A well balanced and entropy conservative discontinuous Galerkin spectral element method for the shallow water equations

Gregor J Gassner, Andrew R Winters and David A Kopriva

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-156863

N.B.: When citing this work, cite the original publication.

Original publication available at:
https://doi.org/10.1016/j.amc.2015.07.014

Copyright: Elsevier
http://www.elsevier.com/
A WELL BALANCED AND ENTROPY CONSERVATIVE DISCONTINUOUS GALERKIN SPECTRAL ELEMENT METHOD FOR THE SHALLOW WATER EQUATIONS

GREGOR J. GASSNER\textsuperscript{1,*}, ANDREW R. WINTERS\textsuperscript{1}, AND DAVID A. KOPRIVA\textsuperscript{2}

\textbf{Abstract.} In this work, we design an arbitrary high order accurate nodal discontinuous Galerkin spectral element type method for the one dimensional shallow water equations. The novel method uses a skew-symmetric formulation of the continuous problem. We prove that this discretisation exactly preserves the local mass and momentum. Furthermore, we show that combined with a special numerical interface flux function, the method exactly preserves the entropy, which is also the total energy for the shallow water equations. Finally, we prove that the surface fluxes, the skew-symmetric volume integrals, and the source term are well balanced. Numerical tests are performed to demonstrate the theoretical findings.

\textbf{Keywords:} skew-symmetric shallow water equations, discontinuous Galerkin spectral element method, Gauss-Lobatto Legendre, summation-by-parts, entropy conservation, well balanced

\section{Introduction}

Discontinuous Galerkin formulations for hyperbolic conservation laws typically approximate the conservative form of the equations. This automatically guarantees that the resulting method preserves the conserved quantities discretely. According to the Lax-Wendroff theorem it follows that such a discretisation approximates the right shock speeds, which is of fundamental importance when simulating such phenomena.

When deriving a discontinuous Galerkin method, one needs to choose several ingredients often with many choices available: Two important choices are the specific polynomial basis functions and the numerical quadrature rule used to approximate the inner products of the Galerkin formulation. This discretisation step has a dramatic impact on the efficiency, accuracy and stability of the method. This is especially true when approximating nonlinear hyperbolic problems, where the flux function depends nonlinearly on the conserved quantities. When using a polynomial of degree $N$ in the DG ansatz, it is clear that the flux function no longer belong to the same polynomial space. It could even be that the flux function isn’t a polynomial at all, for example if the flux terms are rational with respect to the conserved quantities. Prominent examples where this occurs are the Euler equations of gas dynamics and the shallow water equations. In both cases, the
fluxes are rational and hence when using a polynomial ansatz for the conserved variables, are non polynomial. Non polynomial flux functions have a direct consequence on the stability of the discontinuous Galerkin discretisation. If the inner products are not evaluated exactly, aliasing errors are introduced as no precise $L^2$-projection is performed anymore. In the general approximation, integrals are not evaluated analytically but by choosing a numerical quadrature rule based on Gauss-Legendre-type methods. However, such integration rules are all derived from polynomial interpolation and their exactness is only guaranteed when the integrand belongs to the polynomial space. Consequently, it is not possible to avoid aliasing issues in a standard discontinuous Galerkin approximation of hyperbolic problems with non polynomial flux functions. Although it is possible to increase the number of quadrature points so that the integration errors shrink to machine precision, the number of points needed heavily depends on the nonlinearity of the flux function and may vary from problem to problem (e.g. solutions with shocks or smooth solutions only).

In the above sense, there is no rigorous way to prove stability for standard DG discretisations based on the conservative form of hyperbolic problems with non polynomial flux functions. A similar problem exists for other discretisation types such as spectral methods or finite difference methods. In these communities, an interesting strategy to tackle aliasing in nonlinear problems is to resort to a skew-symmetric formulation of the underlying hyperbolic problem, as for instance proposed by Morinishi [20] for the compressible Euler equations. With such formulations, it is possible to derive finite difference methods that are conservative and preserve kinetic energy, e.g. [26, 27] and e.g. [28, 1]. Another strategy is to derive formulations which are entropy conservative or entropy stable respectively, see e.g. Tadmor [31, 33] and e.g. [18, 9, 19, 15]. Tadmor showed however, that entropy conservation (stability) is always related to a skew-symmetric formulation of the problem [32]. Thus, these two approaches are closely linked. In fact, we will show in this work that the skew-symmetric formulation of the shallow water equations yields an entropy conservative discretisation.

The skew-symmetric form of the problem is typically achieved by averaging the conservative form of the equation and the non conservative advection form. This is problematic in the sense that discretisations in this form are not obviously conservative for the conserved quantities. However, as mentioned previously, conservation is an important property for the discretisation to have correct shock speeds. Recently, it was shown that when using diagonal norm summation-by-parts operators to discretise the spatial derivatives, skew-symmetric formulations exactly preserve the conservation quantities [8, 10, 11]. In particular, the derivations in Fisher et al. [8] show that skew-symmetric operators based on summation-by-parts derivative matrices are consistent and conservative in the Lax-Wendroff sense. It was noted e.g. in [10] that a discontinuous Galerkin spectral element (DGSEM) operator with Gauss-Lobatto points satisfies all properties of a diagonal summation by parts operator and allows for skew-symmetric discretisations that are exactly conservative. This result was used in [11] to derive a kinetic energy conserving DGSEM formulation for the compressible Euler equations using the skew-symmetric form introduced by Morinishi [20]. Unfortunately, kinetic energy for the Euler equations is not an entropy function so kinetic energy conservation is not sufficient to obtain nonlinear stability. Nonlinear entropy stability was recently achieved by Carpenter et al. [2] by constructing a special DGSEM approach, not based on a skew-symmetric form, but using an equivalence of every diagonal norm summation-by-parts operator to a staggered grid type finite volume method.

There are many variants of the DG method applied to the shallow water equations available in the literature, e.g. [29, 35, 7, 34, 12, 13, 36, 37] and more general strategies in the context of
hyperbolic systems with nonconservative products, e.g. [25, 6]. In this work, we combine a DG type discretisation with the idea of skew-symmetry to derive an entropy conserving and well balanced approximation for the shallow water equations.

We start with a special skew-symmetric formulation in Sec. 2 for the one dimensional shallow water equations, which we use for our DGSEM formulation in Sec. 3. We prove in Sec. 4 exact conservation in the Lax-Wendroff sense and show that by choosing an appropriate numerical flux function, the method exactly preserves the total energy, which is an entropy function for the shallow water equations. Furthermore, we will show that this novel skew-symmetric DGSEM formulation allows formal proof of the well balanced property. Section 5 demonstrates and underlines our theoretical findings with numerical results. Our conclusions are presented in the last section.

2. A Skew-Symmetric Form of the Shallow Water Equations

The standard form of the shallow water equations in one dimension for smooth solutions is

\[(C) : \quad h_t + (hv)_x = 0 \]

\[(M1) : \quad (hv)_t + (hv^2 + gh^2/2)_x = -ghb_x \]

where \((C)\) and \((M1)\) indicate the continuity equation and the momentum balance respectively. The quantity \(h = h(x,t)\) denotes the water height measured from the bottom topography \(b = b(x)\) with the total height given by \(H = h + b\). The constant \(g = \text{const}\) denotes the gravitational acceleration and \(v = v(x,t)\) is the velocity.

Instead of writing \((gh^2/2)_x\) in flux form, we use the chain rule \(ghh_x\) to get an alternative second form of the momentum equation \((M2)\)

\[(M2) : \quad (hv)_t + (hv^2)_x + gh(h + b)_x = 0. \]

We will see in the derivations below that this form of the gravity acceleration term is important to achieve a well balanced approximation.

