THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Mathématiques & Modélisation

École doctorale Information, Structures, Systèmes (I2S)

Unité de recherche Institut Montpelliérain Alexander Grothendieck (IMAG)

High-order numerical methods for shallow-water flows on unstructured meshes: moving boundaries and wave-structure interactions

Présentée par **Sacha Cardonna**
le 2 octobre 2026

Sous la direction de **Fabien Marche** et **François Vilar**

Rapporteurs

Michael Dumbser	Professore Ordinario	Università degli Studi di Trento, Italy
Jean-Luc Guermond	Full Professor	Texas A&M University, USA

Examinateurs

Jérôme Droniou	Directeur de recherche	CNRS, France
Sergey Gavrilyuk	Professeur des universités	Aix-Marseille Université, France
Maria Kazakova	Maîtresse de conférences	Université Savoie Mont-Blanc, France
Pascal Noble (Président)	Professeur des universités	Université de Toulouse, France

Invités

David Del Rey Fernández	Associate Professor	University of Waterloo, Canada
Bijan Mohammadi	Professeur des universités	Université de Montpellier, France

Direction de thèse

Fabien Marche	Maître de conférences HDR	Université de Montpellier, France
François Vilar	Maître de conférences	Université de Montpellier, France

UNIVERSITÉ DE MONTPELLIER

Institut Montpelliérain Alexander Grothendieck
Faculté des Sciences
Place Eugène Bataillon
34090 Montpellier

Informations Structures Systèmes (I2S)
Doctoral School
Rue de St-Priest
34090 Montpellier

Oh, my life, is changing every day, in every possible way

FROM: ***"Dreams"***, The Cranberries — *Everybody Else Is Doing It, So Why Can't We?* (1993)

Abstract. This thesis investigates the numerical modeling of free-surface flows, aiming to develop a robust, high-order numerical framework to simulate wave-structure interactions in the shallow-water regime. As these fluid-mechanic problems are strongly nonlinear, they present severe mathematical and computational challenges. To tackle them, we investigate several advanced and modern numerical tools to design novel and sophisticated approximation strategies capable of handling complex dynamics, while rigorously preserving, at the discrete level, the fundamental physical properties of the continuous models.

The manuscript is organized into two parts:

(*i*) The first part focuses on the development and analysis of the numerical methods. We design a novel local monolithic Discontinuous Galerkin (DG) and Finite Volume (FV) subcell scheme to approximate the nonlinear shallow-water equations with source terms. The core idea is to reinterpret the high-order DG formulation as an FV method on a sub-grid and blend high-order numerical fluxes with robust, low-order invariant-domain preserving fluxes *a priori* at the subcell level. Doing so, we mathematically guarantee the positivity of the water-height and effectively suppress the Gibbs phenomenon near discontinuities, while exactly preserving motionless steady states. Initially developed for the 1D Saint-Venant equations, the method is also extended to the dispersive Serre-Green-Naghdi equations by employing a discontinuous finite element approach for the elliptic contribution. We then generalize the monolithic framework to 2D unstructured meshes, and subsequently introduce an Arbitrary Lagrangian-Eulerian (ALE) formulation. By computing flows on deforming grids while rigorously respecting a semi-discrete geometric conservation law, this new monolithic DG/FV-ALE scheme provides a stepping stone toward simulating fully free-floating objects.

(*ii*) The second part focuses on the mathematical modeling and numerical simulation of surface obstacles and floating structures interacting with shallow-water flows. At the continuous level, these configurations lead to multi-physic coupled problems: the resulting system combines hyperbolic PDEs in the exterior free-surface region, elliptic equations governing the fluid dynamics beneath the object, and ordinary differential equations dictating the structural motion and enforcing interface transmission conditions. A major challenge is therefore to develop a rigorous discrete framework for this class of problems, to ensure robust, accurate and flexible numerical approximations. To achieve this, the exterior hyperbolic problem is solved using the monolithic subcell scheme developed in the first part. At the wave-structure interface, the coupling is driven by a Dirichlet-to-Neumann (DtN) operator, whose approximation requires to resolve the evolution of the fluid beneath the structure in the interior domain: we handle these problems with advanced high-order non-conforming methods, using a Symmetric Weighted Interior Penalty (SWIP-DG) formulation in 1D, and a Hybrid High-Order (HHO) method in 2D. Moreover, in a particular configuration, this DtN operator can be evaluated analytically using Fourier analysis, providing a rigorous reference to validate the proposed numerical methodology.

Extensive numerical validations are presented throughout the manuscript, demonstrating the accuracy and robustness of the proposed discrete strategies. These results fully contribute to the development of modern and flexible high-order numerical frameworks for nonlinear multi-physics problems and are a step towards practical applications of shallow-water models for the study of floating structures, relying on the effectiveness in capturing extreme nonlinearities and complex fluid-structure dynamics. Hence, this work offers encouraging perspectives for the development of advanced numerical tools for highly nonlinear challenges in coastal engineering.

Keywords. Wave-structure interactions – Shallow-water asymptotics – High-order methods – Arbitrary Lagrangian–Eulerian – Subcell schemes – Unstructured meshes

Résumé. Cette thèse porte sur la modélisation numérique des écoulements à surface libre, visant à développer un cadre d'ordre élevé et robuste pour simuler les interactions vagues-structure en régime d'eaux peu profondes. Ces problèmes étant fortement non-linéaires, ils posent d'importants défis mathématiques et numériques. Pour y faire face, nous nous appuyons sur des stratégies d'approximation modernes capables de traiter ces dynamiques complexes tout en préservant rigoureusement les propriétés physiques fondamentales des modèles continus au niveau discret.

Le manuscrit s'articule autour de deux parties :

(*i*) La première partie se concentre sur les méthodes numériques. Nous concevons un nouveau schéma sous-mailles monolithique local mélangeant Galerkin Discontinu (GD) et Volumes Finis (VF) pour approcher les équations de Saint-Venant avec termes sources. L'idée centrale est de réinterpréter la formulation DG d'ordre élevé comme une méthode VF sur un sous-maillage, et combiner *a priori* des flux d'ordre élevé avec des flux robustes d'ordre bas à l'échelle de la sous-maille, afin de garantir mathématiquement la positivité de la hauteur d'eau et supprimer les oscillations parasites près des discontinuités, tout en préservant exactement les états stationnaires du lac au repos. Initialement développée en 1D, la méthode est étendue aux équations dispersives de Serre-Green-Naghdi et combinée à une approche éléments finis discontinus pour la contribution elliptique. Nous généralisons ensuite ce cadre monolithique aux maillages non-structurés 2D, puis introduisons une formulation Arbitraire Lagrangien-Eulérien (ALE). En calculant les écoulements sur des grilles déformables tout en respectant rigoureusement la loi de conservation géométrique, cette nouvelle méthode monolithique GD/VF-ALE constitue une étape fondamentale vers la simulation d'objets flottants totalement libres.

(*ii*) La seconde partie porte sur la modélisation mathématique et la simulation numérique de structures flottantes interagissant avec des écoulements en eaux peu profondes. Au niveau continu, ces configurations conduisent à des problèmes couplés multi-physiques : le système combine des équations aux dérivées partielles hyperboliques dans la région extérieure à surface libre, des équations elliptiques régissant la dynamique du fluide sous l'objet, et des équations différentielles ordinaires dictant le mouvement de celui-ci et imposant certaines conditions de transmission aux interfaces. Le défi majeur est d'assurer que ces composantes mathématiques fondamentalement différentes communiquent correctement. Pour y parvenir, la région hyperbolique extérieure est calculée via le schéma monolithique de la première partie. À l'interface vague-structure, le couplage physique est piloté par un opérateur de Dirichlet-to-Neumann (DtN), dont l'évaluation nécessite la résolution du domaine fluide intérieur sous la structure. Nous traitons cette région par des méthodes d'éléments finis avancées : Galerkin Discontinu à pénalité intérieure pondérée en 1D, et des méthodes hybride d'ordre élevé (HHO) en 2D. De plus, dans un cas particulier, cet opérateur de DtN peut être obtenu analytiquement par analyse de Fourier, offrant une référence rigoureuse pour valider la méthodologie numérique proposée.

De nombreuses validations numériques sont présentées tout au long du manuscrit, mettant en évidence la précision et la robustesse des stratégies discrètes proposées. Ces résultats contribuent pleinement au développement de cadres numériques d'ordre élevé, à la fois modernes et flexibles, pour le traitement de problèmes multi-physiques non linéaires ; leur pertinence étant ici clairement établie par l'efficacité de ces méthodes à capturer des non-linéarités extrêmes ainsi que des dynamiques fluide-structure complexes. Enfin, ces travaux ouvrent des perspectives particulièrement prometteuses, en participant au développement d'outils numériques avancés pour l'étude de phénomènes fortement non linéaires en ingénierie côtière.

Mots-clés. Interactions vagues-structure – Modèles asymptotiques en eaux peu profondes – Méthodes d'ordre élevé – Arbitraire Lagrangien-Eulérien – Schémas de sous-mailles – Maillages non-structurés

A few words of gratitude — Quelques mots de gratitude

> – *Avec mon fils t'es très croyant, avec ma fille t'es complètement athée.*
> – *Parfait… Avec vous je dois être comment ?*
> – *Moi je crois qu'aux mathématiques, donc on a que ce qu'on mérite.*
> – *Et vous pensez que… Que je mérite d'être là avec vous ?*
> – *Je commence à le croire.*
>
> Sam Lion to Albert Duvivier, IN: ***"Itinéraire d'un enfant gâté"***, Claude Lelouch (1988)

On m'a reproché bien souvent ce cruel manque de concision, ma déférence ou encore la capacité que j'ai à politiser un peu tout ce qui sort de ma bouche : je me suis donc permis de transférer ici tous les potentiels manquements qui auraient jonché ce manuscrit. Il est donc facile de deviner qu'il ne faudra chercher parmi ces lignes ni sobriété, ni impartialité, et s'attendre ainsi à une bonne dose de sensiblerie. Car il est rare d'avoir une tribune où l'on peut remercier les autres, sans forcément devoir les *regarder dans les yeux*, alors autant en profiter.

Tout d'abord, une pensée pour les mathématiques, qui m'ont autant fait rêver qu'elles m'ont agacé, et notamment pour ces dernières années où elles m'ont permis de payer mon loyer. Quand j'ai commencé mes études, c'était pour moi cette voie "*royale*" – oserais-je dire, *parisienne*[1] ? – où l'on se détachait de *l'individu*, et où l'on développait quelque chose qui soit le plus *concis* et *impersonnel* possible ; les gens qui me sont proches savent notamment à quel point je suis tout sauf *concis* ou *impersonnel*[2]. Et pourtant, avoir travaillé pendant trois ans sur un seul et même sujet, ça a été l'occasion de voir passer les saisons un peu trop vite (des pulls en été aux shorts en hiver, toujours décoiffé quoi qu'il arrive), de prendre très (trop ?) régulièrement des leçons d'humilité, mais aussi de réaliser qu'une fois de plus je me trompais, et que les maths étaient une aventure humaine, qu'on pouvait les vivre avec style, panache, et sensibilité. Tout sauf impersonnellement, Fabien Marche et François Vilar m'ont accompagné et m'ont démontré[3] que l'on peut être un chercheur honnête, avec de la personnalité et des convictions, sans pour autant *devoir marcher sur la tête des autres* pour s'élever. J'aurai l'occasion de les remercier dans quelques lignes, mais je leur exprime déjà ma gratitude pour m'assumer[4] et avoir sollicité autant de personnes que j'admire pour composer ce jury.

Je commencerai donc évidemment par remercier Jean-Luc Guermond et Michael Dumbser, rapporteurs de la thèse, pour leurs compliments et leur expertise ; ce sont leurs précieux conseils et pistes d'amélioration qui ont permis de concevoir la meilleure version de ce manuscrit. Sachez à quel point je suis honoré d'avoir pu profiter de votre savoir et de vos immenses compétences sur ce travail.

Comme si deux tels rapporteurs ne suffisaient pas, j'ai la chance de compter sur des examinateurs spécialistes de l'analyse numérique et de mécanique des fluides, qui sont toutes et tous des références de notre communauté. À commencer par les professeurs Pascal Noble et Sergey Gavrilyuk, que j'ai eu la chance de rencontrer en conférence et qui ont de suite montré de l'intérêt à mes travaux encore balbutiants. Vos carrières sont une vraie source d'inspiration pour moi. Dans ces mêmes rassemblements, j'ai eu aussi la chance de croiser la route de Maria Kazakova, d'une gentillesse et d'une humilité rares (puis aussi brillante mathématicienne, rien que ça). Je suis ravi que tu aies accepté de participer à cette défense, et encore merci pour ce verre partagé à Aussois. Finalement, j'ai entendu parler du dernier examinateur[5] durant mon stage de M1, et j'avoue avoir passé quelques nuits à étudier son excellent polycopié de théorie de la mesure puis, un peu plus tard, `HArD::Core`. C'est donc naturellement que je témoigne à Jérôme Droniou ma reconnaissance pour avoir, lui aussi, accepté de donner de son précieux temps pour participer à ce jury.

Mais avant de rédiger ces quelques mots, et avant d'être un *beau bébé* de 300 pages, cette thèse était donc *en gestation*, et il fallait des personnes compétentes pour s'assurer que tout se déroule bien. J'ai ainsi eu la chance d'avoir dans mon comité de suivi individuel Quentin Richard et Arnaud Duran, qui m'ont accompagné durant ces trois années.

Lors de mon premier séminaire d'équipe en tant que doctorant, je me suis assis à côté de Quentin, qui a été le premier de l'équipe à me traiter comme un collègue à part entière, et le premier qui m'a laissé le tutoyer. Arnaud, lui, était déjà un modèle pour moi[6] avant même de le connaître personnellement ; j'ai eu un peu plus tard la chance de profiter de ses compétences scientifiques mais surtout humaines, notamment aux moments où je commençais à envisager la suite de ma carrière.

[1] pas de procès s.v.p., je suis né dans les Yvelines!
[2] voire à quel point je suis sûrement un littéraire plus qu'un scientifique.
[3] ou prouvé, pour ceux qui préfèrent l'anglicisme.
[4] i.e. ne pas avoir trop honte de moi.
[5] je cite "*un brillant prof' de Montpellier parti travailler en Australie*".
[6] sache que c'est grâce à ton manuscrit que j'avais enfin réussi à comprendre le principe des reconstructions hydrostatiques.

Je parle de suite, mais d'après l'*éternel retour* de Nietzsche, il n'y a ni suite ni commencement, et il faut surtout se demander si ce que nous faisons est digne d'être répété à l'infini[7]. Peut-être que je me passerais bien de ces *quelques* bugs dans mon code qui m'ont gâché *quelques* nuits, mais je referais volontiers le reste, ne serait-ce que pour apprendre à nouveau aux côtés de tous ces professeurs qui m'ont formé, accompagné, et qui m'offrent aujourd'hui cette tribune pour exprimer ma gratitude.

Au collège et au lycée, c'est d'abord Jérôme Calmeil[8] qui m'a donné le goût des mathématiques et de la rigueur[9]. J'ai ensuite fait la connaissance d'Isabelle Marty, arrivée à une période où je décrochais complètement, qui a su m'aider à retrouver la confiance que j'avais égarée en chemin. Avec le temps, vous êtes devenue une amie, parmi ces amis qu'on ne voit jamais et avec qui on communique peu, mais qui contribuent de près à la personne que l'on est. J'ai aussi un souvenir particulier de ce temps où j'ai appris à me *déconstruire*, notamment en suivant les cours de littérature de Paloma Belin-Casanova, qui est celle qui réveilla en moi le goût des mots et des arts, sans lequel je serais aujourd'hui sans nul doute un moins bon mathématicien.

C'est un peu plus tard que je me suis égaré une première fois, et que j'ai rejoint pendant une courte année les bancs de la faculté de médecine de Montpellier, où j'ai eu la chance de suivre les cours de Denis Mariano-Goulart et Winston Fairbairn, qui ont fini de me convaincre que j'étais plutôt fait pour *autre chose*.

À la faculté des sciences, j'ai finalement découvert ce qu'étaient les maths et la physique : j'ai eu notamment la chance d'avoir un dialogue privilégié avec Thierry Guillet, le premier qui m'a expliqué en quoi consistait le métier de maître de conférences, et avec Sylvain Brochard[10], le premier à m'avoir mis sur un sujet en autonomie.

Lors de mon master, j'ai eu la chance de faire un travail encadré de recherche avec Michel Bellieud, qui s'est poursuivi l'été suivant. Michel a été d'une gentillesse et d'une disponibilité rares[11], et je me souviendrai de lui comme le premier professeur à m'avoir affirmé que j'étais capable de faire une thèse. Je regrette de ne pas avoir pu garder contact avec lui, mais je lui serai toujours reconnaissant pour le temps qu'il m'a consacré, en espérant qu'il le sache un jour. Merci aussi à Matthieu Hillairet pour son intérêt durant ma dernière année, et pour m'avoir assuré que tout le monde, même des chercheurs chevronnés, souffrait aussi du syndrome de l'imposteur.

À vrai dire, certains de mes professeurs sont devenus mes collègues de travail : comment ne pas mentionner Pascal Azerad, un exemple pour qui veut accorder de la considération à ses étudiants, lui dont je me suis inspiré en essayant de retenir le prénom de tous mes étudiants de TD. Merci de m'avoir traité comme un prof' à part entière, et ce dès le début. Dans la lignée, je souhaite témoigner mon admiration à Jean-François Crouzet, qui en M2 nous a donné un cours d'une rare qualité, et qui avait le don d'écrire mieux au tableau que moi sur LaTeX. Vous êtes un modèle de pédagogie et de gentillesse ; merci aussi pour ces deux discussions, à trois ans d'intervalle, qui se sont étirées un peu plus que prévu.

Je n'oublie pas ma chère Céline Dricot, à qui j'ai pu démontrer mon incapacité à remplir correctement un ordre de mission ; merci pour ta patience et ton temps. Je suis également reconnaissant à tout le reste de l'équipe administrative de l'IMAG, notamment à Sophie Cazanave-Pin, qui m'a pistonné pour avoir un plus grand bureau[12], et à Nathalie Collain, qui a pris de mes nouvelles lorsque j'étais encore en convalescence. Sans votre travail quotidien, nous les matheux serions encore en train d'écrire sur des tablettes en argile.

J'ai aussi eu la chance de voyager à travers le monde grâce à mon travail, afin de pouvoir rencontrer des collègues en participant à des conférences, d'Aussois à Porto, en passant par Montréal et Chania. J'y ai notamment fait la rencontre de David Lannes, théoricien des EDPs-cycliste, aussi humble que brillant ; de Daniel le Roux, qui, d'une gentillesse rare, a tout de suite témoigné beaucoup d'intérêt à mes travaux, et qui a été d'une grande aide pour m'orienter dans mes recherches au Canada ; of Elena Gaburro, who always trusted me, even before I started my Ph.D. ; de Camilla Fiorini, où nous étions les deux seuls à ne pas boire d'alcool à SHARK[13] ; de Raphaël Loubère, qui m'a donné quelques pistes très intéressantes pour la suite et qui a des chemises parfaitement repassées ; de Gauthier Wissocq, aussi amical et accessible que son cas-test Leblanc était joli ; of *Takis* Tsoutsanis, for `UCNS3D`, which I am still exploring to this day ; de Maria Kazolea, pour sa gentillesse à Chania ; of Davide Torlo, whose brilliant papers I studied extensively ; de Victor Michel-Dansac, que je ne pensais pas si jeune. Je tiens également à témoigner à Walter Boscheri combien j'ai apprécié travailler avec lui durant cette semaine en Grèce. Merci pour ta confiance, et j'espère que l'on pourra continuer à bosser ensemble très vite. Finally, I also had the chance to meet Gregor Gassner, and I deeply thank him for sharing his wide expertise with me, his interest in my work and for his very contagious laugh.

Au laboratoire, j'aurai plutôt passé une thèse *en solitaire*. J'ai cependant eu la chance de croiser la route de quelques-uns de mes collègues, avec qui le temps pouvait s'étirer un peu plus. Mes premiers co-bureaux, mes premiers modèles, Aurelio et surtout Tanguy : j'étais triste d'avoir sympathisé avec toi juste avant que tu ne partes, mais tu es resté bien après ton départ, et tu es devenu une vraie référence pour moi, tant sur le plan pro[14] que sur l'accent anglais. J'ai eu l'occasion, plus tard, de commencer des sessions de *prise de nouvelles intensives* avec la géniale Orlane : désolé cependant de n'avoir jamais fait le premier pas. Pour Leonel et Patrick : merci de m'avoir aidé à composer le bureau le plus *gymbro* du labo,

[7] en tout cas, c'est ce que j'en ai compris.
[8] il est très probable qu'il ne se souvienne même pas de moi.
[9] et également, je dois l'admettre, le goût de tacler ses étudiants à longueur de journée.
[10] pour l'anecdote, c'est lui qui m'a parlé de Fabien en premier, que je ne connaissais pas, 6 ans avant le début de ma thèse.
[11] nous avons dû faire une bonne dizaine de visios, qui duraient rarement moins de trois heures.
[12] c'était il y a 2 ans, il y a prescription maintenant non ?
[13] mais pas pour les mêmes raisons !
[14] mille mercis de m'avoir conté tous les potins du labo.

et merci à Joseph de nous avoir rejoints un peu plus tard pour faire varier les plaisirs. Comment aussi ne pas remercier le gang, que dis-je, la Mafia des Libanaises de l'IMAG : Marwa, Christelle et puis Christina, ancienne adversaire devenue une amie, qui me rappelle que la vie, des fois, c'est un peu comme dans *Dragon Ball Z*.

Un soir de juin[15] 2024, une voiture roulait entre Pérols et Montpellier, et à son bord, quatre hommes extrêmement séduisants : votre cher narrateur[16], mais aussi et surtout Ibrahim, Marien et Ali. Je chéris ce moment de route, où j'ai gagné des amis, et quelques souvenirs. Marien, je suis tellement heureux qu'on ait commencé à travailler ensemble, et les remerciements de cette thèse ne failliront pas à la règle, comme on l'a écrit avant moi : "*il est hyper fort et hyper intelligent ce Marien*". Quant à toi Ali, mon frère, je t'aime autant que j'ai détesté le code que tu m'as laissé : je repense d'ailleurs à la fois où, dans une discussion autour de la programmation orientée objet, tu m'as dit "*oui Sacha tu sais je ne suis pas fort pour cela, tu vois la classe* `animal` *? Moi je mets* `voiture` *dedans*". Mais je me souviens aussi de ce soir pendant mon stage, où au téléphone, nous avons réussi à faire marcher mon premier code *Galerkin Discontinu*. Merci pour tout ce temps que tu m'as consacré, et pour ton amitié. Puissiez-vous, Marwa, Hassan, Mahdi et toi, avoir une longue et belle vie.

Je n'oublie pas non plus, quelques années auparavant, ce poster de *Trainspotting*[17] rue du Carré du Roi, trônant au milieu de deux grands tableaux blancs, où Baptiste et moi passions notre vie à faire de tout, sauf des maths. Merci d'avoir été là pour *viser le carbone* avec moi, de m'avoir rappelé parfois qu'on avait *le chien*, de dire bonjour à *Tal*, d'avoir payé quelques PVs, le *Wok* et de m'avoir donné 5.75 en optique ondulatoire. Bonne chance mon frère dans ta nouvelle vie de barman en Australie[18].

Et les étudiants dans tout ça ? "*I study too, you know. How could I teach anyone if I didn't grow and learn day by day?*". Quel plaisir ce fut pour moi d'avoir pu être votre prof. J'ai toujours souhaité enseigner ; mais qui pouvait penser qu'on apprenait tout autant si ce n'est plus auprès de nos élèves ? Me revoilà endetté, mais d'autant plus riche. Je me souviens de chacun de vos prénoms (j'ai essayé en tout cas), de notre complicité, de vos peurs, les doutes, vos réussites et nos échecs ; on a tant partagé ensemble! Merci d'avoir été là pour me changer les idées quand je vacillais, pour avoir ri à chacune de mes blagues, pour vos mots, et votre reconnaissance. Mon salaire était ici : vous voir partir de loin (Benjamin, toi à qui j'ai enseigné chaque année et qui es devenu tout aussi humble que brillant), me laisser vous *mettre des tunnels* (Victor, heureusement que tu n'étais pas pressé non plus), m'avoir fait confiance (Tom, promis si je perce je te prends en post-doc), m'avoir nourri (Gabin & Co, merci pour cette super soirée), devenir votre ami (Samir). Je vous souhaite à toutes et tous de réussir tout ce que vous entreprendrez, et d'avoir autant de chance que moi.

Pouvoir faire ce genre de rencontres aussi facilement est un cadeau : on a souvent tendance, en tant que fonctionnaires[19], à beaucoup se plaindre de ce métier et de notre employeur. Bien que notre système soit loin d'être idéal, je tiens quand même à remercier la France et sa République, pour m'avoir instruit, elle qui a financé mes livres, payé mes bourses, et qui m'a ainsi permis de faire dix ans d'études sans être endetté à vie. J'avoue être un peu triste d'y constater à mon tour les inégalités qui se creusent, et comment le milieu socioculturel fait trop souvent *la* différence dans la réussite individuelle. J'ai eu pour ma part la chance de sauter dans l'ascenseur social alors en pleine marche, sans trop y laisser des dents, et j'espère que notre génération et celles qui suivront auront à cœur d'en réparer le mécanisme.

