Fully well-balanced entropy controlled discontinuous Galerkin spectral element method for shallow water flows: global flux quadrature and cell entropy correction

Yogiraj Mantri, Philipp Öffner, Mario Ricchiuto

Fully well-balanced entropy controlled discontinuous Galerkin spectral element method for shallow water flows: global flux quadrature and cell entropy correction

Yogiraj Mantri[b], Philipp Öffner[c], Mario Ricchiuto[a]

[a] *Inria, Univ. Bordeaux, CNRS, Bordeaux INP, IMB, UMR 5251, 200 Avenue de la Vieille Tour, 33405 Talence cedex, France*
[b] *Vellore Institute of Technology, India*
[c] *Johannes Gutenberg-University, Mainz, Germany*

Abstract

We present a novel approach for solving the shallow water equations using a discontinuous Galerkin spectral element method. The method we propose has three main features. First, it enjoys a discrete well-balanced property, in a spirit similar to the one of e.g. [20]. As in the reference, our scheme does not require any a-priori knowledge of the steady equilibrium, moreover it does not involve the explicit solution of any local auxiliary problem to approximate such equilibrium. The scheme is also arbitrarily high order, and verifies a continuous in time cell entropy equality. The latter becomes an inequality as soon as additional dissipation is added to the method. The method is constructed starting from a global flux approach in which an additional flux term is constructed as the primitive of the source. We show that, in the context of nodal spectral finite elements, this can be translated into a simple modification of the integral of the source term. We prove that, when using Gauss-Lobatto nodal finite elements this modified integration is equivalent at steady state to a high order Gauss collocation method applied to an ODE for the flux. This method is superconvergent at the collocation points, thus providing a discrete well-balanced property very similar in spirit to the one proposed in [20], albeit not needing the explicit computation of a local approximation of the steady state. To control the entropy production, we introduce artificial viscosity corrections at the cell level and incorporate them into the scheme. We provide theoretical and numerical characterizations of the accuracy and equilibrium preservation of these corrections. Through extensive numerical benchmarking, we validate our theoretical predictions, with considerable improvements in accuracy for steady states, as well as enhanced robustness for more complex scenarios.

Keywords: balance laws, general steady equilibria, discontinuous Galerkin spectral element, fully well-balancing, entropy conservation, Gauss-Lobatto integration
2010 MSC: 68Q25, 68R10, 68U05

1. Introduction

Hyperbolic balance laws play a fundamental role in various phenomena in natural science and engineering. A system of balance law is given by the following form

$$\partial_t U + \nabla \cdot F(U) = S(U; \varphi(x)), \tag{1}$$

where U contains the conserved variables, F is the flux function, and S denotes the source term, which depends on the solution as well as on some external data (e.g. bathymetry, friction coefficient map, etc.)

Email addresses: yogiraj.mantri@vit.ac.in (Yogiraj Mantri), poeffner@uni-mainz.de (Philipp Öffner), mario.ricchiuto@inria.fr (Mario Ricchiuto)

dependent on space. The numerical approximation of (1) is a very active research topic, with a lot of fundamental contributions developed to models such as the Euler equations with gravity [25, 40, 75, 77], or shallow water (SW) equations with various sources accounting for topography variations, friction and/or Coriolis forces in Cartesian or curvilinear coordinates [12, 7, 27, 28, 58, 8, 15]. Even in one space dimension, the challenge of devising well-balanced numerical approximations agnostic of the form of the steady state is still open. There is already quite a large literature on the subject, with several different approaches to manage this issue. This paper focuses on the shallow water equations in Cartesian coordinates, including all the effects mentioned above.

The source term in (1) leads to a rich set of possible solutions, and in particular to a large number of different forms of steady equilibria between the source terms and the flux derivatives. Many of such equilibria have some interest in themselves. Many physical applications involve small perturbations of such equilibria. The ability of a numerical method to resolve with enhanced accuracy such steady states is a unanimously acclaimed design criterion usually referred to as well-balanced or full well-balanced depending on whether the property applies to a specific equilibrium, or to all steady states. Giving a full review of the subject is way beyond our scope. Questions relevant to the work of this paper are the following:

- what do we know of the equilibrium we wish to preserve: do we know it already in (some) explicit form, do we know if it belongs to a family of solutions verifying some algebraic constraint, or do we want our method to be agnostic of it, and still be well-balanced in some sense;
- do we seek to preserve the analytical steady state or some approximation of it, and which one;
- can we use some auxiliary problem to improve the well-balanced character of the scheme;
- what is the accuracy aimed for, is the order of the method arbitrary.

Note that we do not address the property of genuinely multidimensional well-balanced, which is related to the notion of the preservation of solenoidal involutions. We also leave out extensions including dry areas, for which the only relevant steady state is the lake at rest, at least in one space dimension.

Concerning the other aspects, some have been discussed widely in literature. For example, when steady (or even unsteady) equilibria are explicitly known one can use a simple and efficient idea due to [18], which consists in evolving a discrete error with respect to the given equilibrium. This boils down to removing from the discrete equations of any scheme the discrete expression corresponding to the application of the scheme itself to the given solution. This idea has been adapted to many discretization approaches and applied to several systems of equations ranging from the Euler equations with gravity [40, 10], to the shallow water on manifolds [15], to the MHD equations [10, 13], to hyperbolic reformulations of the Einstein equations of relativity [35], to cite a few. This approach can be applied to study perturbations of a given equilibria with any scheme, and to any order of accuracy.

A more intricate scenario arises when the relevant equilibrium is not explicitly known, but it can be characterized by a set of (generally nonlinear) algebraic relations that define constant invariants. To put it differently, in this scenario, the equilibrium state can be described through the collection of relationships expressed as $V = V_0$, where V represents a comprehensive set of variables that, ideally, can be leveraged to derive the conserved quantities U. In this case, one tries to devise a discretization consistent in some way with an approximation of the modified form of (1) reading (in one space dimension),

$$\partial_t U + A_V(U, \varphi(x))\partial_x V = 0. \tag{2}$$

This idea has been adapted within a variety of methods going from finite volume to finite elements, and embedded either in the polynomial approximation, and/or in the definitions of the discrete divergence and of the numerical flux and source terms [19, 53, 69, 70, 79, 24, 58, 12, 48]. While not requiring the full a-priori knowledge of the steady state, one still requires the existence and the knowledge of the invariant set $V(U)$. In this situation also, discretizations with very high order of accuracy have been proposed in literature.

When nothing is known of the steady state, two interesting approaches are considered here. The first, introduced in [20], is a discrete generalization of the idea of [18]. As the latter, it is quite general and it

has been applied across various numerical frameworks such as finite elements, finite differences and finite volume [20, 47, 50, 46, 45, 16]. As in [18] the idea is to evolve a perturbation with respect to the equilibrium solution. However, when the latter is unknown the authors of [20] propose to solve locally the auxiliary Cauchy problem $U'(x) = A^{-1}(U)S(U,x)$, where $A(U)$ is the flux Jacobian $\partial_U F$ to obtain an estimate of a local admissible steady state. The admissibility is linked to the match with the mesh data. This method is not an exactly well-balanced, however it has also a clear definition of the notion of discrete equilibrium, associated to the solution of the local Cauchy problem. Its drawback is that it requires the explicit resolution of the latter as an auxiliary problem.
A second approach which is also agnostic of the steady equilibrium, is provided by the class of methods known as global flux schemes. These schemes, initially introduced in the research by [42], were initially employed to develop nonlinear shock-capturing techniques for balance laws in [17] and [34]. More recently, they have been utilized as means to achieve fully well-balanced methods, as seen in works like [25, 23, 57]. This approach relies on the observation that (up to a constant) the non-local operator

$$G(U,x) = F(U(x)) - \int_{x_0}^{x} S(U(s),\varphi(s))ds \tag{3}$$

provides a natural invariant for the steady equations. So recasting (1) as

$$\partial_t U + \partial_x G(U,x) = 0, \tag{4}$$

provides a reasonable path to obtain fully well-balanced schemes. As shown in [6], this approach also has relations with flux difference splitting and residual distribution methods consistently embedding the source integral in the splitting and can thus be related to many classical versions of this idea [72, 11, 76, 63]. It has clear relations with methods which define the discrete source term as the derivative of a steady equilibrium flux, e.g. [62]. While providing very interesting results for many different models and solutions, one of the drawbacks of this approach is that, differently from the previous one, there is no precise characterization of the meaning of the steady solution. For this reason in [16], the authors have chosen to combine the use of (4) with the correction approach of [20] which allows a more clear control to the notion of well-balancing.

In addition to the above constraints, system (1) is usually endowed with an entropy pair $(\eta(U), F_\eta(U))$ verifying the additional constraint [31, 52]

$$\partial_t \eta + \nabla \cdot F_\eta(U) \overset{(\leq)}{=} S_\eta(U;\varphi(x)), \tag{5}$$

where $\eta = \eta(U)$ denotes the mathematical entropy (a convex function), $F_\eta(U)$ is the entropy flux, and $S_\eta(U;\varphi(x))$ represents dissipation/production term. In (5) the equality holds for smooth solutions while weak admissible solutions are characterized by the inequality. A degree of control on the production of entropy is thus a desirable feature in numerical schemes. A numerical method is called entropy conservative if it fulfills the equality in (5) and entropy dissipative if it ensures the inequality for one specific entropy [74]. There exist many different techniques which have been applied to obtain entropy conservation (dissipation), e.g. artificial viscosity and correction techniques [1, 3, 4, 49], the summation-by-parts framework combined with entropy conservative (EC) two-points fluxes (flux differencing) [22, 37, 68], or multi-point approaches via with combination of EC fluxes [39, 56]. Some of this work has been generalized to embed some notion of well-balanced. Here, we refer to [38, 65, 78] for the literature on shallow water equations.

In this paper, we propose a high order discontinuous Galerkin spectral element method (DGSEM) formulation for balance laws which embeds a fully discrete general well-balanced criterion agnostic of the exact steady state. The proposed construction exploits the idea of a global flux formulation to infer an ad-hoc quadrature strategy called here global flux quadrature. This is then used to establish a one-to-one correspondence between the discretization of the local steady Cauchy problem, and a discretization of the non-local integral operator underlying the definition of the global flux. If the discretizations exploit the same data

on the same stencil, the steady solution can be obtained indifferently by means of one of the two methods. This equivalence allows to construct balanced schemes without explicit knowledge of the steady state, and without the need of solving explicitly the local Cauchy problem. In this paper, we use a Gauss-Lobatto DGSEM setting which allows a natural connection to continuous collocation methods for integral equations. Thus we are able to fully characterize the discrete steady solution, and moreover provide a superconvergence result at the collocation points. The notion of entropy control is also included in the construction via appropriately designed artificial viscosity corrections at the cell level following [1, 4, 2, 41, 61]. The accuracy and equilibrium preservation of these corrections are characterized theoretically and numerically.

The paper is organized as follows: The main notation for the shallow water equations are recalled in section §2, where we also recall several families of steady equilibria, depending on the terms included in the sources, and the definition of the mathematical entropy. In section §3, we recall the basics of the DGSEM approach underlying the paper, introducing some of its properties and the notation necessary for the remainder of the paper. The main idea of the paper is discussed in section §4 devoted to the explicit derivation of the global flux quadrature approach, and to the characterization of its properties for the particular case of the Gauss-Lobatto DGSEM method. The entropy correction is studied in section §5, while a simple multidimensional extension is proposed in section §6. Section §7 provides a thorough verification of the theoretical expectations, as well as some applications to challenging cases and to multidimensional problems showing the great potential of the approach proposed. Some conclusive remarks and future perspectives are drawn in section §8.

2. Shallow water equations

This paper focuses on the shallow water (SW) equations, widely used in geophysical applications. Despite their relative simplicity with respect to other systems, they offer an excellent test bed for well-balanced methods due to the variety of source terms they embed. The two dimensional form of the system reads

$$\partial_t \begin{pmatrix} h \\ hu \\ hv \end{pmatrix} + \partial_x \begin{pmatrix} hu \\ hu^2 + p(h) \\ huv \end{pmatrix} + \partial_y \begin{pmatrix} hv \\ huv \\ hv^2 + p(h) \end{pmatrix} = -h \begin{pmatrix} 0 \\ \partial_x \varphi + c_f u + \omega v \\ \partial_y \varphi + c_f v - \omega u \end{pmatrix}, \tag{6}$$

where h denotes the water depth, $\vec{\mathrm{v}} = (u, v)^T$ is the horizontal velocity, $p = gh^2/2$ is hydrostatic pressure with g the gravity acceleration, $\varphi = gb$ denotes the gravitational potential with bottom topography $b(x, y)$, $c_f = c_f(h, \vec{\mathrm{v}})$ is friction coefficient, and ω denotes the Coriolis coefficient. It is customary and useful to introduce the additional variables $\zeta = h + b$, representing the free surface elevation, and the total energy density $E = g\zeta + k$, with $k = u^2/2 + v^2/2$ being the kinetic energy. In this work we will only consider continuous bathymetry, so the data φ in the source is also assumed to be continuous. Preliminary extensions of ideas similar to those discussed here to the more general case are considered in [29].
An interesting variant of (6) is the pseudo-one dimensional rotating SW system proposed in [21]

$$\partial_t \begin{pmatrix} h \\ hu \\ hv \end{pmatrix} + \partial_x \begin{pmatrix} hu \\ hu^2 + p(h) \\ huv \end{pmatrix} = -h \begin{pmatrix} 0 \\ \partial_x \varphi + c_f u + \omega v \\ -\omega u \end{pmatrix}. \tag{7}$$

This is a one dimensional system which has all the richness of the multidimensional one in terms of sources and steady states, as we will shortly discuss.

The shallow water equations are endowed with a convex entropy pair (η, F_η) given by

$$\eta = p(h) + hk\ , \quad F_\eta = hu\,(gh + k). \tag{8}$$

In the case of (7) the associated balance law reads,

$$\partial_t \eta + \partial_x F_\eta \overset{(\leq)}{=} -hu\partial_x \varphi - \underbrace{c_f hu^2}_{\mathcal{D}_f} \tag{9}$$

with $\mathcal{D}_f \geq 0$ implicitly defined above representing the dissipation due to friction. For time independent potentials, the form of the entropy production term allows to introduce a total entropy pair $(\eta_\varphi, F_{\eta_\varphi})$ which is given for (7) via

$$\eta_\varphi = p(h) + hk + h\varphi \,, \quad F_{\eta_\varphi} = hu\,(gh + k + \varphi) = hu\,(g\zeta + k) \tag{10}$$

which satisfies the simpler balance

$$\partial_t \eta_\varphi + \partial_x F_{\eta_\varphi} \overset{(\leq)}{=} -\mathcal{D}_f \leq 0 \tag{11}$$

which reduces to a special conservation law for smooth solutions and in absence of friction. Note that no dissipation or production term are associated to the Coriolis terms, as the x and y contributions to the energy balance cancel identically when dotting the momentum equations by the velocity $(u, v)^T$.

2.1. Partial taxonomy of steady equilibria

We recall here some of the classical equilibria of system (7) and (6), which will be used in the process of numerical validation. We limit the description to five types of solutions, involving three combinations of sources, but many more can be imagined.

Frictionless one dimensional equilibria. This is the most classical case under consideration. It involves two families of steady equilibria. For smooth solutions, the most general form of these equilibria is defined by the relations

$$h(x)u(x) = q_0 \,, \;\; g\zeta(x) + k(x) = E_0, \tag{12}$$

to obtain the invariants q_0, E_0, with q_0 and E_0 being the given values of the volume flux and total energy density. Of course this is also a special solution of (6) if $v = 0$, otherwise a similar one can be defined in the frame of reference with the x axis aligned to the velocity. The moving equilibrium (12) can be determined analytically by solving a cubic equation, given q_0 and E_0 and the bathymetry $b(x)$ (see e.g. [60]).

A very well known particular case is the so-called *lake at rest* state defined by

$$u(x) = 0 \,, \;\; \zeta(x) = \zeta_0. \tag{13}$$

This is the most widely used analytical state to construct well-balanced methods, and first historically studied [11].

Note that at the analytical level, for smooth cases anyways, we can replace (12) by a global flux definition reading:

$$hu = q_0 \,, \;\; Q := hu^2 + p(h) + \int_{x_0}^{x} gh\partial_x b = Q_0. \tag{14}$$

For smooth cases the two definitions coincide.

Frictionless pseudo-one dimensional equilibria with Coriolis effects. This is a more general case of the previous one, including transverse velocity and Coriolis effects. It also involves two families of solutions, depending on whether $u = 0$ or not. When $u = 0$ we have solutions which are defined by

$$hu = 0 \,, \;\; \zeta = \zeta_0 - \frac{1}{g}\int_{x_0}^{x} \omega v(s) ds. \tag{15}$$

In this case the transverse velocity v is essentially a free parameter. This is some sort of generalized analytical lake at rest state, accounting for movement in directions transverse to those of the main bathymetric variations.

For the moving case, combining the first and last equation in (7), with the steady limit of (11) one can show that (16) holds, provided $u \neq 0$.

$$hu = q_0\,, \quad g\zeta + k = E_0\,, \quad v - \omega x = v_0. \tag{16}$$

Equation (16) needs to be solved analytically given values of q_0, E_0 and v_0, and the bathymetry $b(x)$. Particular smooth solutions can also be obtained by defining $b(x)$ for given variations of h and u. An example will be presented in the results section. As before, for smooth cases we can equivalently provide a global flux version of (16), which reads

$$hu = q_0\,, \quad Q := hu^2 + p(h) + \int_{x_0}^{x} (gh\partial_x b + \omega hv) = Q_0\,, \quad V := huv - \int_{x_0}^{x} \omega hu = V_0. \tag{17}$$

For smooth cases the solutions defined by (17) or (16) are equivalent. A detailed discussion for such problems and more can be found in recent papers [33, 26].

One dimensional equilibria with friction. This is a moving equilibrium with friction and slope variations but no Coriolis force. It cannot in general be defined analytically. It is characterized by the relations ($v = 0$ since it is one dimensional)

$$hu = q_0\,, \quad g\zeta + k = E_0 - \int_{x_0}^{x} c_f u\, ds. \tag{18}$$

A trivial particular case for it is $b'(x) = \xi_0$ =const, which admits as exact solution a special state $h = h_0$ and $u = u_0$ verifying the algebraic relations

$$h_0 u_0 = q_0\,, \quad c_f(h_0, u_0)u_0 = -g\xi_0. \tag{19}$$

These can be solved for given values of q_0 and of the slope ξ_0, and the friction law (see e.g. [70]). More general examples obtained integrating (18) are proposed e.g. in [58].

The global flux definition of these states differs a little bit from (18) and is given by

$$hu = q_0\,, \quad Q := hu^2 + p(h) + \int_{x_0}^{x} (gh\partial_x b + c_f(h,u)hu) = Q_0. \tag{20}$$

3. DGSEM discretization: main notation and setting

3.1. Main notation

We consider a tessellation of the d-dimensional spatial domain Ω in non overlapping elements K, obtained as tensor products of 1d elements. In other words in two dimensions $K = K_x \times K_y$ with K_j a segment of width h in the direction j (cf figure 1). As it is classical, we consider in each direction a linear map onto the classical unit reference element $x(\xi) : K_x \mapsto [0, 1]$, and similarly in the other directions. In the multidimensional case we will denote by $\{\xi_j\}_{j=1,d}$ the reference coordinates. On each 1d reference element we consider a standard Gauss-Lobatto (GL) collocated finite element approximation spanned by $\{\phi_i(\xi)\}_{i=0,p}$ degree p one dimensional Lagrange basis functions corresponding to the $p + 1$ GL points $\{\xi_i\}_{i=0,p}$. For a given function u we thus set in 1d

$$u_{\mathrm{h}} := \sum_{j=0}^{p} \phi_j(x(\xi))u_j \tag{21}$$

while in 2d we have the usual tensor product representation

$$u_{\mathrm{h}} := \sum_{i,j=0}^{p} \phi_i(y(\zeta))\phi_j(x(\xi))u_{ij} \tag{22}$$

interpolating the solution on a grid of Gauss-Lobatto points $\{\xi_{ij}\}_{i,j=0,p}$ as in figure 1, with $\xi_{ij} = \xi_i\zeta_j$.

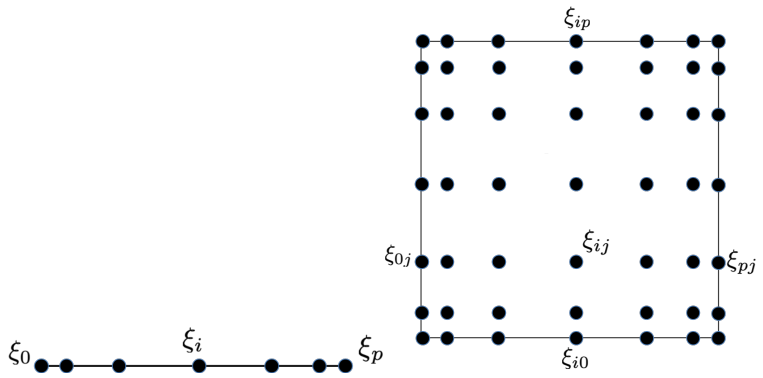


Figure 1: One and two dimensional Gauss-Lobatto nodes. The polynomial approximation is of degree p in each direction separately for the $p+1$ GL points.

