λ then reproduces the momentum equations. We observe that λ itself may be interpreted as a kind of potential.

The first step for the discrete case is the definition of $\mathcal{L}$, or rather the definition of discrete energy. Before employing the expressions in (30) it may be instructive to see the outcome when inserting the simplest possible discrete expressions that could pass for kinetic and potential energy. Employing the expressions in (37) we observe that the arithmetic means will all disappear in the summation process. Hence, we may define a discrete Hamilton's principle according to

$$\mathcal{D} \sum_n \sum_i \sum_j \left(\left[\frac{\gamma_y h^{(u)}}{2m_x}u^2 \right]_{i+\frac{1}{2},j}^{(n+\frac{1}{2})} + \left[\frac{\gamma_x h^{(v)}}{2m_y}v^2 \right]_{i,j+\frac{1}{2}}^{(n+\frac{1}{2})} - \left[\frac{M}{2}\eta^2 \right]_{i,j}^{(n)} \right) = 0. \tag{72}$$

with (13) as constraint. We observe that the constant factor $\Delta x \Delta y \Delta t$ has been omitted. Again we will assume that all variations vanish at the boundaries, whenever needed. Without constraints variational principles of this form are trivial. The corresponding (Euler-Lagrange) difference equations are then obtained simply by requiring that the derivatives with respect to the free node values are all zero. Unfortunately, the presence of the constraint instead make the calculations cumbersome. Anyhow the first step is still simple

$$\sum_n \sum_i \sum_j \left(\left[\frac{\gamma_y h^{(u)}}{m_x}u\mathcal{D}u \right]_{i+\frac{1}{2},j}^{(n+\frac{1}{2})} + \left[\frac{\gamma_x h^{(v)}}{m_y}v\mathcal{D}v \right]_{i,j+\frac{1}{2}}^{(n+\frac{1}{2})} - [M\eta\mathcal{D}\eta]_{i,j}^{(n)} \right) = 0. \tag{73}$$

It is now convenient to multiply the varied continuum equation by $\overline{\lambda}^{xyt}$, instead of λ. This means that λ in principle is specified at the corners of the volume control cell, see figure 1. Alternatively, some extra field variables, needed in the "summation by parts" process, had to be defined. Summation over the grid yields

$$\sum_n \sum_i \sum_j \left[\overline{\lambda}^{xyt} M \delta_t \mathcal{D}\eta + \overline{\lambda}^{xyt} \delta_x(h^{(u)}\gamma_y \mathcal{D}u) + \overline{\lambda}^{xyt} \delta_y(h^{(v)}\gamma_x \mathcal{D}v) \right]_{i,j}^{(n+\frac{1}{2})} = 0. \qquad (74)$$

Next, we need to make a discrete version of integration by parts. Recasting the sequence of summation the the innermost sum of the first term may be rewritten by (54) and (55) according to

$$\sum_n \left[\overline{\lambda}^{xyt} M \delta_t \mathcal{D}\eta \right]_{i,j}^{(n+\frac{1}{2})} = \sum_n \left[M\mathcal{D}\overline{\eta}^t \delta_t \overline{\lambda}^{xy} \right]_{i,j}^{(n+\frac{1}{2})} + \text{b.t.} = \sum_n \left[M\mathcal{D}\eta \delta_t \overline{\lambda}^{xyt} \right]_{i,j}^{(n)} + \text{b.t.} \qquad (75)$$

where b.t. denotes boundary contributions that eventually will disappear due to zero variations at the "summation" boundaries. With x and y, respectively, in role of t the latter two terms of (74) can be handled in the same manner and the constraint becomes

$$\sum_n \sum_i \sum_j \left(\left[M\delta_t \overline{\lambda}^{xyt} \mathcal{D}\eta \right]_{i,j}^{(n)} + \left[\delta_x \overline{\lambda}^{xyt} h^{(u)}\gamma_y \mathcal{D}u \right]_{i+\frac{1}{2},j}^{(n+\frac{1}{2})} + \left[\delta_y \overline{\lambda}^{xyt} h^{(v)}\gamma_x \mathcal{D}v \right]_{i,j+\frac{1}{2}}^{(n+\frac{1}{2})} \right) = 0. \qquad (76)$$

Subtracting this from (73) we obtain the discrete counterpart to (71)

$$\left[\frac{u}{m_x} = \delta_x \overline{\lambda}^{xyt} \right]_{i+\frac{1}{2},j}^{(n+\frac{1}{2})}, \quad \left[\frac{v}{m_y} = \delta_y \overline{\lambda}^{xyt} \right]_{i,j+\frac{1}{2}}^{(n+\frac{1}{2})}, \quad \left[\eta = -\delta_t \overline{\lambda}^{xyt} \right]_{i,j}^{(n)}. \qquad (77)$$