Starting from \((C)\) and \((M2)\), we derive the advection form \((A)\) of the momentum equation \((M2)\) by using the chain rule for the first two terms and subtracting the continuity equation \((C)\) multiplied by the velocity \(v\)

\[(A) : \quad hv + (hv)v_x + gh(h + b)_x = 0. \]

We get the skew-symmetric form \((S)\) of the momentum equation for the advection terms containing the velocity \(v\) by averaging the standard form \((M2)\) and the advection form \((A)\)

\[(S) : \quad \frac{1}{2}[(hv)_t + hv] + \frac{1}{2}[(hv^2) + (hv)v_x]/2 + gh(h + b)_x = 0. \]

We note that \((S)\) is indeed skew-symmetric for the advection terms

\[
\frac{1}{2} [\partial_t(h \cdot v) + h \partial_t v] + \frac{1}{2} [\partial_x ((hv) \cdot v) + (hv) \partial_x v] + gh(h + b)_x = 0.
\]

The skew-symmetric form is well suited to derive the kinetic energy balance of the system by simply multiplying \((S)\) with the velocity \(v\) and rearranging terms to get

\[(K) : \quad k_t + (hv^3/2)_x + g(hv)(h + b)_x = 0, \]
where $k = \frac{hv^2}{2}$ is the kinetic energy. We get the potential energy balance by multiplying (C) with $g (h + b)$

$$ (P) : \quad p_t + g (h + b) (hv)_x = 0, $$

where $p = gh^2/2 + ghb$ is the potential energy. The total energy balance results when we add the kinetic and potential energy balance

$$ e_t + (h v^3/2)_x + g (hv) (h + b)_x + g (h + b) (hv)_x = 0, $$

where $e = k + p = h v^2/2 + gh^2/2 + gbb$. This can be further simplified to

$$ (E) : \quad e_t + (h v^3/2 + g (hv) (h + b))_x = 0, $$

which shows that the total energy $e$ is a conserved quantity. In fact, it can be directly seen that the total energy is an entropy function of the one dimensional shallow water equations with the entropy flux $f = h v^3/2 + g (hv) (h + b)$. The associated entropy variables are given by $q_1 := \frac{\partial}{\partial h} e = g (h + b) - v^2/2$ and $q_2 := \frac{\partial}{\partial v} e = v$. Thus, by multiplying the continuity equation (C) with the first entropy variable $q_1$ and adding the momentum equation (M2) multiplied by $q_2$, we get the total energy conservation (E). We note that instead of adding the scaled momentum equation (M2), we can equivalently add the skew-symmetric form (S) multiplied by $v$ to the scaled continuity equation.

Summarising this section, we use the standard form of the continuity equation (C) and the skew-symmetric form (S) of the momentum as a base for our DGSEM discretisation

$$ h_t + (hv)_x = 0 $$

and show in the following that we can reproduce discrete and consistent versions of (M2), (K), (P) and (E).

### 3. Discontinuous Galerkin Spectral Element Method

The first step of a spectral element discretisation is to subdivide the computational domain into non-overlapping elements $G$ and map each of them to a reference space

$$ x(\xi) = \bar{x} + \frac{\Delta x}{2} \xi, \quad \xi \in [-1, 1], $$

where $\bar{x}$ is the element center and $\Delta x$ is the element length. The mapping is used to transform the spatial derivative with respect to the physical coordinate $x$ to the derivative operator with respect to the reference coordinate $\xi$

$$ \frac{\partial}{\partial x} = \frac{\partial \xi}{\partial x} \frac{\partial}{\partial \xi} = \frac{2}{\Delta x} \frac{\partial}{\partial \xi}. $$

Thus, the transformed skew-symmetric shallow water equations are

$$ J h_t + (hv)_\xi = 0, $$

$$ J [(hv)_t + hh]_\xi + [(hv^2)_\xi + (hv) v]_\xi/2 + g h (h + b)\xi = 0, $$

where $J := \frac{\Delta x}{2}$ denotes the Jacobian of the element mapping.
For each element, the conservative variables are approximated by a polynomial of degree $N$ in the reference coordinate $\xi$. Spectral elements are defined by a nodal approximation at Legendre-Gauss-Lobatto (GL) points $\{\xi_j\}_{j=0}^N$ and the associated Lagrange basis functions

$$\ell_j(\xi) = \prod_{i=0, i \neq j}^N \frac{\xi - \xi_i}{\xi_j - \xi_i}, \quad j = 0, ..., N,$$

which satisfy the cardinal property

$$\ell_j(\xi_i) = \delta_{ij},$$

where $\delta_{ij}$ denotes Kronecker’s symbol with $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ for $i \neq j$.

Thus, we get an element-wise approximation for each conservative variable in an element $G$

$$u_k(x, t) |_{G} \approx u_k(\xi, t) = \sum_{j=0}^N u_{jk}(t) \ell_j(\xi), \quad k = 1, 2,$$

where $\{u_{jk}(t)\}_{j=0}^N$ are the time dependent nodal degrees of freedom of the conservative variables $u_1 = h$ and $u_2 = hv$. We follow the collocation principle and compute derived quantities at the corresponding GL point from the values of the conservative variables, e.g. the velocity field $v := (v_0, ..., v_N)^T$ has nodal components

$$v_j = \frac{u_2^j}{u_1^j}, \quad j = 0, ..., N.$$

We further introduce the square or powers of such a nodal vector in a pointwise sense, e.g.

$$v^2 := (v_0^2, ..., v_N^2)^T.$$

As a direct consequence, we approximate each quantity by a polynomial of the same degree $N$, even if it depends nonlinearly on the conserved quantities $u_1$ and $u_2$. This construction principle drastically increases efficiency and simplifies the method, however it is clear that the collocation of all quantities introduces aliasing errors which may be severe and cause aliasing driven instabilities. In contrast to standard DGSEM formulations based on the conservative formulation, a DGSEM formulation based on the skew-symmetric formulation of the shallow water equations controls these errors by effectively cancelling them out such that nonlinear (entropy) stability can be achieved.

Similar to the argument presented in [11], we use a strong form DGSEM approximation of the skew-symmetric shallow water system (11), as this allows a one to one translation of the continuous form into a discrete form by replacing the continuous derivative $\partial_\xi$ by its discrete counterpart. However, we note, that for DGSEM-GL, the strong form and weak form are algebraically equivalent [16] and this decision is only to facilitate the derivations below.

For the strong form, we need the discrete derivatives at the nodes, which we get by computing

$$u'(\xi_i) = \sum_{j=0}^N u^j \frac{\partial \ell_j}{\partial \xi} \bigg|_{\xi = \xi_i}, \quad i = 0, ..., N,$$

where $'$ denotes the discrete derivative of the (derived) quantity $u$. We write this in compact matrix vector notation as

$$u' = D u,$$
where the derivative matrix is given by

\[ D_{ij} = \left. \frac{\partial \ell_j}{\partial \xi} \right|_{\xi = \xi_i}, \quad i, j = 0, \ldots, N. \tag{23} \]

By using the GL points to construct the Lagrange basis functions \( \ell \) and the derivative matrix \( D \) respectively, we obtain the following relationship for all polynomial degrees \( N \), see e.g. [10, 2, 11]

\[ Q := M D, \tag{24} \]

with

\[ Q + Q^T = B, \quad \text{or} \quad (MD) + (MD)^T = B, \tag{25} \]

where \( B \) and \( M \) are diagonal matrices

\[ M = \text{diag} [\omega_0, \ldots, \omega_N], \quad B = \text{diag} [-1, 0, \ldots, 0, 1], \tag{26} \]

with \( \{\omega_j\}_{j=0}^N \) denoting the GL quadrature weights. Typically, the matrix \( M \) is called the mass matrix (or norm matrix) and \( B \) the boundary matrix. We will denote all operators with capital letters and physical quantities using lower case letters. The relationship of the derivative operator (25) is called the summation-by-parts property and is common in the finite difference community, e.g. [17, 5, 4, 22, 30, 24, 23].