Remark 1 (Nodal approximation for the data)**.** *As already mentioned in section §2, we only consider continuous data, and in particular for continuous bathymetry. Since for continuous $\phi(x)$ the discrete spectral element projection ϕ_h will boil down to standard nodal interpolation on Gauss-Lobatto points. This leads to an approximation with* no jumps at element boundaries *despite of the locality of the overall approximation, and of the fact that jumps may occur for the solution U_h. Initial work including discontinuous solutions as well as data (e.g. bathymetry) with jumps is discussed in [29]. The extension to the current DGSEM framework is ongoing.*

3.2. DGSEM for 1d conservation laws

We consider for the moment the approximation of solutions of

$$\partial_t U + \partial_x F(U) = 0. \tag{23}$$

On a generic element K, we start from the standard DG variational form

$$|K| \int_0^1 \phi_i(\xi)\partial_t U_{\mathrm{h}} - \int_0^1 \partial_\xi \phi_i(\xi) F_{\mathrm{h}} + (\phi_i \hat{F}_{\mathrm{h}}(U_{\mathrm{h}}, U_{\mathrm{h}}^+))_{\xi=1} - (\phi_i \hat{F}_{\mathrm{h}}(U_{\mathrm{h}}, U_{\mathrm{h}}^+))_{\xi=0} = 0. \tag{24}$$

Following the classical spectral element approach, the quadrature used is the one associated to the GL nodes also used for the nodal approximation. The resulting semi-discrete equations can be written in matrix form as [55, 54],

$$\frac{\mathrm{d}\mathbf{U}}{\mathrm{d}t} - \widetilde{D}_x^T \mathbf{F} + \mathcal{M}^{-1}\mathcal{B}\widehat{\mathbf{F}} = 0, \tag{25}$$

where $\mathcal{M} = \mathrm{diag}(\{w_i\}_{i=0,p})$ with w_i the Gauss-Lobatto quadrature weights[1], $\widetilde{D}_x = \mathcal{M} D_x \mathcal{M}^{-1}$ with $(D_x)_{ij} = \partial_\xi \phi_j(\xi_i)$, and with $\mathcal{B} = \mathrm{diag}(-1, \ldots, 1)$ the matrix sampling boundary values. The arrays $\mathbf{U}$, $\mathbf{F}$, and $\widehat{\mathbf{F}}$ respectively contain nodal values of the solution, the flux and the numerical flux.

3.3. Summation-by-parts, strong form, and residual distribution

The nodal DGSEM method is known to enjoy a summation-by-parts (SBP) property mimicking integration-by-parts on the discrete level. In terms of the operator notation focusing on a reference element, this means that

$$\mathcal{M}D_x + D_x^T\mathcal{M} = \mathcal{B} \Longleftrightarrow D_x^T\mathcal{M} = \mathcal{B} - \mathcal{M}D_x \Longleftrightarrow D_x^T = \mathcal{B}\mathcal{M}^{-1} - \mathcal{M}D_x\mathcal{M}^{-1} \Longleftrightarrow D_x^T = \mathcal{M}^{-1}\mathcal{B} - \widetilde{D}_x$$

[1]Note that the GL quadrature is of degree $2p-1$ for $p+1$ points. Therefore, the last entry of the mass matrix is lumped, cf. [43, 66].

By applying summation-by-parts on the second term of (25), the semi-discrete equation (25) can be recast to

$$\frac{\mathrm{d}\mathbf{U}}{\mathrm{d}t} + D_x\mathbf{F} + \mathcal{M}^{-1}\mathcal{B}(\widehat{\mathbf{F}} - \mathbf{F}) = 0. \tag{26}$$

In other words, we can recast the DGSEM semi-discrete equations using a residual or a fluctuation distribution form. Therefore, we multiply the mass matrix $\mathcal{M}$ again and consider the $i-th$ degree of freedom, i.e. the i-th GL point. The scheme is given by,

$$w_i\frac{dU_i}{dt} + \Phi_i + \Psi_i^L + \Psi_i^R = 0, \tag{27}$$

with cell and face residuals arising for the GL quadrature of

$$\begin{gathered}\Phi_i := \int_K \phi_i \partial_x F_\mathrm{h}, \\ \Psi_i^L := [\phi_i(\hat{F}_\mathrm{h} - F_\mathrm{h})]_{\xi=0}\ ,\ \ \Psi_i^R := [\phi_i(\hat{F}_\mathrm{h} - F_\mathrm{h})]_{\xi=1}.\end{gathered} \tag{28}$$

Consistency and accuracy conditions can be characterized in terms of the properties of the split residuals, as discussed in thorough detail in [5, 6]. More importantly, from the point of view of preservation of steady states, this form is very interesting. Indeed, due to the nodal Gauss-Lobatto approximation, continuous solutions with the constant flux $F_\mathrm{h} = F_0$ are exact discrete steady states.

Note that for general numerical fluxes of the form

$$\hat{F}_\mathrm{h} = \alpha F_\mathrm{h}^+ + (1-\alpha)F_\mathrm{h} + \mathcal{D}(U_\mathrm{h}^+ - U_\mathrm{h}), \tag{29}$$

where α can be a scalar or a matrix weight, and $\mathcal{D}$ a positive definite matrix, the DGSEM semi-discrete equation (25) can also be written as

$$\frac{\mathrm{d}\mathbf{U}}{\mathrm{d}t} + \widetilde{D}_x\mathbf{F} + \mathcal{M}^{-1}\mathcal{B}(\alpha[\![\mathbf{F}]\!]) + \mathcal{M}^{-1}\mathcal{B}(\mathcal{D}[\![\mathbf{U}]\!]) = 0 \tag{30}$$

having introduced the interface jumps $[\![\cdot]\!] = (\cdot)^+ - (\cdot)^-$.

4. Discrete full well-balancing and global flux quadrature

We now consider the numerical approximation of a (hyperbolic) system of balance laws

$$\partial_t U + \partial_x F(U) = S(U;\varphi(x)). \tag{31}$$

A steady state $U^*(x)$ of (31) can be described by the non-linear ODE

$$F'(U^*)(x) = S(U^*;\varphi(x)). \tag{32}$$

At the continuous level, (32) can be equivalently expressed as a non-local integral equation

$$F(U^*)(x) = F_0 + \int_{x_0}^{x} S(U^*;\varphi)(s) \tag{33}$$

for a given initial integration point x_0, for example the left-end of the domain, and with appropriate definition of the initial state F_0. Integral relations as (33) are the most general way of defining reference solutions, as shown by some of the definitions of section §2.1. At the discrete level, however, given a high order strategy to discretize the local equation (32), it is in general not true that the discrete equations for all point (or average) values correspond to a discretization of (33). We thus give the following definition in the spirit of [20] and subsequent works.

Definition 2 (Discrete fully well-balanced)**.** *Consider a scheme discretized using some mesh data (point values, cell averages, etc.). Assume that using the same data we can construct a discrete approximation of any continuous exact steady state with enhanced accuracy compared to the scheme, either in terms of error convergence rate, or of error magnitude, or both. If this approximation also satisfies exactly the stationary algebraic equations obtained from the discretization of* (31)*, then we say that the scheme is discretely fully well balanced. In other words, a discretely fully well-balanced scheme is superconvergent for any given continuous exact steady state.*

The definition above assumes that given a set of point/averaged values, we are able to construct with the same data both a local approximation of the balance law, reducing to a discretization of (32) at steady state, as well as a consistent discretization of (33), and that the two are equivalent at steady state.

Fully well-balanced schemes verifying this definition have been proposed e.g. in [47, 50]. In these works, the authors exploit the corrected approach recalled in the introduction, initially proposed in [20], combined with a collocation method to formulate the discrete integral equations. This approach thus requires the explicit computation of U^*. In this work, we proceed differently, and show that for a specific approximation of the integral of the source arising in the finite element statement the above equivalence holds. This provides an approach which does not necessarily require the evaluation of U^*.

We start from the global formulation in which we replace the source by the derivative of an unknown flux R in the DG variational form

$$\int_K \phi_i S = \int_K \phi_i \partial_x R. \tag{34}$$

We then evaluate R from some approximation of the integral relation

$$R(U;\varphi(x)) = r_0 + \int_{x_0}^{x} S(U;\varphi)(s), \tag{35}$$

where x_0 denotes the left end of the spatial domain, and r_0 the left value of the source flux. Note that this value is essentially an integration constant not affecting the solution. It is set to $r_0 = 0$ in the following. Introducing the global flux $G = F - R$, we then simply apply the standard DGSEM approach to the global flux form of (31)

$$\partial_t U + \partial_x G(U;\varphi(x)) = 0. \tag{36}$$

This is obviously a good setting to obtain consistency with (33), since at steady state $F = R$. The details of this consistency depend on how $R(U;\varphi(x))$ is evaluated at mesh nodes, or otherwise said on how (35) is approximated. In the DGSEM setting, the most natural approach is to introduce discrete elemental approximations in the local polynomial space for all the quantities involved, and in particular for S_h, R_h, and φ_h, as described in section §3. Then, in each element K we compute the nodal values $\{R_i\}_{i=0,p}$ as

1. Set $R_0 = r^-$
2. For $i = 1, p$ set $R_i = R_{i-1} - \int_{x_{i-1}}^{x_i} S_\mathrm{h}(x)$

with r^- a local integration constant to be defined.

Using the spectral element expansion for S_h we can recast the previous expressions in a compact form by introducing the following $(p+1)\times(p+1)$ integration matrix $\mathcal{I}$

$$\mathcal{I}_{jk} := \int_0^{\xi_j} \phi_k(\xi)\, d\xi. \tag{37}$$

With the previous definition we now have

$$\mathbf{R} = \mathbf{R}^- - \mathrm{h}\,\mathcal{I}\mathbf{S}, \tag{38}$$

where now $\mathbf{R}$ is the array of nodal values of R, $\mathbf{R}^-$ has entries all equal to r^-, and $\mathbf{S}$ contains the nodal values of the source. The semi-discrete global flux DGSEM equations can now be readily written in matrix form (cf. (26)) as

$$\frac{\mathrm{d}\mathbf{U}}{\mathrm{d}t} + \widetilde{D}_x\mathbf{G} + \mathcal{M}^{-1}\mathcal{B}(\widehat{\mathbf{G}} - \mathbf{G}) = 0 \tag{39}$$

with $\mathbf{G} = \mathbf{F} - \mathbf{R}$, or equivalently in the fluctuation form (27) by setting

$$\Phi_i := \int_K \phi_i \partial_x G_\mathrm{h} \ , \quad \begin{cases} \Psi_i^L := [\phi_i(\hat{G}_\mathrm{h} - G_\mathrm{h})]_{\xi=0}, \\ \Psi_i^R := [\phi_i(\hat{G}_\mathrm{h} - G_\mathrm{h})]_{\xi=1}. \end{cases} \tag{40}$$

To proceed further we use an explicit form of the numerical fluxes similar to (29), namely

$$\hat{G}_\mathrm{h} = \alpha G_\mathrm{h}^+ + (1-\alpha)G_\mathrm{h} + \mathcal{D}(U_\mathrm{h}^+ - U_\mathrm{h}) \tag{41}$$

Making explicit use of (38) and of (29) we can write global flux DGSEM as

$$\frac{\mathrm{d}\mathbf{U}}{\mathrm{d}t} + \widetilde{D}_x\mathbf{F} + \mathcal{M}^{-1}\mathcal{B}(\alpha[\![\mathbf{F}]\!]) + \mathcal{M}^{-1}\mathcal{B}(\mathcal{D}[\![\mathbf{U}]\!]) = \mathrm{h}\,\widetilde{D}_x\mathcal{I}\mathbf{S} + \mathcal{M}^{-1}\mathcal{B}(\alpha[\![\mathbf{R}]\!]). \tag{42}$$

The last term only involves jumps in the source flux, depending on the definition of the local integration constant r^-. Consistently with Remark 1, in this work we have not accounted for this jump. We refer to [29] for a way to include this contribution. Neglecting the jump is equivalent to the choice $r^- = (R_p)^{K^-}$, the last value of the left neighbouring element. This leads to the following *DGSEM formulation with global flux quadrature of the source*:

$$\frac{\mathrm{d}\mathbf{U}}{\mathrm{d}t} + \widetilde{D}_x\mathbf{F} + \mathcal{M}^{-1}\mathcal{B}(\alpha[\![\mathbf{F}]\!]) + \mathcal{M}^{-1}\mathcal{B}(\mathcal{D}[\![\mathbf{U}]\!]) = \mathrm{h}\,\widetilde{D}_x\mathcal{I}\mathbf{S}. \tag{43}$$

The only remaining ingredient is the definition of the nodal value of the source.

The above formula shows that with our choices the global flux DGSEM approach boils down to a very specific quadrature of the source with weights provided for each nodal degree of freedom by the integration tableau $\mathcal{I}$. In particular, the new scheme only requires modifying the mass matrix in front of the source term. For this reason we find it more appropriate, instead of a global flux method, to speak of a DGSEM method with *global flux quadrature*. For a given function f, global flux quadrature corresponds to the approximation

$$\int_K \phi_i f_\mathrm{h}\, dx = \mathrm{h}\,(\widetilde{D}_x\mathcal{I}\mathbf{f})_i \tag{44}$$

The formula above shows that the global flux quadrature is actually a fully local method. Note that this locality is also true for (42), as long as one can express the jumps in the source flux $[\![\mathbf{R}]\!]$ as a function of the source term itself. Preliminary work in this sense is discussed in [29].

4.1. Discrete equilibria and connection with collocation methods

For smooth solutions, the DGSEM formulation with global flux quadrature of the source has a neat connection with continuous collocation methods for ODEs. Summarised by the following proposition.

Proposition 3 (Global flux Gauss-Lobatto DGSEM: discrete steady states)**.** *Provided that the mapping $F(U)$ is invertible and that the inverse $U(F)$ is bounded and uniquely defined, then the global flux Gauss-Lobatto DGSEM equations* (43) *admit a nodally continuous discrete steady state $U^* = U^*(F)$ obtained upon integration with the fully implicit continuous collocation RK-LobattoIIIA method of the nonlinear ODE* (32) *with $S(U^*;\varphi) = S(U^*(F);\varphi)$, and with initial condition U_0 satisfying $F(U_0) = F_0$ with F_0 given.*

Proof. Note that the entries of the matrix $\mathcal{I}$ associated to the Gauss-Lobatto points/basis is by definition the integration tableau of the $p+1$ stages fully implicit RK-LobattoIIIA ODE solver [64]. So the steady state defined in the proposition can be written element by element as

$$\mathbf{F} = \mathrm{h}\mathcal{I}\mathbf{S},$$

which by using the definition of the source flux and the initial condition implies $F_i - R_i = F_0 \ \forall\, i$. The condition $[\![\mathbf{R}]\!] = 0$ boils down (only at steady state) to $[\![\mathbf{F}]\!] = 0$, from which the continuity of U^* follows from the properties of the map $U(F)$. Putting all this together shows that U^* so defined is a solution of the steady discrete equations (43). □

The proposition states simply that at steady state the DGSEM method with global flux quadrature provides a direct approximation of (33) in which S is replaced by the piecewise Lagrange polynomial through the Gauss-Lobatto interpolation points. There are three very important aspects which need to be addressed:

- the implication of the above proposition on the accuracy of the discrete steady state U^*;
- the validity of the hypotheses (e.g. on the map $U(F)$) required for the analysis to be true;
- and the implication of this analysis on the implementation of the method, and some clarifications with respect to the relations with schemes having similar properties proposed in [20, 47, 50].

Concerning the first aspect, as a corollary of the last proposition we have the following property.

Corollary 4 (Global flux Gauss-Lobatto DGSEM: superconvergence at steady state)**.** *Consider a system of conservation laws for which one can exhibit a flux linearization reading*

$$F(U) - F(V) = A(U,V)(U-V) \tag{45}$$

with A be a matrix with entries uniformly bounded with respect to its arguments, and diagonalizable with eigenvalues $\{\lambda_j^A\}_{j\geq 1}$. Provided that there exist a bounded strictly positive constant C_A such that

$$\lambda^A_{\min} := \min_{j\geq 1} |\lambda_j| \geq C_A > 0, \tag{46}$$

then under the hypotheses of Proposition 3, the discrete steady state U^* is nodally superconvergent, and in particular the element endpoint values of U^* are approximations of order $2p$ of any smooth exact steady solution, while the internal nodal values have accuracy $h^{\min(p+2,2p)}$.

Proof. There are two parts to the proof. The first is to invoke the properties of the LobattoIIIA method to argue that $F(U^*)$ has an accuracy of order h^{2p} at the elements endpoints and $\mathrm{h}^{\min(p+2,2p)}$ at the internal nodes. One can refer to e.g. Theorem 7.10 in [51] (see also [64]) for the proof. The second part is to bound the error on the solution. To this end, we use the following

$$(F(U_i^*) - F(U_i^{\text{exact}}))^2 = \left(A(U_i^*, U_i^{\text{exact}})(U_i^* - U_i^{\text{exact}})\right)^2 \geq (\lambda^A_{\min})^2 (U_i^* - U_i^{\text{exact}})^2$$

which using (46) readily leads to

$$\|U_i^* - U_i^{\text{exact}}\| \leq \frac{1}{C_A} \|F(U_i^*) - F(U_i^{\text{exact}})\|$$

and thus the result. □

The corollary above shows the main added value of the global flux quadrature: the consistency with direct integration of the ODE by means of a Gauss-Collocation method, by which we inherit its accuracy, and in particular, superconvergent properties. This can be interpreted now in two ways. The first is to say, following the approach of [20, 47, 50], that the global flux Gauss-Lobatto DGSEM scheme is *exactly and*

fully well-balanced with respect to all solutions U^* characterized by Proposition 3 and Corollary 4. We can otherwise understand the corollary as the scheme having enhanced accuracy, with orders h^{2p} for element boundaries and $\mathrm{h}^{\min(2p,p+2)}$ for internal nodes, *for all smooth steady states*.

Concerning now the validity of the hypotheses used, some of them have some simple and clear justifications. For example, linearizations of the type (45) are known and are exhibited by many systems of balance laws. For the shallow water equations, the matrix A can be obtained by means of a conservative Roe-like linearization of the Jacobian of the flux $\partial_U F$, see e.g. [11] (and [73] §3.1.1 for a multidimensional generalization). Concerning condition (46), for the classical shallow water equations it implies that the analysis only holds far from critical points. We will verify that in practice the method still performs well across such points.

While there may be a strategy to avoid (46) and prove the above result in a more general setting, we are going to see shortly that critical points also play a role for the strongest hypothesis made, which is the one on $F(U)$ being invertible, and $U(F)$ being bounded and unique. For several hyperbolic systems in 1D precise conditions can be provided to characterize this issue. For the shallow water equations, the study of such conditions is classical. One can easily show that by setting $F = (q, M)$ in the non-trivial case $q^2 > 0$ given admissible data verifying $M > 3\sqrt[3]{gq^4}/2$, one can always compute two non-negative values of the depth corresponding to these data, a unique sub-critical one and a unique super-critical one. So the inversion can be performed as soon as one knows the nature of the flow. The case $M = 3\sqrt[3]{gq^4}/2$ defines a critical solution (we refer e.g. to [23] for details). This shows that the inversion of the mapping requires a-priori knowledge of the super- or sub-critical of the flow. So if one was able to apply the ODE solver using the flux as main unknown, flows with transition from sub- to super-critical may be problematic.

This brings us to the last aspect of the discussion. The use of the global flux quadrature allows to make a direct link between the DGSEM method and continuous collocation RK methods applied to the steady ODE. As already mentioned, other methods exploit collocation ODE integrators e.g. in [47, 50]. These methods use locally the ODE integration to construct a reference discrete steady solution and correct the scheme so that it is exact in correspondence of such solution. In both cases, exact well-balancing is only obtained in correspondence of certain steady states as e.g. the lake at rest (cf. below). Otherwise, the schemes preserve only approximate solutions corresponding to those obtained by the ODE integrators. The strong point of both, the method proposed here and those in the references, is that nothing of these steady solutions needs to be known a-priori. For our proposed method however note that the solution of the ODE integrator is **not** needed. The analogy is in fact just a tool to characterize the steady states. As we will see in the numerical experiments, for example there is absolutely no issue in applying the method to trans-critical problems with excellent results, although the current proofs do not apply to such cases. The fact that the solution of the ODE is not required to implement the method is a net advantage of the scheme studied in this paper, compared to those proposed in [47, 50]. Conversely, its drawback is that there is no flexibility in the choice of the collocation method. In other words, once the polynomial degree is fixed, so is the best possible approximation of the steady state. For the scheme in [47, 50] the two ingredients are somehow independent, which is an advantage.

4.2. Modified evaluation of the nodal bathymetric source to get exact lake at rest states

The DGSEM approach with global flux quadrature described in the previous section requires the definition of the nodal values of the source. For the quasi-1D shallow water system (7), if $\mathbf{u} = (u, v)^T$ denotes the velocity vector, and $\mathbf{u}^\perp = (v, -u)^T$ its orthogonal, both the friction term $c_f u$, and the Coriolis terms $\omega \mathbf{u}^\perp$ can be easily evaluated nodally. The tricky part is how to define the value of the term $h\partial_x \varphi = gh\partial_x b$ related to the potential effects due to bathymetric variations.