Je suis également très reconnaissant envers le Canada, qui a accepté de m'accueillir sans même me connaître. J'y suis allé pour la première fois il y a maintenant dix ans, et c'est l'image que j'en ai conservée qui m'a souvent redonné du courage quand j'en manquais, et que le doute m'effleurait. Je raconte souvent cette anecdote, mais au Québec je me suis retrouvé à souvent déjeuner chez l'habitant. Très loin d'une vie pleine de fastes, le père de famille coupait du bois, la mère travaillait de la maison, et la grand-mère s'endormait au bout de la table, pendant que les enfants jouaient dessous. Je me souviens m'être dit "*ça peut être aussi simple que ça la vie ?*". Je ne sais pas si j'aurai moi-même les compétences nécessaires pour bûcheronner proprement ; en attendant j'ai hâte de découvrir ses terres, d'être agréablement surpris, parfois profondément déçu, et de payer la boîte de *Vache qui rit* 25$.

Speaking of this wonderful country, it is therefore the right place to thank David Del Rey Fernández, invited member for the defense, who immediately showed interest in my work, and who is giving me this amazing opportunity to continue my training as a young mathematician in Waterloo, Ontario. Finding a postdoc nowadays is rather difficult, especially on the other side of the world, and I feel truly honored that someone as kind, talented, and busy as David put his trust in me[20]. Merci d'ailleurs une nouvelle fois à toi Vincent Perrier, pour ta gentillesse à Montréal et pour m'avoir mis en contact avec David. I hope the three of us will have the chance to work together very soon!

Le second membre invité de ce jury fait tout sauf office de figuration dans cette liste. J'ai rencontré Bijan Mohammadi en dernière année de licence ; il portait un t-shirt *Matlab* et je me souviens qu'il nous avait présenté un exercice sur la minimisation d'une fonctionnelle[21]. C'est grâce à lui que j'ai appris à coder, et c'est d'ailleurs le premier enseignant qui nous a souligné l'importance de savoir (bien) programmer pour un mathématicien appliqué. Professeur Mohammadi a eu environ trois carrières différentes (bientôt quatre ?), et a constamment brillé en tant que chercheur. Mais trop peu ont l'occasion de souligner à quel point il a toujours eu *la vision* pour ses étudiants, et toujours prodigué les meilleurs

[15] était-ce mai, novembre, ici ou là ?
[16] au cas où vous n'auriez pas encore compris que j'avais un sacré melon.
[17] choose your future, choose life.
[18] ou bien est-ce tourneur-fraiseur à Ouagadougou ?
[19] pourtant, dans notre cas, loin d'être les plus mal lotis.
[20] and even replies to my emails on weekends!
[21] $\mathcal{J}(\boldsymbol{x}) := \frac{1}{2}\langle \mathbb{A}\boldsymbol{x}, \boldsymbol{x}\rangle - \langle \boldsymbol{b}, \boldsymbol{x}\rangle$, celles et ceux qui ont suivi ses cours savent qu'on faisait cet exo chaque année.

conseils. Je repense souvent à ce rendez-vous avec Fabien et François, à un moment de ma première année où j'étais totalement la tête sous l'eau – ironique quand on fait une thèse sur les vagues – et où je songeais à abandonner. Il se trouve que, par pur hasard, il passait dans le couloir, nous a rejoints dans ce bureau, et a, lui aussi, su trouver les *mots justes*. Ceci n'est qu'un exemple parmi tant d'autres qui me permet d'affirmer que, sans avoir croisé sa route, je ne serais jamais allé au bout de ce manuscrit. Même si ma gratitude doit peut-être le gêner, je tiens à lui dire à travers ces lignes qu'elle a toujours été sincère, et qu'il m'a littéralement permis de m'offrir une vie meilleure. Merci d'avoir accepté de composer ce jury, et de m'accompagner dans ce moment.

Last but not least.

François, ça fait plus de trois ans qu'on travaille ensemble, et que je suis toujours autant impressionné par ton humilité, alors que tu es vraiment *très fort*. Je savais déjà, en tant qu'étudiant, que tu étais un excellent pédagogue, mais en stage j'ai vite appris qu'aucun détail ne t'échappait : tu étais capable de voir où étaient les problèmes dans mon code juste en regardant une oscillation. J'ai pourtant bien souvent essayé de te truander, mais tu as une très (trop?) bonne mémoire[a] et j'ai appris, en mûrissant, que si tu me demandais de faire quelque chose, c'était sûrement pour une bonne raison. Je me souviens des quelques fois où j'arrivais dans ton bureau, je prenais une craie, j'écrivais des trucs au tableau en monologuant pendant dix minutes. Tu restais silencieux. À la fin, tu te tournais vers ton écran, et tu disais juste "*non*". Puis tu me rejoignais, pour me réexpliquer une énième fois les mêmes choses, avec la même patience et le même entrain. J'adore le fait que tu saisisses toujours les occasions qui se présentent pour donner à ton interlocuteur l'impression qu'il comprend quelque chose. Je suis heureux d'avoir récemment constaté que nous avions plusieurs convergences – d'ordre arbitraire – et ainsi que l'on avait encore beaucoup à partager[b].

[a] notamment les fois où j'esquivais certains trucs que tu me demandais de faire, mais que tu t'en souvenais quand même.

[b] promis je ne t'ai pas espionné, mais j'ai constaté certaines fois qu'on partageait d'autres points communs que l'analyse numérique.

Fabien, sache que je ne connais personne d'aussi *polymathe* que toi. Je commencerai en évoquant une certaine frustration, parce que j'aurais aimé que chacune de nos digressions dure éternellement ; j'ai tant appris à tes côtés. Merci pour ton exigence, pour ces mails parfois *légèrement assassins* mais toujours *foncièrement bienveillants*, qui m'ont permis de ne jamais refaire deux fois la même erreur. Nous n'avons pas souvent eu l'occasion de travailler ensemble en présentiel, je garde cependant le souvenir de nos échanges hebdomadaires à deux heures du matin[a] pour savoir quel commit avait cassé le code. Je te suis très reconnaissant pour m'avoir laissé m'ouvrir à toi très tôt, et pour avoir fait de même ; merci de m'avoir appris à assumer et gérer ma sensibilité. Je suis très heureux d'avoir reçu tes conseils, et d'avoir échangé avec cet interlocuteur qui savait faire la part des choses, écouter des idées de tout bord, avec qui le mot *dialogue* reprenait alors tout son sens. Je suis aussi très admiratif de ta capacité à faire passer les notions d'*égalité* et de *vérité* avant tout, pour celles et ceux qui suivront, même quand cela dérange, même quand cela implique d'y laisser des plumes. J'espère que la distance ne nous empêchera jamais de continuer à nous rapprocher[b].

[a] c'est sûrement durant mes insomnies à tes côtés que j'ai été le plus productif : si j'avais fait des nuits correctes, ce manuscrit serait bien plus court.

[b] et en espérant que ton prochain séjour au Canada pour venir me voir se passe mieux que le premier à Banff.

Pour toutes ces raisons, et bien d'autres, vous êtes des modèles pour moi, et je n'aurais jamais pu espérer avoir une meilleure direction de thèse que la vôtre. Merci de m'avoir choisi pour ce projet ambitieux, puis pour m'avoir formé, accompagné, et fait de moi un meilleur homme en sortie que je ne l'étais à l'entrée. Je suis fier de ce que nous avons accompli ensemble, et d'avoir, à vos côtés, appris *à marcher avant de savoir courir*. Puis-je rester dans ces sillons que vous avez tracés pour moi.

Comme le lecteur l'aura remarqué, j'avais beaucoup de choses à dire ; j'espère que les mots que j'ai choisis étaient les bons. Que l'on sache en tout cas que je suis à jamais reconnaissant d'avoir la chance de rédiger ces dernières lignes en bonne santé, quand je pense à toutes ces personnes qui auraient sûrement aimé n'avoir comme problème que de trouver de bonnes notations. Merci à ce tas d'os et de chair qui me soutient (environ neuf fois sur dix), et qui me permet de me mouvoir librement dans $\mathbb{R}^3 \times [0, T)$ – l'espace et le temps – pour faire des maths, enseigner, pratiquer du sport, écouter de la musique, briser de la vaisselle. Merci à cette *force invisible* qui n'a jamais le même nom, pour toute cette chance.

Avant de conclure, je souhaite adresser une pensée à Françoise Krasucki, que j'ai rencontrée durant mon master. Françoise, vous adoriez parler. Cela tombait bien, car moi aussi ; nous avions l'habitude d'étirer ce créneau du jeudi soir à discuter, et à refaire le monde. Sans ostentation, sans jugement ni condescendance, vous êtes la première à m'avoir *regardé dans les yeux*, à une époque où je n'étais ni assidu, ni compétent. J'aimerais avoir l'œil suffisamment aiguisé pour vous distinguer dans l'assemblée le jour de la soutenance, mais certaines choses nous dépasseront toujours, il faut croire.

Pour certains, la vie est parfois un peu injuste. Sans prétention à pouvoir y changer quoi que ce soit, je dédie malgré tout ces quelques années à toutes celles et ceux qui sont partis *avant d'avoir les cheveux blancs*. Pour J.-C., C., M., mais aussi à tous les gens qui restent, ne les oublions pas, ceux qui ont cette triste chance de vivre aussi pour ceux qui sont loin ; à mon père, à ma mère, et pour C.C.-C.

Pour toutes les places visitées où on ne s'rait jamais allés
Et celles où on est allés sans jamais vraiment visiter
FROM: ***"Merci ben !"***, Les Cowboys Fringants – *Pub Royal* (2024)

Contents

General introduction **1**
1. Physical context and modeling framework 1
1.1 Some applications driving development 1
1.2 The hyperbolic shallow-water regime 2
1.3 Weakly dispersive regimes 3
1.4 Source terms, friction and admissibility 3
1.5 Wave-structure interactions modeling 4
2. Numerical state-of-the-art for water-wave equations 6
2.1 Finite volume methods 7
2.2 High-order finite element methods 8
2.3 Well-balancing and discrete positivity preservation 9
2.4 Limiting strategies: the *a priori* paradigm 9
2.5 Limiting strategies: the *a posteriori* paradigm 10
2.6 Monolithic approaches 11
2.7 Moving domains and the ALE framework 12
3. Structure of the manuscript 13

I. A new high-order framework for shallow-water asymptotics **18**

Chapter 1. An introduction to subcell monolithic DG/FV scheme on one-dimensional nonlinear shallow-water equations **19**
1.1 Continuous and discrete settings 20
1.1.1 General shallow-water model in one space dimension 20
1.1.2 Discrete setting 21
1.2 Local-subcell monolithic DG/FV for homogeneous system 25
1.2.1 A hierarchy of numerical fluxes 25
1.2.2 DG and local-subcell DG/FV equivalency 27
1.2.3 Local subcell monolithic method and blending coefficients 27
1.3 Local-subcell approximations of bathymetry source term 30
1.4 Numerical validations 36
1.4.1 Smooth sinusoidal solutions 36
1.4.2 Dam-break over flat bottoms 38
1.4.3 Well-balancing with discontinuous bathymetry 42
1.4.4 Transcritical flow over a bump 42
1.4.5 Carrier & Greenspan's solutions 44
1.4.6 Periodic oscillations in a parabolic basin 47
1.4.7 Vacuum occurrence wave over a step 51

Chapter 2. Symmetric Interior Penalty DG/FV subcell formulation for 1D Serre-Green-Naghdi equations **53**
2.1 Some remarks about Serre-Green-Naghdi equations 54
2.1.1 Pre-balanced formulation 54
2.1.2 Symmetric discrete formulation 56
2.2 SWIP-DG discretization 57
2.2.1 Discrete bilinear form 57
2.2.2 Discrete Laplace and gradient operators 57
2.2.3 Semi-discrete DG–SWIP-DG formulation 58

2.2.4 Time discretization 59
2.3 Combined DG/FV–SWIP-DG fully-discrete formulations 59
2.4 Numerical results 61
2.4.1 Oscillating dam-break 61
2.4.2 Solitary-waves over a piecewise-linear bottom 62

Chapter 3. Local monolithic DG/FV subcell scheme for 2D nonlinear shallow-water equations on unstructured grids 65
3.1 Continuous and discrete settings 66
3.1.1 Nonlinear shallow-water model in two space dimension 66
3.1.2 Discrete framework 67
3.2 Discrete formulations and flux reconstruction 71
3.2.1 Classical DG formulation 72
3.2.2 Equivalency between DG and a subcell FV-like scheme 72
3.3 Local subcell monolithic DG/FV scheme 76
3.3.1 High and low-order topography-dependent numerical fluxes 76
3.3.2 Blending high and low-order fluxes 78
3.3.3 Source term discretization 78
3.4 Calibrating the blended fluxes 79
3.4.1 Blended and first-order Riemann intermediate states 80
3.4.2 Blending coefficient smoothening 80
3.4.3 Water-height positivity (PAD criterion) 81
3.4.4 Relaxed subcell numerical admissibility detection (SubNAD criterion) 82
3.4.5 Well-balancing for motionless steady-states 83
3.5 Numerical simulations 85
3.5.1 Steady vortex with smooth topography 85
3.5.2 Well-balancing 86
3.5.3 Dam-breaks in a channel 90
3.5.4 Circular dam-breaks 95
3.5.5 Carrier-Greenspan solutions 95
3.5.6 Wave-island interactions 101
3.5.7 Tsunami on an island with conical obstacles 108

Chapter 4. Subcell Arbitrary–Lagrangian Eulerian scheme for shallow-water flows on unstructured moving meshes 113
4.1 Continuous and discrete settings 114
4.1.1 The 2D pre-balanced nonlinear shallow-water equations 114
4.1.2 Discrete framework 114
4.2 Arbitrary Lagrangian–Eulerian description 117
4.2.1 Grid velocity and geometric conservation law 117
4.2.2 Adapted basis functions 118
4.2.3 ALE integral formulation 121
4.3 Discrete formulations and ALE extension 121
4.3.1 DG formulation and DG-ALE formulation 122
4.3.2 Monolithic DG/FV-ALE subcell formulation 122
4.4 Computation of the ALE grid velocity 124
4.4.1 From boundary motion to mesh motion 124
4.4.2 On the choice of the mesh velocity 125
4.4.3 Finite element formulations for the grid velocity 126
4.4.4 Reduction to explicit affine mapping for pseudo-1D configurations 129
4.5 Local subcell monolithic DG/FV–ALE scheme 132
4.6 Conservation properties of the ALE formulation 134
4.6.1 Semi and fully-discrete geometric conservation laws 135
4.6.2 Well-balancing property 139
4.7 Numerical illustrations 140
4.7.1 Well-balancing on a moving unstructured mesh 140
4.7.2 Periodic wave generation by an oscillating piston 142
4.7.3 Wave run-up over a sloping beach generated by an impulsively stopped piston 143
4.7.4 Wave generation by an oscillating cylindrical wavemaker 147

II. Waves interacting with floating structures in shallow-water regimes **154**

Chapter 5. Numerical modeling of a (pseudo)-one-dimensional wave-structure interactions problem with floating object **155**
5.1 Shallow-water free-surface flows 156
5.2 Floating structure in a shallow-water flow 156
5.3 Discrete formulation 161
5.3.1 SWIP-DG method for the elliptic problem on Bernoulli pressure 163
5.3.2 General SSP-RK discretization of the ODEs for the coupled system 164
5.3.3 Local subcell monolithic DG/FV method for the exterior region 165
5.3.4 Semi-discrete DG/SWIP-DG formulation 165
5.3.5 Fully-discrete monolithic DG/FV–SWIP-DG–SSP-RK formulation 166
5.4 Numerical illustrations 166
5.4.1 Geometry of the floating object 167
5.4.2 Single-wave interaction with an obstacle over an emerged topography 168
5.4.3 Wave generation by a prescribed vertical motion 169
5.4.4 Wave interaction with a freely heaving floating structure 175
5.4.5 Return to equilibrium test with analytical reference 180

Chapter 6. Coupled subcell monolithic DG/FV–HHO scheme for a new two-dimensional wave-structure interactions model **185**
6.1 Wave-structure interaction model 186
6.1.1 Description as a partly congested nonlinear shallow-water system 186
6.1.2 A mixed IBVP in the exterior region for irrotational data 188
6.1.3 An equivalent formulation 189
6.1.4 A special configuration: the infinite pontoon 189
6.2 Discrete formulations 194
6.2.1 Approximation framework 194
6.2.2 SSP-RK discretization of the ODE for the interface potential trace 200
6.2.3 Fast Fourier Transform computation of the DtN operator 201
6.2.4 HHO formulation of the interior elliptic problem 202
6.2.5 Local monolithic DG/FV subcell method for the exterior region 203
6.2.6 Semi-discrete DG–HHO formulation 205
6.2.7 Fully-discrete monolithic DG/FV–HHO–SSP-RK formulation 205
6.3 Numerical experiments 206
6.3.1 Geometry of the structures 206
6.3.2 Wave-pontoon interactions and DtN solver comparison 207
6.3.3 Incident wave interactions with a fixed cylindrical obstacle 210

Conclusion and perspectives **217**

Appendices **224**

Appendix A. Implementation aspects **225**
A.1 Semi-discrete DG formulation of hyperbolic systems 225
A.2 Data structures and design choices 226
A.2.1 The (unstructured) mesh object 226
A.2.2 Quadrature rules 227
A.2.3 Discrete basis functions 228
A.2.4 Modal and nodal representations objects 228
A.3 Monolithic DG/FV subcell framework 229
A.3.1 Subcell semi-discrete reformulation 229
A.3.2 Subcell geometry and topology 230
A.3.3 Inherited DG components and additional objects 232
A.4 SSP-RK–DG/FV-subcells solver workflow 232
A.5 Extension to Arbitrary Lagrangian-Eulerian (ALE) framework 235
A.5.1 Dynamic mesh and time-dependent geometry 235
A.5.2 Grid velocity computation and $\mathbb{P}^1$ finite element solvers 235

A.5.3 Interpolation at Gauss points 237
A.5.4 Time-dependent quadrature and basis functions 237
A.6 SSP-RK–ALE–DG/FV-subcells solver workflow 238

Appendix B. Positivity-preserving first-order Riemann intermediate states 241

Appendix C. Extension of model (6.15)-(6.18) towards heaving floating structures 243

Appendix D. Well-posedness of the HHO scheme for the wave-structure model 247

Appendix E. HHO algebraic formulation and static condensation 251

Appendix F. Entropy considerations 255
E.1 Subcell discrete inequality 256
E.2 Subcell semi-discrete inequality 257
E.3 Cell semi-discrete inequality 257

Appendix G. Continuous settings and functional spaces 263
F.1 Smooth function spaces 263
F.2 Classical differential operators 263
F.3 Lebesgue spaces 264
F.4 Distributions and weak derivatives 265
F.5 Sobolev spaces 265
F.6 Fourier transform 266
F.7 Some fundamental results 267

List of Figures 269

Bibliography 289

General introduction

"L'expérience est la source unique de la vérité : elle seule peut nous apprendre quelque chose de nouveau ; elle seule peut nous donner la certitude. Voilà deux points que nul ne peut contester [...] *L'hypothèse ainsi renversée a-t-elle donc été stérile ? Loin de là, on peut dire qu'elle a rendu plus de services qu'une hypothèse vraie ; non seulement elle a été l'occasion de l'expérience décisive, mais on aurait fait cette expérience par hasard, sans avoir fait l'hypothèse, qu'on n'en aurait rien tiré ; on n'y aurait rien vu d'extraordinaire ; on n'aurait catalogué qu'un fait de plus sans en déduire la moindre conséquence* [...] *Maintenant à quelle condition l'usage de l'hypothèse est-il sans danger ? Le ferme propos de se soumettre à l'expérience ne suffit pas ; il y a encore des hypothèses dangereuses ; ce sont d'abord, ce sont surtout celles qui sont tacites et inconscientes. Puisque nous les faisons sans le savoir, nous sommes impuissants à les abandonner."*

Henri Poincaré, FROM: ***"La Science et l'Hypothèse"***, 1902.

Contents

1. Physical context and modeling framework 1
   - 1.1 Some applications driving development 1
   - 1.2 The hyperbolic shallow-water regime 2
   - 1.3 Weakly dispersive regimes 3
   - 1.4 Source terms, friction and admissibility 3
   - 1.5 Wave-structure interactions modeling 4
2. Numerical state-of-the-art for water-wave equations 6
   - 2.1 Finite volume methods 7
   - 2.2 High-order finite element methods 8
   - 2.3 Well-balancing and discrete positivity preservation 9
   - 2.4 Limiting strategies: the *a priori* paradigm 9
   - 2.5 Limiting strategies: the *a posteriori* paradigm 10
   - 2.6 Monolithic approaches 11
   - 2.7 Moving domains and the ALE framework 12
3. Structure of the manuscript 13

1. Physical context and modeling framework

The numerical prediction of wave-structure interactions is a central problem in coastal engineering, environmental fluid mechanics, and geophysical flow modeling. This difficulty is already present in the nature of water waves themselves: as Feynman once told his students, water waves are often used as the simplest example of waves, although they are in fact "*the worst possible example because they are in no respects like sound and light, they have all the complications that waves can have*". Such problems appear for instance in tsunami propagation, harbor protection, wave energy devices, and coastal inundation. In all these situations, one has to describe the propagation of water waves in domains where fixed, floating, or partially immersed structures may strongly modify the flow. The difficulty is that these flows are most of the time strongly nonlinear, may involve complex bottom variations, and can generate steep fronts, wet-dry transitions, or large-scale free-surface motions. The numerical method must therefore be accurate enough to capture all the wave details, but also robust enough to remain stable in difficult regimes.

1.1 Some applications driving development

Several application areas have strongly motivated the development of accurate and robust numerical methods for shallow-water models. We briefly mention three of them, since they are representative of the main difficulties addressed in this manuscript.



Kanagawa-oki nami-ura (神奈川沖浪裏) , "The Great Wave off Kanagawa" by Katsushika Hokusai (1760–1849).

Tsunami modeling and prediction. Large tsunami events, such as the 2004 Indian Ocean tsunami and the 2011 Tohoku tsunami, have shown the importance of fast and reliable numerical prediction tools. In this context, one has to simulate waves over very large distances, using complex bathymetric data, and then predict arrival times, wave amplitudes, and possible inundation zones near the coast. Operational models must therefore be efficient enough to run quickly after an earthquake, but also robust enough to handle strong topographic variations and wet-dry transitions. This has motivated the development of well-balanced and positivity-preserving shallow-water solvers, often combined with adaptive meshes and high-performance computing tools [75, 87, 129, 147, 166, 195, 207, 252, 255].

Wave energy conversion. Wave energy converters (WEC) are floating or fixed devices designed to extract energy from ocean waves. Their efficiency depends strongly on the interaction between the incoming waves and the structure, including effects such as added mass, radiation damping, and nonlinear wave response. Linear potential-flow models are useful in many preliminary studies, but they may become insufficient when large-amplitude waves, dispersive effects, or near-resonant regimes are involved. In this setting, nonlinear shallow-water or weakly dispersive models provide an intermediate level of description: they are less expensive than full CFD models, but can still include important nonlinear effects in the wave-structure coupling [13, 35, 116, 144, 198, 214, 289].

Nearshore hydrodynamics and coastal inundation. Coastal flooding is another important application. Storm surges, large tides, sea-level rise, and intense rainfall events may all lead to inundation in nearshore areas. Numerical models are used to predict water levels, flooded zones, and flow velocities, which are essential quantities for risk assessment and coastal planning. These simulations usually involve irregular bathymetry, moving dry areas, and sometimes fixed or partially immersed structures. The shallow-water equations are well adapted to many of these regimes, but they require numerical methods that can remain stable on complex domains and under strong nonlinear effects. Recent developments, including high-order models for coastal flows [134, 230, 247, 271], show the relevance of accurate and efficient solvers for these applications.

1.2 The hyperbolic shallow-water regime

Many geophysical and coastal flows take place in a regime where the horizontal length scale is much larger than the vertical one. This is the case for tsunami waves, tidal flows, storm surges, or long waves propagating in shallow coastal areas. In such configurations, the typical water depth is small compared with the wavelength of the perturbation. This motivates the shallow-water, or long-wave, regime, described by the small parameter

$$\mu := \frac{H_0^2}{\lambda^2} \ll 1,$$

where H_0 is a characteristic water depth and λ is a characteristic wavelength. When μ is small, the vertical dynamics can be simplified, and the leading-order behaviour of the flow is mainly governed by depth-averaged horizontal quantities.

Starting from the incompressible Euler equations with a free-surface, one can derive reduced models by using an asymptotic expansion with respect to μ. This procedure is now well understood and rigorously presented in modern references such as [181]. The idea is to expand the unknowns in powers of μ, integrate the vertical

dependence, and obtain depth-averaged equations. If only the leading-order terms are kept, while all terms of order $\mathcal{O}(\mu)$ and higher are neglected, one obtains the *Nonlinear Shallow Water* (NSW) equations.