Although we use the derivative matrix, we note that the operators \( M, Q, D \) and \( B \) are based on a weak Galerkin formulation where the test and basis functions are chosen as the Lagrange basis functions. The inner products of the Galerkin formulation are approximated by the same quadrature rules as used for the definition of the nodal points, i.e. the GL quadrature with the points \( \{\xi_j\}_{j=0}^N \) and weights \( \{\omega_j\}_{j=0}^N \). For instance, assuming two interpolants of functions \( u(\xi) \) and \( w(\xi) \)

\[ (I_N u)(\xi) = \sum_{j=0}^N w^j \ell_j(\xi) \quad \text{and} \quad (I_N w)(\xi) = \sum_{j=0}^N w^j \ell_j(\xi), \tag{27} \]

with nodal coefficient vectors \( u \) and \( w \) and multiplying the summation-by-parts relation (25) with \( w^T \) from the left and \( u \) from the right, we get

\[ w^T (MD) u + w^T (MD)^T u = w^T Bu, \quad \text{or} \quad w^T M (D u) + (D w)^T M u = w^T Bu, \tag{28} \]

which directly translates to

\[ \langle (I_N u)', w \rangle_N + \langle u, (I_N w)' \rangle_N = u w|_{-1}^{+1} = u(1) w(1) - u(-1) w(-1), \tag{29} \]

with discrete inner products \( \langle \cdot, \cdot \rangle_N \), where integrals are evaluated with the same GL quadrature points used to define the interpolants.

Thus, the following discretisation in matrix vector form has an equivalent interpretation as a inner product Galerkin formulation, see [10] for details.
We first look to discretise the continuity equation (C). We insert our polynomial ansatz and replace the spatial derivative by our discrete derivative to get
\begin{equation}
J \partial_t u_1 + D f_1 = S \left[ f_1^* - f_1 \right],
\end{equation}
where the mass flux is computed as
\begin{equation}
f_1 = h v,
\end{equation}
with \( h = \text{diag}(h) \). We have further introduced element interface penalty terms that account for the discontinuous nature of our piecewise polynomial ansatz. The surface term matrix is
\begin{equation}
S := -M^{-1} B = \text{diag} \left( \begin{array}{cccc}
\frac{1}{\omega_0}, & 0, & \ldots, & 0, \frac{-1}{\omega_N}
\end{array} \right).
\end{equation}
We note for the discretisation of the continuity equation, the standard DGSEM-GL results, which is the same as the flux reconstruction method with correction function \( g_2 \) called “lumping with Gauss-Lobatto” \([14]\). In fact, equation (30) again directly translates into inner product form. When multiplied from the left by the mass matrix \( M \) and using the discrete inner products with GL points, we get
\begin{equation}
J \langle \partial_t u_1, \ell_i \rangle_N + \langle (I_N f_1)', \ell_i \rangle_N = -\left[ f_1^* - f_1 \right] \ell_i \bigg|^{+1}_{-1}, \quad i = 0, \ldots, N.
\end{equation}
The remaining terms \( f_1^* \) or \( f_1^* \) respectively contain the numerical flux functions at entries \( j = 0 \) and \( j = N \), which resolve the discontinuous interface data and thus typically depend on the values from the left and from the right side of the element interface. We note that the choice of the numerical flux function has an impact on the accuracy, the stability and well balancedness of the resulting method. We derive a particular numerical flux \( f_1^{*, ec} \) in Sec. 4.2, for which exact entropy conservation and well balancedness is guaranteed.

The discretisation of the skew-symmetric momentum equation (S) is more delicate, as it is not in obvious conservative form and has a nontrivial temporal derivative term. We follow the same approach as presented in \([11]\). Each term in (15) is translated into its discrete version and a penalty-like interface term is introduced, with up to now unknown interface functions \( g_2 \) and \( g_2^* \)
\begin{equation}
J \left( \partial_t h v + h \partial_t v + \frac{1}{2} D h v^2 + h v D v \right) + g h D (h + b) = S \left[ g_2^* - g_2 \right].
\end{equation}
We choose the ansatz for the right hand side as this mimics the typical structure of a DG surface term, e.g. as can be seen in (30). Equivalently an inner product notation for (34) reads as
\begin{equation}
\frac{J}{2} \langle \partial_t (h v) + h \partial_t v, \ell_i \rangle_N + \frac{1}{2} \langle (I_N h v^2)' + h (I_N v)' + h v D v \rangle_N + g h D (h + b) \langle \ell_i \rangle_N = -\left[ g_2^* - g_2 \right] \ell_i \bigg|^{+1}_{-1}, \quad i = 0, \ldots, N,
\end{equation}
where we use that \( h \) and \( h v \) are polynomials of degree \( N \) by our ansatz.

We derive the expression of \( g_2 \) and \( g_2^* \) in Sec. 4.1 such that exact conservation of the momentum \( h v \) is guaranteed. The term \( g_2^* \) is strongly connected to the numerical flux function \( f_2^* \) and, thus again, its choice has an impact on the properties of the resulting methods. We also derive an expression in Sec. 4.2 for \( f_2^* \) so that well balancedness and entropy conservation is achieved.
4. Properties of the Skew-Symmetric DGSEM

In the following proofs and derivations we always assume that time is continuous, so that the time derivative operator follows the typical rules such as the product rule. As a consequence, the numerical results presented below are influenced by the temporal discretisation. However, by choosing the explicit time step small enough, the influence is negligible, which is also demonstrated.

4.1. Mass and Momentum Conservation. It is a priori not clear that the method described in Sec. 3 conserves momentum as it is not based on the conservative form of the problem. However, according to the Lax Wendroff theorem, this is a fundamental requirement to correctly predict shock speeds.

In Fisher et al. [8] the authors proved the following very important properties of a diagonal norm SBP operator, such as the DGSEM-GL operator. First, they showed that $\frac{D}{\text{Bang}}f$ can be recast into a consistent subcell flux-difference formulation and for such a discretisation the Law-Wendroff theorem holds. As a consequence, the DGSEM-GL discretisation (30) of the continuity equation is conservative in the Lax-Wendroff sense. Furthermore, they proved that

$$\alpha D\beta + \beta D\alpha,$$

with $\alpha$ and $\beta$ being the point wise nodal representations of functions $\alpha$ and $\beta$, is a consistent and conservative approximation of the derivative of the corresponding flux $(\alpha \beta)_x$. If we subtract the discretisation of $(\alpha \beta)_x$ from (36), we get the error in the associated product rule

$$\varepsilon_{\alpha\beta} := -D\alpha \beta + \alpha D\beta + \beta D\alpha.$$

In [11], the author showed that the discrete mean value of terms with such a structure is exactly zero, i.e.

$$-1^T M D\alpha \beta + 1^T M \alpha D\beta + 1^T M \beta D\alpha = 0,$$

for all vectors $\alpha, \beta \in \mathbb{R}^{N+1}$. The multiplication by $1^T M$ with $1 = (1, ..., 1)^T$ is equivalent to integrating the equation over the reference element and using the associated GL quadrature rule to approximate the integral, e.g.

$$1^T M \alpha = \sum_{j=0}^{N} \alpha_j \omega_j = (I_N \alpha, 1)_N \approx \int_{-1}^{1} \alpha(\xi) d\xi.$$

Similar to the skew-symmetric DGSEM for the Euler equations presented in [11], we show in the following that the volume terms of the skew-symmetric momentum discretisation can be recast as a combination of the standard flux form approximation and a product error term of the structure (37). This directly proves exact conservation for the skew-symmetric DGSEM for the shallow water equations, due to the previously mentioned finding for diagonal norm summation-by-parts operators.
We start with the skew-symmetric discretisation of the momentum equation (34) and add the
discretised continuity equation (30) multiplied point wise by the velocity half \( \frac{1}{2} \).