In this work, we have used two approaches. The first is a straightforward analytical evaluation of $\partial_x b(x)$:

$$S_l = -h_l \begin{pmatrix} 0 \\ g\partial_x b(x_l) + c_f u_l + \omega v_l \\ -\omega u_l \end{pmatrix}. \tag{47}$$

To be exactly well-balanced for the hydrostatic equilibrium, we adapt an idea proposed in [80]. Denoting by ζ the free surface level $\zeta = h + b$, we set

$$S_l = -\begin{pmatrix} 0 \\ g\zeta_l \partial_x b_{\mathrm{h}}(x_l) - \partial_x p_{\mathrm{h}}(b)(x_l) + c_f h_l u_l + \omega h_l v_l \\ -\omega h_l u_l \end{pmatrix}, \tag{48}$$

where now b_{h} is the finite element expansion built using the nodal values of the bathymetry, and similarly for $p_{\mathrm{h}}(b) = g(b^2/2)_{\mathrm{h}}$. With this definition we can prove the following simple result.

Proposition 5 (Exact well-balanced for lake at rest)**.** *If the interpolation polynomial b_h is continuous, then the DGSEM formulation with global flux quadrature* (43)*, and with nodal source given by* (48) *is exactly well-balanced for lake at rest states $\zeta = \zeta^* = const$, and $\mathbf{u} = 0$.*

Proof. First of all note that since b_{h} is continuous (cf. Remark 1.) and so is ζ^* which is constant, h will also be continuous. Since also $h\mathbf{u} = 0$ is constant, then we have $[\![\mathbf{U}]\!] = 0$ and $[\![\mathbf{F}]\!] = 0$ at all element boundaries. The remaining term can be written as

$$\frac{d\mathbf{U}}{dt} = -\widetilde{D}_x(\mathbf{F} - \mathbf{F}_0 - \mathrm{h}\mathcal{I}\mathbf{S})$$

with $\mathbf{F}_0$ be the array containing the values $F(U_0)$ with U_0 containing the value of the unknowns in the first Gauss-Lobatto point within the element. Because of the hypotheses made, only the hydrostatic terms remain in $\mathbf{F}$ and $\mathbf{S}$. For the source, using the identity $\partial_x f_{\mathrm{h}} = \sum_{l=0,p} \partial_x \phi_l(x) f_l = \sum_{l=0,p} \phi_l(x) \partial_x f_{\mathrm{h}}(x_l)$, we can easily show that

$$\begin{aligned}(\mathrm{h}\mathcal{I}\mathbf{S})_i =& \int_{x_0}^{x_i} g\zeta^* \sum_{l=0,p} \phi_l(x_l)\partial_x b_{\mathrm{h}}(x_l) - \int_{x_0}^{x_i} g \sum_{l=0,p} \phi_l(x_l)\partial_x p_{\mathrm{h}}(b)(x_l) \\ =& g\zeta^* \int_{x_0}^{x_i} \partial_x b_{\mathrm{h}} - \int_{x_0}^{x_i} \partial_x p_{\mathrm{h}}(b) = g\zeta^*(b_i - b_0) - g\left(\frac{b_i^2}{2} - \frac{b_0^2}{2}\right).\end{aligned}$$

Using the fact that $b_i + h_i = \zeta^* = b_0 + h_0$ we deduce $(\mathrm{h}\mathcal{I}\mathbf{S})_i = p_{\mathrm{h}}(h_i) - p_{\mathrm{h}}(h_0)$, and thus the result. □

This second approach will be referred to in the results section as *modified global flux quadrature*, being obtained with a modification of the nodal source terms allowing to guarantee the exact preservation of lake at rest states.

5. Entropy control via cell corrections: flux vs energy conservation

We consider now the issue of the compatibility between the consistency with constant global flux, underpinning notion of the previous section, and the consistency with a given pair entropy/entropy flux. In absence of friction, smooth solutions of the quasi one dimensional shallow water equations (7) embed an additional conservation for the total entropy η_φ with flux F_{η_φ} defined in (10). Defining the total energy density $E = g\zeta + k$, we can readily show that analytical steady states of (7) (without friction) are characterized by the invariants

$$\begin{aligned} hu &= q_0, \\ E &= E_0, \\ v - \omega x &= v_0. \end{aligned} \tag{49}$$

These three invariants are compatible with both the steady equations and a constant distribution of the entropy flux F_{η_φ}. The same equations, in global flux form provide a steady state described by the invariance

of the global flux components:

$$\begin{aligned} hu &= q_0, \\ hu^2 + p + r_u &= Q_0, \\ huv + r_v &= V_0, \end{aligned} \tag{50}$$

having denoted by r_u and r_v the components of the source flux arising from the second and third equations. These three relations are by construction compatible with the discrete full well-balanced property, and the related superconvergence property of Corollary 4. At the continuous level all is fine.

At the discrete level, however, the global flux quadrature is exactly consistent only with (50), and only within its error with the first. In other words, when looking at the preservation of discrete steady states, we may have two different situations depending on the selected data initialization strategy:

1. We use exact analytical expressions, satisfying (49), to initialize the data. In this case, we may use the analytical expressions for the entropy fluxes to construct entropy conservative/dissipative schemes. Scheme (43) will still be discretely fully well-balanced within the conditions of Corollary 4.
2. We choose to initialize with the global flux solution U^* of Proposition 3. In this case, the resulting accuracy will still be the one of Corollary 4, but scheme (43) would be exactly well-balanced with respect to this initial steady state. For this initialization, however using the standard analytical expressions for the entropy fluxes, would break this property.

The objective of the following subsections is to propose a modification of the schemes which allows to solve this dichotomy, and provide a degree of control on the entropy production of the method independently on the initialization procedure.

5.1. Cell entropy correction method

There exist several well established techniques to embed DGSEM methods with degree of control of the entropy evolution. We refer to the introduction for a short overview. The approach used here is based on the idea of [1, 4] to introduce local corrections in the form of a symmetric positive definite (SPD) bilinear term with a free coefficient allowing to impose a desired constraint in terms of entropy balance.

To briefly describe the method, we start from the frictionless case. We consider the fluctuation form (27) (with (40)) to which we add the correction term:

$$w_i \frac{dU_i}{dt} + \Phi_i + \Psi_i^{\mathrm{L}} + \Psi_i^{\mathrm{R}} + \alpha_K \mathcal{D}_i^K = 0 \ , \quad \mathcal{D}_i^K := \int_K \partial_x \phi_i \, A_0 \partial_x W_{\mathrm{h}}, \tag{51}$$

where W is the appropriate set of entropy variables such that

$$\partial \eta = W^t \partial U, \tag{52}$$

and A_0 is the inverse of the Hessian $A_0^{-1} = \partial_{UU}\eta = \partial W / \partial U$. To obtain the entropy balance we need only to dot (51) by W_i^t and sum over $i = \{0, p\}$. The aim is to specify the correction term $\alpha_K \mathcal{D}_i^K$, in particular α_K that conservation is not violated, but entropy conservation is additionally ensured.

This readily leads to the cell entropy evolution equation

$$|K| \frac{d\bar{\eta}_K}{dt} + \Phi_\eta^K + \alpha_K \|\partial_x W\|^2_{L^2_{A_0}(K)} = 0 \tag{53}$$

having introduced the average cell entropy $\bar{\eta}_K$, and cell entropy production of the scheme Φ_η^K:

$$\begin{aligned} \bar{\eta}_K &:= \frac{1}{|K|} \int_K \eta_{\mathrm{h}}, \\ \Phi_\eta^K &:= \sum_{i=0,p} W_i^t \left(\Phi_i + \Psi_i^{\mathrm{L}} + \Psi_i^{\mathrm{R}}, \right) \end{aligned} \tag{54}$$

and norm $\|g\|_{L^2_{A_0}(K)} := \left(\int_K g\, A_0 g\right)^{1/2}$. From (53), we can see that, unless the solution is locally constant, the entropy production associated to the correction term multiplied by α_K is strictly positive. This allows to use the free parameter to gain direct control on entropy production. Let $\hat{F}_\eta$ denote a consistent numerical entropy flux. Its calculation and specification will be part of the next subsection 5.2. For now, we set

$$\Psi_\eta^K := \oint_{\partial K} \hat{F}_\eta, \tag{55}$$

reducing in 1D to

$$\Psi_\eta^K = (\hat{F}_\eta)_K^R - (\hat{F}_\eta)_K^L. \tag{56}$$

Also, following [5, 6], for a given smooth exact solution $U^e(x,t)$, and given smooth compactly supported test function $v(x,t)$ define the consistency error

$$\mathcal{E} := \sum_K \sum_{i\in K} v_i \left\{ w_i \frac{dU_i^e}{dt} + \Phi_i(U_\mathrm{h}^e) + \Psi_i^R(U_\mathrm{h}^e) + \Psi_i^L(U_\mathrm{h}^e) + \alpha_K(U_\mathrm{h}^e)\mathcal{D}_i^K(U_\mathrm{h}^e) \right\}. \tag{57}$$

with U_h^e the projection of $U^e(x,t)$ on the local finite element space. The proposition below characterizes the construction used here, following [1, 4].

Proposition 6 (Entropy correction: local/global entropy balance, consistency)**.** *The corrected DGSEM scheme obtained from* (51) *setting*

$$\alpha_K = \frac{\Psi_\eta^K - \Phi_\eta^K}{\|\partial_x W_h\|^2_{L^2_{A_0}(K)}} \tag{58}$$

verifies the elemental entropy balance

$$|K|\frac{d\bar{\eta}_K}{dt} + \Psi_\eta^K = 0 \tag{59}$$

and, for homogeneous or periodic boundary conditions, the global entropy conservation equation

$$\sum_K |K|\frac{d\bar{\eta}_K}{dt} = 0. \tag{60}$$

Moreover, the scheme verifies a consistency estimate of the type $|\mathcal{E}| = \mathcal{O}(h^{p+1})$.

Proof. Properties (58) and (59) are a straightforward consequence of (57). Concerning the consistency estimate, the proof follows e.g. Section 3.2 in [5]. It is reported in appendix with the necessary definitions from the reference. □

The above proposition shows that the correction approach allows to readily recover a local entropy balance law with a simple choice of the scalar coefficient α_K. Moreover, the correction does not spoil the formal consistency of the scheme which remains of order h^{p+1}. However, as discussed in the beginning of Section §5, the well-balanced properties of the corrected scheme depend on the properties of the numerical entropy flux $\hat{F}_\eta$.

Remark 7 (Division by zero)**.** *For uniform flows,* (58) *gives a zero divided by zero singularity. In practice, the denominator is modified to avoid this singularity as* $\max(\|\partial_x W_h\|^2_{L^2_{A_0}(K)}, \epsilon_K)$*, with* ϵ_K *constant, and set to* 10^{-8} *in the numerical experiments. The interested reader can refer to [41, 4, 2] for similar modifications.*

Remark 8 (Friction and dissipation)**.** *In presence of friction (or other dissipative relaxation terms), the above construction needs to be modified by including the effects of these terms in the definition of the entropy balance* Ψ_η^K *which becomes*

$$\Psi_\eta^K := \oint_{\partial K} \hat{F}_\eta + \mathcal{D}_f \tag{61}$$

with $\mathcal{D}_f$ *a consistent approximation of the dissipation*

$$\mathcal{D}_f \approx 2 \int_K h c_f k \geq 0 \tag{62}$$

with k *the kinetic energy. In this case, Proposition 6 still holds, but* (60) *becomes*

$$\sum_K |K| \frac{d\bar{\eta}_K}{dt} = -\sum_K \mathcal{D}_f \leq 0 \tag{63}$$

5.2. Entropy fluxes compatible with global flux quadrature

We propose here two possible definitions of the numerical entropy flux required in (55) and (56). The first is given by

$$\hat{F}_\eta(U_\mathrm{h}^+, U_\mathrm{h}^-) = \lambda F_\eta(U_\mathrm{h}^+) + (1-\lambda) F_\eta(U_\mathrm{h}^-) \tag{64}$$

with $\lambda \geq 0$ some scalar coefficient, and $F_\eta(U)$ the *analytical entropy flux*. For numerical tests in Section §7, we choose $\lambda{=}\frac{1}{2}$. We then define the local entropy balance as (we consider the one dimensional case for simplicity)

$$\begin{aligned} \Psi_\eta^K =& \hat{F}_\eta(U_\mathrm{h}^+, U_\mathrm{h}^-)^R - \hat{F}_\eta(U_\mathrm{h}^+, U_\mathrm{h}^-)^L + 2 \sum_{i=0,p} w_i c_{fi} h_i k_i \\ =& \lambda [\![F_\eta]\!]^R + \int_K \partial_x F_\eta + (1-\lambda)[\![F_\eta]\!]^L + 2 \sum_{i=0,p} w_i c_{fi} h_i k_i, \end{aligned} \tag{65}$$

where the second identity is trivially obtained by adding and removing the internal values of F_η on the element's boundaries. Following the discussion from the beginning of Section §5, this definition is compatible with analytical steady states verifying (49). It is however not exactly compatible with the global flux solutions of the DGSEM scheme with global flux quadrature, which verify instead (50).

To obtain a definition which is exactly compatible the global flux quadrature approach proposed in this work, we modify (65) as

$$\Psi_\eta^K = \lambda [\![F_\eta]\!]^R + \int_K W_\mathrm{h}^t \partial_x G_\mathrm{h} + (1-\lambda)[\![F_\eta]\!]^L \tag{66}$$

This definition embeds all the non-differential effects in the global flux term in the middle, and it is compatible with global flux solutions verifying (50). In particular, the resulting corrected scheme is exactly consistent with the discrete solutions U^* of Proposition 3.

6. A 2D extension

We briefly discuss here a possible extension to the two dimensional shallow water model (6). This extension is based on a tensor product implementation of the one dimensional DGSEM scheme, as well as on a dimensionally split generalization of the notion of global flux. The study of the genuinely multidimensional case is object of current work and left out of this paper. We refer to section §3.1 for the notation, and to [55, 43, 44] for more details on the implementation.

We start by rewriting (6) as

$$\partial_t U + \partial_x F_x + \partial_y F_y = S\,, \tag{67}$$

having set

$$F_x = \begin{bmatrix} hu \\ hu^2 + p(h) \\ huv \end{bmatrix}\,,\quad F_y = \begin{bmatrix} hv \\ huv \\ hv^2 + p(h) \end{bmatrix}\,. \tag{68}$$

The dimension by dimension generalization used here is similar to the one proposed initially in [42], and boils down to adding an appropriately defined diagonal flux tensor to the conservative flux. In other words, we recast (6) as System (67) can be written succinctly as

$$\partial_t U + \partial_x G_x + \partial_y G_y = 0 \tag{69}$$

where

$$G_x = F_x + [0\ R_x\ 0]^T\,,\quad G_y = F_y + [0\ 0\ R_y]^T\,, \tag{70}$$

having defined

$$R_x := \int_x h(\partial_x\varphi + c_f u + \omega v)d\xi\,,\quad R_y := \int_y h(\partial_y\varphi + c_f v - \omega u)d\eta. \tag{71}$$

The above definition allows to construct direction by direction fully well-balanced schemes which preserve global fluxes/equilibrium variables G_x in the x direction, and G_y in the y direction.

We discretize (69) with a classical tensor DGSEM strategy using the Gauss-Lobatto points of the right picture on figure 1. All computations done, and using a notation similar to that of the previous sections, within each element K we obtain $(p+1)^2$ equations for the degrees of freedom U_{ij} which can be compactly written as

$$\begin{aligned} \frac{d\mathbf{U}}{dt} &+ \widetilde{\mathbf{D}}_x\mathbf{F}_x + \mathcal{M}_x^{-1}\mathcal{B}_x(\alpha[\![\mathbf{F}_x]\!]_x) + \mathcal{M}_x^{-1}\mathcal{B}_x(\mathcal{D}[\![\mathbf{U}]\!]_x) \\ &+ \widetilde{\mathbf{D}}_y\mathbf{F}_y + \mathcal{M}_y^{-1}\mathcal{B}_y(\alpha[\![\mathbf{F}_y]\!]_y) + \mathcal{M}_y^{-1}\mathcal{B}_y(\mathcal{D}[\![\mathbf{U}]\!]_y) = \mathrm{h}\widetilde{\mathbf{D}}_x\mathbf{I}_x\mathbf{S}_x + \mathrm{h}\widetilde{\mathbf{D}}_y\mathbf{I}_y\mathbf{S}_y \end{aligned} \tag{72}$$

where now the array $\mathbf{U}$ contains line ordered values of the solution at the Gauss-Lobatto collocation points, $\mathbf{F}_x$ and $\mathbf{F}_y$ the flux values at collocation points, while $\widetilde{\mathbf{D}}_x$ and $\widetilde{\mathbf{D}}_y$ denote mono-dimensional derivative matrices with a structure depending on the numbering used for $\mathbf{F}_x$ and $\mathbf{F}_y$. If $\mathbf{F}_x$ is line ordered, and $\mathbf{F}_y$ is ordered by columns, both derivative matrices have a block diagonal structure with block entries given by the corresponding one dimensional matrices. As before with $\mathcal{M}$ we denote the one dimensional mass matrices in each direction, and $\mathcal{B}_{x/y}$ are the matrices selecting values on the left/right and top/bottom element boundaries, while $[\![\cdot]\!]_{x/y}$ denote the horizontal/vertical jumps at the left/right and top/bottom element boundaries. Finally, the right hand side contains the source terms, integrated using one dimensional global flux quadrature involving the derivative matrices, and block matrices $\mathbf{I}_{x/y}$ whose entries are the same as the one dimensional ODE collocation methods, and whose structure depends on the ordering of the entries of the $\mathbf{S}_{x/y}$ arrays which are values at collocation points of the sources

$$S_x := -[0\ h(\partial_x\varphi + c_f u + \omega v)\ 0]^T\,,\quad S_y := -[0\ 0\ h(\partial_y\varphi + c_f v - \omega u)]^T \tag{73}$$

The entropy correction in 2D is evaluated with natural extensions of the one dimensional formulas, see e.g. [41] and references therein for details.

7. Numerical results

We have thoroughly tested the scheme proposed to verify all the theoretical properties presented in the previous sections, as well as to investigate its application to resolution of complex solutions and to test its robustness. We also provide results on 2D flows which show the advantage of the global flux quadrature, despite of the simplicity of the extension proposed.

7.1. Verification of Corollary 4

We start by verifying the theoretical predictions on the superconvergence of the scheme. We performed two sets of tests. The first consists in computing explicitly the discrete steady solution (or global flux solution) U_{h}^* by solving, element by element, the nonlinear algebraic equations associated to the condition $G_{\mathrm{h}} = G_0$. The second set of tests involves initializing the solution with the exact analytical steady state, and letting the scheme evolve it until a given finite time. In both cases we report the grid convergence for finite element approximations of degree $p = 1, 2, 3, 4$. For the second test, we compare the global flux quadrature to the classical approach in which the source term is integrated following a classical DGSEM implementation as

$$\int_K \phi_i S_{\mathrm{h}} \to \mathcal{M}\mathbf{S}\,.$$

Only moving equilibria are considered, since the lake at rest is exactly preserved using (48). Small perturbations of the latter are studied later.

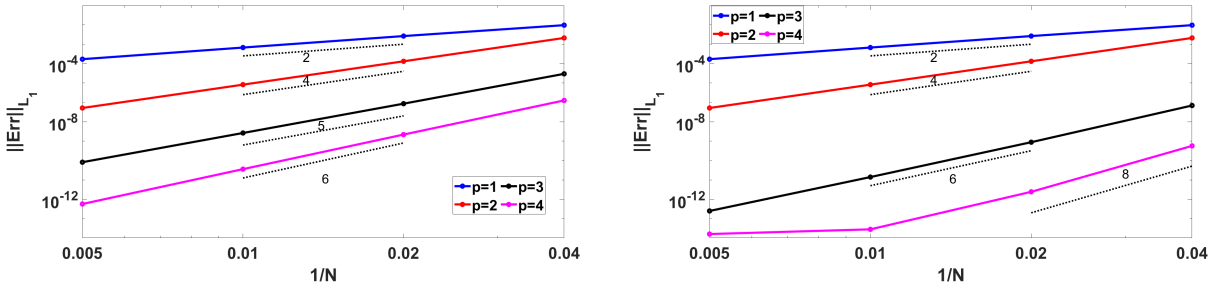
Frictionless one dimensional equilibria. We start from the first family of equilibria from section §2.1, analytically defined by (12). We compute solutions for two classical smooth states: a sub-critical one with $(q_0, E_0) = (4.42\mathrm{m}^2/\mathrm{s}, 22.05535\mathrm{m}^2/\mathrm{s}^2)$; a super-critical one for which $(q_0, E_0) = (4.42\mathrm{m}^2/\mathrm{s}, 28.8971\mathrm{m}^2/\mathrm{s}^2)$ (the gravity constant is set to $g = 9.80665\mathrm{m}/\mathrm{s}^2$ here). The computational domain is 25m long, $x \in [0, 25]$, and the bottom topography is taken as (we omit the units to avoid clutter)

$$b(x) = \begin{cases} 0.2 - 0.05(x-10)^2 & x > 8 \text{ and } x < 12 \\ 0 & \text{else} \end{cases}\,. \tag{74}$$

As already mentioned, first we study the difference between the discrete steady state corresponding to constant global fluxes (global flux solution) and the analytical one. The constant global flux solution V_0 is obtained such that we get the state (q_0, E_0) on the left boundary of the domain, i.e. $V_0 = (4.42\mathrm{m}^2/\mathrm{s}, 29.3815\mathrm{m}^3/\mathrm{s}^2)$ for sub-critical flow and $V_0 = (4.42\mathrm{m}^2/\mathrm{s}, 31.7365\mathrm{m}^3/\mathrm{s}^2)$ for super-critical flow. The boundary conditions are also such that the solution satisfies the respective steady state on the boundary. Following the predictions of Corollary 4, we compute two versions of the discrete L_1-error

$$\|\mathrm{Err}\|_{L_1} := \frac{1}{M} \sum_{j=1,M} |U_j^* - U^e(x_j)|$$

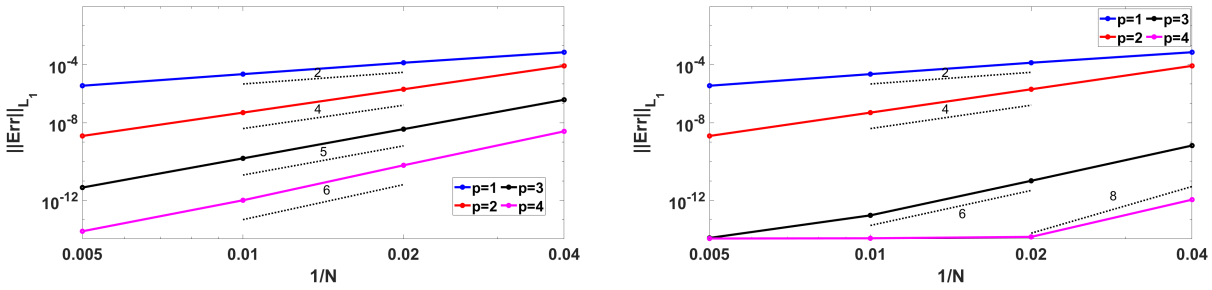
the first with j running over all the degrees of freedom of all the elements, the second running over only the element end-points. The logarithmic plots for $1/N$-L_1 error for h with $p = 1, 2, 3, 4$ and $N = 25, 50, 100, 200$ are as shown in figure 2 and figure 3 for sub- and super-critical case respectively.