An important point is that this derivation does not require the wave amplitude to be small compared with the depth. If a denotes a typical wave amplitude, the parameter

$$(\textit{fully nonlinear}) \quad \varepsilon := \frac{a}{H_0} = \mathcal{O}(1),$$

is allowed to be of order one. This means that the NSW equations are fully nonlinear: they can describe waves whose amplitude is comparable with the water depth. This is one reason why they are widely used in coastal applications, including tsunami modeling and inundation studies [68, 110, 296].

Historically, the NSW equations were introduced by Barré de Saint-Venant in 1871 [77], through a depth-averaging argument. Their modern derivation [181] gives a more precise asymptotic justification, but the physical idea is the same: replacing the full vertical structure of the flow by a depth-averaged model. Despite their simplicity, NSW equations are able to reproduce many important phenomena, such as tidal flows, dam-break waves, flood propagation, tsunami run-up, and wave transformation over varying bathymetry. Their validity is of course limited, but in many long-wave regimes they provide a good compromise between physical accuracy and computational cost. The reader will find more details about the models in General shallow-water model in one space dimension and Nonlinear shallow-water model in two space dimension.

In the following, some numerical methods will be illustrated on the homogeneous two-dimensional nonlinear shallow-water equations, written in conservative form as

$$\partial_t \boldsymbol{v} + \nabla_{\boldsymbol{x}} \cdot \mathbb{F}(\boldsymbol{v}) = 0, \tag{1}$$

where $\boldsymbol{v} = (H, \boldsymbol{q})^\top \in \mathscr{H}_b^+$ denotes the vector of conservative variables living in the convex set of admissible states

$$\mathscr{H}_b^+ := \{(H, \boldsymbol{q}) \in \mathbb{R}^3 \mid H \geq 0\},$$

and $\mathbb{F}(\boldsymbol{v})$ is the corresponding flux function. In this simplified setting, the system can be viewed as a prototypical hyperbolic system of conservation laws.

1.3 Weakly dispersive regimes

The main limitation of the NSW model is that it neglects dispersive effects. In the asymptotic derivation, all terms of order μ and higher are discarded. These terms are responsible for the dependence of the wave speed on the wavelength. This is what is meant by dispersion: different components of a wave signal may travel at different speeds, depending on their wavelength or frequency. As a consequence, an initially localized wave can change shape during its propagation. In the full water-wave problem, waves of different wavelengths do not all travel at the same speed. This frequency dispersion can be important, especially for shoaling waves, solitary waves, or wave groups approaching the coast. In such situations, the NSW equations may give a correct large-scale description, but they cannot reproduce the detailed transformation of the wave profile.

Keeping terms of order μ leads to weakly dispersive shallow-water models. The most classical examples are the *Boussinesq-type equations* (Bt) [43] and the *Serre-Green-Naghdi equations* (SGN) introduced by Green and Naghdi [128] and Serre [245]. Bt models are usually derived in a weakly nonlinear regime, with $\varepsilon \ll 1$ and a balance between nonlinearity and dispersion. SGN equations, on the other hand, keep the fully nonlinear scaling $\varepsilon = \mathcal{O}(1)$ while adding weak dispersive corrections. They are therefore more appropriate for large-amplitude long waves than weakly nonlinear Bt models [7, 99, 137, 274, 276].

Many variants of weakly dispersive models have been proposed, depending on the target regime and on the desired accuracy of the dispersion relation. Classical works such as [209, 225] improved Bt models for practical coastal computations. Other contributions developed fully nonlinear SGN-type models and related asymptotic systems [36, 56, 57, 99, 100, 137, 184]. For example, the Peregrine model [232] and the Abbott model [1] belong to the weakly nonlinear family, while SGN variants are designed for fully nonlinear weakly dispersive flows.

1.4 Source terms, friction and admissibility

In coastal and geophysical applications, the bottom is not flat. Bathymetry may vary slowly, as on a sloping beach, or very sharply, as near submarine ridges, channels, or coastal structures. These variations have a strong influence on wave propagation. When a tsunami approaches the coast, for example, the decrease of water depth

modifies the wavelength and increases the amplitude of the wave. More generally, bathymetry can produce shoaling, refraction, diffraction, focusing effects, or wave reflection. It is therefore essential to include bottom variations in the model and to discretize them carefully.

In the NSW equations, bathymetry appears through the algebraic source term

$$\boldsymbol{B}[b] := -\mathrm{g}H(\boldsymbol{x},t)\nabla_{\boldsymbol{x}}b(\boldsymbol{x}),$$

where $b(\boldsymbol{x})$ denotes the bottom elevation. This term is one of the main numerical difficulties of the model: a lake at rest over a non-flat bottom should be an equilibrium solution of the system, given by the following velocity and free-surface profiles:

$$\boldsymbol{u} \equiv 0, \qquad H + b = \text{constant}. \tag{2}$$

A numerical method should preserve this equilibrium exactly (or at least up to machine precision). Otherwise, the scheme may create artificial velocities even though the exact solution is motionless. This is the well-balancing problem, which will be discussed later in more detail. Let us remark that some models take into account the time-dependence of the bathymetry, for instance to describe sediment transport or moving obstacles. In such cases, the source term may also include a contribution of the form $\partial_t b$, which can be even more difficult to handle numerically [50, 58, 74, 203].

Bottom friction is another important physical effect. It can be added through various forms, one of them being the source term

$$\boldsymbol{R}[\mathrm{n_f},\gamma] := -\mathrm{n_f}\frac{\boldsymbol{q}(\boldsymbol{x},t)\|\boldsymbol{q}(\boldsymbol{x},t)\|}{H(\boldsymbol{x},t)^{\gamma}},$$

where $\mathrm{n_f}$ is a friction coefficient and γ depends on the chosen friction law. This term dissipates kinetic energy and can modify the amplitude and shape of propagating waves. In long-time simulations, or in shallow regions, the effect of friction may become significant. From the numerical point of view, it must be treated without destroying stability, positivity of the water-height, asymptotic behaviors, or the well-balanced structure of the scheme [30, 63, 64, 67, 79, 80, 192, 210].

Another constraint is the physical admissibility of the shallow-water state. The conserved variables have a direct meaning: $H(\boldsymbol{x},t)$ is the water depth and $\boldsymbol{q}(\boldsymbol{x},t) = H(\boldsymbol{x},t)\boldsymbol{u}(\boldsymbol{x},t)$ is the depth-integrated horizontal discharge. Water depth cannot be negative, and this condition is particularly important near dry areas, where wet regions with $H > 0$ meet dry regions with $H = 0$. Such interfaces appear naturally in dam-break problems, run-up and run-down on beaches, and flood simulations. A numerical method must therefore keep the water depth non-negative while still capturing the motion of the shoreline with enough accuracy. This point is one of the numerical difficulties discussed in the next section.

1.5 Wave-structure interactions modeling

The modeling problems considered in this manuscript are not restricted to wave propagation over a varying bottom. A central motivation is the interaction between shallow-water flows and surface-structures. Depending on the physical configuration, the structure may be fixed, partially immersed, floating, or allowed to move with constraints (heaving) or even freely for a complete description. In each case, the presence of the object modifies the flow through free-boundary problems, pressure effects and geometric constraints. This is already a difficult problem in a fixed computational domain, and it becomes even more difficult when the boundary of the fluid region may evolve in time.

The modeling of waves interacting with a floating or partly immersed object is a complex fluid-structure interaction: it requires us to describe the coupled evolution of an incompressible fluid and a solid body. In addition to the usual difficulties related to free surface flows, one also has to deal with the geometry of the contact between the fluid and the structure. In general, the wetted part of the body is not fixed and may change with time, which leads to an additional free-boundary difficulty. This is one of the reasons why the derivation of reduced models, and their numerical approximation, remains challenging.

Classical mathematical descriptions of floating body problems go back to the work of John [163]. In this linear framework, the fluid is assumed to be inviscid and irrotational, so that the hydrodynamics can be described by a potential flow model. The motion of the solid is also assumed to have small amplitude around a prescribed equilibrium position. Although this setting is too restrictive for many applications, it already shows one of the main difficulties of wave-structure interaction problems: the need to prescribe suitable transmission conditions between different fluid regions. In the exterior region, the free surface evolves without direct contact with the

structure, whereas in the interior region, located below the body, the fluid exerts a pressure force on the wetted surface.

This linear approach has been developed in several directions, especially for offshore structures [226], floating WEC [190], and floating breakwaters [174, 278]. When combined with *Boundary Element Methods* (BEM) in the frequency domain, linear potential theory also provides the basis of standard hydrodynamic solvers such as WAMIT or ANSYS Aqwa. These methods are widely used because they give fast predictions of added mass, radiation damping, diffraction effects, and wave-induced motions, at least when the assumptions of linear theory remain valid. Other simplified approaches are also based on linear or weakly nonlinear assumptions: for instance, semi-analytic descriptions such as the point absorber approximation for WEC neglect part of the diffraction effects and use simplified potential representations [44]. Another related strategy consists in using the time-dependent mild-slope equation to describe wave propagation over slowly varying bathymetry or in the presence of floating devices [21]. Compared with Reynolds Averaged Navier-Stokes simulations, or with fully nonlinear potential flow solvers with nonlinear free surface conditions, these models are much cheaper and often give reliable leading-order estimates for small to moderate sea states.

However, the range of validity of linear models is limited. When the wave amplitude increases, or the mean water-height decreases, nonlinear effects may become important. This is the case, for instance, for WEC operating close to resonance, or for floating breakwaters located in nearshore regions. In shallow-water regimes, bathymetric effects may also play a central role. Phenomena such as shoaling, refraction, nonlinear energy transfers, and dissipation cannot be neglected anymore; these effects are not naturally captured by classical linear-BEM approaches, especially in intermediate or shallow-water configurations.

Several nonlinear strategies have therefore been proposed. A first class of methods relies on fully nonlinear potential flow models, often coupled with nonlinear BEM in the time domain. This makes it possible to keep a potential description of the fluid while accounting for nonlinear free surface effects and, in some cases, variable bathymetry [139]. Such models remain less expensive than full CFD computations, but they are still more demanding than depth-integrated shallow-water or Bt models.

A second strategy, largely less investigated up to recent years, is to replace the full water-wave equations by asymptotic models accounting for some specific flow regimes. This gives a compromise between oversimplified linear theories and expensive three-dimensional solvers. Such depth-integrated models may be used to describe the exterior free surface evolution, while another reduced description is imposed below or near the structure. For instance, in [173], floating breakwaters are modeled with a Bt system discretized by a finite-difference method. The flow below the breakwater is treated as a confined flow, and the pressure field beneath the structure is obtained through the implicit resolution of a Laplace problem. Reduced asymptotic models have also been used in near-ship configurations. In [66], the Kadomtsev-Petviashvili equations are used to describe shallow-water wave generation by ships, while Bt equations have been applied to model near-ship wave interactions in [161, 162, 287]. More recently, Bt formulations have also been used for floating structures with vertical side walls, for instance in [41], where the heave motion of straight-sided floating bodies is considered.

A more systematic formulation of the nonlinear floating body problem was introduced by Lannes in [182]. Instead of writing the model only in terms of the velocity potential, the unknowns are the free surface elevation and the horizontal discharge. This formulation is particularly useful for deriving reduced models, since it naturally separates the exterior free-surface region from the constrained interior region below the body. In the interior region, the pressure exerted on the structure appears as a Lagrange multiplier associated with the constraint imposed by the solid. One important consequence is that this pressure can be determined from a lower-dimensional elliptic problem. This gives a natural way to couple NSW or Bt equations with the motion of a floating body.

From the modeling point of view, these configurations often lead to transmission problems between different regions. In the exterior region, the flow may be governed by shallow-water equations. Under or near the structure, additional constraints may be imposed, such as a prescribed free surface, a pressure unknown, or a coupling with the motion of a rigid body. The resulting systems are therefore generally not pure conservation laws. They combine hyperbolic equations, source terms, coupling or free-boundary conditions, and sometimes elliptic or ordinary differential components. Such mixed structures partly explain why wave-structure interaction problems are numerically demanding [19, 20, 34, 125, 173, 174, 182, 227].

This point of view has motivated several recent works on NSW models with floating structures. For instance, the local well-posedness of an axisymmetric shallow-water floating body problem is studied in [34]. In [125], a congested shallow-water formulation is used to describe the coupling between a free surface flow and a moving floating body. Other recent studies focus on the mathematical and numerical structure of NSW models with

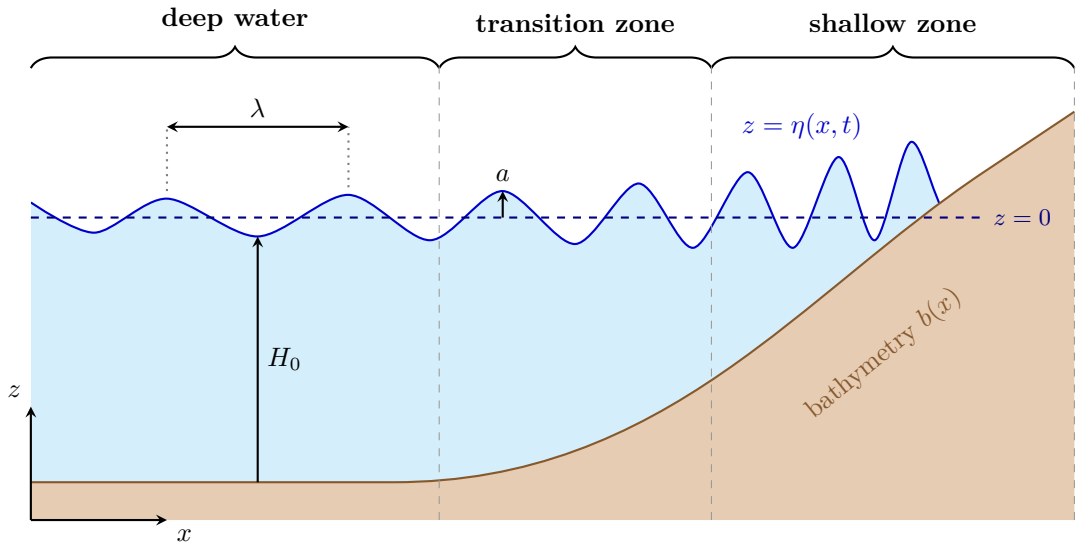


Figure 1: Illustration of the different coastal wave regimes over a varying bathymetry.

partially immersed obstacles, including the role of moving contact lines, constrained regions, and pressure coupling [19,20,186,187,251]. These contributions show that NSW models are also useful to provide a rigorous framework to understand the coupling mechanisms between the exterior flow, the interior constrained flow, and the solid motion.

For moving boundaries, one also has to choose how the computational domain is represented. A first possibility is to keep a fixed mesh and to describe the object inside it, as in immersed or embedded boundary methods. Another possibility is to move the mesh itself so that it follows the boundary of the fluid domain. This second point of view leads to Arbitrary Lagrangian–Eulerian (ALE) methods, in which the grid velocity is introduced explicitly in the equations. The ALE framework will be discussed later from the numerical side, but it is useful to already emphasize its role here: it provides a natural way to keep a body-fitted description of moving boundaries, while avoiding a purely Lagrangian motion of the mesh.

Only few works have been devoted to the development of ALE methods for shallow-water flows, and even less for wave-structure interaction problems [10,113,141,142,205]. This is also one of the motivations of the present thesis, which explores this direction in a high-order framework.

2. Numerical state-of-the-art for water-wave equations

The numerical approximation of NSW equations brings together several difficulties which are not easy to satisfy at the same time. The system is nonlinear and hyperbolic: steep gradients, shocks, hydraulic jumps and discontinuities appear in finite time. It also contains source terms, mainly due to bathymetry and friction as mentioned previously, which have to be discretized consistently with the fluxes. In addition, the water-height must remain non-negative, and steady states such as the lake at rest should be preserved by the numerical method. These requirements become even more difficult to achieve when one wants to use high-order approximations, unstructured meshes and moving domains, as in wave-structure interactions problems.

Ideally, one would like to build a numerical method *simultaneously* satisfying the following properties:

- *High-order accuracy* to capture smooth solutions and small-scale wave structures,
- *Efficient shock-capturing features* to handle discontinuities, hydraulic jumps and strong nonlinear effects,
- *Preservation of positivity* to ensure physical admissibility, i.e. the numerical solution stays in $\mathcal{H}_b^+$,
- *Well-balancing* to exactly preserve lake at rest steady states,
- *Efficient treatment of source terms* such as bottom topography and friction effects,

♯ *Suitability for unstructured meshes* to deal with complex geometries and realistic domains, possibly with moving boundaries.

Achieving all these properties in a single framework remains difficult, even with modern techniques. Classical methods often provide some of them, but usually at the price of losing others. Robust low-order methods are generally too diffusive, while high-order methods are accurate in smooth regions but require additional stabilization near discontinuities or dry fronts. One of the objectives of this thesis is to investigate a monolithic hybrid strategy which combines several of these properties in the same discretization framework, while keeping in mind that some trade-offs remain unavoidable.

Since the sixties, many numerical methods have been developed to approximate weak solutions of the NSW equations. These include finite volume methods (FV) [4, 8, 22, 111, 117, 272], finite element methods (FE) [18, 221, 231, 258], spectral methods [41, 138, 154, 172, 208, 228], and residual distribution methods [10, 239, 240]. We briefly recall below the main numerical ideas that are relevant for the present work, with a particular focus on the tension between robustness and high-order accuracy.

2.1 Finite volume methods

Godunov-type FV methods, introduced in the context of conservation laws and later widely used for shallow-water flows [4, 42, 257], are among the most classical discretizations for the NSW equations. They are relatively cheap, conservative by construction, and well adapted to discontinuous solutions. This makes them very efficient for sharp-fronted flows, transcritical regimes, dam-break problems, and wet-dry transitions. We refer for instance to [8, 29, 79, 200, 224, 241, 257] among many others. A large part of the FV literature for NSW equations is also devoted to the treatment of topography source terms and to the preservation of steady states, see for instance [11, 49, 59, 119, 170, 200, 206, 216, 223]. Early approaches to balancing the topographic source include the upwind discretization of Bermúdez and Vázquez [23] and the source-aware Riemann solvers of Le Roux [189], which treat the bottom level together with the flow variables (H, q) in an augmented interface problem. A distinct semi-implicit finite-difference line of work is represented by Casulli's two-dimensional shallow-water scheme [60].

On a cell ω_c, a semi-discrete FV scheme updates the mean-value $\overline{\boldsymbol{v}}_h^c$ through the formula

$$\frac{\mathrm{d}\overline{\boldsymbol{v}}_h^c}{\mathrm{d}t} = -\frac{1}{|\omega_c|} \sum_{\Gamma \subset \partial\omega_c} \ell_\Gamma \boldsymbol{\mathcal{F}}_\Gamma^*,$$

where $\boldsymbol{\mathcal{F}}_\Gamma^*$ is a numerical flux through the edge Γ of length ℓ_Γ, leading to natural conservation properties. For hyperbolic systems such as the NSW equations, the numerical flux is usually built from an approximate Riemann solver.

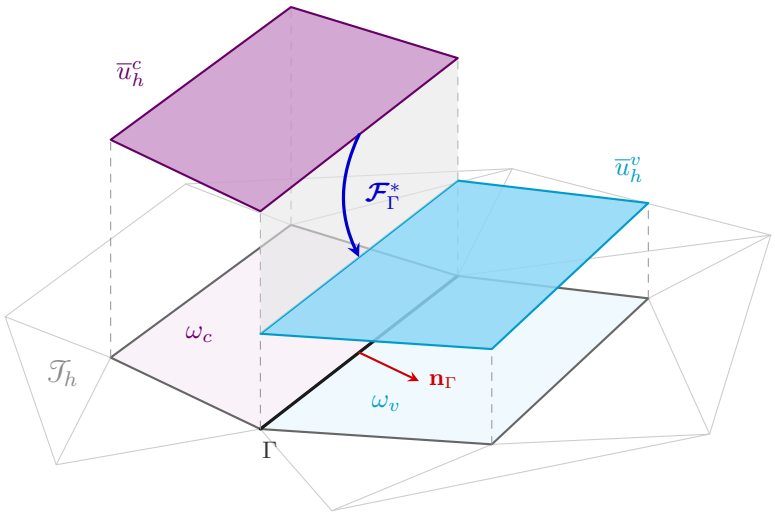


Figure 2: Schematic representation of a standard FV method on an unstructured mesh.

Several first-order FV schemes are robust for shallow-water computations: for an admissible initial condition $(H_0, \boldsymbol{q}_0) \in \mathscr{H}_b^+$, they preserve the positivity of the water-height under a classical CFL condition. FV schemes

are also naturally conservative and may satisfy discrete entropy or energy inequalities when the numerical flux is properly chosen.

One of the main drawbacks is accuracy; FV schemes are usually too diffusive, especially when smooth waves have to be propagated over long times or distances. To alleviate this, reconstruction techniques may be introduced, leading to second-order or higher-order FV methods, see for instance, among others, [26, 27, 31, 32, 48, 159, 191, 215–217, 234].

As an example, in the *Monotonic Upstream-centered Scheme for Conservation Laws* (MUSCL) approach, piecewise constant cell averages are replaced by piecewise linear reconstructions, allowing one to recover formal second-order accuracy in smooth regions while keeping the conservative FV structure. In practice, slope limiters are also used to avoid the creation of new extrema and to control oscillations near discontinuities. These reconstructions and limiters may classically reduce the accuracy near smooth extrema, and make the simultaneous preservation of positivity, well-balancing and high-order accuracy more challenging, but one of the main difficulties comes from the larger reconstruction stencils.

2.2 High-order finite element methods

From the late 1980s and especially during the 2000s, another strategy became increasingly important: using high-order methods, relying on the use of polynomials of arbitrary degree k inside each element.

Among these methods, the Discontinuous Galerkin (DG) methods have played a central role. The first DG method for first-order PDEs was introduced in 1973 [238], in the context of steady neutron transport. The first analysis for steady first-order PDEs was then given in 1974 [193, 194]. The error estimate is later improved in 1986 [165], with a convergence order of $k + 1/2$ in the L^2 norm, and a method for fluid flow problems [61, 62]. Around the same period, DG methods are extended to time-dependent hyperbolic PDEs in [65, 72, 73], using explicit Runge–Kutta time discretizations. A convergence proof toward the entropy solution is then obtained in [158].

On each element ω_c, the numerical solution is approximated by a local polynomial:

$$\boldsymbol{v}_{h|\omega_c}(\boldsymbol{x}, t) = \sum_{m=0}^{\dim \mathbb{P}^k} \underline{\boldsymbol{v}}_m^c(t)\varphi_m(\boldsymbol{x}),$$

where $\{\varphi_m\}_m$ denotes a local basis (for instance a Legendre basis or a Lagrange basis), and the coefficients $\{\underline{\boldsymbol{v}}_m^c\}_m$ are the local degrees of freedom.

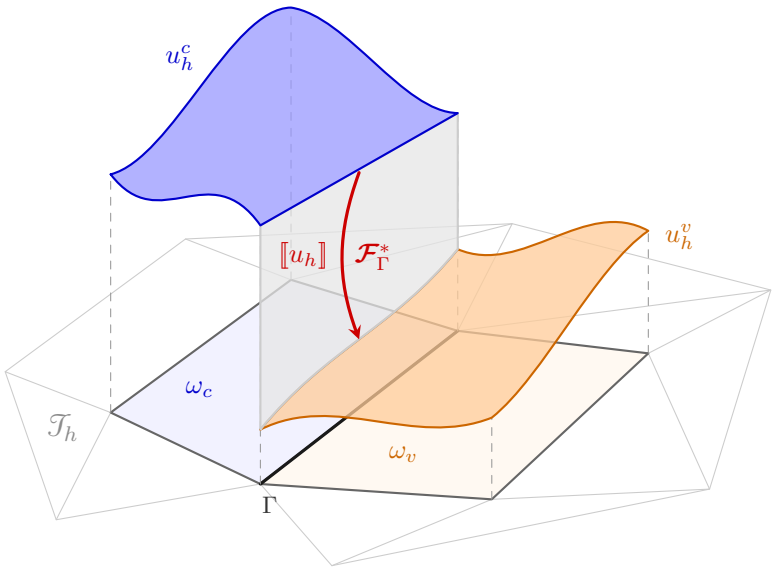


Figure 3: Schematic representation of DG approximation u_h across two neighboring control volumes ω_c and ω_v.

The benefits of DG are numerous: compact stencil, useful for parallel computing, boundary conditions imposed weakly through the numerical fluxes, natural framework for p-adaptivity, moving boundaries, or local mesh refinement, among others.

However, these advantages come with difficulties: DG methods are prone tothe so-called Gibbs phenomenon, leading to non-physical states and crashes of the computation. The corresponding positivity issue for the NSW equations may be summarized as follows: even if the cell-average $\overline{H}_h^c$ is positive, the underlying polynomial H_h may take negative values inside the element as a positive mean value gives almost no pointwise control on the extrema of the polynomial. Hence, extending positivity-preserving low-order constructions to high-order methods is a difficult task. In particular, [284,286] showed that classical high-order well-balanced reconstructions may lose the positivity property enjoyed by their first-order counterparts.

Several DG methods have been developed for the NSW equations since the early 2000s, see for instance [3,6,24, 112,123,169,171,196,220,244,244,253,280,282,283]. Their efficiency in smooth regimes is now well established, and their use for realistic shallow-water flows requires additional stabilization mechanisms, such as limiters, artificial viscosity, or subcell corrections.