(40)
\[
\frac{1}{2} \left[ \partial_t (b v) + h \partial_x v + \frac{1}{\nu} \partial_t u_1 \right] + \frac{1}{2} \left[ D_h v^2 + h v \partial_x \nu + \nu \partial_x L_1 \right] + g h D (h + b) = S \left[ \left( g^*_2 + \frac{1}{2} v L_1^* \right) - \left( g_2 + \frac{1}{2} v L_1 \right) \right].
\]

We note that the matrices \( \frac{1}{\nu} \) and \( S \) commute, as they are both diagonal. Assuming time continuity,
the first term reduces to

(41)
\[
\frac{1}{2} \left[ \partial_t (b v) + h \partial_x v + \frac{1}{\nu} \partial_t u_1 \right] = J \partial_t u_2.
\]

The second term can be rewritten as

(42)
\[
\frac{1}{2} \left[ D_h v^2 + h v \partial_x \nu + \nu \partial_x L_1 \right] = D_h v^2 + \frac{1}{2} \left[ -D_h v^2 + h v \partial_x \nu + \nu \partial_x h \right] =: D_h v^2 + \varepsilon_{h,v^2}.
\]

Note that the additional nonlinear correction term \( \varepsilon_{h,v^2} \) is exactly of the structure (37), which is
thus fully consistent to the corresponding flux term as discussed above. This correction term is the
error of the discretisation of the product rule \(-(h v^2)_x + (hv)v_x + v(hv)_x\) scaled by \(1/2\). The third
term is similarly reformulated as

(43)
\[
g h D h = \frac{g}{2} D h^2 + \frac{g}{2} \left[ -D h^2 + h D h + \nu \partial_x h \right] =: \frac{g}{2} D h^2 + \varepsilon_{h,h^2},
\]

where again the additional correction term \( \varepsilon_{h,h^2} \) is fully consistent and represents the error of the
discrete product rule \(-(h^2)_x + 2 h h_x\) scaled by \(g/2\). Summarising, we get the following intermediate
formulation

(44)
\[
J u_2 + D_h f_3 + g h D h + \varepsilon_{h,v^2} + \varepsilon_{h,h^2} = S \left[ \left( g^*_2 + \frac{1}{2} v L_1^* \right) - \left( g_2 + \frac{1}{2} v L_1 \right) \right],
\]

where \( f_3 = h v^2 + \frac{1}{2} h^2 \). The result (44) shows that we have an evolution of the discrete nodal values
of the momentum \( hv \), where all the volume terms are in conservative and consistent form. This
further shows that we can use the standard DGSEM-GL formulation of the conservative form (M2)
of the momentum balance and simply add the nonlinear correction terms \( \varepsilon_{h,v^2} \) and \( \varepsilon_{h,h^2} \) to recover the
skew-symmetric form. Thus, the skew-symmetric DGSEM for the shallow water equations can be implemented in an existing DGSEM-GL code with minimal impact by just adding local terms
to the formulation.

What remains is the definition of the auxiliary surface terms \( g_2 \) and \( g_2^* \). The definition follows
directly from the requirement that the formulation is exactly conservative for a constant bottom
topography \( b \). In this case, the fluxes need to exactly balance with the volume fluxes \( f_3 \), which gives

(45)
\[
f_3 = \frac{1}{2} g_2 + \frac{1}{2} v L_1 \quad \text{or} \quad g_2 = f_3 - \frac{1}{2} v L_1 = h v^2 + \frac{q}{2} h^2 - \frac{1}{2} v h v = \frac{1}{2} h v^2 + \frac{q}{2} h^2.
\]

Consequently, a point wise interpretation of the auxiliary surface term gives its continuous expression

(46)
\[
g_2 = \frac{1}{2} \left( h v^2 + g h^2 \right).
\]

Following the same reasoning, it must be that the numerical flux terms are consistent to the numerical
flux function \( f_3^* \)

(47)
\[
f_3^* = \frac{1}{2} g_2^* + \frac{1}{2} v L_1^* \quad \text{or} \quad g_2^* := f_3^* - \frac{1}{2} v L_1^*.
\]
Replacing the formulas (46) and (47) into the intermediate skew-symmetric formulation, we obtain the final version of the discrete momentum balance

\[
J \frac{d}{dt} \mathbf{u}_2 + D f_2 + g \hat{a} \hat{b} + \varepsilon_{\text{hv}}^2 + \varepsilon_{\text{b}} = S^* \left[ f_2^* - f_2 \right],
\]

which demonstrates even more clearly that the only difference to a standard DGSEM-GL in the conservative formulation are the nonlinear correction terms \( \varepsilon_{\text{hv}}^2 \) and \( \varepsilon_{\text{b}} \). We note that in the numerical results Sec. 5 we use this particular form of the skew-symmetric formulation as the additional implementation impact is very small.

Concluding this section, we have shown that with an appropriate choice of the auxiliary surface terms \( g_2 \) and \( g_2^* \) the skew-symmetric formulation is conservative for the mass and the momentum in the Lax-Wendroff sense.

### 4.2. Entropy Conservation

Up to now, the numerical interface fluxes \( f_2^* \) and \( f_2^* \) are not specified. We show in this subsection that the choice is important to achieve exact entropy conservation. For this, we closely follow the derivations presented in Sec. 2 for the continuous shallow water equations by mimicking each step to get the discrete entropy evolution.

However, as a first step, we look at the balance of the kinetic energy by multiplying the skew-symmetric form (34) (which is algebraically equivalent to the form (48)) with the velocity \( \mathbf{v} \)

\[
J \frac{1}{2} \left[ \mathbf{v} \partial_t (\hat{a} \mathbf{v}) + (\hat{a} \mathbf{v}) \partial_t \mathbf{v} \right] + \frac{1}{2} \left[ \mathbf{v} D \hat{a} v^2 + \hat{a} \mathbf{v}^2 D \mathbf{v} \right] + g \hat{a} \hat{b} = S^* \left[ g_2^* - g_2 \right].
\]

Assuming again continuity in time, the first term reduces to

\[
J \frac{1}{2} \left[ \mathbf{v} \partial_t (\hat{a} \mathbf{v}) + \hat{a} \mathbf{v} \partial_t \mathbf{v} \right] = J \partial_t \mathbf{k},
\]

where the kinetic energy is given by \( \mathbf{k} = \frac{1}{2} \hat{a} \mathbf{v}^2 \). The second term can be recast into

\[
D f_{\text{kin}} + \frac{1}{2} \left[ -D \hat{a} v^3 + \hat{a} \mathbf{v} D \mathbf{v} + \hat{a} \mathbf{v}^2 D \mathbf{v} \right] = D f_{\text{kin}} + \varepsilon_{\text{hv}^3},
\]

where \( f_{\text{kin}} = \frac{1}{2} D \hat{a} v^3 \) is the advection related flux and \( \varepsilon_{\text{hv}^3} \) is, again, a nonlinear correction term that is consistent and conservative to the flux \( f_{\text{kin}} \) and represents the error of the discrete product rule \(- (h v^3)_x + (h v^2)_x v + (h v^2)_x v \) scaled by the factor 1/2. Summarising, the skew-symmetric evolution of the discrete kinetic energy is

\[
J \partial_t \mathbf{k} + D f_{\text{kin}} + \varepsilon_{\text{hv}^3} + g \hat{a} \hat{b} = S^* \left[ g_2^* - g_2 \right],
\]

which is a consistent approximation to the continuous kinetic energy balance of the shallow water equations

\[
J (h v^2/2)_x + (h v^3/2)_x + g (h v) (h + b)_x = 0.
\]