Elimination of $\overline{\lambda}^{xyt}$ now leads to our standard discrete momentum equation (13), even though we started with a Lagrangian that is not consistent with the energy in (30). But, what will substitution of (30) into the Lagrangian lead to ? We may readily find out. The variation of the terms, containing a geometric mean, may now be rewritten by (56)

$$\frac{\gamma_y h^{(u)}}{2m_x} \sum_n [\mathcal{D}(u^{(*2)})]^{(n)} = \frac{\gamma_y h^{(u)}}{m_x} \sum_n [\overline{u}^{L_t} \mathcal{D}u]^{(n+\frac{1}{2})} + \text{b.t.} \qquad (78)$$

where we have omitted spatial sums and indices and the mean is defined in (11). A corresponding relation applies for the v term. We then arrive at an equation similar to (73), save that that non-varied velocities are replaced by temporal means. The steps of the above derivation will then repeat to yield a modified (77) where the velocities are replaced by their temporal means. This leads to momentum equations like (38).

C   The elastic string analogy

Equations that are equivalent to (2) and (3) can be derived from a series of different physical systems, with the elastic string, or membrane in two dimensions, as the most well known. The description of an elastic string is obtained from (2) and (3) by deletion of v and the y dependence and elimination of u. Moreover, η is to be interpreted as the transverse displacement of the string,

h as a variable elasticity factor and (1) must be replaced by the appropriate transformation equation for an elastic string. The energy quantities of the differential equations now become:

$$E = \frac{1}{2}\left(\frac{\partial \eta}{\partial t}\right)^2 + \frac{1}{2}h\left(\frac{\partial \eta}{\partial x}\right)^2, \quad U = -h\frac{\partial \eta}{\partial x}\frac{\partial \eta}{\partial t}. \tag{79}$$

The two terms in E are kinetic and elastic potential energy respectively, while U is the effect of elastic work on a cross section of the string.

Assuming constant map factor we obtain

$$\delta_t^2 \eta - \delta_x(h\delta_x \eta) = 0, \tag{80}$$

from (2) and (3). Averaging (80) with respect to t, multiplying by $\delta_t \eta$ and invoking (51) we find

$$\delta_t\left(\frac{1}{2}(\delta_t \eta)^{(*2)}\right) = \delta_t \eta \delta_x(h\delta_x \overline{\eta}^t). \tag{81}$$

Using (48), once again, we transform the right hand side into total differences and find the discrete counterpart to (79)

$$E = \frac{1}{2}(\delta_t \eta)^{(*2)} + \frac{1}{2}\overline{h(\delta_x \eta)^2}^x, \quad U = -h\delta_x \overline{\eta}^t \delta_t \overline{\eta}^x. \tag{82}$$

For a slowly varying h we still obtain the correct amplification from $\langle U \rangle$=const. as well as $\langle U \rangle = c_g \langle E \rangle$.

References

[1] A. Arakawa. A potential enstrophy and energy conservation scheme for the shallow water equations. *J. Comp. Phys.*, 1:119–143, 1966.

[2] A. Arakawa and V. R. Lamb. A potential enstrophy and energy conservation scheme for the shallow water equations. *Monthly Weather Review*, 109:18–36, 1981.

[3] H. P. Langtangen and G. Pedersen. Computational models for weakly dispersive nonlinear water waves. *Comp. Meth. Appl. Mech. Engrg.*, 160:337–358, 1998.

[4] D. K. Lilly. Numerical solutions for the shape-preserving two-dimensional thermal convection element. *J. Atmospheric Sciences*, 21:83–98, 1964.

[5] C. C. Mei. *The Applied Dynamics of Ocean Surface Waves*. World Scientific, 1989.

[6] Espelid T. O., Berntsen J., and Barthel K. Conservation of energy for schemes applied to the propagation of shallow-water inertia-gravity waves in regions with varying depth. *Int. J. Numer. Meth. Engng*, 49:1521–1545, 2000.

[7] Pedersen G. P. An optical theory for discrete media. *Wave Motion*, 32:79–92, 2000.

[8] G. Pedersen. Three-dimensional wave patterns generated by moving disturbances at transcritical speeds. *J. Fluid Mech.*, 196(39-63), 1988.

[9] D. H. Peregrine. Equations for water waves and the approximation behind them. In R. E. Meyer, editor, *Waves on beaches*, pages 357–412. Academic Press, New York, 1972.

[10] D. H. Peregrine. Interaction of water waves and currents. *Adv. Appl. Mech.*, 16:10–117, 1976.

[11] G. Strang and G. J. Fix. *An analysis of the finite element method.* Prentice-Hall series in automatic computation. Prentice-Hall, 1973.

[12] G. B. Whitham. *Linear and nonlinear waves.* Pure & Applied Mathematics. John Wiley & sons, 1974.

[13] T. Y. Wu. Long waves in ocean and coastal waters. *Proc. ASCE, J. Eng. Mech. Div.*, 107:501–522, 1981.