2.3 Well-balancing and discrete positivity preservation

Specifically considering NSW equations, another important difficulty comes from the source terms discretization. For the non-homogeneous NSW equations, the numerical method has to approximate both the hyperbolic fluxes and the bathymetry source term. Formally, one may separate the system into

$$(\textit{hyperbolic part}) \quad \partial_t \boldsymbol{v} + \nabla_{\boldsymbol{x}} \cdot \mathbb{F}(\boldsymbol{v}, b) = 0, \qquad (\textit{source part}) \quad \partial_t \boldsymbol{v} = \boldsymbol{B}[b](\boldsymbol{v}),$$

but a steady state $\boldsymbol{v}_{\text{ss}}$ should satisfy the balance

$$\nabla_{\boldsymbol{x}} \cdot \mathbb{F}(\boldsymbol{v}_{\text{ss}}, b) = \boldsymbol{B}[b](\boldsymbol{v}_{\text{ss}}).$$

At the discrete level, this balance is not satisfied using a splitting strategy. More precisely, if the flux gradient and the source term are not approximated in a compatible way, the numerical residual does not vanish, even starting from an exact steady state.

The most classical example is the lake at rest equilibrium, given earlier in (2). This property is essential in long-time simulations, because even very small spurious currents may accumulate and pollute the solution. Several well-balanced FV schemes, such as [11], usually achieve this by reconstructing variables adapted to the equilibrium. Instead of reconstructing directly $(H, H\boldsymbol{u})$, one reconstructs quantities such as $(\eta, \boldsymbol{q})$ with the free-surface elevation $\eta = H + b$, which is constant at rest, and then recovers the conservative variables from this reconstruction.

This idea is very effective at low order, but it becomes more delicate in a high-order setting. If one reconstructs a high-order polynomial approximation of η and combines it with a high-order bathymetry representation, the resulting water-height $H_h = \eta_h - b_h$ may become negative at some points, even if the cell-average remains positive. Thus, the well-balanced reconstruction may conflict with the positivity requirement. This is one of the main reasons why designing high-order schemes for the NSW equations is difficult: one has to preserve steady states, keep the solution admissible, and still retain high-order accuracy away from singular regions.

For this reason, many modern shallow-water schemes are designed relying on a combination of three constraints: conservation, well-balancing, and positivity. Conservation guarantees the correct exchange of mass and momentum between cells. Well-balancing prevents the creation of artificial waves over non-flat topography. Positivity ensures that the numerical solution remains physically admissible. These three properties are not independent, and this is precisely what makes the construction of robust high-order schemes highly non-trivial.

2.4 Limiting strategies: the *a priori* paradigm

Because of these difficulties, high-order schemes usually need some form of limiting or stabilization. The first classical family of methods follows an *a priori* strategy: before computing the next time step, the polynomial solution is inspected inside each element, and it is modified if it is considered dangerous. In practice, this decision is based on a troubled-cell detector, namely a criterion intended to identify cells where oscillations, loss of admissibility, or large non-physical variations may appear [236]. The difficulty is that this detection is done before the new solution is actually known, so that one has to decide in advance where the high-order approximation should be corrected.

A first example is given by slope limiters [33, 71, 72]. Their principle is to reduce the high-order part of the polynomial approximation in a cell where the solution is detected as troubled. For instance, if $\overline{\boldsymbol{v}}_h^c$ denotes the

cell-average of the conservative variables in an element ω_c, and if $\boldsymbol{v}_h^c$ is the original high-order polynomial, a simple limited polynomial can be written as

$$\widehat{\boldsymbol{v}}_h^c(\boldsymbol{x},t) = \overline{\boldsymbol{v}}_h^c(t) + \Theta\left(\boldsymbol{v}_h^c(\boldsymbol{x},t) - \overline{\boldsymbol{v}}_h^c(t)\right),$$

where $\Theta \in [0,1]$ is chosen so as to reduce the oscillations and keep the polynomial within an acceptable range. This construction preserves the cell-average, which is essential for conservation, but it also removes part of the high-order content of the approximation. This is useful near discontinuities, but it can become too diffusive near smooth extrema. Moreover, the distinction between a true oscillation and a smooth peak is often delicate and may require additional thresholds or user-defined parameters [71, 72, 236].

Moment limiters follow a similar idea, but act more directly on the polynomial coefficients [33, 45, 90, 156, 157, 175, 197, 288]. Instead of limiting only a slope, one progressively damps the moments associated with higher-order modes. This gives a more flexible way to stabilize multi-dimensional high-order polynomials, since one can reduce the high-frequency part of the approximation while keeping the low-order structure. However, the method still relies on decisions about which modes should be modified and how strongly they should be damped. It is therefore effective, but not completely automatic [33, 45, 175, 288].

Another important family is *Weighted Essentially Non-Oscillatory* (WENO) limiting [17, 97, 160, 199, 204, 237, 294, 295]. In this case, the troubled DG polynomial is not simply scaled. It is replaced by a new polynomial obtained from a nonlinear reconstruction based on several neighboring stencils. The WENO weights are designed to give more importance to smooth stencils and less importance to stencils crossing discontinuities. This makes WENO-type limiters very effective for shock-capturing. The price to pay is a heavier implementation, especially on unstructured meshes, where the construction of stencils, reconstruction matrices, and nonlinear weights is more involved [17, 237, 294, 295].

Artificial viscosity methods take another route [114, 131, 172, 212, 229, 260]. Instead of modifying the polynomial directly, they add a dissipative term to the numerical operator. This additional viscosity is usually small or zero in smooth regions, and becomes active near shocks or strong gradients. Such methods can be quite robust since viscosity damps the non-physical oscillations generated by the high-order approximation. However, they also introduce extra diffusion, and the amount of viscosity must be scaled carefully with the mesh size and the polynomial degree k. If it is too small, oscillations remain; if it is too large, shocks and contact waves are smeared [131, 229, 260].

All these *a priori* strategies have been very useful, but they share some limitations. First, most of them require tuning parameters, such as troubled-cell thresholds, TVB constants, or viscosity coefficients, and these parameters may depend on the test case [72, 229, 236]. Second, they may activate in smooth regions containing local extrema, where the solution is regular but the detector interprets the variation as dangerous [33, 71, 72, 175]. Third, when they are applied to well-balanced reconstructions, they may destroy the precise cancellation between fluxes and source terms, unless additional constraints are imposed [107, 224]. Finally, classical limiters usually act on the whole polynomial in a troubled element, even when the problematic region is much smaller than the cell itself [55, 76, 148, 249, 264]. This may remove high-order information in parts of the element where the solution was actually smooth.

2.5 Limiting strategies: the *a posteriori* paradigm

A different point of view is provided by *a posteriori* correction methods. Instead of deciding before the time step which cells should be limited, one first computes a candidate high-order solution at time t^{n+1}. This candidate is then checked against some admissibility criteria. If it is admissible, it is accepted. If not, a more robust local correction is applied [70, 88, 89]. This changes the logic of the method: the detector no longer tries to predict where the high-order scheme may fail, but checks whether it has actually failed after the update.

The *Multi-dimensional Optimal Order Detection* (MOOD) method is one of the most representative examples of this philosophy [28, 70, 88, 89, 89]. Starting from a high-order candidate solution, the method checks whether the computed values satisfy prescribed admissibility properties, such as physical bounds, discrete maximum principles, or positivity constraints. If the candidate is accepted, the high-order update is kept. If it is rejected, the local approximation is recomputed with a lower-order scheme, possibly by decreasing the polynomial degree from $\mathbb{P}^k$ down to $\mathbb{P}^0$, until an admissible solution is obtained. The main advantage is that the correction is only applied where the candidate solution has effectively failed. Moreover, the criteria can be formulated in terms of physical admissibility, rather than through user-tuned smoothness thresholds [70, 88, 89]. This philosophy has also been extended to more complex shallow-water models, including dispersive or turbulent regimes [46, 47, 47].

Closely related are FV-subcell shock-capturing methods [76, 101, 103, 118, 153, 213, 243, 249, 264]. The idea is simple from a geometric point of view: when a DG element is detected as troubled, it is subdivided into smaller subcells, and a robust FV scheme is applied on this local subgrid. The resulting subcell averages are then used to reconstruct or update the DG polynomial. In this way, the low-order method is only used where robustness is needed, while the high-order DG approximation is kept elsewhere. This construction is attractive because FV schemes are naturally conservative and can be designed to preserve positivity [140, 153, 213]. It also gives a more surgical correction than classical element-wise limiting, since the troubled structure can be resolved at the subcell scale [101, 103, 153].

These subcell correction ideas have also played an important role in the development of *Arbitrary High Order DERivatives Discontinuous Galerkin* (ADER-DG) schemes [40, 46, 104, 105, 291]. Contrary to RK-DG methods, where high-order accuracy in time is obtained through several Runge–Kutta stages, ADER methods use a local space-time predictor and produce a one-step update: one first constructs, inside each element ω_c, a local space-time polynomial predictor

$$\boldsymbol{v}_h^{c,\star} \in \mathbb{P}^k(\omega_c) \otimes \mathbb{P}^l([t^n, t^{n+1}]),$$

which approximates the conservative state inside $\omega_c \times [t^n, t^{n+1}]$. This predictor is computed locally, without communication with neighboring elements. The coupling between elements only appears afterward, through the numerical fluxes in the one-step DG update

$$\int_{\omega_c} \boldsymbol{v}_h^{n+1}\varphi_h \,\mathrm{d}\boldsymbol{x} = \int_{\omega_c} \boldsymbol{v}_h^{n}\varphi_h \,\mathrm{d}\boldsymbol{x} - \int_{t^n}^{t^{n+1}} \left\{ \int_{\partial\omega_c} \mathbb{F}^*((\boldsymbol{v}_h^{c,\star})^-, (\boldsymbol{v}_h^{c,\star})^+)\varphi_h \,\mathrm{d}s - \int_{\omega_c} \mathbb{F}(\boldsymbol{v}_h^{c,\star}) \cdot \nabla_{\boldsymbol{x}}\varphi_h \,\mathrm{d}\boldsymbol{x} \right\} \mathrm{d}t.$$

This framework fits naturally with *a posteriori* correction. One computes a high-order ADER-DG candidate solution $\boldsymbol{v}_h^{*,n+1}$, checks its admissibility, and replaces it locally by a robust FV subcell update if the test fails. In smooth regions, the compact high-order ADER-DG update is retained; near shocks or dry fronts, the finite volume subcell scheme provides the required robustness.

A more recent variant, closer to the point of view followed in this thesis, aims at making the subcell correction even more localized [46, 140, 264, 267]. Instead of replacing the whole element by a subcell finite volume evolution, one identifies more precisely where the correction is needed at the subcell level. The objective is to modify only the problematic subregions while leaving the high-order polynomial information unchanged elsewhere. This makes it possible to preserve fine wave structures and reduce unnecessary diffusion, especially when the shock or wet-dry transition occupies only a small part of the element [140–142, 264, 267].

The main advantage of the *a posteriori* paradigm is therefore its selectivity. The high-order scheme is used by default, and the robust correction is introduced only when the computed candidate solution violates the admissibility criteria. If the fallback scheme satisfies a maximum principle or a positivity property, then the corrected method can inherit this robustness in the troubled regions [70, 88, 89, 101, 103, 153]. However, these methods also have some drawbacks. These are harder to implement than the classical high-order schemes, and the recomputation of troubled cells may increase the cost, especially when many cells are rejected. The projection between DG polynomials and subcell averages must also be handled carefully, otherwise oscillations may be reintroduced. Finally, well-balancing is not automatic in such correction procedures and must still be built into the numerical fluxes, reconstructions, or source-term discretization.

2.6 Monolithic approaches

The idea behind monolithic limiting is rooted in older flux-correction strategies. In the classical *Flux-Corrected Transport* (FCT) approach of Boris and Book and its multidimensional extension by Zalesak, a robust low-order transport step is corrected by limited antidiffusive fluxes, so that one recovers higher accuracy while keeping non-oscillatory or positivity properties [38, 290], see also the comprehensive overview provided in the book by Kuzmin et al. [180]. This point of view was later reformulated in more abstract finite element settings, in particular through the invariant-domain preserving and convex limiting framework of Guermond, Popov and Tomas [133]. In a different but closely related direction, the work of Carpenter et al. on entropy stable SBP and spectral collocation discretizations showed that stability can also be designed directly at the operator level, by choosing fluxes and interface terms which satisfy discrete conservation and entropy estimates [53, 115].

Monolithic approaches have recently been proposed as another way to get the best of both worlds between accuracy and robustness. The idea is not to compute a high-order solution and then limit it afterwards, nor to replace it locally after a failed admissibility check. Instead, the high-order and low-order contributions are coupled directly in the discrete operator. In practice, this often amounts to writing the numerical update as

a convex combination between a high-order flux, which gives accuracy in smooth regions, and a more robust low-order flux, which is used to enforce properties such as positivity, maximum principles, or entropy stability. The stabilization is therefore not added as an external correction, but is rather built directly into the flux or residual itself.

The local subcell monolithic DG/FV method introduced in [265] follows this idea. The starting point is a FV-like interpretation of a high-order DG scheme on an underlying subcell partition. At each subcell interface, the method blends a reconstructed DG subcell flux, which represents the high-order contribution at the subcell level, with a first-order FV subcell flux, which is more diffusive but satisfies stronger admissibility properties. In a schematic form, one may view the resulting flux as the convex combination defined on each subcell interface Γ_{mp} by

$$\widetilde{\mathbb{F}}_{mp} = \mathbb{F}^{\mathrm{FV}}_{mp} + \Theta_{mp}(\mathbb{F}^{\mathrm{DG}}_{mp} - \mathbb{F}^{\mathrm{FV}}_{mp}),$$

where the coefficient $\Theta_{mp} \in [0, 1]$ is chosen locally so that the updated subcell state remains in the prescribed admissible set. This gives a way to keep conservation and high-order accuracy where possible, while adding just enough low-order correction where the high-order update would become unsafe. A major point of [265] is that this can be done in a fully *a priori* manner, without MOOD-type recomputations and without relying on heuristic smoothness indicators.

Similar ideas have also been developed in other high-order frameworks. In [179, 219], continuous Galerkin methods are combined with subcell flux limiting techniques in order to enforce positivity and local maximum principles by convex limiting of the fluxes. In the DG spectral element context, related constructions have been proposed in [201, 242, 243, 281]. There, the use of *Summation-by-Parts* (SBP) quadrature and Gauss–Lobatto nodes gives a natural set of nodal or subcell degrees of freedom, and the limiting is again inserted directly into the semi-discrete operator, often through graph-based diffusion or flux correction terms. These methods share the same general objective: keep the good accuracy and low dissipation of high-order schemes in smooth regions, while forcing the update to remain in a physically admissible set near steep gradients or discontinuities.

There are nevertheless important differences between these approaches. Many *SBP-DG Spectral Element methods* (SBP-DGSEM) are naturally built on tensor-product elements and Gauss–Lobatto collocation, which makes them very efficient on structured or Cartesian-type meshes, but less direct on fully unstructured grids [201, 242, 243]. Some extensions to triangular elements exist, for instance [281], but the low-order correction is then often interpreted through graph operators acting on the degrees of freedom rather than through a genuine geometric subcell mesh. By contrast, the framework of [265, 267] keeps an explicit subcell structure. This is useful for the present work, because subcell averages and subcell fluxes have a clear FV meaning, and the construction can be applied on general unstructured meshes and arbitrary cell subdivisions. This also makes the approach well suited to NSW systems, where positivity, conservation, entropy stability, source terms and dry fronts may have to be handled all together.

From this point of view, monolithic subcell methods can be seen as an attempt to reduce the gap between robust first-order FV schemes and accurate but fragile high-order DG or Continuous Galerkin (CG) methods. The low-order scheme is no longer used only as a last-resort correction. It is part of the numerical method from the beginning, and its contribution is controlled locally through blending coefficients. This gives a flexible way to prescribe the admissible set, to adapt the numerical diffusion in space and time, and to keep most of the high-order approximation untouched when no correction is needed. Of course, several questions remain open, such as the best choice of subcell partitions on complex meshes, the treatment of strongly coupled source terms, or the extension to implicit and semi-implicit solvers. In the present manuscript, this monolithic point of view is used as the main numerical basis for NSW models with bathymetry, friction, wet-dry interfaces, and moving domains.

2.7 Moving domains and the ALE framework

All the previous difficulties already appear on fixed meshes. In wave-structure interactions problems, one often has to face an additional issue: the computational domain itself may change in time. This happens, for example, when a floating body moves, when a boundary is prescribed as a wavemaker, or when an immersed or partially immersed object imposes a moving interface on the fluid domain. A first possible approach is to keep a fixed mesh and represent the object inside it. This is the general philosophy of immersed or embedded boundary methods. These methods are attractive when the geometry changes a lot, or when remeshing would be too expensive, because the moving boundary is handled inside a fixed computational domain; see [235] for the immersed boundary method and [248] for the shifted boundary method.

Another possibility is to move the mesh itself so that it follows the boundary of the fluid domain. This is the idea of the Arbitrary Lagrangian–Eulerian (ALE) framework. The ALE description lies between an Eulerian method, where the mesh is fixed, and a Lagrangian method, where the mesh moves with the material velocity. In a purely Lagrangian shallow-water method, the grid velocity is the fluid velocity, which gives a natural tracking of material interfaces but may quickly produce strong mesh distortion [262, 263, 266, 268, 269]. In an ALE method, the grid velocity is chosen independently of the fluid velocity. This gives more freedom: the mesh can follow the moving boundary, while its interior motion can be designed to preserve a reasonable mesh quality [91, 92, 146, 149].

Historically, ALE methods were first introduced in a finite difference setting by Hirt, Amsden and Cook [146], and then developed in finite element formulations for incompressible flows and fluid-structure interaction problems [91, 93, 149]. General presentations can be found in [92]. One often distinguishes indirect and direct ALE strategies. In an indirect ALE method, one first performs a Lagrangian step, then remeshes and projects the solution onto the new mesh [25, 176, 202]. In a direct ALE method, the grid velocity is introduced directly in the governing equations, typically through a modified numerical flux [39, 102]. This direct form is particularly relevant here, because the motion of a floating or immersed structure can be translated into a mesh velocity and inserted directly into the shallow-water fluxes.

In the ALE framework, the physical position is usually written as a time-dependent mapping from a reference coordinate, $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{\xi}, t)$, and the grid velocity is defined by $\boldsymbol{w}^g(\boldsymbol{x}, t) = \partial_t \boldsymbol{x}(\boldsymbol{\xi}, t)$. For a conservation law, this leads to a modified flux in which the physical transport is corrected by the motion of the grid. In a schematic form, the ALE flux is

$$\mathbb{G}(\boldsymbol{v}) = \mathbb{F}(\boldsymbol{v}) - \boldsymbol{w}^g \otimes \boldsymbol{v}.$$

This formula is one of the main reasons why ALE methods are attractive for hyperbolic problems: the mesh motion can be included directly at the flux level, while keeping a conservative structure on the moving control volumes.

For shallow-water equations, ALE-DG methods are attractive because they combine high-order accuracy, geometric flexibility, and a natural weak treatment of boundary conditions. However, moving meshes introduce an additional consistency constraint: the numerical method must not create artificial waves only because the grid is moving. This is the role of the Geometric Conservation Law (GCL), which expresses the compatibility between the time variation of the cell volumes and the flux of the grid velocity through their boundaries [136, 254]. At the continuous level, this compatibility can be written as

$$\partial_t \mathsf{J} = \mathsf{J} \nabla_{\boldsymbol{x}} \cdot \boldsymbol{w}^g,$$

where J is the Jacobian of the ALE mapping. If the corresponding property is not respected at the discrete level, even a uniform or steady solution may be perturbed by the mesh motion. This issue is well known in ALE computations and is also connected to time stability [205, 233]. In the shallow-water context, Zhang et al. [293] proposed positivity-preserving and well-balanced ALE-DG schemes able to preserve still-water and moving-water equilibria, under suitable discrete GCL conditions.
The remaining difficulty is that the flows considered in this manuscript are not only smooth moving-boundary problems. They may contain shocks, hydraulic jumps, dry fronts, and strong interactions with topography or structures. Standard space-time or DG-ALE methods can be very effective for smooth deforming-domain problems [259–261], but additional stabilization is needed when the solution becomes strongly nonlinear. This is where the ALE framework naturally meets the subcell correction ideas discussed above: the moving-mesh formulation provides the correct geometric description of the domain, while the local DG/FV subcell mechanism gives the robustness needed near discontinuities and dry zones.

This last point is central for the present work. The objective is not only to combine a high-order shallow-water solver with an ALE mapping, but to do so in a way that preserves the main properties required on fixed meshes: conservation, positivity, well-balancing, and shock-capturing. In other words, the ALE extension should not be viewed as a separate technical layer but rather as a part of the same numerical problem. The mesh motion changes the geometry, the fluxes, the stability condition and the conservation properties, and it must therefore be coupled consistently with the monolithic DG/FV subcell discretization.

3. Structure of the manuscript

The manuscript is organized in two main parts. The first one is devoted to the construction of monolithic DG/FV subcell schemes for shallow-water type models, while the second one focuses on wave-structure interactions problems.

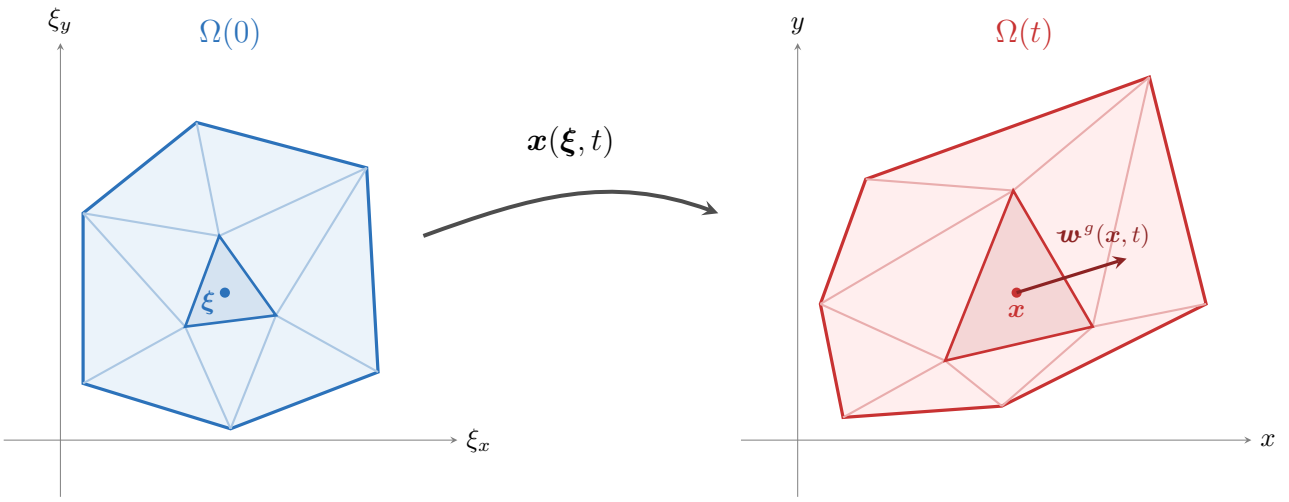


Figure 4: Schematic representation of the Arbitrary Lagrangian–Eulerian (ALE) mapping on a simplicial unstructured mesh. The fixed reference domain $\Omega(0)$ (left) with coordinates $\boldsymbol{\xi}$ is mapped continuously in time onto the moving physical domain $\Omega(t)$ (right) with coordinates $\boldsymbol{x}(\boldsymbol{\xi}, t)$. The mapping purely displaces the nodes while preserving straight edges, generating a deformed triangular element (red) whose trajectory is tracked by the grid velocity $\boldsymbol{w}^g$.

Part I. A new high-order framework for shallow-water asymptotics

This contains the first group of contributions and is devoted to the design of numerical schemes for shallow-water type equations. These led to two submitted publications [51, 52].

- Chapter 1 starts with the one-dimensional NSW equations. This chapter introduces the monolithic DG/FV subcell strategy, built step by step, with a particular attention to topography source terms, positivity and well-balancing. This chapter plays the role of an entry point for the whole numerical part of the thesis.

- Chapter 2 then considers the one-dimensional Serre-Green-Naghdi equations. This model adds dispersive effects through an elliptic correction. The numerical strategy combines the monolithic DG/FV method for the hyperbolic part with a SWIP-DG discretization for the elliptic problem. This chapter is shorter, but it introduces an important idea used later in the manuscript: the monolithic subcell correction can be viewed as an operator acting on the hyperbolic part of a larger coupled system, and can be coupled easily with auxiliary problems. This point will be useful when the fluid equations are coupled with elliptic contributions in the wave-structure interactions chapters.

- Chapter 3 extends the monolithic DG/FV framework to the two-dimensional NSW equations on unstructured meshes. This is a natural but non-trivial step, since the subcell geometry, the flux reconstruction and the admissibility constraints have to be treated in a fully multidimensional setting. The chapter shows that the same philosophy can be kept in two space dimensions, while preserving the main structure of the one-dimensional method. Several test cases are used to check the robustness of the scheme in more demanding configurations.