As a second step, we derive the discrete balance equation for the potential energy \( p = g h^2/2 + g h b \) by multiplying the discrete continuity equation (30) by the factor \( g (\hat{a} + \hat{b}) \)

\[
J g (\hat{a} + \hat{b}) \partial_t \mathbf{w}_1 + g (\hat{a} + \hat{b}) D f_1 = g S (\hat{a} + \hat{b}) \left[ f_1^* - f_1 \right].
\]

With the assumption of time continuity this reduces to

\[
J \partial_t p + g (\hat{a} + \hat{b}) D g \mathbf{v} = g S (\hat{a} + \hat{b}) \left[ f_1^* - f_1 \right].
\]
Now we derive the discrete balance for the total energy \( e = k + p \) by adding the discrete kinetic energy balance (52) and the discrete potential energy balance (55)

\[
J \partial_t e + D f_{\text{kin}} + g h v D (h + b) + g (h + b) D h v = \text{s.t.,}
\]

where \( \text{s.t.} \) represents the surface terms. We focus first on the volume terms and then investigate the surface terms. We realise that the last two terms on the left hand side of (56) can be recast to

\[
g h v D (h + b) + g (h + b) D h v = D f_{\text{pot}} + g \left[ -D h v (h + b) + h v D (h + b) + g (h + b) D h v \right] = D f_{\text{pot}} + g h v (h + b),
\]

where we introduced the flux \( f_{\text{pot}} = g h v (h + b) \) and the typical correction term \( g h v (h + b) \) corresponding to the error in the discrete product rule. Summarising these derivations, we get the discrete evolution of the total energy as

\[
J \partial_t e + D \left( f_{\text{kin}} + f_{\text{pot}} \right) + g h v^3 + g h v (h + b) = \text{s.t.}
\]

For our diagonal norm summation-by-parts operator, the nonlinear correction terms are consistent and conservative to the fluxes \( f_{\text{kin}} \) and \( f_{\text{pot}} \). Thus, ignoring the surface terms for the moment, we have now that the discrete total energy is exactly preserved by our skew-symmetric DGSEM-GL. Formulation (58) is further a consistent approximation of the continuous total energy conservation (E)

\[
J e_t + f_\xi = 0,
\]

with \( e = h v^2 / 2 + g h^2 / 2 + g h b \) and \( f = h v^3 / 2 + g h v (h + b) \). As discussed above, the total energy plays a special role as it is also an entropy function for the shallow water equations. Thus, as a consequence of these derivations, the volume terms of the skew-symmetric formulation are discretely entropy conservative.

What is remaining is the choice of proper interface numerical flux functions \( f_1^* \) and \( f_2^* \) such that the entropy is also conserved when using multiple elements. For this analysis it is more suited to use the associated weak form of the equation, as in this case we get exact conservation of the quantity when all surface terms at a single interface cancel out.

We use the summation-by-parts property (25) to replace the derivative matrix in (58) by

\[
D = M^{-1} Q = M^{-1} \left( B - Q^T \right) = -S - M^{-1} Q^T,
\]

where we introduced the surface matrix \( S \), which includes a negative sign in its definition (32). The discrete weak evolution equation of the entropy reads

\[
J \partial_t e + \left( -S - M^{-1} Q^T \right) f + g h v^3 + g h v (h + b) = \text{s.t.},
\]

\[
J \partial_t e - M^{-1} Q^T f + g h v^3 + g h v (h + b) = \text{s.t.} + S f,
\]

where \( f = f_{\text{kin}} + f_{\text{pot}} = \frac{1}{2} h v^3 + g h v (h + b) \) is the nodal vector of the entropy flux \( f \). Computing the surface term, we get

\[
s.t. + S f = S + \left[ g h v^3 - g g_2 \right] + g S (h + b) \left[ f_1^* - f_1^* \right] + S f
\]

\[
= S \left[ g (h + b) f_1^* + v g_2^* - \left( g (h + b) f_1^* + v g_2^* + f \right) \right].
\]
We insert the definitions of \( g_2 \) and \( g_2^* \), see Eqs. (46) and (47) respectively, and rearrange the terms to

\[
\begin{aligned}
s.t. + \mathbb{S} f &= \mathbb{S} \left[ \left( (g(h + b) - \frac{1}{2} v^2) f_1^* + f_2^* \right) - \left( \frac{1}{2} h v^2 + g h v (h + b) + \frac{g}{2} h^2 v \right) + f \right] \\
&= \mathbb{S} \left[ \left( \frac{q_1}{q_1} f_1^* + f_2^* \right) - \left( f + \frac{g}{2} h^2 v \right) + f \right],
\end{aligned}
\]

(63)

where we inserted the definitions of the entropy variables \( q_1 = g(h + b) - v^2/2 \), \( q_2 = v \) and the entropy flux \( f \). Summarising, we get the following discrete evolution equation for the total energy (or entropy)

\[
J \partial_t c - M^{-1} Q^T \mathbb{M} f + \mathbb{S} f + z_{hv} + z_{hv(h+b)} = \mathbb{S} \left[ \left( \frac{q_1}{q_1} f_1^* + f_2^* \right) - \left( f + \frac{g}{2} h^2 v \right) + f \right].
\]

(64)

To obtain the evolution of the discrete mean value of the entropy we integrate (64) over the reference element and use the numerical quadrature associated to the GL points to replace the exact integral. This however is equivalent to multiplying (64) with \( \mathbb{I}^T M \)

\[
J \partial_t \mathbb{I}^T M c - \mathbb{I}^T Q^T \mathbb{M} f + \mathbb{I}^T Q^T \mathbb{S} f + \mathbb{I}^T Q^T \mathbb{S} z_{hv(h+b)} = \mathbb{I}^T M \mathbb{S} \left[ \left( \frac{q_1}{q_1} f_1^* + f_2^* \right) - \left( f + \frac{g}{2} h^2 v \right) + f \right].
\]

(65)

Due to the consistency of the derivative matrix \( \mathbb{D} \) in the sense that the derivative of a constant is exactly zero, we further have

\[
0 = \mathbb{D} \mathbb{1} = \mathbb{S} \mathbb{1} = \mathbb{M}^{-1} \mathbb{Q} \mathbb{1} \quad \Leftrightarrow \quad \mathbb{Q} \mathbb{1} = 0.
\]

(66)

With the definition of the surface matrix \( \mathbb{S} \) we have

\[
J \partial_t \mathbb{I}^T M c = \left[ \left( (q_1)_0 f_1^* \right)^+ - \left( (q_2)_0 f_2^* \right)^+ \right] - \left( f_0 + \frac{g}{2} h^2 v_0 \right) + f_0 - \left[ \left( (q_1)_N f_1^* \right)^+ - \left( (q_2)_N f_2^* \right)^+ \right] - \left( f_N + \frac{g}{2} h_N^2 v_N \right).
\]