(a) All nodes

(b) End nodes

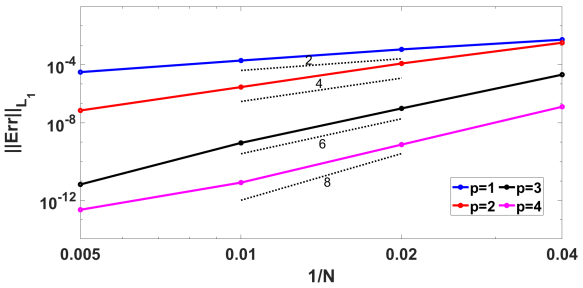
Figure 2: Frictionless one dimensional flow: sub-critical case. Error of the global flux solution for polynomial spaces of degree $p = 1, 2, 3, 4$, measured with all nodes (left) and only end nodes (right).



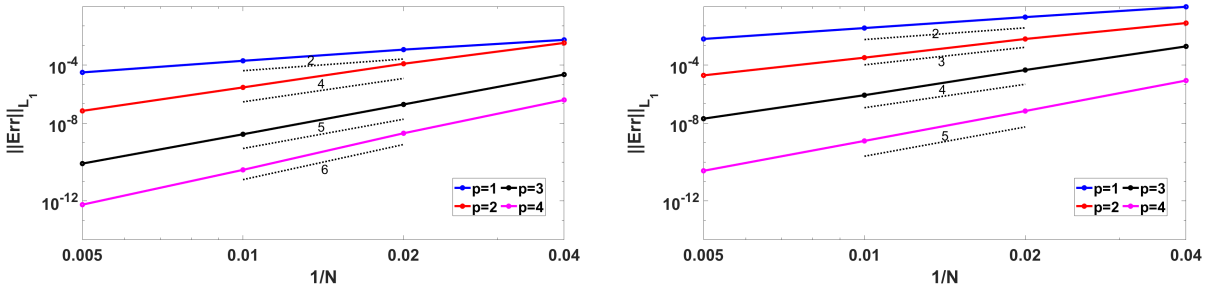
(a) All nodes (b) End nodes

Figure 3: Frictionless one dimensional flow: super-critical case. Error of the global flux solution for polynomial spaces of degree $p = 1, 2, 3, 4$, measured with all nodes (left) and only end nodes (right).

We can see that the convergence rates from figure 2 and figure 3 are as predicted in Corollary 4, i.e. by using all nodes to measure the L_1 error the convergence is of order $p + 2$, whereas by using only the end nodes, the order of convergence is $2p$. It is to be noted that with $p = 1$, nodes for Gauss Lobatto are at the ends of the cells, and hence the order of convergence is 2, even in figure 2(a) and figure 3(a). It can also be noted that with $p = 4$ the error at endpoints reaches machine accuracy on coarse meshes, which of course affects the slopes which appear somewhat reduced.



(a) WB, end nodes

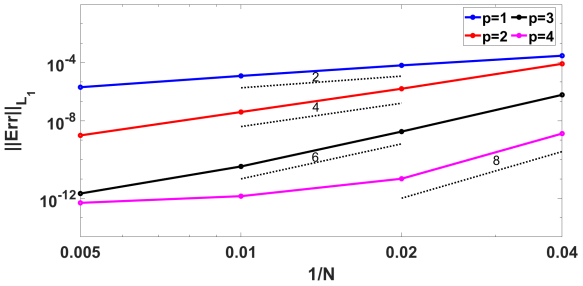


(b) WB, all nodes (c) NWB

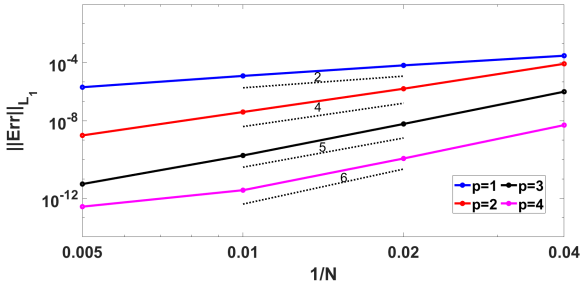
Figure 4: Frictionless one dimensional flow: sub-critical case. Error at finite time for polynomial spaces of degree $p = 1, 2, 3, 4$. Top: end-nodes error. Bottom: global flux quadrature (left) and non-well-balanced method (right).

Next, we initialize the solution at degrees of freedom using the analytical value, and we run both the global flux quadrature based and non-well-balanced schemes up to $T = 2$s. Note that this is an entirely different exercise, as we now let the scheme perturb the exact equilibrium, and we measure that the magnitude of

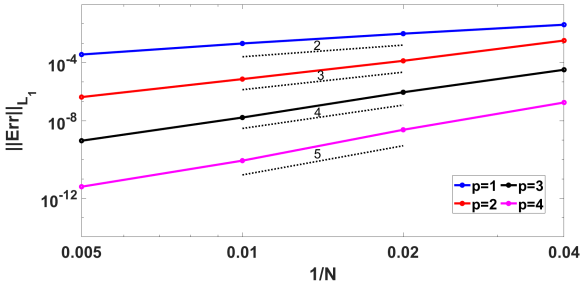
this error remains within the accuracy foreseen by Corollary 4. The logarithmic plots for $1/N$- L_1 error with $p = 1, 2, 3, 4$ and $N = 25, 50, 100, 200$ with error for well-balanced schemes measured using both all and end nodes and non-well-balanced scheme with all-nodes is as shown in figures 4 and 5.



(a) WB, end nodes



(b) WB, all nodes



(c) NWB

Figure 5: Frictionless one dimensional flow: super-critical case. Error at finite time for polynomial spaces of degree $p = 1, 2, 3, 4$. Top: end-nodes error for the global flux quadrature method. Bottom: total error for the global flux (left) and non-well-balanced method (right).

The results show that, without computing the global flux solution, the proposed method still delivers superconvergent results following the predictions of Corollary 4. Moreover, we can see that the gain in accuracy with respect to the analytical solution on a given mesh compared to the classical DGSEM implementation is of more than two orders of magnitude. This brings the error of higher order approximation very rapidly to very low values, allowing to resolve very small perturbations as we will see shortly.

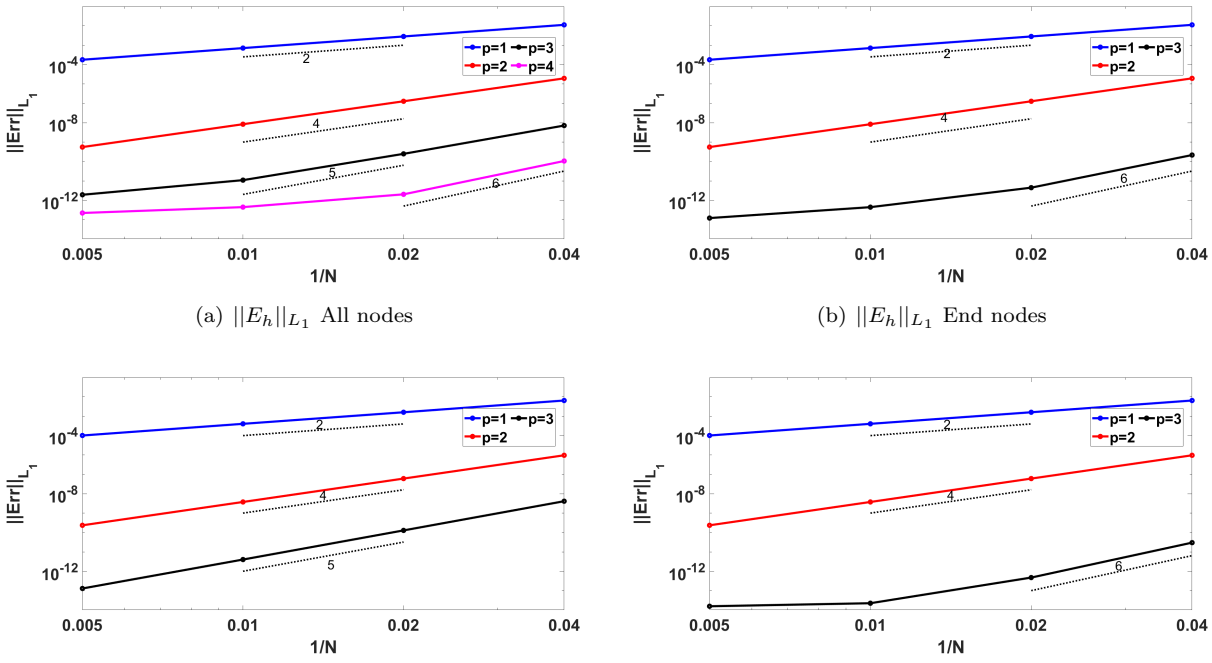
Frictionless pseudo-one dimensional equilibria with Coriolis effects. We consider next the solutions defined by (16). We choose the manufactured state also used in [32] given within $x \in [0, 1]$m by (here the gravity is set to $g = 1\text{m/s}^2$ and Coriolis coefficient $\omega = 1\text{s}^{-1}$) (we omit units for simplicity):

$$h = e^{2x}, \quad hu = 1, \quad hv = -\omega x e^{2x} \tag{75}$$

with bathymetry

$$b(x) = -\frac{1}{2}\omega^2 x^2 - e^{2x} - \frac{1}{2}e^{-4x}$$

We perform the same comparisons done for the previous case. The discrete global flux steady state is calculated using the values of h, hu, hv on the left boundary. We report in figure 6 the error convergence plots for the global flux solutions, using the discrete norms involving all the degrees of freedom, as well as only the endpoints. For completeness we report both the solutions for h and for the transverse momentum which is not constant.



(a) $||E_h||_{L_1}$ All nodes

(b) $||E_h||_{L_1}$ End nodes

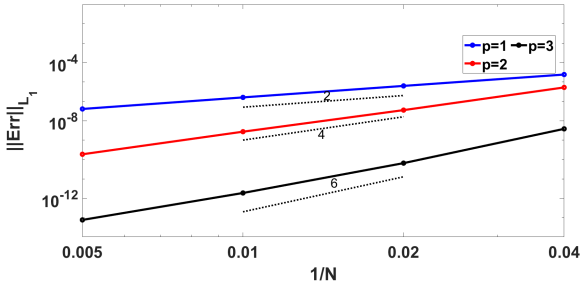
(c) $||E_{hv}||_{L_1}$ All nodes

(d) $||E_{hv}||_{L_1}$ End nodes

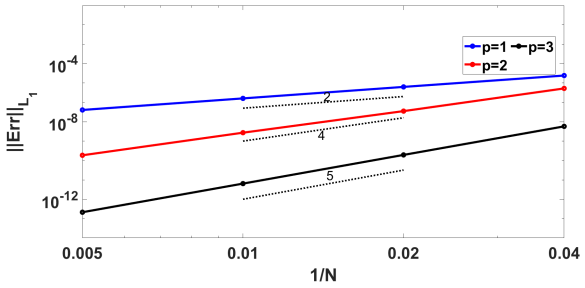
Figure 6: Pseudo-1D equilibria with Coriolis effects. Error convergence of the global flux solution with $p = 1, 2, 3, 4$ for depth (top) and transverse momentum (bottom). Error at all the degrees of freedom (left), and only elements endpoints (right).

As before, the convergence rates confirm Corollary 4. For this case, with $p = 4$ the errors reach machine accuracy extremely fast, already on the coarsest mesh for most cases. These cases are not shown in the plots for clarity. The $p = 3$ errors show a similar trend at end-nodes. This makes these schemes in practice almost exactly well-balanced with the analytical state.

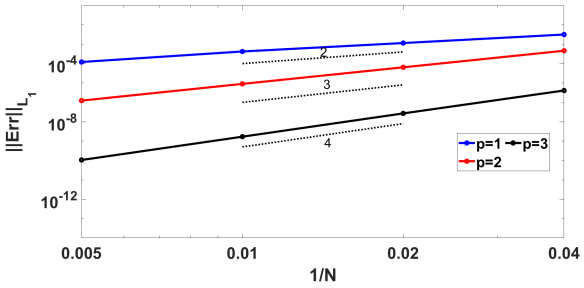
Next we run the well-balanced and non-well-balanced schemes with initial condition given by analytical equilibrium (75). The L_1 error at $T = 1$ with all nodes for both schemes is as shown in figure 7 for the depth, and in figure 8 for the transverse momentum. We only report the results for $p = 1, 2, 3$. We omit the $p = 4$ results as for which the global flux quadrature method is within machine accuracy of the exact solution. We can see that the results once again verify the rates predicted by Corollary 4, with convergence order of $p + 2$ for the total error, and $2p$ for the end points error when using the global flux quadrature. As for the previous case, the comparison with the non-well-balanced DGSEM implementation shows a gain in error of several orders of magnitude.



(a) $||E_h||_{L_1}$ WB, all nodes

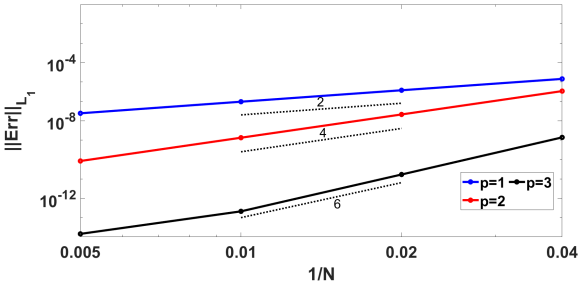


(b) $||E_h||_{L_1}$ WB

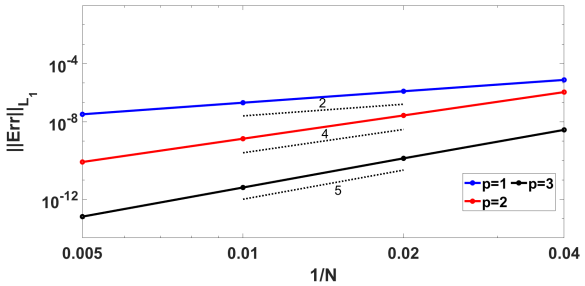


(c) $||E_h||_{L_1}$ NWB

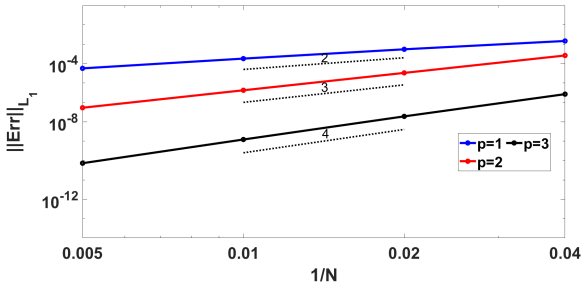
Figure 7: Pseudo-1D equilibria with Coriolis effects. Error convergence of the global flux solution with $p = 1, 2, 3, 4$ for the depth. Top: End-nodes error for the global flux quadrature method. Bottom: Total error for the global flux (left) and non-well-balanced method (right).



(a) $||E_{hv}||_{L_1}$ WB all nodes



(b) $||E_{hv}||_{L_1}$ WB all nodes



(c) $||E_{hv}||_{L_1}$ NWB

Figure 8: Pseudo-1D equilibria with Coriolis effects. Error convergence of the global flux solution with $p = 1, 2, 3, 4$ for the tranverse momentum. Top: End-nodes error for the global flux quadrature method. Bottom: Total error for the global flux (left) and non-well-balanced method (right).

7.2. Perturbations of steady states

7.2.1. Tests setup and solution visualization

In the following sections we consider a classical exercise consisting in studying the evolution of small perturbations of steady state equilibria. The initial condition for all the tests is given by

$$\begin{aligned} h(0,x) =& h^*(x) + \xi e^{-\frac{(x-x_0)^2}{100}} \\ u(0,x) =& u^*(x) \end{aligned} \tag{76}$$

where $h^*(x)$ and $u^*(x)$ are the steady distributions of water depth and speed.

The definition of the last two plays a critical role. Very often in literature one finds reference to *well prepared* initial conditions without clearly specifying what exactly this preparation of the initial data consists of. Here we consider two cases. The first is the use of the analytical solution, if known. This choice is independent of the scheme, and moreover allows to test the schemes wrt the preservation of the exact equilibrium which is the ideal situation.

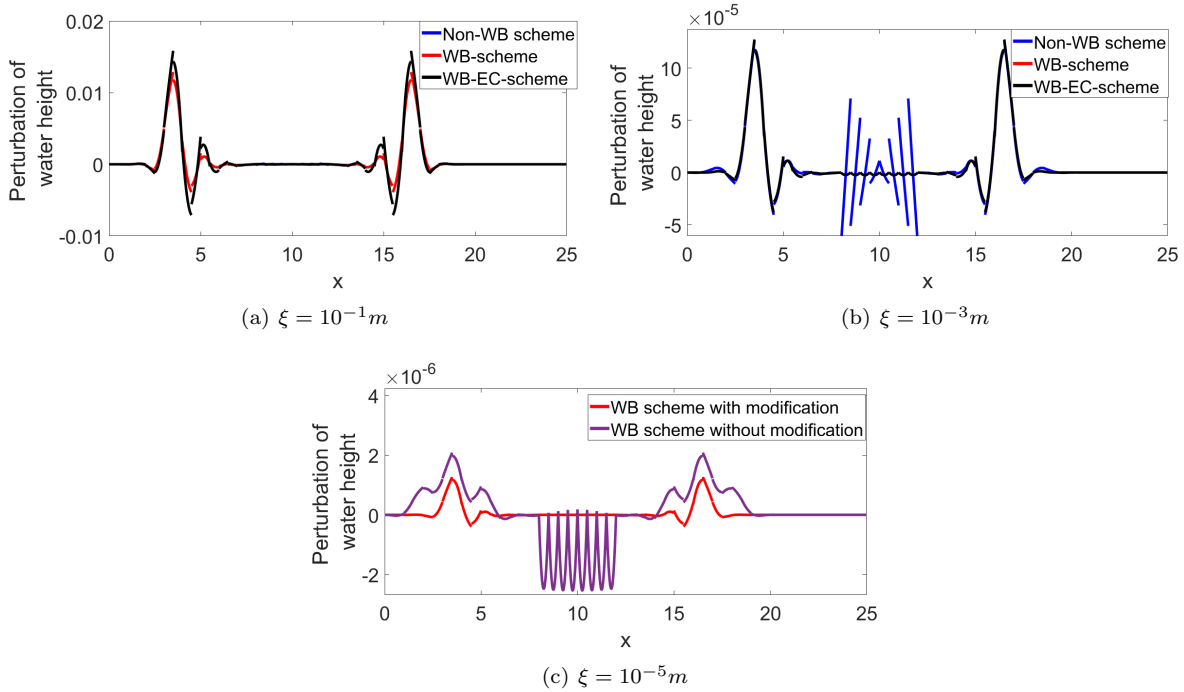
Another possible choice is to use a well defined and well behaved discrete approximation of the exact equilibrium. For example, in the works [20, 47, 50] the authors use as a natural choice the enhanced discrete approximation also employed to modify the polynomial reconstruction of the scheme. This initialization is done by construction of the discrete steady state of the well balanced scheme proposed in the references. A similar procedure can be used in our case, and would consist in using the solution provided by the ODE integrator LobattoIIIA of Proposition 3 and Corollary 4. These are reasonable choices, however slightly lacking objectivity when comparing different schemes, as they involve the discrete steady solution of one scheme in particular, which is thus favoured by this choice. An example is reported in the appendix to confirm this fact. In practice, in the following tests $h^*(x)$ and $u^*(x)$ are given by the exact analytical steady

values.