- Chapter 4 introduces an Arbitrary Lagrangian–Eulerian extension of the previous method. The objective is to prepare the numerical framework for moving-domain problems. The chapter recalls the basic ALE formulation, the role of the grid velocity, and the importance of the geometric conservation law. A particular point is the construction of the mesh velocity, which is also obtained through an auxiliary elliptic problem. The resulting ALE DG/FV subcell scheme is then tested on moving meshes. This chapter also gives a first numerical basis for future simulations involving moving floating structures.

Part II. Waves interacting with floating structures in shallow-water regimes

The second group of contributions is devoted to wave-structure interactions problems. The point of view is different from the previous part: the shallow-water equations are no longer considered alone, but are coupled with a constrained region below or inside a structure. This leads naturally to mixed systems, involving hyperbolic equations in the exterior fluid region and elliptic problems in the interior region. The results displayed in this part will be gathered in a set of two future publications.

♯ Chapter 5 studies a one-dimensional floating structure model. The exterior fluid is governed by shallow-water equations, while the region below the object is described through an elliptic pressure equation. The structure may have a prescribed motion or may move according to Newton's second law. The numerical method combines the monolithic DG/FV scheme for the exterior fluid and a SWIP-DG method for the elliptic pressure problem. This chapter makes several links with the previous numerical chapters: the shallow-water part relies on Chapter 1, while the elliptic discretization uses tools introduced in Chapter 2. The numerical tests also compare different implementations and recover results consistent with the literature. An extension to pseudo-one-dimensional geometries is also presented, relying on two-dimensional monolithic DG/FV scheme and Hybrid High-Order (HHO) methods depicted in Chapter 6.

♯ Chapter 6 finally moves to genuinely two-dimensional wave-structure interactions models. The structure is now fixed, but the geometry is no longer reduced to a one-dimensional setting. Two types of objects are considered: a straight-edged pontoon and a cylindrical obstacle. In this case, the coupling requires a numerical approximation of the interior elliptic problem. Depending on the geometry, this is done either through analytic or Fourier-based tools, or through a HHO method. The chapter shows how the DG/FV subcell method for the exterior flow introduced in Chapter 3 can be coupled with this interior elliptic solver. The test cases are designed to check that the method gives coherent wave reflection, transmission and redistribution around the structure, both on flat and non-flat bottoms.

Some reading suggestions

The manuscript has been organized so that it can be read progressively, but also in a more selective way depending on the interests of the reader. Each contribution chapter (except Chapter 1) uses tools introduced earlier in the manuscript, and the structure is close to a sequence of papers building on previous ones. In this sense, Part I can be read as the numerical core of the thesis. Readers mainly interested in the monolithic DG/FV subcell method may focus first on the one-dimensional shallow-water chapter, where the main ideas are introduced in the simplest setting, and then move to the two-dimensional extension on unstructured meshes. The Serre-Green-Naghdi chapter can be read as a first extension towards a coupled hyperbolic-elliptic problem, while the ALE chapter should be read after the two-dimensional shallow-water chapter, since it relies on the same fixed-grid discretization before adding mesh motion.

Part II is more hybrid in nature. It may be read by colleagues interested in modeling issues related to wave-structure interaction, but also by readers interested in the numerical coupling between hyperbolic equations and elliptic problems. The first chapter of this part introduces the pseudo-one-dimensional floating-object model and uses several numerical tools already presented in Part I. The second chapter then moves to a genuinely two-dimensional setting, where the contact line is a curve and where the elliptic part is treated with HHO methods. For this reason, Part II is written to be accessible both from the modeling side and from the numerical analysis side, while still keeping clear links with the previous numerical chapters.

Several Appendices complete the manuscript. They gather technical details and additional material that are useful for the work, but would interrupt the flow of the main text. In particular, one appendix is devoted to implementation aspects: it describes some of the data structures and coding choices used to implement DG methods, subcell methods, and ALE subcell schemes. This appendix was included because the implementation is a non-negligible part of this thesis, and so we wanted to give it a proper place in the manuscript. It is of course not designed to give an *optimal way* of coding these methods, but mainly meant to be useful for readers who would like to build similar solvers, or simply understand how these numerical methods are translated into code. The reader will also find complementary remarks on the HHO scheme developed in the thesis, as well as the derivation of an additional model for heaving structures. The appendices further include some entropy-related results and a final collection of background material gathering the continuous notation, the main functional spaces, the differential operators, and several useful analytical results used throughout the manuscript.

Finally, let us remark that some notations are local to each chapter, especially for the numerical discretizations. This choice was made to keep each chapter readable and to avoid forcing a single notation on models that are not exactly identical. However, the continuous notation and the main mathematical objects introduced in Appendix G are kept consistent throughout the manuscript. Some test cases also contain links to animations of the numerical simulations whenever possible, in order to give a more direct view of the dynamics shown in the figures. All the presented simulations are also available directly on the following link: sachacardonna.github.io/codes/.



FROM: ***"Travelers"***, Andrew Prahlow – *Outer Wilds Original Soundtrack* (2019)

Part I

A new high-order framework for shallow-water asymptotics

You'll have to trust me, just don't believe me
Reap what you sow, Comeback Kid, says hello
FROM: ***"Comeback Kid"***, Kasabian – *For Crying Out Loud* (2017)

CHAPTER 1

An introduction to subcell monolithic DG/FV scheme on one-dimensional nonlinear shallow-water equations

The main results presented here, gathered with the extension proposed in Chapter 2, have been submitted [51].

This chapter introduces the first building block of the numerical framework developed in this thesis: a high-order local subcell monolithic DG/FV discretization for the one-dimensional nonlinear shallow-water equations with bathymetry source terms. Starting from the homogeneous system, the construction is progressively enriched in order to handle non-flat topographies, source terms, dry or nearly dry configurations, and lake-at-rest steady states. This one-dimensional setting is deliberately used as a reference framework: it allows us to expose the main mechanisms of the method in a clear form before extending them to more complex models and geometries in the following chapters.

The main contribution is the design of a convex property preserving subcell strategy in which the high-order DG update is locally recast as a finite-volume balance on subcells. At each subcell interface, high-order reconstructed fluxes are blended with robust low-order finite-volume fluxes through locally computed coefficients. This *a priori* convex blending is designed to enforce essential admissibility properties, and in particular the strict positivity of the water-height, while recovering the high-order DG scheme in smooth regions. A second important contribution concerns the discretization of the bathymetry source term. Since well-balancing is not automatic in the original subcell formulation, a dedicated two-scale hydrostatic reconstruction is introduced, combining reconstructed states at both subcell and cell levels. This allows the method to exactly preserve the lake-at-rest equilibrium while retaining the local subcell structure of the scheme. The numerical experiments then assess convergence, robustness near discontinuities, positivity preservation and the ability of the method to compute nontrivial steady and unsteady shallow-water configurations without spurious oscillations.

Contents

1.1 Continuous and discrete settings 20
1.1.1 General shallow-water model in one space dimension 20
1.1.2 Discrete setting 21
1.2 Local-subcell monolithic DG/FV for homogeneous system 25
1.2.1 A hierarchy of numerical fluxes 25
1.2.2 DG and local-subcell DG/FV equivalency 27
1.2.3 Local subcell monolithic method and blending coefficients 27
1.3 Local-subcell approximations of bathymetry source term 30
1.4 Numerical validations 36
1.4.1 Smooth sinusoidal solutions 36
1.4.2 Dam-break over flat bottoms 38
1.4.3 Well-balancing with discontinuous bathymetry 42
1.4.4 Transcritical flow over a bump 42
1.4.5 Carrier & Greenspan's solutions 44
1.4.6 Periodic oscillations in a parabolic basin 47
1.4.7 Vacuum occurrence wave over a step 51

Because life is so brief and time is a thief when you're undecided
And like a fistful of sand, it can slip right through your hands
FROM: ***"Young Turks"***, Rod Stewart – *Tonight I'm Yours* (1981)

1.1 Continuous and discrete settings

In this section, we introduce the continuous and discrete settings for the approximation of nonlinear shallow-water equations with source terms, which will be used in the following sections to introduce the proposed local subcell monolithic DG/FV method. We also introduce some useful notations and definitions that will be used throughout the chapter.

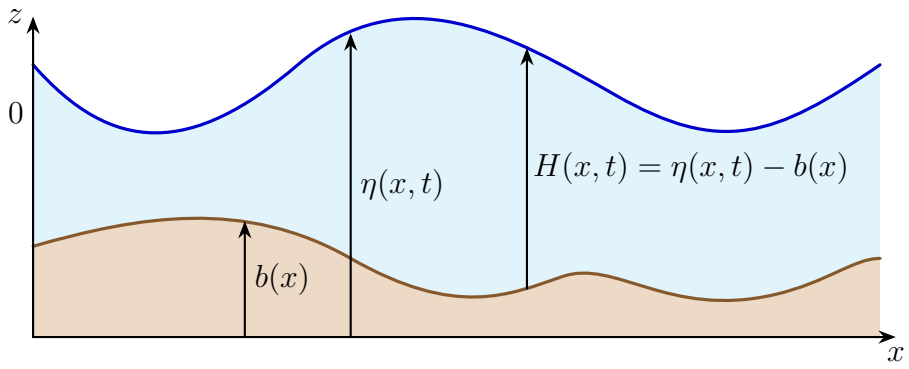


Figure 1.1: Main notations for prebalanced shallow-water equations. The free-surface elevation is denoted by η, the bathymetry by b and the water-height by H. The horizontal velocity is denoted by u and the horizontal discharge by q.

1.1.1 General shallow-water model in one space dimension

Given a smooth parameterization b of the bathymetry variations, denoting by H the water-height, u the horizontal (depth-averaged) velocity and $q = Hu$ the horizontal discharge, and defining the total free-surface elevation as $\eta := H + b$, see Figure 1.1, we consider the classical nonlinear shallow-water model which may be commonly expressed as systems of partial differential equations:

$$\begin{cases} \partial_t \eta + \partial_x q = 0, \\ \partial_t q + \partial_x \left(uq + \frac{1}{2}\mathrm{g}(\eta^2 - 2\eta b) \right) = S[\tau](\boldsymbol{v}) + \mathfrak{D}[H,b](\boldsymbol{v}), \end{cases}$$

or in a more compact form

$$\partial_t \boldsymbol{v} + \partial_x \boldsymbol{F}(\boldsymbol{v}, b) = \boldsymbol{S}[\tau](\boldsymbol{v}) + \boldsymbol{\mathfrak{D}}[H,b](\boldsymbol{v}), \tag{1.1}$$

supplemented with some initial-data $\boldsymbol{v}(\cdot,0) := \boldsymbol{v}^0$ and where:

- $\mathbb{R} \times \mathbb{R}_+ \ni (x,t) \mapsto \boldsymbol{v}(x,t) := (\eta, q)(x,t) \in \mathscr{H}_b^+$ gathers the chosen flow description variables, with
$$\mathscr{H}_b^+ = \{(\eta, q) \in \mathbb{R}^2 \mid H := \eta - b \geq 0\},$$
- $\mathscr{H}_b^+ \times \mathbb{R} \ni (\boldsymbol{v}, b) \mapsto \boldsymbol{F}(\boldsymbol{v}, b) := \left(q,\ uq + \frac{1}{2}\mathrm{g}(\eta^2 - 2\eta b)\right)^\top \in \mathbb{R}^2$ is the (nonlinear) flux function, given here in the pre-balanced form,
- $\mathscr{H}_b^+ \times \mathbb{R} \ni (\boldsymbol{v}, \tau) \mapsto \boldsymbol{S}[\tau](\boldsymbol{v}) := (0, S[\tau](\boldsymbol{v}))^\top \in \mathbb{R}^2$ may stand for any algebraic source term, generally relying on some external data, denoted by τ to keep a general formulation, and which may be, in the following, composed of:
  - a topography-related source term $\boldsymbol{B}[b](\boldsymbol{v}) := (0, -\mathrm{g}\eta\, \partial_x b)^\top$,
  - a pressure-related source term $\boldsymbol{P}[\underline{\mathrm{p}}](\boldsymbol{v}) := \left(0, -\frac{1}{\rho} H \partial_x \underline{\mathrm{p}}\right)^\top$, with $\underline{\mathrm{p}}$ the pressure at the free-surface, ρ the fluid density,
  - or some nonlinear bed-friction stresses term of the form $\boldsymbol{R}[\mathrm{n_f}](\boldsymbol{v}) := \left(0, -\mathrm{n_f}^2 \frac{q|q|}{(\eta - b)^\gamma}\right)^\top$, where $\mathrm{n_f}$ is a suitable drag coefficient modeling the bottom-roughness.

Note also that other algebraic source terms accounting for instance for Coriolis effects or surface stresses may also be addressed, possibly with some *ad-hoc* adaptation,

– $\mathscr{H}_b^+ \ni \boldsymbol{v} \mapsto \mathfrak{D}[H,b](\boldsymbol{v}) := (0, \mathscr{D}[H,b](\boldsymbol{v}))^\top \in \mathbb{R}^2$ stands for higher-order differential operator acting on the flow variables $\boldsymbol{v}$, which may be used to account for some weakly dispersive effects. In our work, we investigate the later case by considering the Serre-Green-Naghdi (SGN) model by setting:

$$\mathscr{D}[H,b](\boldsymbol{v}) := -(1+\mathscr{T}[H,b])^{-1}\mathscr{Q}[H,b](\boldsymbol{v}) + \mathrm{g}H\partial_x\eta, \tag{1.2}$$

where $\mathscr{T}[H,b]$ is a linear operator (with nonlinear terms with respect to H and its derivatives) defined for any $w \in \mathsf{H}^2(\mathbb{R})$, as follows:

$$(1+\mathscr{T}[H,b])w := \Big(H + \mathcal{S}_1[H,b]^\star H \mathcal{S}_1[H,b] + \mathcal{S}_2[b]^\star H \mathcal{S}_2[b]\Big)\frac{w}{H}.$$

In this previous definition, $\mathcal{S}_1[H,b]^\star$ and $\mathcal{S}_2[b]^\star$ stand for the respective adjoint operators of:

$$\mathcal{S}_1[H,b]w := \frac{H}{\sqrt{3}}\partial_x - \frac{\sqrt{3}}{2}\partial_x b, \qquad \mathcal{S}_2[b]w := \frac{1}{2}\partial_x b,$$

and where the nonlinear operator $\mathscr{Q}[H,b]$ is defined for all sufficiently smooth scalar-valued functions w as follows:

$$\mathscr{Q}[H,b](\boldsymbol{v}) := \mathrm{g}H\partial_x\eta + H(-2\mathscr{R}_1[H,b]((\partial_x w)^2) + \mathscr{R}_2[H,b]((w\partial_x)^2 b)), \tag{1.3}$$

with

$$\mathscr{R}_1[H,b]w := -\frac{1}{3H}\partial_x(H^3 w) - \frac{H}{2}w\partial_x b, \qquad \mathscr{R}_2[H,b]w := \frac{1}{2H}\partial_x(H^2 w) + w\partial_x b.$$

This source term is set to zero for the NSW equations considered here (and will therefore be non-zero in the SGN case considered in Chapter 2).

Remark 1.1.1 (*Pressure-related source term*). *In the classical Saint-Venant framework, the pressure at the free-surface is assumed to be purely hydrostatic and equal to the atmospheric pressure, which is taken as constant. As a consequence,* $\underline{\mathrm{p}}$ *does not depend on* x *and therefore* $\partial_x \underline{\mathrm{p}} = 0$*. The pressure-related source term* $\boldsymbol{S}_p[\underline{\mathrm{p}}]$ *then only becomes relevant when modeling external forcings such as wind-induced surface pressure variations or wave-structure interactions effects, see Chapter 5 for instance.*

Remark 1.1.2 (*Friction-related source term*). *In the friction-related source term, the coefficient* $\mathrm{n_f}$ *will be constant in what follows but for practical applications, it is generally space-dependent and comes from measurements and sediment roughness estimations, or even a simultaneous combination of these two effects. We recall that stability issues generally occur when* $\eta \to b$ *and some suitable numerical treatment may be applied, like the one proposed in [30, 106] for instance, but we do not focus on such asymptotic behavior in this work.*

1.1.2 Discrete setting

Let us now introduce the discrete setting for the approximation of the previous system of PDEs, which will be used in the following sections to introduce the proposed local subcell monolithic DG/FV method. We also introduce some useful notations and definitions that will be used throughout the chapter.

Mesh and related polynomial approximation

Let $\Omega \in \mathbb{R}$ denote the computational domain, with boundary $\partial\Omega$. We consider the following partition $\mathcal{T}_h := \{\omega_1, \dots, \omega_{N_{\mathrm{el}}}\} \subset \Omega$ of open disjoint elements ω of boundary $\partial\omega$ such that

$$\overline{\Omega} = \bigcup_{\omega \in \mathcal{T}_h} \overline{\omega}.$$

This partition is characterized by the mesh size $h := \max_{\omega_i \in \mathcal{T}_h} h_i$ where h_i is the size of ω_i. For a given element ω_i, we also note $\omega_i := [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ and x_i its barycenter. For any integer $k \geq 0$, let us define the broken-polynomials space:

$$\mathbb{P}^k(\mathcal{T}_h) := \{v \in L^2(\Omega) \mid \forall \omega \in \mathcal{T}_h,\ v_{|\omega} \in \mathbb{P}^k(\omega)\},$$

where $\mathbb{P}^k(\omega)$ denotes the space of polynomials in ω of total degree at most k, with $\dim \mathbb{P}^k(\omega) = k+1$. Piecewise polynomial functions belonging to some broken-polynomial space $\mathbb{P}^k(\mathcal{T}_h)$ are denoted with a subscript h in the following, and for any $v_h \in \mathbb{P}^k(\mathcal{T}_h)$, and for all $\omega_i \in \mathcal{T}_h$, we use the shortcut $v_h^i \coloneqq v_{h|\omega_i}$. For any $\omega_i \in \mathcal{T}_h$, a basis for $\mathbb{P}^k(\omega_i)$ is denoted by

$$\mathbf{\Psi}_{\omega_i} \coloneqq \{\psi_j^i\}_{j\in[\![1,k+1]\!]}, \tag{1.4}$$

and for $v_h^i \in \mathbb{P}^k(\omega_i)$, $\underline{v}^i \coloneqq \left\{\underline{v}_j^i,\ j \in [\![1,k+1]\!]\right\}$ stand for the *local* degrees of freedom associated with ω_i, introduced as follows:

$$v_h^i \coloneqq \sum_{j=1}^{k+1} \underline{v}_j^i \psi_j^i.$$

We also denote by $\underline{v} \coloneqq \left\{\underline{v}^i,\ i \in [\![1,N_{\text{el}}]\!]\right\}$ the *global* degrees of freedom of v_h. For $\omega_i \in \mathcal{T}_h$, we denote by π_i^k the L^2-orthogonal projector onto $\mathbb{P}^k(\omega_i)$ and π_h^k the global L^2-orthogonal projector onto $\mathbb{P}^k(\mathcal{T}_h)$, gathering all the local L^2 projectors π_i^k. We also define the broken gradient operator $\nabla_h^k \colon \mathbb{P}^k(\mathcal{T}_h) \to \mathbb{P}^{k-1}(\mathcal{T}_h)$ such that, for all $v_h \in \mathbb{P}^k(\mathcal{T}_h)$:

$$(\nabla_h^k v_h)_{|\omega} \coloneqq \partial_x(v_{h|\omega}), \qquad \forall \omega \in \mathcal{T}_h.$$

We also make use of the following notations for the exterior scalar-products of functions $v, w \in L^2(\mathcal{T}_h)$ and $\mu, \nu \in L^2(\partial\mathcal{T}_h)$:

$$\left(v,w\right)_{\mathcal{T}_h} \coloneqq \sum_{\omega_i\in\mathcal{T}_h} \left(v,w\right)_{\omega_i} = \sum_{\omega_i\in\mathcal{T}_h} \int_{\omega_i} v(x,t)w(x,t)\,\mathrm{d}x, \qquad \left\langle \mu,\nu\right\rangle_{\partial\mathcal{T}_h} \coloneqq \sum_{\omega_i\in\mathcal{T}_h} \left\langle \mu,\nu\right\rangle_{\partial\omega_i} = \sum_{\omega_i\in\mathcal{T}_h} [\mu\nu]_{i-\frac{1}{2}}^{i+\frac{1}{2}}.$$

Seeking for approximate solutions of (1.2), the next sections are devoted to the computation of $\boldsymbol{v}_h \coloneqq (\eta_h, q_h)$, such that for any time value, $\boldsymbol{v}_h(\cdot,t) \in \mathbf{P}^k(\mathcal{T}_h) \coloneqq (\mathbb{P}^k(\mathcal{T}_h))^2$. For any given $\boldsymbol{v}_h \in \mathbf{P}^k(\mathcal{T}_h)^2$, and for the sake of simplicity, we generally use the shortcuts $\boldsymbol{v}_{i\pm\frac{1}{2}}^{\mp} \coloneqq v_h^i(x_{i\pm\frac{1}{2}})$ to describe its left and right traces at the mesh interfaces $x_{i\pm\frac{1}{2}}$.

Sub-mesh and related discrete operators

Let us divide the initial mesh into subcells: for any given mesh element $\omega_i \in \mathcal{T}_h$, we introduce a sub-partition $\mathcal{T}_{\omega_i}^k$ into $k+1$ open disjoint subcells:

$$\overline{\omega}_i \coloneqq \bigcup_{m=1}^{k+1} \overline{S_m^i},$$

where the subcell $S_m^i \coloneqq [\widetilde{x}_{m-\frac{1}{2}}^i, \widetilde{x}_{m+\frac{1}{2}}^i]$ is of size $|S_m^i| = |\widetilde{x}_{i+\frac{1}{2}}^i - \widetilde{x}_{i-\frac{1}{2}}^i|$, with the convention $\widetilde{x}_{\frac{1}{2}}^i = x_{i-\frac{1}{2}}$ and $\widetilde{x}_{k+\frac{3}{2}}^i = x_{i+\frac{1}{2}}$. When considering a sequence of neighboring element ω_{i-1}, ω_i and ω_{i+1}, we may use the convenient convention $S_0^i \coloneqq S_{k+1}^{i-1}$ and $S_{k+2}^i \coloneqq S_1^{i+1}$.

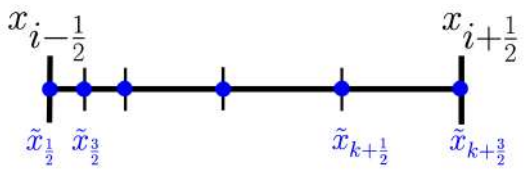


Figure 1.2: Subcell decomposition of ω_i through $k+2$ flux points.

For any given $\omega_i \in \mathcal{T}_h$, let us introduce the set of piecewise-constant functions on the sub-mesh:

$$\mathbb{P}^0(\mathcal{T}_{\omega_i}^k) \coloneqq \{v \in L^\infty(\omega_i) \mid \forall S_m^i \in \mathcal{T}_{\omega_i}^k,\ v_{|S_m^i} \in \mathbb{P}^0(S_m^i)\},$$

together with the subcell-indicator functions

$$\{\mathbb{1}_m^i \mid m \in [\![1,k+1]\!]\},$$

such that $\mathbb{1}^i_m(x) \coloneqq 1$ if $x \in S^i_m$, and 0 otherwise, allowing to define the mean-value projector $\overline{\pi}^k_i$ onto $\mathbb{P}^0(\mathcal{T}^k_{\omega_i})$ such that

$$\mathbb{P}^k(\omega_i) \ni v^i_h \mapsto \overline{\pi}^k_i v^i_h \coloneqq \sum_{m=1}^{k+1} \overline{v}^i_m \mathbb{1}^i_m =: \overline{v}^i_h \in \mathbb{P}^0(\mathcal{T}^k_{\omega_i}),$$

where

$$\left\{\overline{v}^i_m \mid m \in [\![1, k+1]\!]\right\} =: \overline{v}^i,$$

is the collection of piecewise-constant components defined as the mean-values of v^i_h on the subcells belonging to the subdivision $\mathcal{T}^k_{\omega_i}$, and called *sub-mean values*. In order to reach a global description of subcells and related operators, we gather the local subcell partitions and define the associated global approximation space as follows:

$$\widetilde{\mathcal{T}}_h \coloneqq \bigcup_{\omega_i \in \mathcal{T}_h} \mathcal{T}^k_{\omega_i}, \qquad \mathbb{P}^0(\widetilde{\mathcal{T}}_h) \coloneqq \bigtimes_{\omega_i \in \mathcal{T}_h} \mathbb{P}^0(\mathcal{T}^k_{\omega_i}).$$

The global projector onto $\mathbb{P}^0(\widetilde{\mathcal{T}}_h)$, obtained by gathering the local ones $\overline{\pi}^k_i$, is denoted by $\overline{\pi}^k_h$. Note that $\overline{\pi}^k_h$ can be extended to larger functional spaces (typically $\mathsf{H}^s(\mathbb{R})$, $s > 1 + d/2$ for the initial data, and $\mathscr{C}^\infty_{\mathrm{b}}(\mathbb{R})$ for the bathymetry) and denoted $\widetilde{\pi}^k_h$. In what follows, piecewise-constant functions belonging to $\mathbb{P}^0(\widetilde{\mathcal{T}}_h)$ are denoted with a $\overline{\ }$ and a subscript h, and for any $\overline{v}_h \in \mathbb{P}^0(\widetilde{\mathcal{T}}_h)$, and for all $S^i_m \in \widetilde{\mathcal{T}}_h$, we obviously have the identity $\overline{v}^i_m = \overline{v}_{h|S^i_m}$. Let us also adopt the following vectorial shortcuts: $\mathbf{P}^0(\widetilde{\mathcal{T}}_h) \coloneqq (\mathbb{P}^0(\widetilde{\mathcal{T}}_h))^2$ and $\mathbf{P}^k(\mathcal{T}_h) \coloneqq \mathbf{P}^k(\mathcal{T}_h)^2$.