(67)

We now look at a single interface to determine the numerical fluxes \( f_1^* \) and \( f_2^* \) such that the surface contributions exactly cancel out, which guarantees exact preservation of the discrete entropy mean value. We note that only the \( * \) quantities, i.e. the numerical flux functions, are unique at a surface. The other quantities in (67) are defined locally in an element. Thus with respect to a single interface, we get values from left and from the right. It is common to introduce the jump and the average at the interface as operators. For this, we consider an interface \( m + \frac{1}{2} \) between the elements \( m \) and \( m + 1 \) and define the jump respectively the average from a nodal DGSEM quantity \( a \) at the interface as

\[
[a] := a_{m+1}^m - a_N^m \quad \text{and} \quad \{a\} := \frac{1}{2}(a_{m+1}^m + a_N^m).
\]

(68)

If we sum equation (67) over all elements, the surface terms combine at every single interface to

\[
\text{interface contribution} = - \left[ [q_1] f_1^* \right]^+ - \left[ [q_2] f_2^* \right]^+ + \left[ f + \frac{g}{2} h^2 v \right] - \left\{ f \right\}.
\]

(69)
omitting the scaling by the corresponding entry of the surface matrix $S$. Our goal is to choose the numerical flux functions so that this interface contribution vanishes

$$ f + \frac{g}{2} h^2 v - \|q_1\| f_1^* + 1 - \|q_2\| f_2^* + 1 = \|f\|. $$

By realising that $(q_1, q_2) \circ (f_1, f_2) := q_1 f_1 + q_2 f_2 = f + \frac{g}{2} h^2 v$ for the one dimensional shallow water equations, this condition is equivalent to

$$ a b = \{a\} [b] + \{a\} \{b\} \quad \text{and} \quad a^2 = 2 \{a\} [a], $$

to get the following condition on the numerical flux functions $f_1^*$ and $f_2^*$

$$ \|q_1\| f_1^* + 1 + \|q_2\| f_2^* + 1 = \left[ \frac{g}{2} h^2 v \right] $$

Note that the jumps in height $[h]$ and velocity $[v]$ are independent variables, we require that the coefficients of the jump terms, i.e. the terms in the brackets, are independently zero. Thus, the first term yields

$$ f_{1,ec}^* = \{v\} \{h\}. $$

Inserting this in the second term yields

$$ f_{2,ec}^* = \{v\}^2 \{h\} + \frac{1}{2} g \{h^2\}. $$

We note that (74) and (75) match the numerical flux function derived for an entropy conservative finite volume discretisation of the shallow water equations, e.g. Fjordholm et al. [9].

Summarising these derivations, by choosing the numerical flux function

$$ f^*_{ec} = \begin{pmatrix} f_{1,ec}^* \\ f_{2,ec}^* \end{pmatrix} = \begin{pmatrix} \{v\} \{h\} \\ \{v\}^2 \{h\} + \frac{1}{2} g \{h^2\} \end{pmatrix}, $$

we exactly conserve the entropy when using the skew-symmetric DGSEM. We thus call this discretisation ECDGSEM.
4.3. **Well Balancedness of the ECDGSEM.** In the last part of this section, we show that the ECDGSEM is well balanced. For this, we consider the 'lake at rest' problem: we assume that the velocity $v$ is equal to zero and that the total height $h + b$ is constant. For such initial conditions, the solutions should be exactly steady state.

From the discrete continuity equation (30) with the entropy conserving flux (74) and the assumption that $v = 0$, we immediately see that

$$(77) \quad J \partial_t u_1 = -D f_1 + S \left[ f^{*,*ec}_1 - I_1 \right] = 0,$$

as the flux $f_1$ as well as the numerical flux $f^{*,*ec}_1$ depend on the velocity $v$. Also, by definition, the first nonlinear correction term in the momentum equation reduces to zero

$$(78) \quad 2h v^2 = 0.$$ 

The momentum flux reduces to

$$(79) \quad f_2 = g \frac{h^2}{2},$$

and so (48) reduces to

$$(80) \quad D f_2 + 2h v^2 + g h Db = g h Dh + g h Db = g h Db(h + b) = 0.$$ \text{due to the consistency of the derivative matrix } D \text{ and that } h + b \text{ is assumed to be constant for each GL point. Summarising our derivations so far, we get for the "lake at rest" initial conditions}

$$(81) \quad J u_2 = S \left[ f^{*,ec}_2 - f_2 \right].$$

Remaining is to check if the surface terms vanish, for which we consider the contribution of a single interface

$$(82) \quad \text{interface contribution} = \left( -f_2^{*,ec} \right)_{+1} + (f_2)_N = \left( -f_2^{*,ec} \right)_{+1} + (f_2)_0 = \left[ f_2 \right] = \frac{g}{2} \left[ h^2 \right],$$

where we omit the scaling by the matrix $S$.

We use the linearity and the rules of the jump operator (72) to get

$$(83) \quad \frac{g}{2} \left[ h^2 \right] = g \left\{ \left\{ h \right\} \left\{ h \right\} \right\} \left\{ \left\{ h \right\} \left\{ \left\{ h \right\} \right\} \left\{ h + b \right\} = 0, \right.$$ 

where we use the fact that the bottom topography $b$ is smoothly approximated across element interfaces by our GL ansatz, i.e. $[b] = 0$. As according to the initial conditions of the "lake at rest" problem, the total height is constant, i.e. $[h + b] = 0$, it follows consequently that

$$(84) \quad [h] = 0,$$

which we must satisfy when initialising our DG polynomials at time $t = 0$. With this, the discrete time derivative of the momentum reduces to

$$(85) \quad J \partial_t u_2 = 0.$$ 

Summarising our findings, the ECDGSEM is well balanced if the bottom topography is approximated smoothly across element interfaces, i.e. $[b] = 0$ and if the initialisation of the quantity $h$ is such that it is also continuous across element interface, i.e. $[h] = 0$ for $t = 0$ for all interfaces.

In total, we get the following
Theorem 1. (ECDGSEM) If we use the skew-symmetric DGSEM for the shallow water equations

\[ J \partial_t u_1 + D f_1 = S \left( f_1^* - f_1 \right), \]
\[ J \partial_t u_2 + D f_2 + g h D b + s_{hv^2} + s_{h^2} = S \left( f_2^* - f_2 \right), \]

where

\[ f_1 = h v, \]
\[ f_2 = h v^2 + \frac{g}{2} h^2, \]

and

\[ s_{hv^2} = \frac{1}{2} \left[ -D h v^2 + h v D v + v D h v \right], \]
\[ s_{h^2} = \frac{g}{2} \left[ -D h^2 + 2 h D h \right], \]

in combination with the special numerical flux

\[ f^* = f^{*, ec} = \left( \frac{\{v\}}{\{h\}} \right)^2 \left( \frac{\{h\}}{\{v\}} + \frac{1}{2} g \{h^2\} \right), \]

we get discrete exact conservation of the mass, momentum and total energy. The total energy is an entropy function for the shallow water equations, hence we exactly preserve the entropy of the system and therefore call this discretisation ECDGSEM.

Furthermore, if we approximate the bottom topography \( b \) such that it is continuous across element interfaces and use an initial condition such that the jumps in the water height \( h \) are zero for all interfaces, the ECDGSEM is well balanced.