Also note that in all plots we visualize cell by cell the actual high order finite element polynomial instead of some other interpolation of the nodal values. We feel that this representation is the most faithful to the nature of the high order discontinuous finite element approach of the paper.

7.2.2. Lake at rest

For the first case we consider $u(0,x) = 0$ and $h^*(x) + b(x) = 2m$ where b is as given in (74), and the spatial domain is $x \in [0,25]m$. The boundary conditions are given by $h(t,0) = h(t,25) = 2\text{m}$ and $q(t,0) = q(t,25) = 0\text{m}^2/\text{s}$. We consider three values for the amplitude of the perturbation $\xi = 10^{-1}m$, $\xi = 10^{-3}m$, and $\xi = 10^{-5}m$. We set $x_0 = 10$ in (76), and plot the solution at $T = 1.5$. The tests in this section are performed on a mesh with $N = 50$ cells. We compare the non-well-balanced scheme, and the well-balanced scheme using (48) with and without entropy correction. Note that for this case the choice of the analytical entropy flux or of the modified one to define the cell correction has no effect. This aspect will be studied later in the results section. The results obtained here with $p = 2$ is as shown in figure 9.



(a) $\xi = 10^{-1}m$

(b) $\xi = 10^{-3}m$

(c) $\xi = 10^{-5}m$

Figure 9: Perturbation of the lake at rest. Top: NWB DGSEM (blue) and global flux quadrature approach with (black) and without (red) cell entropy correction for $p = 2$. Bottom: Global flux quadrature without entropy correction using the basic nodal source definition (47) (magenta) and the modified one (48) (red)

.

From the results in figure 9, one can see that with $p = 2$ the largest amplitude perturbation is well resolved by all methods. However, as we go lower with the amplitude of the perturbation we see the spurious oscillations of the non-well-balanced scheme. For the last case, the results of the latter cannot even be plotted on the same scale since the amplitude of the error with respect to the steady solution is several orders of magnitude higher than the perturbation. For completeness, we also report in the same figure a comparison of the global flux quadrature method with the straightforward definition of the nodal source (47), and with the modified one (48). To compare the two and highlight the advantage of the modified

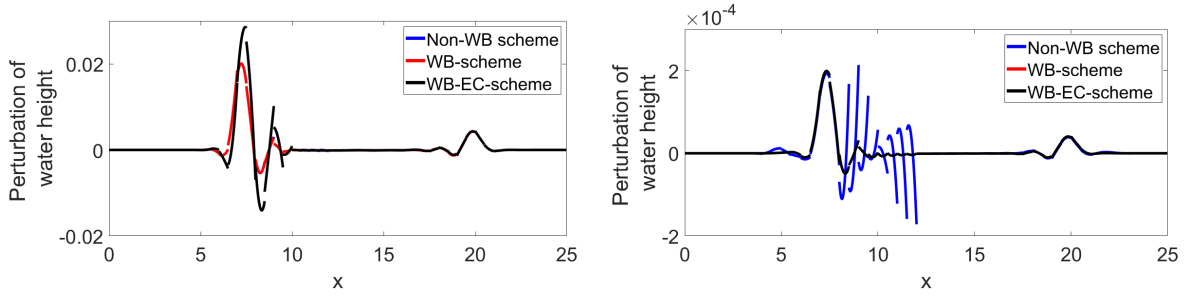
formulation we have to use a much smaller amplitude perturbation of $\xi = 10^{-5}$m. The results for the larger perturbations are almost superimposed.

7.2.3. Frictionless one dimensional moving equilibria

We consider again the equilibria defined by (12) setting respectively $(q_0, E_0) = (4.42\text{m}^2/\text{s}, 22.05535\text{m}^2/\text{s}^2)$, and $(q_0, E_0) = (4.42\text{m}^2/\text{s}, 28.8971\text{m}^2/\text{s}^2)$. The bathymetry is given in (74), and the spatial domain is $x \in [0, 25]m$. The boundary conditions are given by $h(t, 0) = h(t, 25) = 2\text{m}, q(t, 0) = q(t, 25) = 4.42\text{m}^2/\text{s}$ for sub-critical flow and $h(t, 0) = h(t, 25) = 0.66\text{m}, q(t, 0) = q(t, 25) = 4.42\text{m}^2/\text{s}$ for super-critical flow. As in the previous case, we consider the evolution of perturbations of different amplitudes of the analytical steady depth for $p = 2$. We compare the non well-balanced DGSEM, with the one using global flux quadrature with and without cell entropy correction.

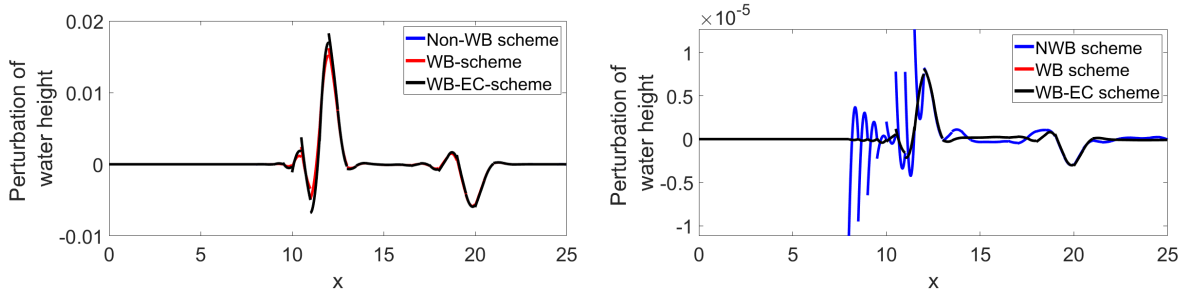
For the sub-critical case we have considered perturbations of order $\xi = 10^{-1}m$, and $\xi = 10^{-3}$m, initially set at $x_0 = 10$. The results at $T = 1.5$s are plotted on figure 10 in terms of $h(x, t) - h^*(x)$. We perform a similar exercise for the supercritical flow, only considering $\xi = 2 \times 10^{-2}$m, and $\xi = 10^{-5}$m. The initial perturbation is this time set at $x_0 = 6.25$. The results are reported in figure 11.

In the same way, we also perform a test for trans-critical flow, with $(q_0, E_0) = (1.53m^2/s, 11.0863m^2/s^2)$ and perturbations of of order $\xi = 10^{-1}m$, and $\xi = 10^{-3}$m, initially set at $x_0 = 6.25$. It is to be noted that in this case the flow for $x < 10$ is sub-critical and is super-critical for $x > 10$ with a critical point at $x = 10$. The boundary conditions for (h, hu) are calculated accordingly with respect to the given steady state. The results for the perturbation $h(x, t) - h^*(x)$ at $T = 1.5$s are shown in figure 12.



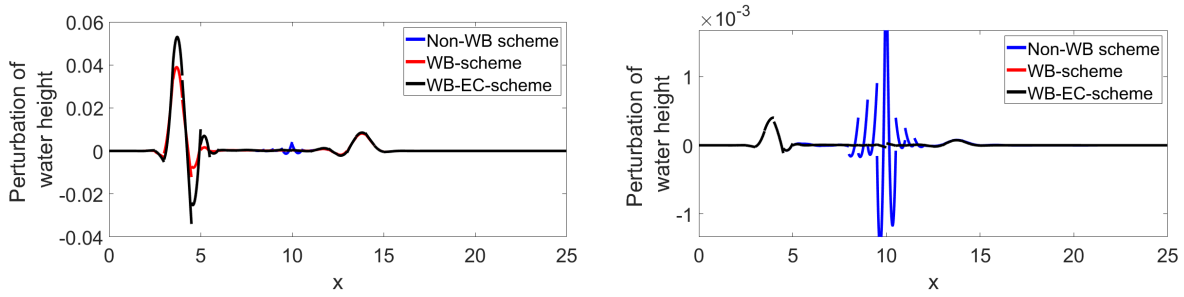
(a) $\xi = 10^{-1}$ (b) $\xi = 10^{-3}$

Figure 10: Perturbation of frictionless 1d sub-critical equilibrium. Perturbation plot for the NWB DGSEM (blue), and for the DGSEM with global flux quadrature with (black) and without (red) cell entropy correction for $p = 2$.



(a) $\xi = 2 * 10^{-2}$ (b) $\xi = 10^{-5}$

Figure 11: Perturbation of frictionless 1d super-critical equilibrium. Perturbation plot for the NWB DGSEM (blue), and for the DGSEM with global flux quadrature with (black) and without (red) cell entropy correction for $p = 2$.



(a) $\xi = 10^{-1}$ (b) $\xi = 10^{-3}$

Figure 12: Perturbation of frictionless 1d trans-critical equilibrium. Perturbation plot for the NWB DGSEM (blue), and for the DGSEM with global flux quadrature with (black) and without (red) cell entropy correction for $p = 2$.

For all three cases, sub-, trans- and super-critical flow the behaviour observed is similar to the one seen for the lake at rest state. For $p = 2$ relatively large perturbations are not affected by the well-balanced nature of the scheme, however as the amplitude of the perturbations decreases, the non-balancing error leads to spurious artefacts of amplitude larger than the physical waves.

For completeness we also report a $p-$convergence study for the sub-critical with $\xi = 10^{-3}$m case using the global flux quadrature scheme with entropy correction. The result is reported in figure 13 showing a classical behaviour: passing from $p = 1$ to $p = 2$ almost removes the phase error; passing to $p = 3$ and $p = 4$ allows to remove both phase, and amplitude errors.

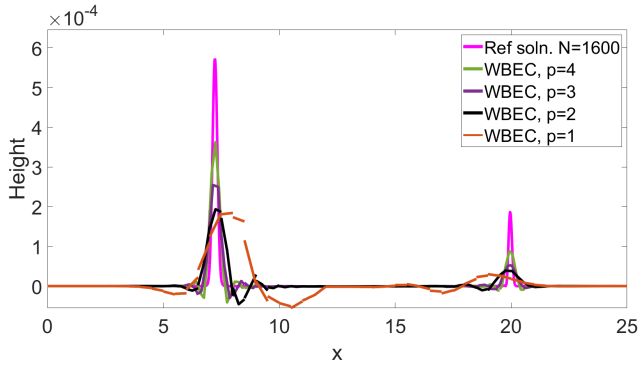


Figure 13: Perturbation of frictionless 1d sub-critical equilibrium. p-convergence study of the perturbation at $T = 1.5$s.

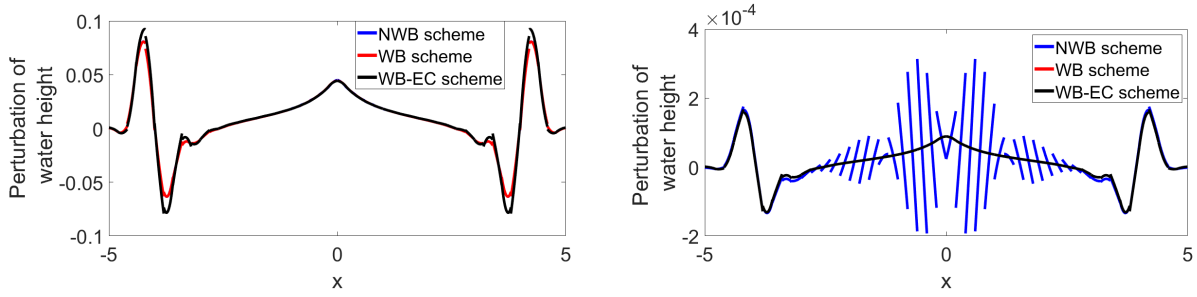
7.2.4. Pseudo-one dimensional equilibrium at rest with transverse Coriolis effects

We consider here the first configuration of the two steady states discussed in section §2.1, given analytically by the relations (15). In particular, on the domain $[-5, 5]$ we consider the transverse velocity field

$$v(x) = \frac{gx}{2} e^{-x^2}. \tag{77}$$

and set $g = 1\text{m/s}^2$, and $\omega = 2\text{s}^{-1}$. The values for $h^*(x)$ are calculated analytically from (15) with $\zeta_0 = 2$m. The steady state is then used to determine the boundary conditions for U at $x = -5, 5$. We consider perturbations with $\xi = 0.5$m, and $\xi = 10^{-3}$m, initially centered at $x_0 = 0$. The perturbation $h - h^*(x)$ at $T = 2$s is studied. We compare the global flux quadrature DGSEM with $p = 2$ with the corresponding non well-balanced formulation. Results are plotted in figure 14. As for the previous cases the largest perturbation

is evolved in a similar manner by all the schemes. However, only the well-balanced formulations correctly capture small amplitude variations as shown by the right plot in the figure.

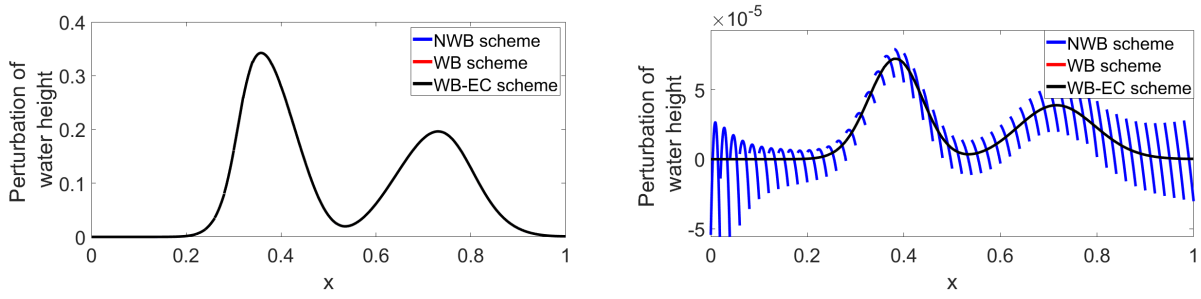


(a) $\xi = 0.5$ (b) $\xi = 10^{-3}$

Figure 14: Pseudo-1d equilibrium at rest with transverse Coriolis effects. Perturbation computed with the $p = 2$ DGSEM using the non well-balanced (blue), and the global flux quadrature forms with (black) and without (red) entropy correction.

7.2.5. Frictionless pseudo-one dimensional equilibria with Coriolis effects

We repeat the above test for the steady solution given by (75). We add to the analytical equilibrium studied in section §7.1 perturbations of amplitudes $\xi = 0.5$m, and $\xi = 10^{-4}$m at $x_0 = 0.5$. We compute the evolution with $p = 2$ until $T = 0.1$s with all the different DGSEM formulations with $p = 2$. In particular, figure 15 compares the perturbation $h - h^*(x)$ obtained with the global quadrature formulations and with the non-well-balanced scheme. We can see just as the other examples that the non-well-balanced scheme fails to capture a small perturbation, which is not the case with the global flux quadrature approach.



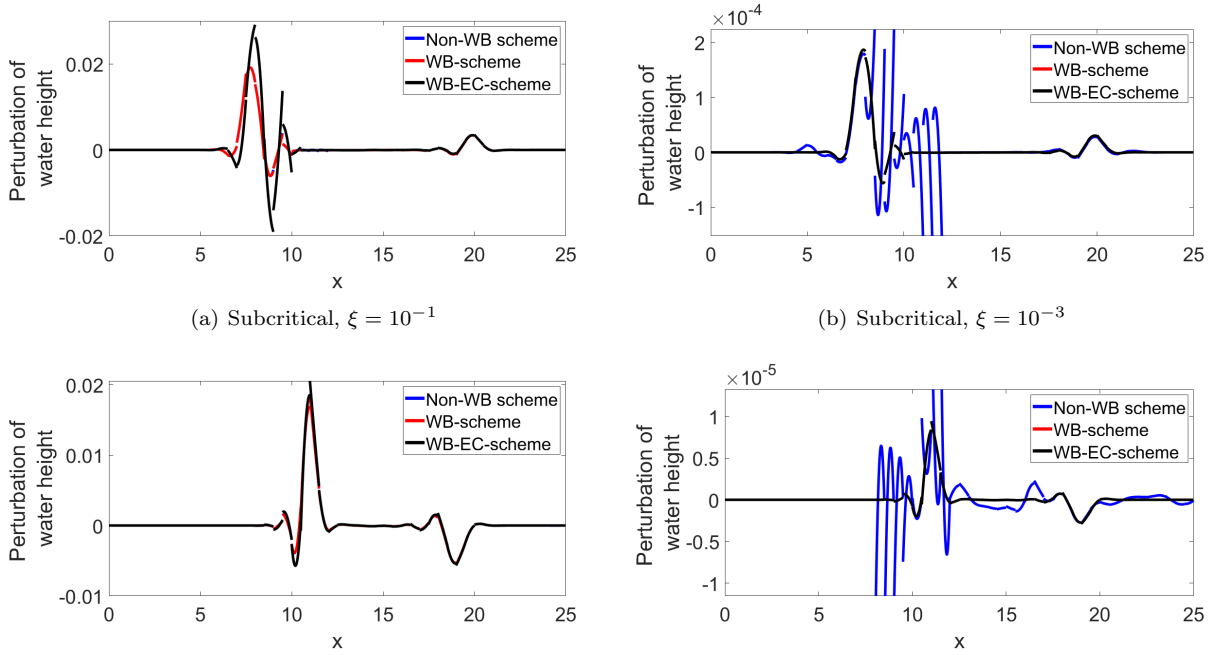
(a) $\xi = 0.5$ (b) $\xi = 10^{-4}$

Figure 15: Pseudo-1d equilibrium with Coriolis effects. Perturbation computed with the $p = 2$ DGSEM using the non well-balanced (blue), and the global flux quadrature forms with (black) and without (red) entropy correction.

7.2.6. Moving equilibria with friction

As a last case we consider a variation of the sub- and super-critical flows given by (12) and already investigated, only with the addition of friction effects. In this case, the analytical solution is characterized by (18), and a reference solution can be computed e.g. evaluating the integrals on a refined mesh. We consider the same sub-critical and super-critical states as before on the domain $[0, 25]$m, and bathymetry (74). The boundary conditions are calculated from the steady state solution of h, hu at $x = 0$ and $x = 25$. For simplicity we use constant values of the friction coefficient, set to $c_f = 0.03$ in the sub-critical case, and $c_f = 0.05$ in the super-critical case. As before we add perturbations of different amplitudes. To best visualize the difference between all the schemes, we have studied this time $\xi = 10^{-1}$m and $\xi = 10^{-3}$m in the sub-critical case. For the super-critical equilibrium we have set $\xi = 2 \times 10^{-2}$m, and $\xi = 10^{-5}$m. We compute the solution with $p = 2$ until $T = 1.5$s. We compare the solutions obtained with the different schemes in

figure 16. The results lead to similar conclusions as for all the other equilibria, and source term types.



(a) Subcritical, $\xi = 10^{-1}$

(b) Subcritical, $\xi = 10^{-3}$

(c) Supercritical, $\xi = 2 \times 10^{-2}$

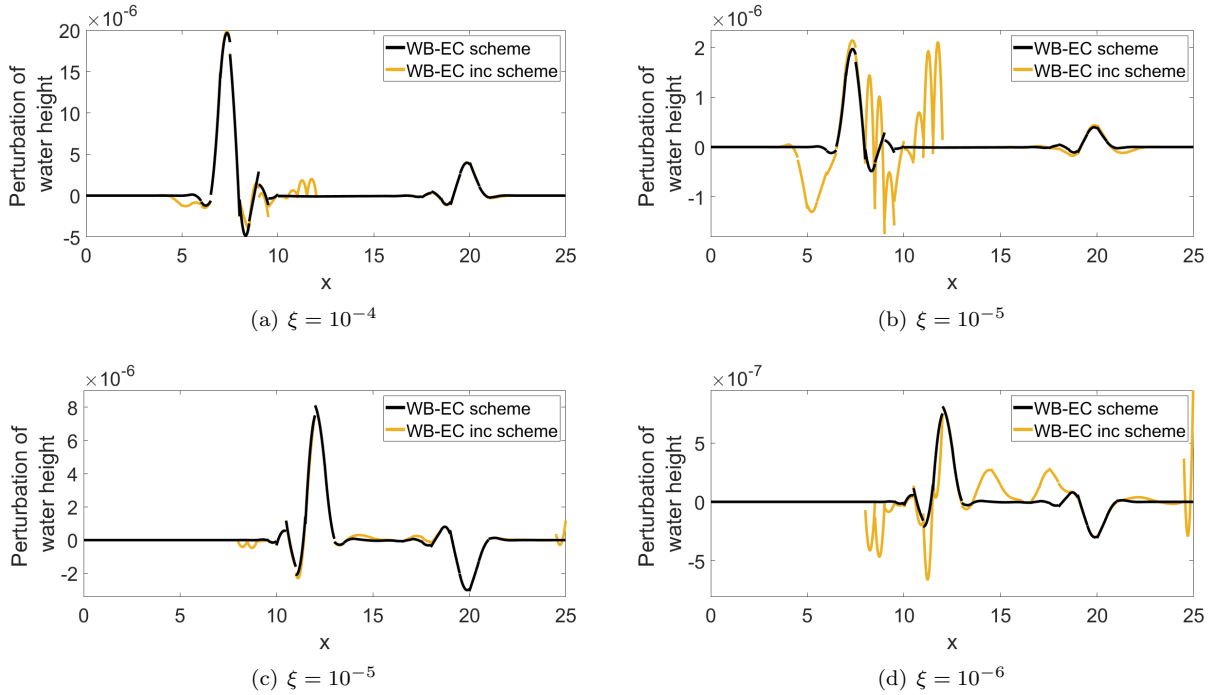
(d) Supercritical, $\xi = 10^{-5}$

Figure 16: Perturbation of moving equilibria with friction. Top: sub-critical case. Bottom: super-critical case. Comparison of the DGSEM with $p = 2$ global quadrature formulations with (black) and without (red) cell entropy correction, with the non well-balanced one (blue).