Remark 1.1.3 (*Switching between representations*). *It can be shown that $\overline{\pi}^k_i : \mathbb{P}^k(\omega_i) \to \mathbb{P}^0(\mathcal{T}^k_{\omega_i})$ is one-to-one and onto. For implementation purpose, the associated local transformation matrix $\mathbb{P}_i \in \mathcal{M}_{k+1\times k+1}(\mathbb{R})$ is defined as:*

$$(\mathbb{P}_i)_{(m,p)\in[\![1,\,k+1]\!]^2} = \frac{1}{|S^i_m|}\int_{S^i_m} \psi^i_p \,\mathrm{d}x, \tag{1.5}$$

and we have the transformation relations:

$$\mathbb{P}_i\, \underline{v}^{\,i} = \overline{v}^i \quad \textit{and} \quad \mathbb{P}_i^{-1}\, \overline{v}^i = \underline{v}^i. \tag{1.6}$$

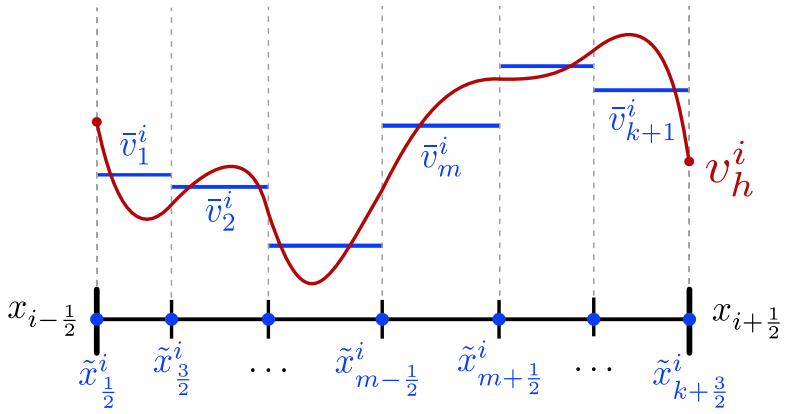


Figure 1.3: Piecewise polynomial function v^i_h and associated sub-mean values.

Time discretization and time-marching algorithm

For a given final time $T > 0$, we also introduce a general partition of the time domain $0 =: t^0 < t^1 < \cdots < t^n < \cdots < t^N \coloneqq T$, and the adaptative time step $\Delta t^n \in \mathbb{R}^*_+$ such that, for any $n \geq 0$, $t^{n+1} = t^n + \Delta t^n$.

Introducing a time-stepping algorithm leads to introduce some new time-related notations, with $\boldsymbol{v}^n_h \coloneqq \boldsymbol{v}_h(\cdot, t^n)$ and $\overline{\boldsymbol{v}}^n_h \coloneqq \overline{\pi}^k_h \circ \boldsymbol{v}^n_h =: (\overline{\eta}^n_h, \overline{q}^n_h)$ being the corresponding piecewise-constant function defined on the sub-mesh. We also denote by

$$\overline{\boldsymbol{v}}^{i,n} \coloneqq \{\overline{\boldsymbol{v}}^{i,n}_m \mid m \in [\![1, k+1]\!]\},$$

the collection of sub-mean values associated with ω_i at time t^n and we obviously have $\overline{\boldsymbol{v}}^{i,n}_m = \overline{\boldsymbol{v}}^n_{h|S^i_m}$. Note that when no confusions are possible, the superscripts i (cell-related) and n (time-related) may be suppressed, for the

sake of simplicity. The time-marching algorithms used during all the thesis are either first-order explicit Euler schemes or higher-order SSP-RK schemes [126]. A general s–stage explicit Runge–Kutta method advances $\boldsymbol{v}_h^n$ to $\boldsymbol{v}_h^{n+1}$ using coefficients $(\mathsf{a}_{ij}, \mathsf{b}_i)$ that define the *Butcher tableau* of the scheme:

$$\begin{array}{c|cccc} \mathsf{c}_1 & 0 & & & \\ \mathsf{c}_2 & \mathsf{a}_{21} & 0 & & \\ \mathsf{c}_3 & \mathsf{a}_{31} & \mathsf{a}_{32} & 0 & \\ \vdots & \vdots & \vdots & \ddots & \\ \mathsf{c}_s & \mathsf{a}_{s1} & \mathsf{a}_{s2} & \dots & 0 \\ \hline & \mathsf{b}_1 & \mathsf{b}_2 & \dots & \mathsf{b}_s \end{array}$$

the coefficients satisfying $\mathsf{c}_i = \sum_{j=1}^{i-1} \mathsf{a}_{ij}$ for all $i \in [\![1, s]\!]$.

Then, considering the semi-discrete formulation written in operator form

$$\partial_t \boldsymbol{v}_h + \mathcal{N}_h(\boldsymbol{v}_h) = 0,$$

where $\mathcal{N}_h$ stands for the nonlinear operator allowing to compute the approximated residuals, and considering for instance the third-order SSP-RK scheme, we advance from time level n to $(n+1)$ with the third-order explicit scheme as follows:

$$\begin{aligned} \boldsymbol{v}_h^{n,1} &= \boldsymbol{v}_h^n - \Delta t^n \mathcal{N}_h(\boldsymbol{v}_h^n), \\ \boldsymbol{v}_h^{n,2} &= \frac{1}{4}(3\boldsymbol{v}_h^n + \boldsymbol{v}_h^{n,1}) - \frac{\Delta t^n}{4}\mathcal{N}_h(\boldsymbol{v}_h^{n,1}), \\ \boldsymbol{v}_h^{n+1} &= \frac{1}{3}(\boldsymbol{v}_h^n + 2\boldsymbol{v}_h^{n,2}) - \frac{2\Delta t^n}{3}\mathcal{N}_h(\boldsymbol{v}_h^{n,2}), \end{aligned}$$

where $\boldsymbol{v}_h^{n,1}$ and $\boldsymbol{v}_h^{n,2}$ are the solution obtained at intermediate stages, and the initial data $\boldsymbol{v}_h^0$ is yet to be defined (see Section 1.3 and the consistency with the bathymetry variations approximation).

CFL condition and time step computation

The time step Δt^n for our monolithic scheme is computed adaptively using the following CFL condition:

$$\Delta t^n := \frac{\min\limits_{\omega_i \in \mathcal{T}_h} \left(\dfrac{h_i}{2k+1}, \min\limits_{S_m^i \in \mathcal{T}_{\omega_i}^k} |S_m^i| \right)}{\sigma}, \qquad \sigma := \max_{\omega_i \in \mathcal{T}_h} \max_{m \in [\![1, k+1]\!]} \left(|\overline{u}_m^i| + \sqrt{g \overline{H}_m^i} \right). \tag{1.7}$$

Remark 1.1.4 (*Velocity regularization*). *Concerning the computation of velocity $u := q/H$ (either from polynomial values or piecewise constant subcell values), we use the following velocity regularization from [177] (omitting the indices for the sake of clarity):*

$$u^{\mathrm{reg}} = \frac{\sqrt{2} H q}{\sqrt{H^4 + \max(H^4, H_{\mathrm{reg}}^4)}} = \frac{q}{H} \frac{\sqrt{2}}{\sqrt{1 + \max(1, H_{\mathrm{reg}}^4 / H^4)}} \quad (H > 0), \tag{1.8}$$

where $H_{\mathrm{reg}} > 0$ is a small water-height threshold chosen a priori according to mesh scale and quality. There are several alternative ways to desingularize the quantity q/H, the simplest one being

$$u_{\mathrm{alt}}^{\mathrm{reg}} = \frac{2Hq}{H^2 + \max(H^2, H_{\mathrm{reg}}^2)} = \frac{q}{H} \frac{2}{1 + \max(1, H_{\mathrm{reg}}^2 / H^2)} \quad (H > 0). \tag{1.9}$$

One might try to reduce the regularization effect by taking a very small value of H_{reg} in the simpler formula. However, this may fail in practice: dividing small values of Hq by $H^2 + H_{\mathrm{reg}}^2$ can introduce significant numerical errors. These errors may affect the accuracy of u and can also slow down the computation. Still, this kind of regularization may sacrifice accuracy for robustness. For this reason, it would be useful to design a velocity regularization better suited to high-order methods, in order to prevent the time step from becoming too small and to avoid tuning threshold values for each benchmark. One possible attempt is to use a polynomial transition below the prescribed water-height threshold $H_{\mathrm{poly}} > 0$, while keeping the exact velocity above this threshold:

$$u_{\mathrm{poly}}^{\mathrm{reg}} = \begin{cases} \dfrac{q}{H} & H \geq H_{\mathrm{poly}}, \\ q\left(\dfrac{3H}{H_{\mathrm{poly}}^2} - \dfrac{2H^2}{H_{\mathrm{poly}}^3} \right) & 0 \leq H < H_{\mathrm{poly}}. \end{cases} \tag{1.10}$$

For fixed q, this regularization is $\mathscr{C}^0$ on $[0,+\infty)$ and is $\mathscr{C}^1$ at the matching point $H = H_{\text{poly}}$, since

$$\frac{3H_{\text{poly}}^2 - 2H_{\text{poly}}^2}{H_{\text{poly}}^3} = \frac{1}{H_{\text{poly}}}, \qquad \frac{3H_{\text{poly}} - 4H_{\text{poly}}}{H_{\text{poly}}^3} = -\frac{1}{H_{\text{poly}}^2}.$$

Moreover, it satisfies $u_{\text{poly}}^{\text{reg}} \to 0$ as $H \to 0$, so that the velocity does not blow up close to vacuum. This construction preserves the exact value q/H away from dry regions, and replaces the singular division by a polynomial expression near vacuum. However, it was not used in the numerical experiments presented in this thesis, since it led to stability issues in our high-order computations. We believe that more work is needed to obtain a robust and accurate regularization strategy, as it could also be useful for other hyperbolic models involving nearly singular ratios, for instance in compressible Euler equations close to vacuum, where the velocity $u = m/\rho$ and the internal energy require divisions by very small densities.

1.2 Local-subcell monolithic DG/FV for homogeneous system

We first introduce the local-subcell monolithic DG/FV strategy in the simpler case of the homogeneous nonlinear shallow-water equations, namely without bathymetry source terms and without any higher-order differential contribution. In this setting, the physical flux reduces to

$$\mathscr{H}_b^+ \ni \boldsymbol{v} \mapsto \boldsymbol{F}_{\text{hom}}(\boldsymbol{v}) \coloneqq \left(q,\ uq + \frac{\mathrm{g}H^2}{2}\right)^\top \in \mathbb{R}^2.$$

This preliminary setting is not only a matter of presentation: it allows us to isolate the genuinely hyperbolic part of the method and to focus on the core mechanism of the proposed discretization, namely the local blending between high-order DG reconstructed fluxes and robust first-order FV fluxes at the subcell level.

Even in this homogeneous framework, the main numerical difficulties already arise. High-order DG approximations may generate spurious oscillations near discontinuities or steep gradients, and the positivity of the water-height must still be preserved at the discrete level. The homogeneous case is therefore a natural starting point to introduce the hierarchy of fluxes, the DG-to-subcell-FV equivalence, and the construction of local blending coefficients designed to enforce suitable convex admissibility properties. It also keeps the algebraic structure clear for readers not familiar with the topic, since the numerical fluxes are conservative, no hydrostatic reconstruction is needed, and the admissibility constraints can be stated directly in terms of the subcell mean values. This gives a clear view of how the monolithic DG/FV correction acts: the high-order DG method is recovered wherever the blending coefficient is equal to one, while the scheme locally falls back to a more robust FV update when the high-order flux threatens stability or admissibility.

1.2.1 A hierarchy of numerical fluxes

In what follows, we manipulate a collection of various numerical fluxes defined on mesh or sub-mesh interfaces that we find useful to gather their definitions in one place, for further referencing:

– *Global Lax-Friedrichs (LF) flux.* This flux is used throughout this work and is defined as follows:

$$(\mathscr{H}_b^+)^2 \ni (\boldsymbol{v}_l, \boldsymbol{v}_r) \mapsto \boldsymbol{\mathcal{F}}(\boldsymbol{v}_l, \boldsymbol{v}_r) \coloneqq \frac{1}{2}\left(\boldsymbol{F}_{\text{hom}}(\boldsymbol{v}_r) + \boldsymbol{F}_{\text{hom}}(\boldsymbol{v}_l) - \sigma(\boldsymbol{v}_r - \boldsymbol{v}_l)\right).$$

Let us emphasize that any other consistent and positive numerical fluxes, including the local LF (Rusanov), the HLL/HLL-C or kinetic-based fluxes, may be used. We retain a single global LF parameter because it gives the same low-order flux and dissipation coefficient at every subcell interface, simplifying the convex-state construction and its analysis. Exploratory tests with Rusanov and HLL fluxes did not produce substantial changes in the numerical results reported here.

– *Lowest-order subcell FV fluxes* $\{\boldsymbol{\mathcal{F}}^{\text{FV}}\}_{m\in[\![1,k]\!]}$. These first-order fluxes may only be evaluated at the *subinterfaces* between neighboring subcells, within a mesh element, and are given by

$$\boldsymbol{\mathcal{F}}^{\text{FV}}_{m+\frac{1}{2}} \coloneqq \boldsymbol{\mathcal{F}}(\overline{\boldsymbol{v}}_m, \overline{\boldsymbol{v}}_{m+1}), \qquad \boldsymbol{\mathcal{F}}^{\text{FV}}_{m-\frac{1}{2}} \coloneqq \boldsymbol{\mathcal{F}}(\overline{\boldsymbol{v}}_{m-1}, \overline{\boldsymbol{v}}_m),$$

where $\overline{\boldsymbol{v}}_m$ is the subcell mean value of $\boldsymbol{v}_h$ on subcell S_m^i, $m \in [\![1,k]\!]$, and $\boldsymbol{\mathcal{F}}$ is the LF flux defined above, with

$$\sigma \coloneqq \max_{\omega_i\in\mathcal{T}_h} \max_{m\in[\![1,k+1]\!]} \left(|\overline{u}_m^i| + \sqrt{\mathrm{g}\overline{H}_m^i}\right).$$

For further use, let us also recall the definition of the related *intermediate Riemann state* defined as follows:

$$\boldsymbol{v}^\star := \frac{1}{2}(\boldsymbol{v}_l + \boldsymbol{v}_r) - \frac{1}{2\sigma}(\boldsymbol{F}_{\text{hom}}(\boldsymbol{v}_r) - \boldsymbol{F}_{\text{hom}}(\boldsymbol{v}_l)), \tag{1.11}$$

– *Higher-order DG fluxes* $\boldsymbol{\mathcal{F}}^{\text{DG}}_{i\pm\frac{1}{2}}$. These fluxes may only be evaluated at the *interfaces* between two neighboring mesh elements as follows:

$$\boldsymbol{\mathcal{F}}^{\text{DG}}_{i\pm\frac{1}{2}} := \boldsymbol{\mathcal{F}}\left(\boldsymbol{v}^-_{i\pm\frac{1}{2}}, \boldsymbol{v}^+_{i\pm\frac{1}{2}}\right),$$

where $\boldsymbol{v}^\pm_{i\pm\frac{1}{2}}$ are the left and right traces of (polynomial) $\boldsymbol{v}_h$ at mesh interface $i \pm \frac{1}{2}$,

– *Reconstructed subcell fluxes* $\{\widehat{\boldsymbol{F}}^i_{m+\frac{1}{2}}\}_{m\in[\![0,k+1]\!]}$. These fluxes are introduced below within the equivalency formulation (1.16) and aim at locally mimicking the behavior of an *in-cell* FV scheme, giving exactly and equivalently the DG approximated solution, and are defined as follows:

$$\widehat{\boldsymbol{F}}^i_{m+\frac{1}{2}} = \boldsymbol{F}^i_h(\widetilde{x}^i_{m+\frac{1}{2}}) - C^{i-\frac{1}{2}}_{m+\frac{1}{2}}\left(\boldsymbol{F}^i_h(x_{i-\frac{1}{2}}) - \boldsymbol{\mathcal{F}}^{\text{DG}}_{i-\frac{1}{2}}\right) - C^{i+\frac{1}{2}}_{m+\frac{1}{2}}\left(\boldsymbol{F}^i_h(x_{i+\frac{1}{2}}) - \boldsymbol{\mathcal{F}}^{\text{DG}}_{i+\frac{1}{2}}\right), \tag{1.12}$$

with $\boldsymbol{F}^i_h := \pi^k_i \boldsymbol{F}(v^i_h)$ corresponding to the L^2 projection of the interior flux, and where the coefficients $C^{i\pm\frac{1}{2}}_{m+\frac{1}{2}}$ are given by:

$$C^{i-\frac{1}{2}}_{m+\frac{1}{2}} := \sum_{j=m+1}^{k+1} \phi^i_j(x_{i-\frac{1}{2}}), \qquad C^{i+\frac{1}{2}}_{m+\frac{1}{2}} := \sum_{j=1}^{m} \phi^i_j(x_{i+\frac{1}{2}}),$$

where $\boldsymbol{\Phi}_{\omega_i} := \{\phi^i_m\}_{m\in[\![1,k+1]\!]}$ is referred as the *sub-resolution basis functions* that reads as the L^2 projection of the subcell indicator function i.e.

$$\phi^i_m = \pi^k_i \mathbb{1}^i_m, \qquad m \in [\![1, k+1]\!], \tag{1.13}$$

– *Blended subcell fluxes* $\{\widetilde{\boldsymbol{F}}^i_{m+\frac{1}{2}}\}_{m\in[\![0,k+1]\!]}$. These fluxes are defined as convex combinations of the lowest-order subcell FV fluxes $\boldsymbol{\mathcal{F}}^{\text{FV}}$ and the *reconstructed* subcell fluxes $\widehat{\boldsymbol{F}}^i_{m+\frac{1}{2}}$. For $\omega_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \in \mathcal{T}_h$, we introduce:

$$\widetilde{\boldsymbol{F}}^i_{m+\frac{1}{2}} := \boldsymbol{\mathcal{F}}^{\text{FV}}_{m+\frac{1}{2}} + \Theta_{m+\frac{1}{2}}\left(\widehat{\boldsymbol{F}}^i_{m+\frac{1}{2}} - \boldsymbol{\mathcal{F}}^{\text{FV}}_{m+\frac{1}{2}}\right),$$

with $\Theta_{m+\frac{1}{2}} \in [0,1]$. The computation of such *blending coefficients* to ensure different desired properties will be detailed below.

Remark 1.2.1 (*Correction coefficients in reconstructed fluxes*). *It is actually possible to explicitly compute such coefficients: let us set* $\boldsymbol{J} \in \mathbb{R}^{k+1}$ *to be the vector defined as*

$$(\boldsymbol{J}_m)_{m\in[\![1,k+1]\!]} = (-1)^{m+1}\binom{k+m}{m}\binom{k+1}{m},$$

where $\binom{p}{m}$ *stands for the binomial coefficient* $\binom{p}{m} := \frac{p!}{m!(p-m)!}$. *Let us note that vector* $\boldsymbol{J}$ *only depends on the degree of approximation* k*, and not on the flux points position. By introducing* $\{\widetilde{\xi}_{m+\frac{1}{2}}\}_{m\in[\![1,k+1]\!]}$ *the flux points counterpart in the referential element* $[0,1]$ *as*

$$\widetilde{\xi}_{m+\frac{1}{2}} = \frac{\widetilde{x}_{m+\frac{1}{2}} - x_{i-\frac{1}{2}}}{x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}},$$

the correction coefficients finally write

$$C^{i-\frac{1}{2}}_{m+\frac{1}{2}} = 1 - \left(\widetilde{\xi}_{m+\frac{1}{2}}, \left(\widetilde{\xi}_{m+\frac{1}{2}}\right)^2, \ldots, \left(\widetilde{\xi}_{m+\frac{1}{2}}\right)^{k+1}\right)^\top \cdot \boldsymbol{J},$$

further details being available in [264, 265]. It should be noted that while this definition of $\{C^{i\pm\frac{1}{2}}_{m+\frac{1}{2}}\}_{m\in[\![0,k+1]\!]}$ *ensures equivalence with DG schemes, other choices naturally lead to different schemes. For example, if one*

sets these constants to zero, except for the first and last being one, one would recover the Spectral Volume (SV) method, [143, 273]. Even though the interior flux, referred to here as the reconstructed flux, is continuous within the cells as in the SV methods, these two methods are quite distinct. In SV methods, the interior flux is merely the flux function applied to the polynomial solution, i.e. $\boldsymbol{F}(\boldsymbol{v}_h, b_h)$*. Here, the reconstructed flux can be seen as a specific approximation of the* L^2 *projection of such a function, with its boundary values prescribed as the DG numerical fluxes. Therefore, this approach may be more comparable to the Flux Reconstruction method, also known as the CPR method (Correction Procedure through Reconstruction), with which we share this reconstructed flux framework. For more insights on the analogy between the present theory and Flux Reconstruction schemes, see for instance [5, 121, 150, 151, 270] or the dedicated paragraph in [264].*

Remark 1.2.2 (*Sub-partition points*). *The choice of the sub-partition points i.e.* $\{\widetilde{x}_{m+\frac{1}{2}}\}_{m\in[\![0,k+1]\!]}$*, has already been discussed in [264]. It appeared that, regarding the reformulation of DG schemes into subcell FV methods, the cell decomposition into subcells does not come into account, as any choice would lead to the same piecewise polynomial solution. However, for the the blending procedure introduced here, the sub-division does have a slight impact. Indeed, the use of a non-uniform sub-partition, for instance by means of the Gauss-Lobatto points, leads to better results compared to a uniform sub-division. This is more likely the manifestation of the Runge phenomenon in the context of histopolation, as the histopolation basis functions underlying the sub-mean value representation, are more oscillatory for a uniform cell sub-partition. Consequently, in the remainder, we make use of Gauss-Lobatto points to define the sub-partition points* $\{\widetilde{x}_{m+\frac{1}{2}}\}_{m\in[\![0,k+1]\!]}$*.*

1.2.2 DG and local-subcell DG/FV equivalency

We recall, see for instance [140], that a straightforward semi-discrete global DG approximation for the homogeneous NSW equations may read as follows: find $\boldsymbol{v}_h \coloneqq (\eta_h, q_h) \in \mathbf{P}^k(\mathcal{T}_h)^2$ such that, for all $\varphi \in \mathbb{P}^k(\mathcal{T}_h)$:

$$\left(\partial_t \boldsymbol{v}_h, \varphi\right)_{\mathcal{T}_h} - \left(\boldsymbol{F}(\boldsymbol{v}_h), \partial_x \varphi\right)_{\mathcal{T}_h} + \left\langle \varphi, \boldsymbol{\mathcal{F}}^{\mathrm{DG}} \right\rangle_{\partial\mathcal{T}_h} = 0, \tag{1.14}$$

which locally translates as: for all $\omega_i \in \mathcal{T}_h$, find $\boldsymbol{v}_h^i \coloneqq (\eta_h^i, q_h^i) \in (\mathbb{P}^k(\omega_i))^2$ such that, for all $\varphi \in \mathbb{P}^k(\omega_i)$

$$\int_{\omega_i} \partial_t \boldsymbol{v}_h^i \varphi \,\mathrm{d}x - \int_{\omega_i} \boldsymbol{F}(\boldsymbol{v}_h^i) \partial_x \varphi \,\mathrm{d}x + \left[\varphi \boldsymbol{\mathcal{F}}^{\mathrm{DG}}\right]_{i-\frac{1}{2}}^{i+\frac{1}{2}} = 0. \tag{1.15}$$

We also recall an essential result of [265], which states the equivalence between DG and a FV-like formulation on the sub-mesh:

Proposition 1.2.3 (*DG and local-subcell FV equivalence*). *For any given* $\omega_i \in \mathcal{T}_h$*, the local DG formulation* (1.15) *can be equivalently written as a FV-like scheme defined on the sub-mesh, as follows:*

$$\frac{\mathrm{d}\overline{\boldsymbol{v}}_m^i}{\mathrm{d}t} = -\frac{1}{|S_m^i|}\left(\widehat{\boldsymbol{F}}_{m+\frac{1}{2}}^i - \widehat{\boldsymbol{F}}_{m-\frac{1}{2}}^i\right), \qquad \forall m \in [\![1, k+1]\!], \tag{1.16}$$

where the reconstructed fluxes $\widehat{\boldsymbol{F}}_{m\pm\frac{1}{2}}^i$ *are defined in* (1.12) *and* $\overline{\boldsymbol{v}}_m^i \coloneqq (\overline{\pi}_i^k \boldsymbol{v}_h^i)_{|S_m^i}$*.*

While an *a posteriori* framework is studied in [140], relying on the same general formulation (A.2) and the equivalent local translation at the subcell level (1.16), we investigate here the construction of suitable *a priori* estimated blending coefficients, ensuring a synergistic combination of (higher-order) DG and (lower-order) FV fluxes in a robust way.