Remark 1. The inner product form of the ECDGSEM reads as

\[ J \langle \partial_t u_1, \ell_i \rangle_N + \langle (I_N f_1)', \ell_i \rangle_N = - |f_1^* - f_1|^{+1}_N, \]
\[ J \langle \partial_t u_2, \ell_i \rangle_N + \langle (I_N f_2)', \ell_i \rangle_N + \langle g h (I_N b)', \ell_i \rangle_N + \langle (I_N s_{hv^2}), \ell_i \rangle_N + \langle (I_N s_{h^2}), \ell_i \rangle_N = - |f_2^* - f_2|^{+1}_N, \]

where

\[ \langle (I_N s_{hv^2}), \ell_i \rangle_N = \frac{1}{2} \left( (I_N hv^2)' + hv (I_N v)' + (I_N v)' hv, \ell_i \right)_N \]
\[ \langle (I_N s_{h^2}), \ell_i \rangle_N = \frac{g}{2} \left( (I_N h^2)' + 2(I_N h) h', \ell_i \right)_N \]

for \( i = 0, \ldots, N \).

5. Numerical Results

In this section, we numerically verify the theoretical findings for the ECDGSEM. We have implemented the formulation summarised in Thm. 1 with a low storage five stage fourth order accurate Runge-Kutta time integrator of Carpenter and Kennedy [3]. We note that for an existing DGSEM-GL code, it is only necessary to implement the nonlinear correction terms (88)/(91) and the entropy conserving numerical flux (89).
5.1. Convergence. Here, a problem with an exact solution is used as a verification tool to assess the accuracy of the ECDGSEM scheme. Specifically, we take the smooth functions

\[ h(x, t) = 2 + \cos(x) \cos(t), \quad v(x, t) = \frac{\sin(x) \sin(t)}{h}, \]

(92) to be the analytical solution of the shallow water equations with the bottom topography

\[ b_1(x) = \begin{cases} \sin \left( \frac{\pi x}{4} \right) & \text{if } |x - 10| \leq 2, \\ 0 & \text{otherwise}. \end{cases} \]

(93) Note that the solution (92) introduces additional source terms for the continuity and momentum equations of the form

\[ \begin{bmatrix} s_1(x, t) \\ s_2(x, t) \end{bmatrix} = \begin{bmatrix} h_v + hv_t - b_1 v_t + h_x v^2 + 2hv v_x - (b_1)_x v^2 - 2v v_x + hh_x - b_1 h_x \end{bmatrix}. \]

(94) The physical domain is taken to be \( \Omega = [0, 20] \). The exact solution (92) is used to prescribe the initial condition and the boundary conditions for the convergence test. The gravitational constant is set to \( g = 1 \) for this test.

We plot the computed height \( h \) and velocity \( v \) in Fig. 1 at the final time \( T = 1.0 \). For the discretisation, we use 10 regular elements and a polynomial degree of \( N = 8 \). The time step is taken such that the spatial error dominates, \( \Delta t = 1/1000 \).

![Figure 1](image.png)

**Figure 1.** (left): The computed total water height \( h + b \). The bottom topography is shown in grey. (right): The computed velocity \( v \).

In Table 1 results of a grid convergence study for an even and odd choice of polynomial degree is presented. As observed in previous work [11, 10], we get optimal convergence order for the even degrees \( (N = 4) \), but suboptimal convergence for odd polynomial degrees \( (N = 3) \) in all our tests.
Table 1. Experimental order of convergence EOC for $N = 3$ and $N = 4$.

<table>
<thead>
<tr>
<th># elements</th>
<th>$L_\infty$ error</th>
<th>EOC</th>
<th># elements</th>
<th>$L_\infty$ error</th>
<th>EOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N = 3$</td>
<td></td>
<td></td>
<td>$N = 4$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>5.45E-03</td>
<td>—</td>
<td>20</td>
<td>1.55E-04</td>
<td>—</td>
</tr>
<tr>
<td>40</td>
<td>6.16E-04</td>
<td>3.14</td>
<td>40</td>
<td>5.29E-06</td>
<td>4.87</td>
</tr>
<tr>
<td>80</td>
<td>7.92E-05</td>
<td>2.95</td>
<td>80</td>
<td>1.84E-07</td>
<td>4.85</td>
</tr>
<tr>
<td>160</td>
<td>9.86E-06</td>
<td>3.01</td>
<td>160</td>
<td>5.36E-09</td>
<td>5.10</td>
</tr>
</tbody>
</table>

Fig. 2 shows exponential convergence of the spatial approximation with 10 elements until $N = 27$ with $T = 1$, where time integrator errors become dominant. When the value of $\Delta t$ is halved, the error in the approximation is reduced by a factor of 16, as expected for the fourth order time integration scheme used. In this convergence study, the odd even behaviour of the polynomials can also be clearly observed especially for higher values of $N$.

5.2. **Mass and momentum conservation.** If we consider a constant bottom topography we know that the scheme will conserve mass and momentum. To test the conservation properties, we consider a dam break problem with a flat bottom topography and initial data

$$h(x, 0) = \begin{cases} 
 2.0 & \text{if } x < 10 \\
 1.5 & \text{if } x > 10 
\end{cases} \quad u(x, 0) = 0.$$
However, to get exact conservation of the total mass, we artificially set the boundary conditions to periodic. Again the computational domain is $\Omega = [0, 20)$. We use 50 elements and a polynomial degree of $N = 4$. The error in the discrete mass and momentum at time $T = 20$ is shown in Fig. 3. For the computation we set CFL = 0.5 and find that the errors are on the order of machine precision.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{l_infinity_error.png}
\caption{Plot of the $L_\infty$ error of the discrete mass and discrete momentum over time up to $T = 20$.}
\end{figure}

5.3. Total energy conservation. The scheme also exactly preserves the total energy, a fact we demonstrate for a flat topography as well as sine topography (93) and a polynomial topography
\cite{9, 38}

\begin{equation}
    b_2(x) = \begin{cases} 
        1 - \frac{1}{4}(x - 10)^2 & \text{if } |x - 10| \leq 2, \\
        0 & \text{otherwise.}
    \end{cases}
\end{equation}

We measure the change of total energy
\begin{equation}
    \Delta e(T) := |e_{int}(t = 0) - e_{int}(T)|,
\end{equation}
where $e_{int}$ is the total energy
\begin{equation}
    e = \frac{1}{2}(hv^2 + gh^2) + ghb
\end{equation}
integrated with the associated GL quadrature rule over the whole domain. The results for the same setup used in Sec. 5.2 are presented in Table 2.

<table>
<thead>
<tr>
<th>N</th>
<th>b(x) = 0</th>
<th>b₁(x)</th>
<th>b₂(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>1.70E-13</td>
<td>5.12E-13</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>3.97E-12</td>
<td>3.81E-12</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
<td>1.87E-12</td>
<td>8.52E-13</td>
</tr>
</tbody>
</table>

Table 2. The error Δe(20) in the total energy (98).

Because the total energy of the system is an entropy function for the shallow water equations [9], this test also shows that the scheme is entropy conserving.

5.4. Well-balancedness for still water equilibrium. We show that the approximation satisfies the “lake at rest” condition for three bottom topographies on the domain Ω = [0, 20]. The sine bump function \( b₁(x) \) is used to fully explore the error, if any, associated with the source term discretisation, because for polynomial functions the derivatives can be exact if the order \( N \) is large enough.

We take the analytical solution for the lake at rest problem to be

\[
h + b = 2, \quad u = 0,
\]

and run the computation up to a final time \( T = 20 \) with a CFL number of 0.5. The value \( N \) is the polynomial order in each element. We use 100 elements for this test case. In Table 3 we collect the \( L_∞ \) errors of the approximation of \( h + b \) at the final time for three topographies. In each case the error of the ECDGSEM is on the order machine precision.