7.3. Entropy control vs well-balancing

We have so far considered entropy correction terms always consistent with the initialization used. In this section we check the impact of this consistency, as well as the effect of the adding the cell correction on the fully discrete evolution of the entropy.

To begin with, we consider a slight variation of the study of the previous section, and we perturb the initial data given by the global flux solutions for the moving one dimensional sub-critical and supercritical flows without friction. We consider small perturbations with amplitudes 10^{-4}m and 10^{-5}m for the sub-critical case, and 10^{-5}m and 10^{-6}m for the super-critical one. The results, in figure 17, show that for small amplitude perturbations properly accounting for the initialization strategy is important. Despite of the fact that the superconvergence of the global flux solution to the analytical one, the use of analytical entropy fluxes in this case spoils the well-balanced character of the scheme, leading to spurious waves comparable to the physical ones. Note that this effect is a much weaker effect than the one observed with the non well-balanced method. However, in some cases it may be necessary to have machine accuracy preservation of a physically relevant initial state, and initializing with the global flux solution may be relevant. In these cases, the use of (66) in the entropy correction is necessary.



(a) $\xi = 10^{-4}$ (b) $\xi = 10^{-5}$

(c) $\xi = 10^{-5}$ (d) $\xi = 10^{-6}$

Figure 17: Frictionless 1D moving equilibria. Perturbations of the discrete global flux solution in the sub-critical (top) and super-critical (bottom) case. Comparison between using consistent (black) and inconsistent (yellow) entropy corrections.

We now look at the impact of the entropy correction term on the fully discrete time evolution of the total entropy/energy $\sum_K \int_K \eta_h$. Note that nothing special is done concerning the time integration, so the conservation property holds exactly only in the time continuous case. Fully discrete variants can be obtained by other means, but are not considered here (see e.g.[67, 2, 41]). Time integration is in particular performed with standard RK schemes of the same formal order of the underlying spatial discretization. We will comment on two examples. The first is the perturbation of amplitude $\xi = 10^{-1}$m of the sub-critical and supercritical moving equilibria without friction. The evolution of entropy until time $T = 50$s in the $p = 2$ case for the NWB, and for the global flux quadrature with and without cell entropy correction are shown in figure 18. We can see that in both cases the entropy correction allows to reduce the variation of entropy in time, improving its conservation.

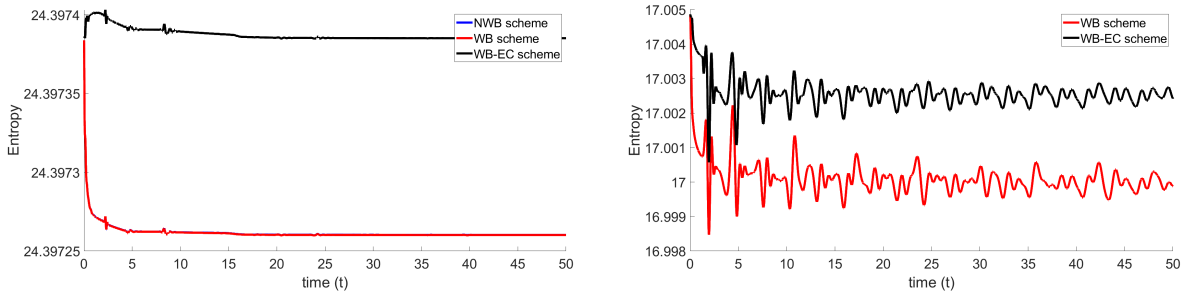


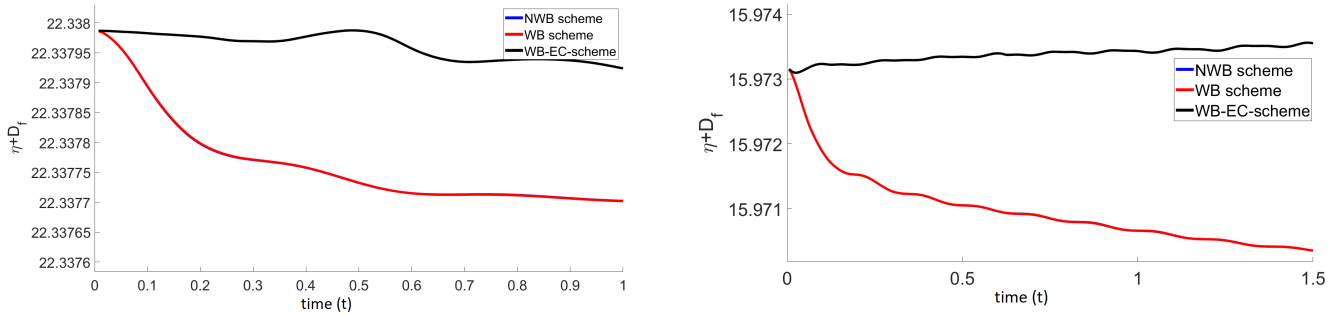
Figure 18: Time discrete evolution of the total energy in the domain for a perturbation of amplitude $\xi = 10^{-1}$m to sub-critical (left) and super-critical frictionless moving 1d equilibrium (right). Global quadrature DGSEM with entropy correction (black) compared to the basic global quadrature DGSEM (red) and to the non well-balanced DGSEM (blue).

As a second example, we consider a perturbation $\xi = 0.1$ of moving equilibria with friction in sub-critical

and supercritical conditions. The entropy evolution is shown in figure 19. In the figures, following Remark 8, we study the evolution of the corrected total entropy balance

$$N_{\text{tot}} := \sum_K \left(|K| \bar{\eta}_\varphi + \int_{t_0}^{t} \mathcal{D}_f \right)$$

which should constant and equal to $\sum_K |K| \bar{\eta}_\varphi(t=0)$. We see from figure 19 that with the entropy correction term allows to obtain good control of the dissipation of the scheme.



(a) Sub-critical, $\xi = 10^{-1}$

(b) Super-critical, $\xi = 10^{-1}$

Figure 19: Perturbation of moving equilibria with friction: evolution of the entropy balance. Left: sub-critical flow. Right: supercritical flow.

7.4. 2d examples and applications

We consider now the application of the scheme proposed to more complex tests. On the one hand we want to investigate the advantage of using the tensor product extension proposed, which is only well-balanced along mesh lines, on genuinely 2D problems. On the other, we want to test the method on problems involving sharp propagating fronts, and complex structures to assess its robustness.

7.4.1. 2d perturbations of 1d equilibria

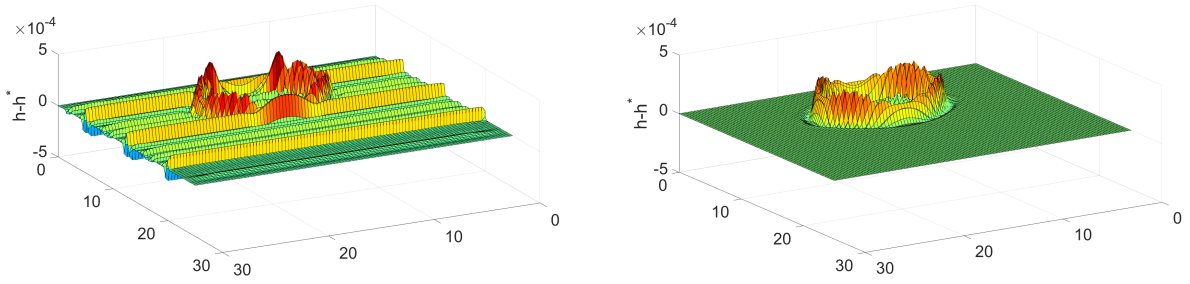
We start from a simple benchmark: the evolution of 2D perturbations to one dimensional exact steady solutions. We consider first the lake at rest with initial and boundary condition with $\zeta_0 = 5.47$m, and $hu = hv = 0$. The bathymetry is given by a series of bumps defined as

$$b(x) = \begin{cases} 0.2 - (x - (4.5k - 0.75))^2/20 & 4.5k - 3 < x < 4.5k + 1.5,\ k = 1,2,3,4,5 \\ 0 & \text{otherwise} \end{cases}$$

on the domain $x \in [0, 25]$ and $y \in [0, 25]$. We add a two dimensional perturbation to this equilibrium flow defined as (see also figure 21(a))

$$h = h^* + 0.05 e^{-100((x-10)^2 + (y-12.5)^2)}. \tag{78}$$

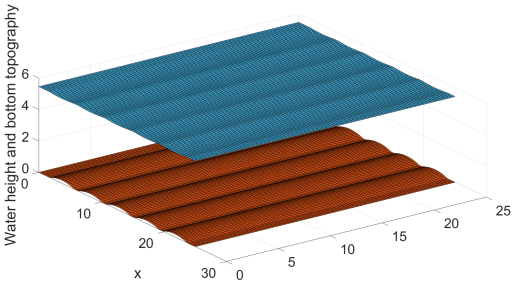
We evolve the solution until $T = 2$s on a 50×50 mesh with both the non-well-balanced and global flux quadrature schemes with $p = 2$. The results are compared on figure 20. As expected, the non well-balanced formulation introduced spurious numerical oscillations of size comparable to the amplitude of the perturbation. These perturbations are completely absent in the well-balanced formulation which preserves the steady state exactly.



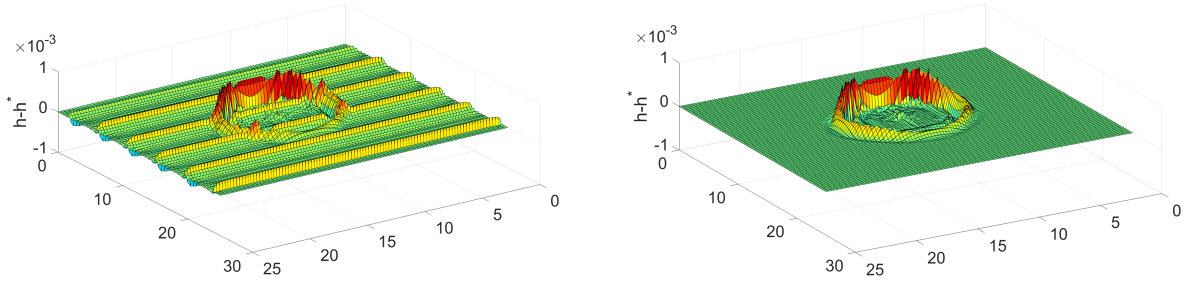
(a) NWB (b) WB

Figure 20: Numerical solution for $h - h^*$ with NWB, WB, WB-EC discontinuous Galerkin scheme

Next we consider a 1D equilibrium where initial condition is calculated using (12) taking $q_0 = 5.6865$ and $E_0 = 54.183738$. As in the previous example, we add the perturbation (78) which is evolved until $T = 2$s on a 50×50 mesh with both the non-well-balanced and global flux quadrature schemes with $p = 2$. The results are shown in figure 21. The NWB scheme is not able to capture the perturbation accurately as the numerical error is as large as the physical perturbation. The global flux quadrature approach allows to obtain a correct preservation of the background flow, and a nice evolution for the perturbation. It may be noted that the results of well-balanced scheme with entropy correction are extremely close, so only the first is shown.



(a) Steady state and bathymetry



(b) NWB (c) WB

Figure 21: Numerical solution for $h - h^*$ with NWB, WB discontinuous Galerkin scheme

2d stationary vortex with bathymetry. This is a genuinely 2D steady state. It is a variation of a test often used in literature, see e.g [9] with added bathymetric effects. The initial condition is given by

$$h^*(r,0)+b(x,y)=1+\xi^2\begin{cases}\frac{5}{2}(1+5\xi^2)r^2 & r\leq\frac{1}{5}\\ \frac{1}{10}(1+5\xi^2)+2r-\frac{3}{10}-\frac{5}{2}r^2 & \\ \qquad+\xi^2[4ln(5r)+\frac{7}{2}-20r+\frac{25}{2}r^2] & \frac{1}{5}<r\leq\frac{2}{5}\\ \frac{1}{5}(1-10\xi^2+20\xi^2ln(2)) & r>\frac{2}{5}\end{cases}\tag{79}$$

$$u^*(x,y,0)=-\xi y\Upsilon(r),\quad v^*(x,y,0)=\xi x\Upsilon(r),\quad \Upsilon(r)=\begin{cases}5 & r\leq\frac{1}{5}\\ \frac{2}{r}-5 & \frac{1}{5}<r\leq\frac{2}{5}\\ 0 & r>\frac{2}{5}\end{cases}$$

where $r=\sqrt{x^2+y^2}$. For this test we consider $\xi=0.1$ and bathymetry defined as

$$b(x,y)=\begin{cases}0.1(1-6.25(x^2+y^2)) & \text{if } (x^2+y^2)<0.16\\ 0 & \text{otherwise}\end{cases}$$

on the domain $[-1,1]\times[-1,1]$. The boundary condition is given by $h=1+\frac{1}{5}\xi^2(1-10\xi^2+20\xi^2ln(2))\text{m}, u=v=0\text{m/s}$ on all the boundaries. As scaled visualization of the initial free surface is reported on the left on figure 22. We report in the same figure the grid convergence with the global flux quadrature and with the non well-balanced method for different degree approximations. Note that (79) only provides C^0 continuity. As shown in [71] this will limit the convergence attainable with high order schemes to roughly second order, which is the slope obtained here. The convergence plot however also shows that, despite the fact that we do not embed a genuinely two dimensional well-balanced criterion, the global flux quadrature formulation still provides a considerable decrease in error especially for $p=2$ and $p=3$. In particular, the errors of the WB $p=2$ are quite comparable with those of the fourth order non well-balanced DGSEM.

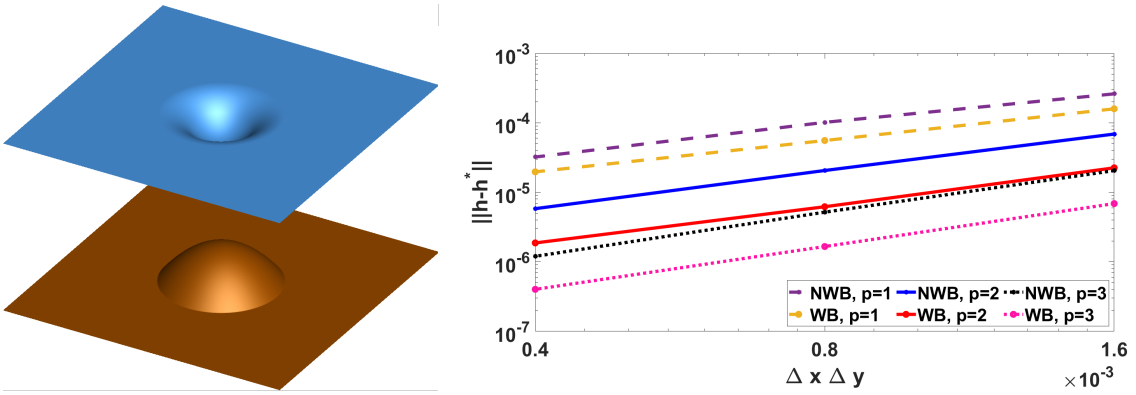
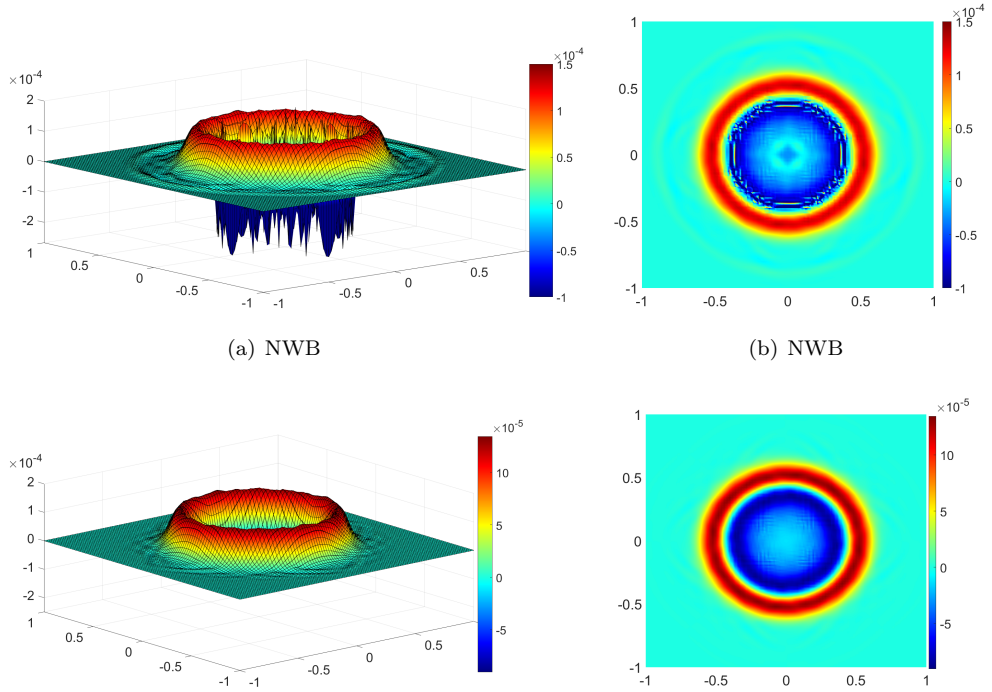


Figure 22: Steady vortex. Left: scaled 3D view of the initial solution (free surface and bathymetry). Right: grid convergence with $p=1,2,3$.

We also test the evolution of a small perturbation added to water height and given by

$$h=h^*+10^{-3}e^{-100(x^2+y^2)}.$$

The perturbation is evolved up to $T=0.05$ with the third order schemes on a 50×50 grid. Snapshots of the perturbation at the final time are reported in figure 23. The NWB scheme introduces spurious oscillations in the solution over time, while the well-balanced method provides a much cleaner resolution of the waves.



(a) NWB (b) NWB

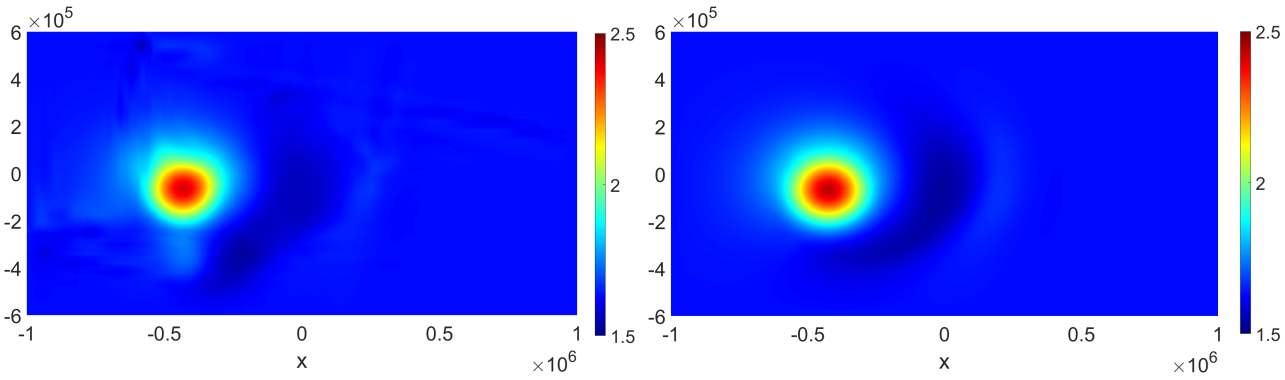
(c) WB (d) WB

Figure 23: Numerical solution for $h - h^*$ with NWB, WB discontinuous Galerkin scheme

Anticyclonic vortex propagation. This is a moving variant of the previous case. It consists of a vortex propagating westward due to the effect of the variation of the Coriolis coefficient which is modified as

$$\omega = \omega_0 + \beta y$$

to model the effects curvature effects using a tangent plane approximations. The domain a rectangular basin of 2000×1200km. The initial condition is given by a Gaussian distribution of the free surface centered at the origin of the domain, prescribed together with a velocity field which is in geostrophic balance. We refer to [21, 59] and references therein for details on the test setup, and values of the different parameters. The solution is computed with the third order schemes until a final time corresponding to 8 weeks on a relatively coarse mesh of 50×30km.



(a) NWB (b) WB

Figure 24: Anticyclonic vortex propagation.