1.2.3 Local subcell monolithic method and blending coefficients

The main idea of the local subcell monolithic method is to substitute the *blended* fluxes for the *reconstructed* fluxes into the local-subcell formulations (1.16), in order to provide some subcell scale flexibility to enforce nonlinear stability properties through suitable and adaptive computations of the local blending coefficients $\Theta_{m+\frac{1}{2}}$. Applying a first-order time-stepping algorithm for the sake of simplicity (any higher-order SSP-RK method may be used as well), we thus consider the local-subcell monolithic formulations:

$$\overline{\boldsymbol{v}}_m^{i,n+1} = \overline{\boldsymbol{v}}_m^{i,n} - \frac{\Delta t^n}{|S_m^i|}\left(\widetilde{\boldsymbol{F}}_{m+\frac{1}{2}}^i - \widetilde{\boldsymbol{F}}_{m-\frac{1}{2}}^i\right), \qquad \forall (m,i) \in [\![1, k+1]\!] \times [\![1, N_{\mathrm{el}}]\!], \tag{1.17}$$

which is of course also conservative at the subcell level, and we are left with the computation of the blending coefficient $\{\Theta_{m+\frac{1}{2}}\}_{m\in[\![0,k+1]\!]}$. To achieve this, let us have a look first to the lowest-order scheme, obtained by uniformly setting $\Theta_{m+\frac{1}{2}}=0$:

$$\overline{\boldsymbol{v}}_m^{i,n+1}=\overline{\boldsymbol{v}}_m^{i,n}-\frac{\Delta t^n}{|S_m^i|}\left(\boldsymbol{\mathcal{F}}^{\mathrm{FV}}_{m+\frac{1}{2}}-\boldsymbol{\mathcal{F}}^{\mathrm{FV}}_{m-\frac{1}{2}}\right),\qquad\forall(m,i)\in[\![1,k+1]\!]\times[\![1,N_{\mathrm{el}}]\!],\tag{1.18}$$

and observe that (1.18) may be seen as a Godunov-like scheme, i.e. a convex combination of flow-states defined at time-step n, as follows:

$$\overline{\boldsymbol{v}}_m^{n+1}=\left(1-\frac{2\Delta t^n}{|S_m|}\sigma\right)\overline{\boldsymbol{v}}_m^n+\frac{\sigma\Delta t^n}{|S_m|}\boldsymbol{v}^{*,-}_{m+\frac{1}{2}}+\frac{\sigma\Delta t^n}{|S_m|}\boldsymbol{v}^{*,+}_{m+\frac{1}{2}},$$

where these $\boldsymbol{v}^{*,\pm}_{m+\frac{1}{2}}$ are called *low-order FV intermediate Riemann flow-states*, and are defined as

$$\boldsymbol{v}^{*,\pm}_{m\mp\frac{1}{2}}:=\overline{\boldsymbol{v}}_m^n\pm\frac{\boldsymbol{\mathcal{F}}^{\mathrm{FV},\pm}_{m\mp\frac{1}{2}}-\boldsymbol{F}(\overline{\boldsymbol{v}}_m^n)}{\sigma}.$$

These intermediate states play a central role in the construction of the blending coefficients. Indeed, under the CFL-like condition (1.7), they belong to the admissible set $\mathscr{H}_b^+$, as shown in Appendix B in the 2D case. The lowest-order FV update therefore provides a robust reference scheme whose admissibility properties are inherited from the convexity of $\mathscr{H}_b^+$. Indeed, the purpose of the blending procedure is then to introduce as much as possible of the high-order reconstructed flux while preserving these convex properties. In other words, the coefficient $\Theta_{m+\frac{1}{2}}$ must be chosen locally so that the corrected intermediate states obtained from the blended fluxes remain admissible. This gives a direct and local criterion for limiting the high-order contribution only where it may compromise stability, while keeping the full DG accuracy in smooth regions.

Now, considering the case of higher-order blended fluxes, we observe in a similar way that (1.33) may be written as a combination of intermediate states defined at time-step n, as follows:

$$\overline{\boldsymbol{v}}_m^{n+1}=\left(1-\frac{2\Delta t^n}{|S_m|}\sigma\right)\overline{\boldsymbol{v}}_m^n+\frac{\sigma\Delta t^n}{|S_m|}\widetilde{\boldsymbol{v}}^{*,-}_{m+\frac{1}{2}}+\frac{\sigma\Delta t^n}{|S_m|}\widetilde{\boldsymbol{v}}^{*,+}_{m+\frac{1}{2}},$$

where the new *blended intermediate Riemann states* $\widetilde{\boldsymbol{v}}^{*,\pm}_{m+\frac{1}{2}}$ are defined as follows:

$$\widetilde{\boldsymbol{v}}^{*,\pm}_{m+\frac{1}{2}}:=\boldsymbol{v}^{*,\pm}_{m+\frac{1}{2}}\pm\Theta_{m+\frac{1}{2}}\left(\frac{\widehat{\boldsymbol{F}}_{m+\frac{1}{2}}-\boldsymbol{\mathcal{F}}^{\mathrm{FV}}_{m+\frac{1}{2}}}{\sigma}\right).$$

Hence, it is enough to seek the blending coefficient values $\Theta_{m+\frac{1}{2}}$ such that the required convex properties are enforced on the *blended states* $\widetilde{\boldsymbol{v}}^{*,\pm}_{m+\frac{1}{2}}$.

Similarly to what has been previously done in the *a posteriori* context, [140, 264], we seek here *a priori* to:

- mimic a local maximum principle at the discrete level (subcell scale), in order to prevent the occurrence of spurious oscillations in the vicinity of singularities: an adaptive *Numerical Admissibility Detection* (NAD), which is applied at the *subcell* level – hence, the *SubNAD* naming of this sensor,
- ensure that the updated water-height remains non-negative at each time-step or sub-step for RK time-stepping, and at the subcell level. This second property, referred to as Physical Admissibility Detection (PAD), should be enforced while taking into account the bathymetry variations and is therefore temporarily postponed to the next section.

Subcell Numerical Admissibility Detection (subNAD)

We aim at mimicking a local maximum principle at the discrete level on the sub-mean values. We impose the updated sub-mean values to be bounded in a suitable local subset of the sub-mean values at the previous time step (or previous sub-step for higher-order RK time-stepping). Note that, as there is no physical local maximum for the conserved variables, we also add the intermediate lowest-order states in this subset, and we aim at enforcing that, for $m\in[\![1,k+1]\!]$:

$$\min\left(\overline{\boldsymbol{v}}_{m-1}^{i,n},\overline{\boldsymbol{v}}_m^{i,n},\overline{\boldsymbol{v}}_{m+1}^{i,n},\boldsymbol{v}^{*,+}_{m-\frac{1}{2}},\boldsymbol{v}^{*,-}_{m+\frac{1}{2}}\right)\le\overline{\boldsymbol{v}}_m^{i,n+1}\le\max\left(\overline{\boldsymbol{v}}_{m-1}^{i,n},\overline{\boldsymbol{v}}_m^{i,n},\overline{\boldsymbol{v}}_{m+1}^{i,n},\boldsymbol{v}^{*,+}_{m-\frac{1}{2}},\boldsymbol{v}^{*,-}_{m+\frac{1}{2}}\right),\tag{1.19}$$

with $\overline{\boldsymbol{v}}_0^{i,n}=\overline{\boldsymbol{v}}_{k+1}^{i-1,n}$ and $\overline{\boldsymbol{v}}_{k+2}^{i,n}=\overline{\boldsymbol{v}}_1^{i+1,n}$.

Remark 1.2.4 (*Relaxation of the SubNAD criterion*). *This criterion should be supplemented with a relaxation procedure to exclude the smooth extrema from the troubled cells, i.e. uniformly enforce* $\Theta_{m+\frac{1}{2}} = 1$ *when such an extremum occurs. Indeed, a numerical solution exhibits a smooth extrema if at least the following linearized version of the variable* η *spatial derivative:*

$$(\partial_x \eta)^{\text{lin}}_{S^i_m}(x) = \overline{\partial_x \eta^i_m} + (x - x_i)\overline{\partial^2_{xx}\eta^i_m},$$

has a monotonous profile, where $\overline{\partial_x \eta^i_m}$ *and* $\overline{\partial^2_{xx}\eta^i_m}$ *are respectively the mean values of* $(\partial_x \eta_h)_{|S^i_m}$ *and* $(\partial^2_{xx}\eta_h)_{|S^i_m}$ *on subcell* S^i_m*. In practice, the relaxation used here works as a vertex-based limiter on* $(\partial_x \eta)^{\text{lin}}_{S^i_m}$*. If the numerical solution can be regarded as smooth in subcell* S^i_m*, the SubNAD criterion is relaxed, i.e. we force* $\Theta^{\text{nad}}_{m-\frac{1}{2}} = \Theta^{\text{nad}}_{m+\frac{1}{2}} = 1$*, allowing the high-order accuracy preservation of smooth extrema. We refer the interested reader to such a relaxation procedure to look at [140, 265] for more details. One of course can do this on Riemann invariants as well, as there is no physical local maximum for the conserved variables on such systems, but it has been proved in the 1D case in [140] that it does not produce better results. A similar relaxation procedure will be detailed in the two-dimensional case in Chapter 3, where we will also consider the possibility to do it on Riemann invariants.*

The following expression of blending coefficient $\Theta^{\text{nad}}_{m+\frac{1}{2}}$ allows to ensure such a (convex) property:

Proposition 1.2.5 (*Subcell numerical admissibility*). *The local-subcell maximum principle* (1.19) *is satisfied as soon as the following inequalities are locally satisfied by the blending coefficients:*

$$\Theta_{m+\frac{1}{2}} \leq \Theta^{\text{nad}}_{m+\frac{1}{2}},$$

with:

$$\Theta^{\text{nad}}_{m+\frac{1}{2}} := \min\left(1, \left|\frac{\sigma}{\Delta \boldsymbol{F}^{(1)}_{m+\frac{1}{2}}}\right| \begin{cases} \min\left(\beta_{m+1} - \eta^{*,+}_{m+\frac{1}{2}},\ \eta^{*,-}_{m+\frac{1}{2}} - \alpha_m\right) & \text{if } \Delta \boldsymbol{F}^{(1)}_{m+\frac{1}{2}} > 0 \\ \min\left(\beta_m - \eta^{*,-}_{m+\frac{1}{2}},\ \eta^{*,+}_{m+\frac{1}{2}} - \alpha_{m+1}\right) & \text{if } \Delta \boldsymbol{F}^{(1)}_{m+\frac{1}{2}} < 0 \end{cases}\right),$$

where $\Delta \boldsymbol{F}^{(1)}_{m+\frac{1}{2}} := \widehat{\boldsymbol{F}}^{(1)}_{m+\frac{1}{2}} - \boldsymbol{\mathcal{F}}^{\text{FV},(1)}_{m+\frac{1}{2}}$ *stands for the first component of the flux (appearing in the mass-conservation equation), and dropping* $\pm$ *indices for sake of clarity, with*

$$\left\{\alpha_m, \beta_m\right\} := \left\{\min(\overline{\eta}_{m-1}, \overline{\eta}_m, \overline{\eta}_{m+1}, \eta^{*,+}_{m-\frac{1}{2}}, \eta^{*,-}_{m+\frac{1}{2}}),\ \max(\overline{\eta}_{m-1}, \overline{\eta}_m, \overline{\eta}_{m+1}, \eta^{*,+}_{m-\frac{1}{2}}, \eta^{*,-}_{m+\frac{1}{2}})\right\}.$$

Proof. We get, from the fully discrete formulation, the following expression:

$$\overline{\boldsymbol{v}}^{n+1}_m = \left(1 - \frac{2\Delta t^n}{|S_m|}\sigma\right)\overline{\boldsymbol{v}}^n_m + \frac{\sigma \Delta t^n}{|S_m|}\widetilde{\boldsymbol{v}}^{*,-}_{m+\frac{1}{2}} + \frac{\sigma \Delta t^n}{|S_m|}\widetilde{\boldsymbol{v}}^{*,+}_{m+\frac{1}{2}} + \Delta t^n \overline{\boldsymbol{S}}^n_m.$$

Focusing on the equation on $\overline{\eta}$, we now have two states $\widetilde{\eta}^{*,\pm}_{m+\frac{1}{2}}$, that can be written as

$$\widetilde{\eta}^{*,-}_{m+\frac{1}{2}} = \eta^{*,-}_{m+\frac{1}{2}} - \Theta_{m+\frac{1}{2}}\left(\frac{\widehat{\boldsymbol{F}}^{(1)}_{m+\frac{1}{2}} - \boldsymbol{\mathcal{F}}^{(1),-}_{m+\frac{1}{2}}}{\sigma}\right), \qquad \widetilde{\eta}^{*,+}_{m+\frac{1}{2}} = \eta^{*,+}_{m+\frac{1}{2}} + \Theta_{m+\frac{1}{2}}\left(\frac{\widehat{\boldsymbol{F}}^{(1)}_{m+\frac{1}{2}} - \boldsymbol{\mathcal{F}}^{(1),+}_{m+\frac{1}{2}}}{\sigma}\right),$$

We can then write Riemann states in a compact form as

$$\widetilde{\eta}^{*,\pm}_{m+\frac{1}{2}} = \eta^{*,\pm}_{m+\frac{1}{2}} \pm \Theta_{m+\frac{1}{2}}\left(\frac{\widehat{\boldsymbol{F}}^{(1)}_{m+\frac{1}{2}} - \boldsymbol{\mathcal{F}}^{(1),-}_{m+\frac{1}{2}}}{\sigma}\right).$$

We already know that

$$\left(\eta^{*,-}_{m+\frac{1}{2}}, \eta^{*,+}_{m+\frac{1}{2}}\right) \in \left[\eta^{\min}_m, \eta^{\max}_m\right] \times \left[\eta^{\min}_{m+1}, \eta^{\max}_{m+1}\right],$$

by the results of Appendix B but we need to ensure that blended intermediate states verify

$$\left(\widetilde{\eta}^{*,-}_{m+\frac{1}{2}}, \widetilde{\eta}^{*,+}_{m+\frac{1}{2}}\right) \in \left[\eta^{\min}_m, \eta^{\max}_m\right] \times \left[\eta^{\min}_{m+1}, \eta^{\max}_{m+1}\right].$$

The computation procedure consist of assuring that non-admissible cases as, for exemple,

$$\eta^{*,-}_{m+\frac{1}{2}} - \frac{\widehat{\boldsymbol{F}}^{(1)}_{m+\frac{1}{2}} - \boldsymbol{\mathcal{F}}^{(1),-}_{m+\frac{1}{2}}}{\sigma} < \eta^{\min}_m \qquad \text{or} \qquad \eta^{*,+}_{m-\frac{1}{2}} + \frac{\widehat{\boldsymbol{F}}^{(1)}_{m-\frac{1}{2}} - \boldsymbol{\mathcal{F}}^{(1),-}_{m-\frac{1}{2}}}{\sigma} > \eta^{\max}_m,$$

do not ever happen, by assuring $\Theta_{m+\frac{1}{2}}$ stay in a certain interval. After calculation, the coefficient preserving local maximum principle writes as

$$\Theta_{m+\frac{1}{2}} \leq \min\left(1, \left|\frac{\sigma}{\Delta \boldsymbol{F}_{m+\frac{1}{2}}}\right| \begin{cases} \min\left(\beta_{m+1} - \eta^{*,+}_{m+\frac{1}{2}},\ \eta^{*,-}_{m+\frac{1}{2}} - \alpha_m\right) & \text{if } \Delta \boldsymbol{F}^{(1)}_{m+\frac{1}{2}} > 0, \\ \min\left(\beta_m - \eta^{*,-}_{m+\frac{1}{2}},\ \eta^{*,+}_{m+\frac{1}{2}} - \alpha_{m+1}\right) & \text{if } \Delta \boldsymbol{F}^{(1)}_{m+\frac{1}{2}} < 0. \end{cases}\right)$$

where $\left[\alpha_m, \beta_m\right] := \mathcal{J}\left(\overline{\eta}_{m-1}, \overline{\eta}_m, \overline{\eta}_{m+1}, \eta^{*,+}_{m-\frac{1}{2}}, \eta^{*,-}_{m+\frac{1}{2}}\right)$ and $\left[\alpha_{m+1}, \beta_{m+1}\right] := \mathcal{J}\left(\overline{\eta}_m, \overline{\eta}_{m+1}, \overline{\eta}_{m+2}, \eta^{*,+}_{m+\frac{1}{2}}, \eta^{*,-}_{m+\frac{3}{2}}\right)$, $\mathcal{J}$ being the admissible envelope associated with a finite family $(a_i)_{i\in[\![1,k]\!]} \in \mathbb{R}^k$ defined as $\mathcal{J}(a_1, \ldots, a_k) := \left[\min_{i\in[\![1,k]\!]} a_i, \max_{i\in[\![1,k]\!]} a_i\right]$. □

Remark 1.2.6 (*Spurious oscillations control*). *We also show in 1.4, through various numerical evidences, that such a constraint effectively allows to get rid of most of the spurious oscillations, while accurately preserving the high-order accuracy wherever needed, as this criterion can be interpreted as a shock sensor, or more generally as a local troubled-zone detector, which identifies the subcell interfaces where additional numerical dissipation is required.*

The issue of enforcing the water-height non-negativity deserves to be investigated while accounting for the possibly varying bathymetry to be fully understood and, as a consequence, is temporarily postponed to the next section.

Entropy considerations in the high-order/low-order blending

Entropy plays an important role in the analysis of nonlinear conservation laws, since it provides an admissibility criterion for weak solutions and a natural measure of dissipation across shocks. From a numerical point of view, entropy stability can be used for robustness. In our setting, entropy considerations can intervene through the construction of adapted blending coefficients: obtaining entropy inequalities at the scale of subcells and cells allows us to control the behaviour of the scheme both locally and at the DG polynomial level. These inequalities guide how much of the first-order flux must be introduced to guarantee stability without degrading accuracy.

During the early stages of this work, we explored in detail the entropy structure of the scheme and although this analysis is mathematically appealing, it significantly increases the technical complexity of the method. In practice, we observed that the usual shock-detection principles and the preservation of standard convex admissibility domains were sufficient to obtain a stable and robust scheme for the applications considered here. For this reason, the complete entropy-based derivations are not used directly in the final formulation and are therefore gathered in Appendix F, where the full discrete inequalities and blending coefficients are presented for completeness. We refer the interested reader to [265] for a detailed presentation of the interest of entropy stability in the local-subcell monolithic framework.

1.3 Local-subcell approximations of bathymetry source term

We investigate now the issue of the discrete approximation, *in the local-subcell framework*, of the *source terms*. We begin with the case of hyperbolic NSW equations with a bathymetry source term and focus on a compatible discretization with the preservation of motionless steady-states, and the admissibility of the discrete solution. In Chapter 2, we will introduce a novel combined local-subcell monolithic DG/FV and SWIP-DG discrete formulation for the SGN equations.

Let us consider the following NSW equations with bathymetry source term:

$$\partial_t \boldsymbol{v} + \partial_x \boldsymbol{F}(\boldsymbol{v}, b) = \boldsymbol{B}[b](\boldsymbol{v}), \tag{1.20}$$

where $\boldsymbol{B}[b]$ is defined in (1.2). For further use, we introduce the following bathymetry approximations:

$$\overline{b}_h := \widetilde{\pi}^k_h b \in \mathbb{P}^0(\widetilde{\mathcal{T}}_h) \qquad \text{and} \qquad b_h := (\overline{\pi}^k_h)^{-1} \overline{b}_h \in \mathbb{P}^k(\mathcal{T}_h). \tag{1.21}$$

A new hierarchy of numerical fluxes and two-scales hydrostatic reconstruction

The occurrence of bathymetry has a global impact on the definition of the numerical fluxes hierarchy previously introduced in the homogeneous case. To account for the bathymetry variations while preserving the accuracy and allowing the resulting global scheme to be *well-balanced*, we have to modify the definition of both high-order and low-order fluxes composing the so-called *blended fluxes*. Such a modification is inspired by the *hydrostatic reconstruction* initially introduced within FV schemes for NSW equations [12] and extended for the pre-balanced form in [107, 200]. We emphasize that the distinctiveness of the proposed approach is that such a reconstruction has to be *performed not only at the cell scale, but also at the subcell scale*, with suitable intermediate reconstructed states, detailed below:

– *Global LF numerical flux and bathymetry variations.* The chosen interface flux has to account for the bathymetry variations from cell to cell or subcell to subcell, and is defined as follows:

$$(\mathscr{H}_b^+)^2 \times \mathbb{R} \ni (\boldsymbol{v}_l, \boldsymbol{v}_r, b) \mapsto \boldsymbol{\mathcal{F}}(\boldsymbol{v}_l, \boldsymbol{v}_r, b) \coloneqq \frac{1}{2}\left(\boldsymbol{F}(\boldsymbol{v}_r, b) + \boldsymbol{F}(\boldsymbol{v}_l, b) - \sigma(\boldsymbol{v}_r - \boldsymbol{v}_l)\right), \tag{1.22}$$

– *Modified higher-order DG fluxes $\boldsymbol{\mathcal{F}}^{\mathrm{DG},\mp}_{i\pm\frac{1}{2}}$ and mesh-scale hydrostatic reconstruction.* The higher-order fluxes are now designed relying on *high-order reconstructed traces* of the bathymetry performed *at the mesh scale* (*i.e* at mesh elements interfaces). Let us define, for any mesh element $\omega_i \in \mathcal{T}_h$:

$$\widehat{b}_{i+\frac{1}{2}} \coloneqq \check{b}_{i+\frac{1}{2}} - \max(0, \check{b}_{i+\frac{1}{2}} - \eta^-_{i+\frac{1}{2}}), \qquad \widehat{b}_{i-\frac{1}{2}} \coloneqq \check{b}_{i-\frac{1}{2}} - \max(0, \check{b}_{i-\frac{1}{2}} - \eta^+_{i-\frac{1}{2}}),$$

with $\check{b}_{i\pm\frac{1}{2}} \coloneqq \max\left(b^-_{i\pm\frac{1}{2}}, b^+_{i\pm\frac{1}{2}}\right)$, and where $b^\pm_{i\pm\frac{1}{2}} \coloneqq b^i_h(x_{i\pm\frac{1}{2}})$ and $\eta^\pm_{i\pm\frac{1}{2}} \coloneqq \eta^i_h(x_{i\pm\frac{1}{2}})$. These reconstructed interface bathymetry values allow to introduce the following *reconstructed* water-height traces values:

$$\check{H}^\pm_{i+\frac{1}{2}} \coloneqq \max(0, \eta^\pm_{i+\frac{1}{2}} - \check{b}_{i+\frac{1}{2}}) \quad \text{and} \quad \check{H}^\pm_{i-\frac{1}{2}} \coloneqq \max(0, \eta^\pm_{i-\frac{1}{2}} - \check{b}_{i+\frac{1}{2}}),$$

together with the corresponding *reconstructed* flow-states $\check{\boldsymbol{v}}^\pm_{i\pm\frac{1}{2}}$, as follows:

$$\check{\boldsymbol{v}}^\pm_{i+\frac{1}{2}} \coloneqq \left(\check{\eta}^\pm_{i+\frac{1}{2}}, \check{q}^\pm_{i+\frac{1}{2}}\right)^\top = \left(\check{H}^\pm_{i+\frac{1}{2}} + \widehat{b}_{i+\frac{1}{2}}, \check{H}^\pm_{i+\frac{1}{2}} \frac{q^\pm_{i+\frac{1}{2}}}{H^\pm_{i+\frac{1}{2}}}\right)^\top,$$
$$\check{\boldsymbol{v}}^\pm_{i-\frac{1}{2}} \coloneqq \left(\check{\eta}^\pm_{i-\frac{1}{2}}, \check{q}^\pm_{i-\frac{1}{2}}\right)^\top = \left(\check{H}^\pm_{i-\frac{1}{2}} + \widehat{b}_{i-\frac{1}{2}}, \check{H}^\pm_{i-\frac{1}{2}} \frac{q^\pm_{i-\frac{1}{2}}}{H^\pm_{i-\frac{1}{2}}}\right)^\top.$$

Using these reconstructed flow-states, some new nonconservative high-order interface fluxes are defined:

$$\boldsymbol{\mathcal{F}}^{\mathrm{DG},-}_{i+\frac{1}{2}} \coloneqq \boldsymbol{\mathcal{F}}\left(\check{\boldsymbol{v}}^-_{i+\frac{1}{2}}, \check{\boldsymbol{v}}^+_{i+\frac{1}{2}}, \widehat{b}_{i+\frac{1}{2}}\right) + \begin{pmatrix} 0 \\ g\check{\eta}^-_{i+\frac{1}{2}}\left(\widehat{b}_{i+\frac{1}{2}} - b^-_{i+\frac{1}{2}})\right) \end{pmatrix}, \tag{1.23}$$

$$\boldsymbol{\mathcal{F}}^{\mathrm{DG},+}_{i-\frac{1}{2}} \coloneqq \boldsymbol{\mathcal{F}}\left(\check{\boldsymbol{v}}^-_{i-\frac{1}{2}}, \check{\boldsymbol{v}}^+_{i-\frac{1}{2}}, \widehat{b}_{i-\frac{1}{2}}\right) + \begin{pmatrix} 0 \\ g\check{\eta}^+_{i-\frac{1}{2}}\left(\widehat{b}_{i-\frac{1}{2}} - b^+_{i-\frac{1}{2}})\right) \end{pmatrix}. \tag{1.24}$$