<table>
<thead>
<tr>
<th>N</th>
<th>b(x) = 0</th>
<th>b₁(x)</th>
<th>b₂(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>1.11E-15</td>
<td>1.33E-15</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>1.39E-14</td>
<td>1.77E-14</td>
</tr>
<tr>
<td>3</td>
<td>9.76E-15</td>
<td>5.39E-14</td>
<td>5.72E-14</td>
</tr>
<tr>
<td>4</td>
<td>1.17E-14</td>
<td>6.99E-14</td>
<td>8.26E-14</td>
</tr>
</tbody>
</table>

Table 3. The \( L_∞ \) error in the total height \( h + b \) for the “lake at rest” at time \( T = 20 \).

Next we check the energy conservation of the ECDGSEM with the sine bottom topography (93) at a very large time \( T = 100 \). We impose the lake at rest initial condition (99). The gravitational constant is set to \( g = 9.812 \), and we impose periodic boundary conditions. The resulting states are shown in Fig. 4. As shown in this figure, the steady state is preserved exactly, even at this large time. Furthermore, the energy vs. time graph in Fig. 4 shows that the errors in total energy are
very small. These errors are due to the discretisation in the time stepping. Thus, the ECDGSEM scheme preserves the steady state as well as energy.

The shallow water equations admit another type of steady state solution often referred to as a general moving water equilibrium [9, 21]:

\[
m := h v = \text{constant}, \quad E := \frac{1}{2}v^2 + g(h + b) = \text{constant},
\]

where \( m \) and \( E \) are the moving water equilibrium variables. The “lake at rest” (99) (or still water steady state) considered in the previous section is a special case of the moving water equilibrium. Preservation of the moving water equilibrium is more difficult than the still water equilibrium and typically requires a special discretisation of the source term, see for example [9, 36].

We will demonstrate that our proposed ECDGSEM scheme maintains the well-balanced property for a moving water subcritical steady state problem with a non-flat bottom topography. This steady state problem is a widely used classical test case for numerical schemes that approximate shallow water equations [9, 21, 37, 36]. For the subcritical test we consider the bottom topography

\[
b_3(x) = \begin{cases} 
0.2 - 0.05(x - 10)^2 & \text{if } |x - 10| \leq 2, \\
0 & \text{otherwise}, 
\end{cases}
\]

on the physical domain \( \Omega = [0, 25] \) with \( g = 9.812 \). The initial condition for the subcritical steady state are

\[
m = 4.42, \quad E = 22.06605,
\]
together with the boundary condition of \( m = 4.42 \) at the upstream, and \( h = 2 \) at the downstream. To convert the initial condition from the equilibrium variables to the conserved variables we apply a Newton method to solve for the water height \( h \) in the formulation (100), for complete details see [21].

We compute the solution of the subcritical steady state up to \( T = 5 \) using 120 elements with polynomial order \( N = 3 \) or \( N = 4 \) in each element. We show in Table 4 the \( L_\infty \) error in the approximation of the water height \( h \) and the discharge \( hv \) to demonstrate that the steady state is maintained up to numerical round-off error. This verifies the desired well-balanced property for the subcritical moving water equilibrium configuration. The water height for the subcritical moving water steady state is shown in Fig. 5.

<table>
<thead>
<tr>
<th>( N )</th>
<th>( h )</th>
<th>( hv )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>8.43E-13</td>
<td>4.67E-13</td>
</tr>
<tr>
<td>4</td>
<td>2.62E-14</td>
<td>9.76E-14</td>
</tr>
</tbody>
</table>

**Table 4.** The \( L_\infty \) error in the water height \( h \) and discharge \( hv \) for the subcritical moving water equilibrium (102) at time \( T = 5 \) simulated with 120 elements.

5.6. **Dam break.** We use the same setup and initial conditions (95), except this time we choose outflow and inflow boundary conditions. The exact solution for this setup consists of a left-going rarefaction and a right-going shock. We present a solution computed with the ECDGSEM as well as study the numerical energy conservation in Fig. 6. The reference for the height \( h \) in the left part of the figure was obtained with a very highly resolved standard stable approximation. The other parameters of the computation are \( N = 4 \) with 100 elements.

The results in Fig. 6 show that the ECDGSEM scheme computes the rarefaction and the shock quite accurately, but at the expense of large post-shock oscillations. These oscillations are to be expected as energy must be dissipated across the shock but the ECDGSEM is basically dissipation free except for the influence of the time integrator, which we demonstrate on the right panel of Fig. 6, where we plot the total energy over time. As shown in the figure, the Runge-Kutta time stepping used creates small energy dissipation errors. These errors are reduced considerably by decreasing the CFL number, and hence the time step. This example and findings are analogous to the one presented in [9].

6. **Conclusions and Outlook**

In this work, we present a novel high order accurate discontinuous Galerkin based discretisation for the one dimensional shallow water equations. We use a skew-symmetric formulation of the problem as a base for our discretisation. The skew-symmetric form is an average of the conservative and the advection form of the equations. Despite being based on a non-conservative problem formulation, we prove that the skew-symmetric DGSEM conserves exactly mass and momentum. We further prove that by selecting appropriate numerical flux functions, the resulting method also
Figure 5. The computed water height for the subcritical moving water equilibrium test problem at time $T = 5$.

An entropy conserving scheme is not consistent for problems involving shocks, as entropy needs to be dissipated at a shock. The presented method is a baseline scheme with zero viscosity. Adding entropy consistent viscosity to this baseline scheme yields an entropy stable DG scheme for the shallow water equations. There are many approaches to include dissipation in a DG type discretisation, such as e.g. Riemann solver type numerical flux functions, slope limiting, filtering or artificial viscosity operators. It is a current research project to determine the form of viscosity which is the most appropriate for entropy conserving DG discretisations involving discontinuities. As an example, we employ ideas from the local Lax-Friedrichs type numerical flux function and replace our entropy...
conserving numerical flux by

\( f^{*,cs} = f^{*,cc} - \frac{1}{2} |\lambda_{max}| \| q \| \),

where \( \lambda_{max} \) is an interface local estimate of the maximal eigenvalue. We use the jump in entropy variables \( q \), to guarantee entropy consistency, in the sense that now the additional term arising in the discrete entropy equation (58) is

\( -\frac{1}{2} \| q \| |\lambda_{max}| \| q \| \),

thus guaranteeing that entropy is decreasing. Figure 7 shows the result of the damn break problem using the same discretisation parameters as in Sec. 5.6, except for the numerical flux function (103) and appropriate inflow/outflow boundary conditions. We see that in comparison to the entropy conserving scheme, Fig. 6, the entropy stable DG discretisation with local Lax-Friedrichs type stabilisation drastically reduces the oscillations. However, it can be seen that especially directly at the shock some oscillations remain, as entropy stability does not guarantee an oscillation free discretisation. Hence, more work is necessary to define a proper additional entropy consistent viscosity for the skew-symmetric DG method presented in this paper.

We finally remark that all theoretical proofs only rely on the summation-by-parts property of the DGSEM operators and that the norm matrix, which is the mass matrix in case of DGSEM, is diagonal. Furthermore, different finite difference blocks can be chosen, as the coupling of the blocks can be done with the same strategy as within the discontinuous Galerkin approach. By choosing the same entropy conserving numerical flux function, we directly get a conservative and well balanced discretisation for all diagonal norm summation-by-parts operators on multiple blocks, such as e.g. SBP finite differences.
Figure 7. (left): Height at $T = 1.5$ in dashed blue computed with the skew-symmetric scheme and the dissipative local Lax-Friedrichs type numerical flux function, resulting into an entropy stable discretisation. Solid black represents the reference solution.

References