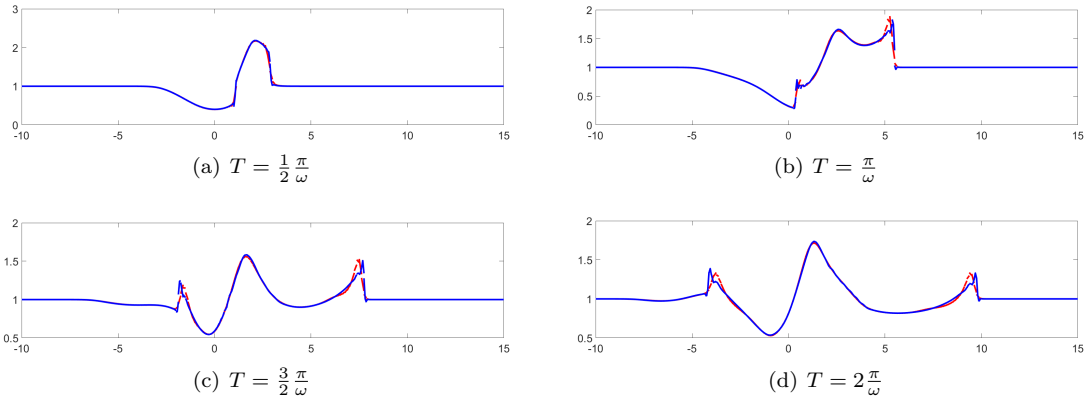
We plot the free surface contour levels obtained with the non well-balanced and global flux quadrature methods on figure 24. The results confirm the observations made for the steady vortex: the well-balanced scheme provides a much cleaner solution, while some spurious waves around the vortex are clearly visible in the non well-balanced result.

Geostrophic adjustment in 1d and 2d. We consider now the simulation of the complex wave dynamics of the geostrophic adjustment in one and two space dimensions. We start from the one dimensional numerical test for geostrophic adjustment used in [14]. The initial conditions are given by

$$h(x,0) = 1, \quad u(x,0) = 0 \tag{80}$$

$$v(x,0) = 2\frac{(1+\tanh(4\frac{x}{L}+2))(1-\tanh(4\frac{x}{L}-2))}{(1+\tanh(2))^2} \tag{81}$$

with $L = 2\text{m}$, and $\omega = 1\text{s}^{-1}$, $g = 1\text{m/s}^2$ in the rotating shallow water equations. The initial disturbance in the momentum leads to two fast inertial gravity waves and shock formation. figure 25 gives the solution for free water surface at time $T = \frac{1}{2}\frac{\pi}{\omega}, \frac{\pi}{\omega}, \frac{3}{2}\frac{\pi}{\omega}, 2\frac{\pi}{\omega}$ computed using a second and third order well-balanced discontinuous Galerkin scheme with a grid of $N = 200$ in the domain $[-10, 15]$m. The boundary condition is given by $(h,u,v)(-10,t) = (h,u,v)(15,t) = (1,0,0)$.



(a) $T = \frac{1}{2}\frac{\pi}{\omega}$ (b) $T = \frac{\pi}{\omega}$

(c) $T = \frac{3}{2}\frac{\pi}{\omega}$ (d) $T = 2\frac{\pi}{\omega}$

Figure 25: Solution with 2nd and 3rd order well-balanced DG scheme at $T = \frac{1}{2}\frac{\pi}{\omega}, \frac{\pi}{\omega}, \frac{3}{2}\frac{\pi}{\omega}, 2\frac{\pi}{\omega}$

From the solution we see that both the schemes are able to capture the two fast waves accurately in time, with sharp front almost free of oscillations despite the absence of any limiter. The third order scheme

is less diffusive and is able to capture the shock front more accurately.

Next, consider the two-dimensional extension of this test, proposed in [21]. In this case the initial conditions are as follows

$$h(x,y,0) = 1 + 0.25 * (1 - \tanh(\frac{\sqrt{2.5x^2 + y^2/(2.5)} - 1}{0.1})) \tag{82}$$

$$u(x,y,0) = v(x,y,0) = 0 \tag{83}$$

on the spatial domain $[-10,10]\text{m} \times [-10,10]\text{m}$. The initial free surface is visualized in figure 26, while figure 27 reports the free surface at times $T = 4, 8, 12, 20\text{s}$ obtained with the third order well-balanced method using a 50×50 mesh.

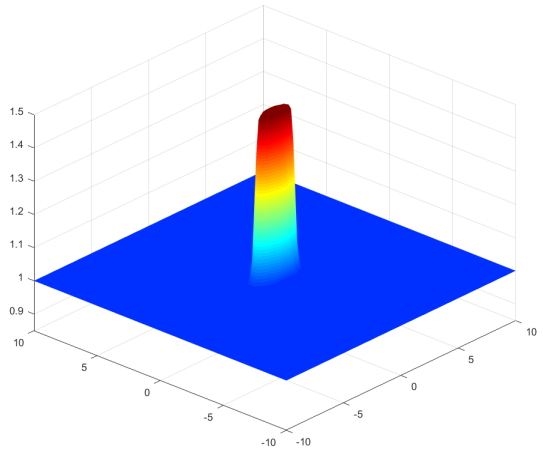
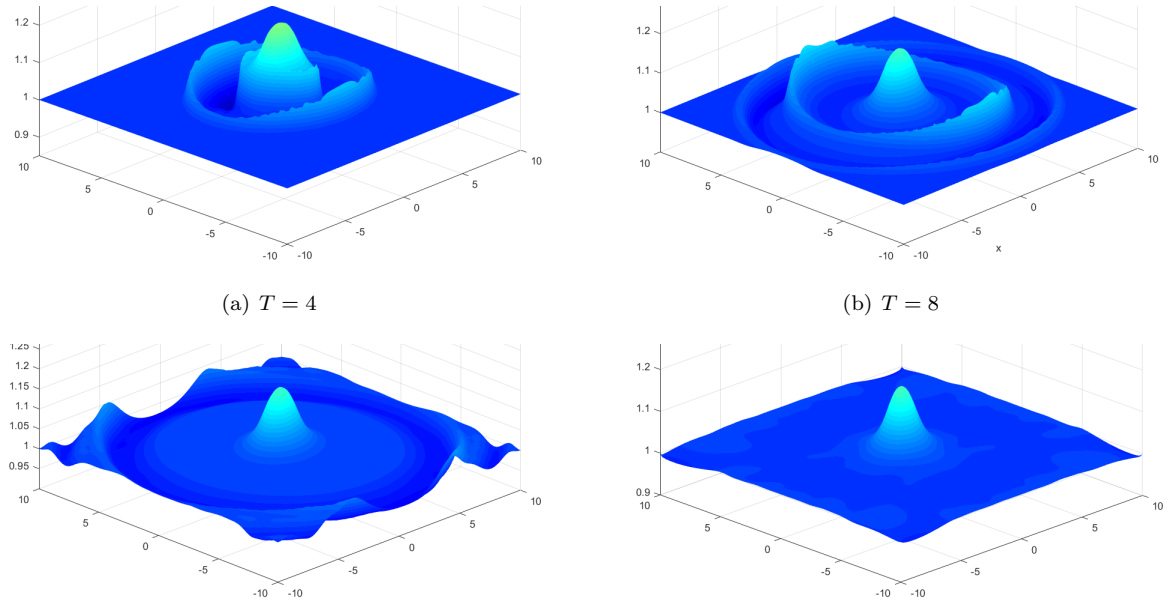


Figure 26: Initial free surface for the 2d geostrophic adjustment problem by [21]



(a) $T = 4$ (b) $T = 8$

(c) $T = 12$ (d) $T = 20$

Figure 27: 2d geostrophic adjustment problem. Water surface at $T = 4, 8, 12, 20$

The solution for this problem is given by two shock waves which move radially away from the center leaving behind some mass which rotates at the initial position. We see that the well-balanced scheme is able to capture these shock waves accurately and the final equilibrium is achieved without any unphysical effects due to the scheme.

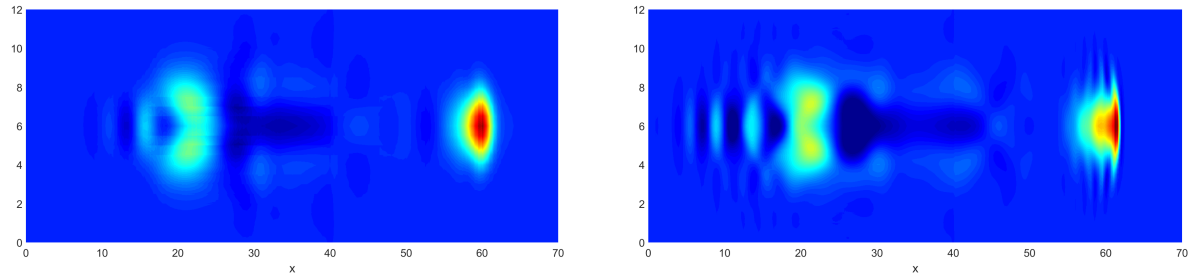
Kelvin front generation on the equatorial β-plane. As a last test, we show the capabilities of the scheme proposed to capture the generation of short secondary waves. The test considered involves the formation and propagation of Kelvin and Rossby waves with a generation of a Kelvin front and secondary Poincaré waves. The set up is the the same as in [59, 36]. The computational domain $[0, 70] \times [0, 12]$ is divided into 140×24 cells. The initial condition and bottom topography are given by,

$$h(x, y, 0) = 2 - b(x, y) + 0.8exp\big(- \frac{(x - 30)^2 + (y - 6)^2}{3}\big)$$

$$b(x, y) = \begin{cases} 0 & x \leq 40 \\ 0.025x - 1 & x > 40 \end{cases}$$

with $g = 1\text{m/s}^2$. As for the moving vortex case, the Coriolis coefficient is modified to model the effects of curvature. In this case, following [59] we set $\omega(y) = y - 6$. Figure 28 shows the solution at $T = 20$ with a second and third order well-balanced discontinuous Galerkin scheme.

The solution consists of the short wavelength Kelvin waves which carry energy eastwards and the long wavelength Rossby waves carrying energy westward. The nonlinear Kelvin waves steepen and eventually break forming a broken wave front propagating eastward which leads to generation of the secondary Poincaré waves. From the solution in figure 28, we see that both the 2nd and 3rd order scheme are able to capture the general physical behavior of both the Kelvin and Rossby waves. However only the third order scheme is able to capture the Poincaré waves. Indeed, the solution obtained with the second order method gives a Kelvin front too diffused to trigger the generation of these secondary waves.



(a) 2nd order

(b) 3rd order

Figure 28: Numerical solution for $h + b$ with 2nd and 3rd order discontinuous Galerkin scheme

8. Conclusions and perspectives

In this work, we have studied a new formulation to construct numerical methods verifying a discrete fully well-balanced criterion agnostic of any specific form of steady equilibrium. In one space dimension, the underlying discrete well-balanced criterion is based on an equivalence between the discontinuous Galerkin spectral element method and a superconvergent Gauss collocation integrator applied to an ODE for the flux. In practice, this property is obtained by means of a simple modification of the integral of the source term. This discrete well-balanced DGSEM approach, which we named global flux quadrature based DGSEM, allows a clear characterization of the discrete steady state, with provable superconvergence estimates to general analytical steady states.

In addition, a cell correction term is proposed to verify cell entropy balance. The accuracy and the consistency of this term with the discrete well-balanced condition proposed is characterized. The numerical benchmarks for the shallow water equations in one space dimension confirm all the theoretical properties on a wide variety of equilibria, involving several different source terms.

Applications on complex problems involving sharp fronts and complex wave patterns confirm the accuracy and show the robustness of the method.

Several perspectives are being explored including the design of non-linear variants allowing to handle genuinely discontinuous solutions, other discretization techniques (e.g. finite volumes), and more general formulations in particular for time-dependent and multidimensional problems.

Appendix A. Proof of the consistency estimate of Proposition 6

We prove the last part of the proposition namely and in particular the consistency estimate $|\mathcal{E}| = \mathcal{O}(h^{p+1})$. To this end, following [5], we consider a smooth exact solution $U^e \in C^p$, and its projection onto the local finite element space U^e_{h}, as well as a smooth compactly supported test function $v \in C^k(\Omega)$, with $k \geq 1$, with v_{h} denoting its projection onto the local finite element space. Proceeding as in [5] we then formally replace U_{h} by U^e_{h} in the scheme. The reminder is a local measure of the consistency error. We now multiply the i-th nodal reminder in each element K by the projected nodal value of the test function v_i, and define the consistency error as in (57) by summing up over all degrees of freedom i, and over all elements K. Then as in [5] first we rewrite the error in weak form and then perform and estimation of each term appearing in the expression. Using the definitions of the fluctuations (40) appearing in the error (57), of the artificial viscosity term $\mathcal{D}_i^K$ (equation (51)), invoking the SBP property of the DGSEM method, and including a rest to account for the inexactness of the underlying Gauss-Lobatto quadrature, we can write the error (57) in

the following form

$$\begin{aligned}\mathcal{E} = \sum_K \int_K v_\mathrm{h} \partial_t U^e_\mathrm{h} - \sum_K \int_K \partial_x v_\mathrm{h} F_\mathrm{h}(U^e_\mathrm{h}) - \sum_K \int_K \partial_x v_\mathrm{h} R^e_\mathrm{h} + \sum_f [\![v_\mathrm{h}]\!] \hat{G}_\mathrm{h}(U^e_\mathrm{h}) \\ + \sum_K \mathcal{R}^{\mathsf{GL}}_K + \sum_K \alpha_K(U^e_\mathrm{h}) \int_K \partial_x v_\mathrm{h} A_0(U^e_\mathrm{h}) \partial_x U^e_\mathrm{h},\end{aligned} \tag{A.1}$$

where the term involving the jump of the test function is a result of the locality of the finite element projection. We now consider the PDE applied to the exact solution $\partial_t U^e + \partial_x G^e = 0$, and in each element we multiply t by v_h and integrate by parts to get

$$\int_K v_\mathrm{h} \partial_t U^e - \int_K \partial_x v_\mathrm{h} G^e + (v_\mathrm{h} G^e)^R - (v_\mathrm{h} G^e)^L = 0,$$

where $G^e = F^e - R^e$ with source flux R^e is obtained by exact integration of (assuming the initial state on the left hand of the domain to be zero)

$$R^e = \int_{x_0}^{x} S(U^e, s). \tag{A.2}$$

Summing up over element and subtracting from (A.1), and omitting the boundary conditions (due to the compactness of v), we can write

$$\begin{aligned}\mathcal{E} = \overbrace{\int_\Omega v_\mathrm{h} \partial_t (U^e_\mathrm{h} - U^e)}^{\mathrm{I}} - \overbrace{\int_\Omega \partial_x v_\mathrm{h} (F_\mathrm{h}(U^e_\mathrm{h}) - F(U^e))}^{\mathrm{II}} - \overbrace{\int_\Omega \partial_x v_\mathrm{h} (R^e_\mathrm{h} - R^e)}^{\mathrm{III}} \\ \underbrace{\sum_f [\![v_\mathrm{h}]\!] (\hat{G}_\mathrm{h}(U^e_\mathrm{h}) - G^e)}_{\mathrm{IV}} + \sum_K \mathcal{R}^{\mathsf{GL}}_K + \sum_K \alpha_K(U^e_\mathrm{h}) \int_K \partial_x v_\mathrm{h} A_0(U^e_\mathrm{h}) \partial_x U^e_\mathrm{h}\end{aligned} \tag{A.3}$$

We now proceed to a term by term estimate. Following [5], we start from noting that for a test function $v \in C^k(\Omega)$ with $k \geq 1$ we can readily claim that

$$\|v_\mathrm{h}\| \leq C_1 < \infty\,, \quad \|\partial_x v_\mathrm{h}\| \leq C_2 < \infty\,, \quad [\![v_\mathrm{h}]\!] \leq C_3 \max(\mathrm{h}^2, \mathrm{h}^{p+1}) \tag{A.4}$$

having omitted the L^2_Ω subscript from the norms to simplify the notation. This allows to bound terms I and II by the approximation error as

$$|\mathrm{I}| \leq \tilde{C}_\mathrm{I}(\Omega, \|v\|)\, \mathrm{h}^{p+1}\,, \quad |\mathrm{II}| \leq \tilde{C}_\mathrm{II}(\Omega, \|\partial_x v\|)\, \mathrm{h}^{p+1} \tag{A.5}$$

The term IV can also be easily controlled by the approximation error by using the form of the numerical flux (41) which allows to write (using $^+$ and $^-$ for quantities on the two sides of an element face)

$$\begin{aligned}|\mathrm{IV}| =& |\sum_f \left\{ \alpha [\![v_\mathrm{h}]\!] ((G^e_\mathrm{h})^+ - G^e) + (1-\alpha) [\![v_\mathrm{h}]\!] ((G^e_\mathrm{h})^- - G^e) + \mathcal{D} [\![v_\mathrm{h}]\!] [\![U^e_\mathrm{h}]\!] \right\}| \\ \leq& \alpha \sum_f |[\![v_\mathrm{h}]\!] ((G^e_\mathrm{h})^+ - G^e)| + (1-\alpha) \sum_f |[\![v_\mathrm{h}]\!] ((G^e_\mathrm{h})^- - G^e)| + \sum_f |\mathcal{D} [\![v_\mathrm{h}]\!] [\![U^e_\mathrm{h}]\!]|.\end{aligned}$$

Now by standard approximation arguments [30], we can claim that on each side of a given face the projection of the exact quantiles provides en error of $\mathcal{O}(\mathrm{h}^{p+1})$ so that the jumps are also of the same order. Since the number of faces is (in 1d) of order h^{-1}, this readily allows to show that

$$|\mathrm{IV}| \leq \tilde{C}_\mathrm{IV}(\Omega, \|v\|)\, \mathrm{h}^{p+1} \mathrm{h}^{\min(1,p)}\,. \tag{A.6}$$

To bound term III, we first consider the estimate of the quadrature rest. Under the the smoothness hypotheses made, for a finite element approximation of degree p in one dimension we can use the exactness of the Gauss-Lobatto formulas for polynomials of degree $2p-1$. To estimate the integration remainder, one can e.g. introduce a local truncated Taylor series of the exact integrand and consider the first term not integrated exactly which can be bounded by an $\mathcal{O}(\mathrm{h}^{2p})$. This leads on each element to an integration rest of order $\mathcal{R}_K^{\mathsf{GL}} = \mathcal{O}(\mathrm{h} \times \mathrm{h}^{2p}) = \mathcal{O}(\mathrm{h}^{2p+1})$. Since the number of elements in one space dimension are of $\mathcal{O}(\mathrm{h}^{-1})$ we deduce that $\sum_K \mathcal{R}_K^{\mathsf{GL}} \leq C_{\mathsf{GL}} \mathrm{h}^{2p}$.

To estimate $(R_\mathrm{h}^e - R^e)(\bar{x})$ for a given point $\bar{x}$, we introduce the set $\mathcal{K}_{\bar{x}}$ of elements such that $x < \bar{x}$ within the element, and we denote by $K_{\bar{x}}$ the element containing the point. We can thus write

$$R_\mathrm{h}^e - R^e = \sum_{K \in \mathcal{K}_x} \left(\int_K S_h - \int_K S(U^e, x) \right) + \int_{x_{K_{\bar{x}}}^L}^{\bar{x}} S_h - \int_{x_{K_{\bar{x}}}^L}^{\bar{x}} S(U^e, x) \tag{A.7}$$

with all integrals computed exactly. As before for all the elements in $\mathcal{K}_{\bar{x}}$ we can use the properties of the Gauss-Lobatto formulas, and bound the first term by a $\mathcal{O}(\mathrm{h}^{2p+1})$. For the second term we can use the approximation error to estimate $S_\mathrm{h} - S$, and we end with

$$|R_\mathrm{h}^e - R^e| \leq C_1^S \mathrm{h}^{2p} + C_2^S \mathrm{h}^{p+2} = \bar{C}^S \min(\mathrm{h}^{2p}, \mathrm{h}^{p+2}). \tag{A.8}$$

We are left with the estimate of the cell correction term, evaluated using sampled values of the exact solution. We start by re-writing the coefficient α_K, re-writing the second in (54), (56), and (58) which under the current hypotheses become (cf. also equation (65))

$$\Phi_\eta^K(U_\mathrm{h}^e) = \int_K (W_\mathrm{h}^e)^T \partial_x G_\mathrm{h}\,, \quad \Psi_\eta^K(U_\mathrm{h}^e) = \int_K \partial_x F_\eta(U_\mathrm{h}^e)\,. \tag{A.9}$$

Noting that

$$\int_K \partial_t \eta_\mathrm{h} = \sum_{i=0,p} w_i \frac{d\eta_i}{dt} = \sum_{i=0,p} w_i W_i^T \frac{dU_i}{dt}, \tag{A.10}$$

we can readily write

$$\Psi_\eta^K(U_\mathrm{h}^e) - \Phi_\eta^K(U_\mathrm{h}^e) = \int_K \left[(\partial_t \eta_\mathrm{h}(U_\mathrm{h}^e) + \partial_x F_\eta(U_\mathrm{h}^e)) - (W_\mathrm{h}^e)^T (\partial_t U_\mathrm{h}^e + \partial_x G_\mathrm{h}(U_\mathrm{h}^e)) \right] + \mathcal{O}(\mathrm{h}^{2p+1}), \tag{A.11}$$

where, following the previous reasoning, the last term is the Gauss-Lobatto quadrature error on the $(W_\mathrm{h}^e)^T(\partial_t U_\mathrm{h}^e$ term. The remaining terms can be easily estimate based on standard approximation arguments following [5]:

$$\int_K (\partial_t \eta_\mathrm{h}(U_\mathrm{h}^e) + \partial_x F_\eta(U_\mathrm{h}^e)) = \int_K (\underbrace{\partial_t \eta_\mathrm{h}(U_\mathrm{h}^e) - \partial_t \eta^e}_{\mathcal{O}(\mathrm{h}^{p+1})} + \underbrace{\partial_x (F_\eta(U_\mathrm{h}^e) - F_\eta^e)}_{\mathcal{O}(\mathrm{h}^p)}) = \mathcal{O}(\mathrm{h}^{p+1}) \tag{A.12}$$

and similarly

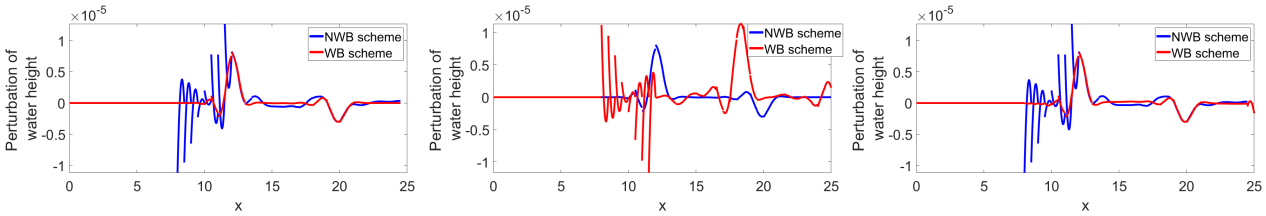
$$\int_K (W_\mathrm{h}^e)^T (\partial_t U_\mathrm{h}^e + \partial_x G_\mathrm{h}(U_\mathrm{h}^e)) = \int_K (W_\mathrm{h}^e)^T \Big(\underbrace{\partial_t U_\mathrm{h}^e - \partial_t U^e}_{\mathcal{O}(\mathrm{h}^{p+1})} + \underbrace{\partial_x (G_\mathrm{h}(U_\mathrm{h}^e) - G^e)}_{\mathcal{O}(\mathrm{h}^p)} \Big) = \mathcal{O}(\mathrm{h}^{p+1}). \tag{A.13}$$

Using the boundedness of $\partial_x v_\mathrm{h}$, of $\partial_x U_\mathrm{h}^e$, and of $\partial_x W_\mathrm{h}^e$, we can estimate

$$\alpha_K(U_\mathrm{h}^e) = \frac{\Psi_\eta^K(U_\mathrm{h}^e) - \Phi_\eta^K(U_\mathrm{h}^e)}{\|\partial_x W_\mathrm{h}^e\|} = \mathcal{O}(\mathrm{h}^{p+1}) \tag{A.14}$$

and

$$\alpha_K(U_\mathrm{h}^e) \int_K \partial_x v_\mathrm{h} A_0(U_\mathrm{h}^e) \partial_x U_\mathrm{h}^e = \mathcal{O}(\mathrm{h}^{p+2}). \tag{A.15}$$



(a) WB discrete steady state (b) NWB discrete steady state (c) Analytical

Figure B.29: Frictionless 1d super-critical equilibrium. Perturbation evolution for the NWB (blue) and WB (red) scheme for $p = 2$. Initial condition : (a) analytical steady state; (b) ODE integrator of Proposition 3; (c) steady state of the NWB scheme

Note that the de-singularization of the value of α_K mentioned in remark 7 guarantees that these scaling are not violated close to regions in which the solution is constant. Using again the fact that in one dimension the number of elements is of $\mathcal{O}(\mathrm{h}^{-1})$, we conclude that

$$|\sum_K \alpha_K(U_\mathrm{h}^e) \int_K \partial_x v_\mathrm{h} A_0(U_\mathrm{h}^e)\partial_x U_\mathrm{h}^e|, = \mathcal{O}(\mathrm{h}^{p+1}) \tag{A.16}$$

which used with the previous estimates leads to the desired result.

Appendix B. Perturbation tests initialization

We consider here an example to show quantitatively why we think the exact equilibrium is to be used in perturbation tests, and in particular to define $h^*(x)$ and $u^*(x)$ in (76). We consider again the perturbation for the test of section §7.2.3 with three different initializations: the discrete solution of the RK-LobattoIIIA collocation method of proposition 3 (corresponding to the discrete steady state of the well-balanced scheme); the discrete steady state of the non-well-balanced standard DGSEM scheme not using global flux quadrature; the analytical steady state. In figure B.29, we show a comparison of the WB and NWB schemes for the three different initial states with a very small perturbation $\xi = 10^{-5}$. As we can see, initializing with the discrete steady state of one of the scheme favours the scheme in question. We can hardly say from the first two pictures in figure B.29 which scheme is better. Only from the perturbation of the analytical initial state in the rightmost picture in the figure we can realize how closely the well-balanced scheme reproduces the exact equilibrium.

Acknowledgments

M. Ricchiuto is a member of the CARDAMOM team, INRIA and University of Bordeaux research center. P.Ö. is thankful for the support received from the Gutenberg Research College, JGU Mainz. Valuable discussions with G. Russo and S. Boscarino on the accuracy of Gauss-Lobatto collocation methods are warmly acnkowledged.

References

[1] R. Abgrall, *A general framework to construct schemes satisfying additional conservation relations. application to entropy conservative and entropy dissipative schemes*, J. Comput. Phys., 372 (2018), pp. 640–666.

[2] R. Abgrall, É. Le Mélédo, P. Öffner, and D. Torlo, *Relaxation deferred correction methods and their applications to residual distribution schemes*, SMAI J. Comput. Math., 8 (2022), pp. 125–160.

[3] R. Abgrall, J. Nordström, P. Öffner, and S. Tokareva, *Analysis of the SBP-SAT stabilization for finite element methods part II: entropy stability*, Commun. Appl. Math. Comput., (2021), pp. 1–23.

[4] R. Abgrall, P. Öffner, and H. Ranocha, *Reinterpretation and extension of entropy correction terms for residual distribution and discontinuous Galerkin schemes: application to structure preserving discretization*, J. Comput. Phys., 453 (2022), p. 24. Id/No 110955.

[5] R. Abgrall and M. Ricchiuto, *High order methods for CFD*, in Encyclopedia of Computational Mechanics, Second Edition, R. d. B. Erwin Stein and T. J. Hughes, eds., John Wiley and Sons, 2017.
[6] R. Abgrall and M. Ricchiuto, *Hyperbolic Balance Laws: Residual Distribution, Local and Global Fluxes*, Springer Nature Singapore, Singapore, 2022, pp. 177–222.
[7] L. Arpaia and M. Ricchiuto, *Well balanced residual distribution for the ALE spherical shallow water equations on moving adaptive meshes*, Journal of Computational Physics, 405 (2020), p. 109173.
[8] L. Arpaia, M. Ricchiuto, A. G. Filippini, and R. Pedreros, *An efficient covariant frame for the spherical shallow water equations: Well balanced DG approximation and application to tsunami and storm surge*, Ocean Modelling, (2021), p. 101915.
[9] E. Audusse, R. Klein, and A. Owinoh, *Conservative discretization of coriolis force in a finite volume framework*, Journal of Computational Physics, 228 (2009), pp. 2934–2950.
[10] J. P. Berberich, P. Chandrashekar, and C. Klingenberg, *High order well-balanced finite volume methods for multi-dimensional systems of hyperbolic balance laws*, Computers & Fluids, 219 (2021), p. 104858.
[11] A. Bermudez and M. Vazquez, *Upwind methods for hyperbolic conservation laws with source terms*, Computers & Fluids, 23 (1994), pp. 1049 – 1071.
[12] C. Berthon and C. Chalons, *A fully well-balanced, positive and entropy-satisfying Godunov-type method for the shallow-water equations*, Math. Comput., 85 (2016), pp. 1281–1307.
[13] C. Birke, W. Boscheri, and C. Klingenberg, *A well-balanced semi-implicit imex finite volume scheme for ideal magnetohydrodynamics at all mach numbers*, 2023.
[14] F. Bouchut, S. J., and V. Zeitlin, *Frontal geostrophic adjustment and nonlinear wave phenomena in one-dimensional rotating shallow water. part 2. high-resolution numerical simulations*, Journal of Fluid Mechanics, 514 (2004), p. 35–63.
[15] M. G. Carlino and E. Gaburro, *Well balanced finite volume schemes for shallow water equations on manifolds*, Applied Mathematics and Computation, 441 (2023), p. 127676.
[16] H. Carrillo, E. Macca, C. Pares, and G. Russo, *Well-balanced adaptive compact approximate taylor methods for systems of balance laws*, 2022.
[17] V. Caselles, R. Donat, and G. Haro, *Flux-gradient and source-term balancing for certain high resolution shock-capturing schemes*, Computers & Fluids, 38 (2009), pp. 16–36.
[18] M. Castro, J. M. Gallardo, J. A. López-GarcÍa, and C. Parés, *Well-balanced high order extensions of godunov's method for semilinear balance laws*, SIAM Journal on Numerical Analysis, 46 (2008), pp. 1012–1039.
[19] M. Castro, A. P. Milanés, and C. Parés, *Well-balanced numerical schemes based on a generalized hydrostatic reconstruction technique*, Mathematical Models and Methods in Applied Sciences, 17 (2007), pp. 2055–2113.
[20] M. J. Castro and C. Parés, *Well-balanced high-order finite volume methods for systems of balance laws*, Journal of Scientific Computing, 82 (2020), p. 48.
[21] M. J. Castro Díaz, J.-A. López, and C. Parés, *Finite volume simulation of the geostrophic adjustment in a rotating shallow-water system*, SIAM J. Sci. Comput., 31 (2008), p. 444–477.
[22] T. Chen and C.-W. Shu, *Review of entropy stable discontinuous Galerkin methods for systems of conservation laws on unstructured simplex meshes*, CSIAM Trans. Appl. Math., 1 (2020), pp. 1–52.
[23] Y. Cheng, A. Chertock, M. Herty, A. Kurganov, and T. Wu, *A new approach for designing moving-water equilibria preserving schemes for the shallow water equations*, J.Sci.Comp., 80 (2019), pp. 538–554.
[24] Y. Cheng and A. Kurganov, *Moving-water equilibria preserving central-upwind schemes for the shallow water equations*, Comm.Math. Sciences, 14 (2016), pp. 1643–1663.
[25] A. Chertock, S. Cui, A. Kurganov, Ş. N. Özcan, and E. Tadmor, *Well-balanced schemes for the euler equations with gravitation: Conservative formulation using global fluxes*, Journal of Computational Physics, 358 (2018), pp. 36–52.
[26] A. Chertock, M. Dudzinski, A. Kurganov, and M. Lukácová-Medvidóvá, *Well-balanced schemes for the shallow water equations with Coriolis forces*, Numer. Math., 138 (2018), pp. 939–973.
[27] A. Chertock, A. Kurganov, X. Liu, Y. Liu, and T. Wu, *Well-balancing via flux globalization: Applications to shallow water equations with wet/dry fronts*, Journal of Scientific Computing, 90 (2022), pp. 1–21.
[28] M. Ciallella, L. Micalizzi, P. Öffner, and D. Torlo, *An arbitrary high order and positivity preserving method for the shallow water equations*, Computers & Fluids, 247 (2022), p. 105630.
[29] M. Ciallella, D. Torlo, and M. Ricchiuto, *Arbitrary high order weno finite volume scheme with flux globalization for moving equilibria preservation*, Journal of Scientific Computing, 96 (2023), p. 53.
[30] P. Ciarlet and P. Raviart, *General lagrange and hermite interpolation in $\mathbb{R}^n$ with applications to finite element methods*, Arch.Ration.Mech.Anal., 46 (1972), pp. 177–199.
[31] C. M. Dafermos, *Hyperbolic conservation laws in continuum physics*, vol. 325 of Grundlehren Math. Wiss., Berlin: Springer, 3rd ed. ed., 2010.
[32] V. Desveaux and A. Masset, *A fully well-balanced scheme for shallow water equations with coriolis force*, 2021.
[33] ———, *A fully well-balanced scheme for shallow water equations with Coriolis force*, Commun. Math. Sci., 20 (2022), pp. 1875–1900.
[34] R. Donat and A. Martinez-Gavara, *Hybrid second order schemes for scalar balance laws*, Journal of Scientific Computing, 48 (2011), pp. 52–69.
[35] M. Dumbser, O. Zanotti, E. Gaburro, and I. Peshkov, *A well-balanced discontinuous galerkin method for the first–order z4 formulation of the einstein–euler system*, 2023.
[36] A. Fedorov and W. Melville, *Kelvin fronts on the equatorial thermocline*, Journal of Physical Oceanography, 30 (2000), pp. 1692–1705.
[37] T. C. Fisher, M. H. Carpenter, J. Nordström, N. K. Yamaleev, and C. Swanson, *Discretely conservative finite-*

difference formulations for nonlinear conservation laws in split form: Theory and boundary conditions, Journal of Computational Physics, 234 (2013), pp. 353–375.
[38] U. S. Fjordholm, S. Mishra, and E. Tadmor, *Well-balanced and energy stable schemes for the shallow water equations with discontinuous topography*, J. Comput. Phys., 230 (2011), pp. 5587–5609.
[39] U. S. Fjordholm, S. Mishra, and E. Tadmor, *Arbitrarily high-order accurate entropy stable essentially nonoscillatory schemes for systems of conservation laws*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 544–573.
[40] E. Gaburro, M. J. Castro, and M. Dumbser, *Well-balanced Arbitrary-Lagrangian-Eulerian finite volume schemes on moving nonconforming meshes for the Euler equations of gas dynamics with gravity*, Monthly Notices of the Royal Astronomical Society, 477 (2018), pp. 2251–2275.
[41] E. Gaburro, P. Öffner, M. Ricchiuto, and D. Torlo, *High order entropy preserving ADER-DG schemes*, Applied Mathematics and Computation, 440 (2023), p. 127644.
[42] L. Gascón and J. Corberán, *Construction of second-order tvd schemes for nonhomogeneous hyperbolic conservation laws*, Journal of Computational Physics, 172 (2001), pp. 261–297.
[43] G. J. Gassner, *A skew-symmetric discontinuous Galerkin spectral element discretization and its relation to sbp-sat finite difference methods*, SIAM Journal on Scientific Computing, 35 (2013), pp. A1233–A1253.
[44] G. J. Gassner, A. R. Winters, and D. A. Kopriva, *Split form nodal discontinuous galerkin schemes with summation-by-parts property for the compressible euler equations*, Journal of Computational Physics, 327 (2016), pp. 39–66.
[45] I. Gómez-Bueno, S. Boscarino, M. Castro, C. Parés, and G. Russo, *Implicit and semi-implicit well-balanced finite-volume methods for systems of balance laws*, Applied Numerical Mathematics, 184 (2023), pp. 18–48.
[46] I. Gómez-Bueno, M. J. Castro, and C. Parés, *High-order well-balanced methods for systems of balance laws: a control-based approach*, Applied Mathematics and Computation, 394 (2021), p. 125820.
[47] I. Gómez-Bueno, M. J. C. Díaz, C. Parés, and G. Russo, *Collocation methods for high-order well-balanced methods for systems of balance laws*, Mathematics, 9 (2021).
[48] L. Gosse, *A well-balanced flux-vector splitting scheme designed for hyperbolic systems of conservation laws with source terms*, Computers & Mathematics with Applications, 39 (2000), pp. 135–159.
[49] J.-L. Guermond, M. Nazarov, B. Popov, and I. Tomas, *Second-order invariant domain preserving approximation of the Euler equations using convex limiting*, SIAM Journal on Scientific Computing, 40 (2018), pp. A3211–A3239.
[50] E. Guerrero Fernández, C. Escalante, and M. J. Castro Díaz, *Well-balanced high-order discontinuous galerkin methods for systems of balance laws*, Mathematics, 10 (2022).
[51] E. Hairer, G. Wanner, and S. Norset, *Solving Ordinary Differential Equations I. Nonstiff problems.*, Springer, Berlin, Heidelberg, 1993.
[52] A. Harten, *On the symmetric form of systems of conservation laws with entropy*, J. Comput. Phys., 49 (1983), pp. 151–164.
[53] G. Hernández-Dueñas and S. Karni, *Shallow water flows in channels*, Journal of Scientific Computing, 48 (2011), pp. 190–208.
[54] J. Hesthaven and T. Warburton, *Nodal Discontinuous Gelerkin Methods. Algorithms, Analysis, and Applications*, Springer, New York, NY, 2008.
[55] D. A. Kopriva and G. Gassner, *On the quadrature and weak form choices in collocation type discontinuous galerkin spectral element methods*, Journal of Scientific Computing, 44 (2010), pp. 136–155.
[56] P. G. LeFloch, J.-M. Mercier, and C. Rohde, *Fully discrete, entropy conservative schemes of arbitrary order*, SIAM Journal on Numerical Analysis, 40 (2002), pp. 1968–1992.
[57] Y. Mantri and S. Noelle, *Well-balanced discontinuous Galerkin scheme for 2 × 2 hyperbolic balance law*, J. Comput. Phys., 429 (2021), pp. 110011, 13.
[58] V. Michel-Dansac, C. Berthon, S. Clain, and F. Foucher, *A well-balanced scheme for the shallow-water equations with topography or manning friction*, Journal of Computational Physics, 335 (2017), pp. 115–154.
[59] A. Navas-Montilla and J. Murillo, *2d well-balanced augmented ader schemes for the shallow water equations with bed elevation and extension to the rotating frame*, Journal of Computational Physics, 372 (2018), pp. 316–348.
[60] S. Noelle, Y. Xing, and C.-W. Shu, *High order well-balanced finite volume weno schemes for shallow water equation with moving water*, J.Comput.Phys, 226 (2007), pp. 29–58.
[61] P. Öffner, *Approximation and Stability Properties of Numerical Methods for Hyperbolic Conservation Laws*, Habilitation, University Zurich, 2020.
[62] C. Parés and C. Parés-Pulido, *Well-balanced high-order finite difference methods for systems of balance laws*, Journal of Computational Physics, 425 (2021), p. 109880.
[63] C. Parés and M. Castro, *On the well-balance property of roe's method for nonconservative hyperbolic systems. applications to shallow-water systems*, ESAIM: Mathematical Modelling and Numerical Analysis, 38 (2004), p. 821–852.
[64] A. Prothero and A. Robinson, *On the stability and accuracy of one-step methods for solving stiff systems of ordinary differential equations*, Math.Comp., 28 (1974), pp. 145–162.
[65] H. Ranocha, *Shallow water equations: split-form, entropy stable, well-balanced, and positivity preserving numerical methods*, GEM. Int. J. Geomath., 8 (2017), pp. 85–133.
[66] H. Ranocha, P. Öffner, and T. Sonar, *Summation-by-parts operators for correction procedure via reconstruction*, J. Comput. Phys., 311 (2016), pp. 299–328.
[67] H. Ranocha, M. Sayyari, L. Dalcin, M. Parsani, and D. I. Ketcheson, *Relaxation Runge-Kutta methods: fully discrete explicit entropy-stable schemes for the compressible Euler and Navier-Stokes equations*, SIAM J. Sci. Comput., 42 (2020), pp. a612–a638.
[68] F. Renac, *Entropy stable DGSEM for nonlinear hyperbolic systems in nonconservative form with application to two-phase*

flows, J. Comput. Phys., 382 (2019), pp. 1–26.
[69] M. Ricchiuto, *On the C-property and Generalized C-property of Residual Distribution for the Shallow Water Equations*, Journal of Scientific Computing, 48 (2011), pp. 304–318.
[70] ———, *An explicit residual based approach for shallow water flows*, J.Comput.Phys., 80 (2015), pp. 306–344.
[71] M. Ricchiuto and D. Torlo, *Analytical travelling vortex solutions of hyperbolic equations for validating very high order schemes*, 2021.
[72] P. Roe, *Upwind differencing schemes for hyperbolic conservation laws with source terms*, in Nonlinear Hyperbolic Problems, C. Carasso, D. Serre, and P.-A. Raviart, eds., Berlin, Heidelberg, 1987, Springer Berlin Heidelberg, pp. 41–51.
[73] D. Sármány, M. Hubbard, and M. Ricchiuto, *Unconditionally stable space–time discontinuous residual distribution for shallow-water flows*, Journal of Computational Physics, 253 (2013), pp. 86–113.
[74] E. Tadmor, *Entropy stability theory for difference approximations of nonlinear conservation laws and related time-dependent problems*, Acta Numerica, 12 (2003), pp. 451–512.
[75] A. Thomann, G. Puppo, and C. Klingenberg, *An all speed second order well-balanced IMEX relaxation scheme for the Euler equations with gravity*, J. Comput. Phys., 420 (2020), p. 25. Id/No 109723.
[76] M. Vázquez-Cendón, *Improved treatment of source terms in upwind schemes for the shallow water equations in channels with irregular geometry*, Journal of Computational Physics, 148 (1999), pp. 497–526.
[77] M. Waruszewski, J. E. Kozdon, L. C. Wilcox, T. H. Gibson, and F. X. Giraldo, *Entropy stable discontinuous Galerkin methods for balance laws in non-conservative form: applications to the Euler equations with gravity*, J. Comput. Phys., 468 (2022), p. 25. Id/No 111507.
[78] A. R. Winters and G. J. Gassner, *A comparison of two entropy stable discontinuous Galerkin spectral element approximations for the shallow water equations with non-constant topography*, J. Comput. Phys., 301 (2015), pp. 357–376.
[79] Y. Xing, *Exactly well-balanced discontinuous galerkin methods for the shallow water equations with moving water equilibrium*, Journal of Computational Physics, 257 (2014), pp. 536–553.
[80] Y. Xing and C.-W. Shu, *High order well-balanced finite volume WENO schemes and discontinuous Galerkin methods for a class of hyperbolic systems with source terms*, J. Comput. Phys., 214 (2006), pp. 567–598.