– *Modified lowest-order subcell FV fluxes $\{\boldsymbol{\mathcal{F}}^{\mathrm{FV},\pm}\}_{m\in[\![1,k]\!]}$ and subcell-scale hydrostatic reconstruction.* The lowest-order fluxes are now defined relying on *reconstructed mean values* of the bathymetry performed *at the subcell scale* (*i.e* at subells interfaces). Let us define, for any subcell S_m:

$$\overline{b}_{m-\frac{1}{2}} \coloneqq \max\left(\overline{b}_{m-1}, \overline{b}_m\right), \qquad \overline{b}_{m+\frac{1}{2}} \coloneqq \max\left(\overline{b}_m, \overline{b}_{m+1}\right),$$

together with corresponding reconstructed values for the water-height as follows:

$$\bar{H}^\pm_m \coloneqq \max(0, \bar{\eta}_m - \overline{b}_{m\pm\frac{1}{2}}).$$

Then, for this given subcell, let us define the three intermediate reconstructed values for the bathymetry

$$\overline{b}^-_{m+1} \coloneqq \overline{b}_{m+\frac{1}{2}} - \max(0, \overline{b}_{m+\frac{1}{2}} - \bar{\eta}_m),$$
$$\overline{b}^\pm_m \coloneqq \overline{b}_{m\pm\frac{1}{2}} - \max(0, \overline{b}_{m\pm\frac{1}{2}} - \bar{\eta}_m),$$
$$\overline{b}^+_{m-1} \coloneqq \overline{b}_{m-\frac{1}{2}} - \max(0, \overline{b}_{m-\frac{1}{2}} - \bar{\eta}_m),$$

that we assemble with the corresponding three intermediate neighboring reconstructed sub-mean flow-states $\overline{\boldsymbol{v}}^{-}_{m-1}, \overline{\boldsymbol{v}}^{+}_{m}, \overline{\boldsymbol{v}}^{-}_{m+1}$ defined from reconstructed surface elevation and water-height as follows:

$$\bar{\eta}^{-}_{m+1} := \bar{H}^{-}_{m+1} + \overline{b}^{-}_{m+1}, \quad \bar{\eta}^{\pm}_{m} := \bar{H}^{\pm}_{m} + \overline{b}^{\pm}_{m}, \quad \bar{\eta}^{+}_{m-1} := \bar{H}^{+}_{m-1} + \overline{b}^{+}_{m-1}.$$

The resulting non conservative low-order fluxes defined upon these reconstructed states are given as

$$\boldsymbol{\mathcal{F}}^{\mathrm{FV},-}_{m+\frac{1}{2}} := \boldsymbol{\mathcal{F}}\left(\overline{\boldsymbol{v}}^{+}_{m}, \overline{\boldsymbol{v}}^{-}_{m+1}, \overline{b}^{+}_{m}\right) + \begin{pmatrix} 0 \\ g\bar{\eta}^{+}_{m}\left(\overline{b}^{+}_{m} - b_h(\widetilde{x}_{m+\frac{1}{2}})\right) \end{pmatrix},$$

$$\boldsymbol{\mathcal{F}}^{\mathrm{FV},+}_{m-\frac{1}{2}} := \boldsymbol{\mathcal{F}}\left(\overline{\boldsymbol{v}}^{+}_{m-1}, \overline{\boldsymbol{v}}^{-}_{m}, \overline{b}^{-}_{m}\right) + \begin{pmatrix} 0 \\ g\bar{\eta}^{-}_{m}\left(\overline{b}^{-}_{m} - b_h(\widetilde{x}_{m-\frac{1}{2}})\right) \end{pmatrix}.$$

– *New reconstructed subcell fluxes* $\{\widehat{\boldsymbol{F}}^{i,\pm}_{m+\frac{1}{2}}\}_{m\in[\![0,k+1]\!]}$. Still locally mimicking the behavior of an "in-cell" FV scheme, we obtain the following new expressions for the *reconstructed* subcell fluxes:

$$\widehat{\boldsymbol{F}}^{i,-}_{m+\frac{1}{2}} := \boldsymbol{F}_h(\widetilde{x}^{i}_{m+\frac{1}{2}}) - C^{i-\frac{1}{2}}_{m+\frac{1}{2}}\left(\boldsymbol{F}_h(x_{i-\frac{1}{2}}) - \boldsymbol{\mathcal{F}}^{\mathrm{DG},+}_{i-\frac{1}{2}}\right) - C^{i+\frac{1}{2}}_{m+\frac{1}{2}}\left(\boldsymbol{F}_h(x_{i+\frac{1}{2}}) - \boldsymbol{\mathcal{F}}^{\mathrm{DG},-}_{i+\frac{1}{2}}\right), \tag{1.25}$$

$$\widehat{\boldsymbol{F}}^{i,+}_{m-\frac{1}{2}} := \boldsymbol{F}_h(\widetilde{x}^{i}_{m-\frac{1}{2}}) - C^{i-\frac{1}{2}}_{m-\frac{1}{2}}\left(\boldsymbol{F}_h(x_{i-\frac{1}{2}}) - \boldsymbol{\mathcal{F}}^{\mathrm{DG},+}_{i-\frac{1}{2}}\right) - C^{i+\frac{1}{2}}_{m-\frac{1}{2}}\left(\boldsymbol{F}_h(x_{i+\frac{1}{2}}) - \boldsymbol{\mathcal{F}}^{\mathrm{DG},-}_{i+\frac{1}{2}}\right). \tag{1.26}$$

– *New blended subcell fluxes* $\left\{\widetilde{\boldsymbol{F}}^{i,\pm}_{m+\frac{1}{2}}\right\}_{m\in[\![0,k+1]\!]}$. The *blended* fluxes at the interfaces of cell ω_i are naturally introduced as:

$$\widetilde{\boldsymbol{F}}^{i,-}_{m+\frac{1}{2}} := \boldsymbol{\mathcal{F}}^{\mathrm{FV},-}_{m+\frac{1}{2}} + \Theta_{m+\frac{1}{2}}\left(\widehat{\boldsymbol{F}}^{i,-}_{m+\frac{1}{2}} - \boldsymbol{\mathcal{F}}^{\mathrm{FV},-}_{m+\frac{1}{2}}\right), \tag{1.27}$$

$$\widetilde{\boldsymbol{F}}^{i,+}_{m-\frac{1}{2}} := \boldsymbol{\mathcal{F}}^{\mathrm{FV},+}_{m-\frac{1}{2}} + \Theta_{m-\frac{1}{2}}\left(\widehat{\boldsymbol{F}}^{i,+}_{m-\frac{1}{2}} - \boldsymbol{\mathcal{F}}^{\mathrm{FV},+}_{m-\frac{1}{2}}\right). \tag{1.28}$$

Local-subcell DG/FV equivalency and monolithic approximation

In order to enforce the well-balancing property for the important class of motionless steady-state solutions, we initially consider the following DG formulation for the NSW equations with bathymetry source term, which relies on the *modified* high-order fluxes (1.23): for all $\omega_i \in \mathcal{T}_h$, find $\boldsymbol{v}^{i}_{h}(\cdot,t) := (\eta^{i}_{h}, q^{i}_{h}) \in (\mathbb{P}^k(\omega_i))^2$ such that, for all $\varphi \in \mathbb{P}^k(\omega_i)$:

$$\int_{\omega_i} \partial_t \boldsymbol{v}^{i}_{h}\varphi \,\mathrm{d}x - \int_{\omega_i} \boldsymbol{F}_h \partial_x\varphi \,\mathrm{d}x + \left[\varphi\boldsymbol{\mathcal{F}}^{\mathrm{DG},\mp}\right]^{i+\frac{1}{2}}_{i-\frac{1}{2}} = \int_{\omega_i} \varphi\, \boldsymbol{B}_h \,\mathrm{d}x, \tag{1.29}$$

with interface fluxes shortcut defined as:

$$\left[\varphi\boldsymbol{\mathcal{F}}^{\mathrm{DG},\mp}\right]^{i+\frac{1}{2}}_{i-\frac{1}{2}} := \varphi_{|x_{i+\frac{1}{2}}}\boldsymbol{\mathcal{F}}^{\mathrm{DG},-}_{i+\frac{1}{2}} - \varphi_{|x_{i-\frac{1}{2}}}\boldsymbol{\mathcal{F}}^{\mathrm{DG},+}_{i-\frac{1}{2}},$$

and $\boldsymbol{F}_h := \pi^k_h \boldsymbol{F}(\boldsymbol{v}_h, b_h)$, $\boldsymbol{B}_h := \pi^k_h \boldsymbol{B}[b_h](\boldsymbol{v}_h)$ the L^2-orthogonal projections of the flux and source term functions onto the piecewise polynomial space $\mathbb{P}^k(\mathcal{T}_h)$ such that for all $\varphi \in \mathbb{P}^k(\mathcal{T}_h)$,

$$\left(\boldsymbol{F}_h, \varphi\right)_{\mathcal{T}_h} = \left(\boldsymbol{F}(\boldsymbol{v}_h, b_h), \varphi\right)_{\mathcal{T}_h}, \qquad \left(\boldsymbol{B}_h, \varphi\right)_{\mathcal{T}_h} = \left(\boldsymbol{B}[b_h](\boldsymbol{v}_h), \varphi\right)_{\mathcal{T}_h}. \tag{1.30}$$

Remark 1.3.1 (*Link between quadrature and pre-balanced formulation*). *We need that the volume and source term integrals in* (1.29) *are exactly computed at motionless steady states. This is achieved, thanks to the pre-balanced formulation, by using any quadrature rule exact for polynomials of degree up to* $2k$*. On the other hand, for a classical NSW formulation, a quadrature rule exact for polynomials of degree up to* $3k$ *is needed. Let us note that we must use the same quadrature rule for the computation of the right hand sides in* (1.30).

Let us also define $\overline{\boldsymbol{B}}_h := \overline{\pi}^k_h \boldsymbol{B}_h$, piecewise constant function associated with the mean-values of $\boldsymbol{B}_h$ on the sub-mesh. Interestingly, such a local modified formulation is still equivalent with a FV-like formulation on the sub-mesh, with a source term computed at the sub-mesh scale:

Proposition 1.3.2 (*Local-subcell DG/FV equivalency with bathymetry source term*). *For all* $\omega_i \in \mathcal{T}_h$*, the local DG formulation* (1.29) *can be equivalently written as a FV-like scheme defined on the sub-mesh as follows:*

$$\frac{\mathrm{d}\overline{\boldsymbol{v}}^{i}_{m}}{\mathrm{d}t} = -\frac{1}{|S^{i}_{m}|}\left(\widehat{\boldsymbol{F}}^{i,-}_{m+\frac{1}{2}} - \widehat{\boldsymbol{F}}^{i,+}_{m-\frac{1}{2}}\right) + \overline{\boldsymbol{B}}^{i}_{m}, \qquad \forall m \in [\![1, k+1]\!], \tag{1.31}$$

where $\overline{\boldsymbol{B}}^{i}_{m} = \overline{\boldsymbol{B}}_{h|S^{i}_{m}}$ *and the reconstructed fluxes are defined in* (1.25)-(1.26).

Proof. We can write, for all $\varphi \in \mathbb{P}^k(\mathcal{T}_h)$,

$$\sum_{\omega_i\in\mathcal{T}_h}\int_{\omega_i}\partial_t \boldsymbol{v}_h\varphi\,\mathrm{d}x-\sum_{\omega_i\in\mathcal{T}_h}\int_{\omega_i}\boldsymbol{F}_h\partial_x\varphi\,\mathrm{d}x+\sum_{\omega_i\in\mathcal{T}_h}\left[\varphi\boldsymbol{\mathcal{F}}\right]_{\partial\omega_i}=\sum_{\omega_i\in\mathcal{T}_h}\int_{\omega_i}\boldsymbol{B}_h\varphi\,\mathrm{d}x,$$

or equivalently, using an integration by parts:

$$\sum_{\omega_i\in\mathcal{T}_h}\int_{\omega_i}\partial_t \boldsymbol{v}_h\varphi\,\mathrm{d}x+\sum_{\omega_i\in\mathcal{T}_h}\int_{\omega_i}\partial_x\boldsymbol{F}_h\varphi\,\mathrm{d}x-\sum_{\omega_i\in\mathcal{T}_h}\left[\varphi(\boldsymbol{F}_h-\boldsymbol{\mathcal{F}})\right]_{\partial\omega_i}=\sum_{\omega_i\in\mathcal{T}_h}\int_{\omega_i}\boldsymbol{B}_h\varphi\,\mathrm{d}x. \tag{1.32}$$

Substituting φ by the sub-resolution basis functions (1.13) into (1.32) gives us the local equations on $\omega_i \in \mathcal{T}_h$:

$$\int_{\omega_i}\partial_t \boldsymbol{v}_h^i\phi_m^i\mathrm{d}x=-\int_{\omega_i}\partial_x\boldsymbol{F}_h^i\phi_m^i\mathrm{d}x+\int_{\omega_i}\boldsymbol{B}_h^i\phi_m^i\mathrm{d}x+\left[\phi_m^i(\boldsymbol{F}_h^i-\boldsymbol{\mathcal{F}})\right]_{\partial\omega_i},\qquad\forall m\in[\![1,k+1]\!].$$

Since $\partial_t\boldsymbol{v}^i,\partial_x\boldsymbol{F}^i,\boldsymbol{B}^i\in(\mathbb{P}^k(\omega))^2$ and using the property of ϕ_m^i, it follows that

$$\partial_t\overline{\boldsymbol{v}}_i=-\frac{1}{|S_m^i|}\left(\left[\boldsymbol{F}_i\right]_{\partial S_m^i}-\left[\phi_m^i(\boldsymbol{F}_i-\boldsymbol{\mathcal{F}})\right]_{\partial\omega_i}\right)+\overline{\boldsymbol{B}}_m^i,\qquad\forall m\in[\![1,k+1]\!],$$

where $\overline{\boldsymbol{v}}_m^i=(\overline{\eta}_m^i,\overline{q}_m^i)^\top$ and $\overline{\boldsymbol{B}}_m^i$ are respectively the mean values of $\boldsymbol{v}^i$ and $\boldsymbol{B}^i$ on the subcell S_m^i, i.e.

$$\overline{\boldsymbol{v}}_m^i:=\frac{1}{|S_m^i|}\int_{S_m^i}\boldsymbol{v}_h^i\mathrm{d}x,\qquad\overline{\boldsymbol{B}}_m^i:=\frac{1}{|S_m^i|}\int_{S_m^i}\boldsymbol{B}_h^i\mathrm{d}x.$$

Let us now introduce the $k+2$ subcells interfaces fluxes $\left\{\widehat{\boldsymbol{F}}_{m+\frac12}^i\right\}_{m\in[\![0,k+1]\!]}$ such that

$$\widehat{\boldsymbol{F}}_{m+\frac12}^i-\widehat{\boldsymbol{F}}_{m-\frac12}^i=\left[\boldsymbol{F}_h^i\right]_{\partial S_m^i}-\left[\phi_m^i(\boldsymbol{F}_h^i-\boldsymbol{\mathcal{F}})\right]_{\partial\omega_i},\qquad\forall m\in[\![1,k+1]\!],$$

such that we have

$$\partial_t\overline{\boldsymbol{v}}_i=-\frac{1}{|S_m^i|}\left(\widehat{\boldsymbol{F}}_{m+\frac12}^i-\widehat{\boldsymbol{F}}_{m-\frac12}^i\right)+\overline{\boldsymbol{B}}_m^i,\qquad\forall m\in[\![1,k+1]\!]. \tag{1.33}$$

The formulation (1.33) can then be seen as a FV-like scheme on subcell S_m^i. □

As in the homogeneous case, the quest for nonlinear stability leads to the introduction of a new local-subcell DG/FV monolithic approximation, written with a simple first-order time-stepping scheme for the sake of simplicity:

$$\overline{\boldsymbol{v}}_m^{i,n+1}=\overline{\boldsymbol{v}}_m^{i,n}-\frac{\Delta t^n}{|S_m^i|}\left(\widetilde{\boldsymbol{F}}_{m+\frac12}^{i,-}-\widetilde{\boldsymbol{F}}_{m-\frac12}^{i,+}\right)+\overline{\boldsymbol{B}}_m^i,\qquad\forall(m,i)\in[\![1,k+1]\!]\times[\![1,N_{\mathrm{el}}]\!], \tag{1.34}$$

where the new *blended* fluxes $\widetilde{\boldsymbol{F}}_{m\pm\frac12}^{i,\pm}$ are defined in (1.27). Non-linear stability therefore relies on the definition of suitable blending coefficients.

Blended coefficients and water-height non-negativity

An admissible approximate solution is defined with respect to the non-negativity of the sub-mean values:

Definition 1.3.3 (*Admissible approximate solution*). *For any discrete time $n\geq 0$, we say that an approximate state $\boldsymbol{v}_h(\cdot,t^n)\in\mathbf{P}^k(\mathcal{T}_h)^2$, or equivalently $\overline{\boldsymbol{v}}_h^n:=\overline{\pi}_h^k\boldsymbol{v}_h(\cdot,t^n)\in(\mathbb{P}^0(\widetilde{\mathcal{T}}_h))^2$, is admissible if $\overline{\boldsymbol{v}}_m^{i,n}$ is such that $\overline{\eta}_m^{i,n}-\overline{b}_m^i\geq 0$, for all $\omega_i\in\mathcal{T}_h$ and for all $S_m^i\in\mathcal{T}_{\omega_i}^k$, i.e. $\overline{\boldsymbol{v}}_m^{i,n}\in\overline{\mathscr{H}}_b^+$, where $\overline{\mathscr{H}}_b^+$ is the set of admissible states defined as follows:*

$$\overline{\mathscr{H}}_b^+=\left\{\overline{\boldsymbol{v}}_h\in(\mathbb{P}^0(\widetilde{\mathcal{T}}_h))^2\mid\overline{H}_h:=\overline{\eta}_h-\overline{b}_h\geq 0\right\}.$$

A definition of uniquely defined interface-related blending coefficients $\Theta_{m+\frac12}^{\mathscr{H}^+}$ that ensure the preservation of water-height non-negativity is provided below:

Proposition 1.3.4 (*Water-height positivity preservation*). *Let us consider the local-subcell monolithic update* (1.34). *Assume that* $\overline{\boldsymbol{v}}_h^n$ *is admissible and that* Δt^n *satisfies the CFL-like condition* (1.7). *For each subcell interface, set*

$$\Delta \boldsymbol{F}_{m+\frac{1}{2}}^{(1),\pm} := \widehat{\boldsymbol{F}}_{m+\frac{1}{2}}^{\pm,(1)} - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}}^{\mathrm{FV},\pm,(1)}.$$

Then $\overline{\boldsymbol{v}}_h^{n+1}$ *is admissible provided that* $\Theta_{m+\frac{1}{2}} \leq \Theta_{m+\frac{1}{2}}^{\mathscr{H}^+}$, *with*

$$\Theta_{m+\frac{1}{2}}^{\mathscr{H}^+} := \min\left(1, \Theta_{m+\frac{1}{2}}^{\mathscr{H}^+,-}, \Theta_{m+\frac{1}{2}}^{\mathscr{H}^+,+}\right),$$

where the two quantities $\Theta_{m+\frac{1}{2}}^{\mathscr{H}^+,\pm}$ *are imposed only when the corresponding constraint is active:*

$$\Theta_{m+\frac{1}{2}}^{\mathscr{H}^+,-} = \frac{\sigma\left(\eta_{m+\frac{1}{2}}^{*,-} - \overline{b}_m\right)}{\Delta \boldsymbol{F}_{m+\frac{1}{2}}^{(1),-}} \quad \text{if } \Delta \boldsymbol{F}_{m+\frac{1}{2}}^{(1),-} > 0, \qquad \Theta_{m+\frac{1}{2}}^{\mathscr{H}^+,+} = \frac{\sigma\left(\overline{b}_{m+1} - \eta_{m+\frac{1}{2}}^{*,+}\right)}{\Delta \boldsymbol{F}_{m+\frac{1}{2}}^{(1),+}} \quad \text{if } \Delta \boldsymbol{F}_{m+\frac{1}{2}}^{(1),+} < 0,$$

and are otherwise set equal to 1.

Proof. The proof is similar to the one of Proposition 1.2.5, ensuring that the inequalities

$$\eta_{m+\frac{1}{2}}^{*,-} - \frac{\widehat{\boldsymbol{F}}_{m+\frac{1}{2}}^{-,(1)} - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}}^{\mathrm{FV},-,(1)}}{\sigma} \geq \overline{b}_m \qquad \text{and} \qquad \eta_{m+\frac{1}{2}}^{*,+} + \frac{\widehat{\boldsymbol{F}}_{m+\frac{1}{2}}^{+,(1)} - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}}^{\mathrm{FV},+,(1)}}{\sigma} \geq \overline{b}_{m+1},$$

are uniformly satisfied. Let us remark that this is ensured only if the 1^{st}-order Riemann intermediate states are also preserving water-height positivity at the discrete level, *i.e* if $\eta_{m+\frac{1}{2}}^{*,-} \geq \overline{b}_m$ and $\eta_{m+\frac{1}{2}}^{*,+} \geq \overline{b}_{m+1}$, which is demonstrated in Appendix B. □

Remark 1.3.5 (*Practical choice of blending coefficients*). *In practice, the formulas defining the blending coefficients provide sufficient conditions only. We therefore choose* $\Theta_{m+\frac{1}{2}} := \min\left(\Theta_{m+\frac{1}{2}}^{\mathscr{H}^+}, \Theta_{m+\frac{1}{2}}^{\mathrm{nad}}\right)$, *which ensures both water-height non-negativity and the control of spurious oscillations near singularities.*

Initial-data approximations

After Proposition 1.3.4, it appears that the approximate initial-data $\boldsymbol{v}_h^0$ has to be admissible in order to initiate a sequence of admissible approximate states, relying on the suitable CFL-condition and blending coefficients. Hence, starting from some physical or idealized (analytical) initial-condition $\boldsymbol{v}^0 := (\eta^0, q^0)$ and bathymetry parameterization b such that $\eta^0 - b \geq 0$, we observe that $\widetilde{\pi}_h^k(\eta^0 - b) \geq 0$, hence $\overline{\boldsymbol{v}}_h^0 := \widetilde{\pi}_h^k \boldsymbol{v}^0$ is admissible. In an equivalent way, we also have that $\boldsymbol{v}_h^0 := (\overline{\pi}_h^k)^{-1}\overline{\boldsymbol{v}}_h^0$ is admissible. From this last formula, we observe that the initial approximate surface-elevation has to be defined is a consistent way with the previous definition of the approximate bathymetry:

$$\eta_h^0 := (\overline{\pi}_h^k)^{-1}\widetilde{\pi}_h^k \eta^0 - b_h. \tag{1.35}$$

Let us also emphasize that when initializing with some data corresponding to a motionless steady-state $\boldsymbol{v}_h = (\eta_h, q_h) := (\eta^e, 0)$, then $\boldsymbol{F}(\boldsymbol{v}_h, b_h)$ belongs to $\mathbf{P}^k(\mathcal{T}_h)^2$ and as a consequence, we have:

$$\boldsymbol{F}_h := \pi_h^k \boldsymbol{F}(\boldsymbol{v}_h, b_h) = \boldsymbol{F}(\boldsymbol{v}_h, b_h).$$

In a similar way, we also have $\boldsymbol{B}_h := \boldsymbol{B}[b_h](\boldsymbol{v}_h)$.

Well-balancing property for motionless steady-states

It is well-known that the preservation of motionless steady-states is a key property for the numerical approximation of shallow-water flows with bathymetry. We show below that the monolithic scheme (1.34) preserves these important steady-states, thanks to the consistency between the source term and initial-data approximations, and to the local-subcell DG/FV equivalency.

Proposition 1.3.6 (*Motionless steady-states preservation*). *The discrete formulation* (1.34) *preserves the motionless steady states:*

$$\forall n \in \mathbb{N}, \qquad \left(\eta_h^n = \eta^e \text{ and } q_h^n = 0\right) \Longrightarrow \left(\eta_h^{n+1} = \eta^e \text{ and } q_h^{n+1} = 0\right).$$

Proof. Let us rewrite the FV-like scheme as follows:

$$\overline{\boldsymbol{v}}_m^{i,n+1} = \overline{\boldsymbol{v}}_m^{i,n} - \frac{\Delta t^n}{|S_m^i|}\Big(\boldsymbol{\mathcal{F}}_{m+\frac{1}{2}}^{i,n,-} + \Theta_{m+\frac{1}{2}}^{i,n}\left(\widehat{\boldsymbol{F}}_{m+\frac{1}{2}}^{i,n,-} - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}}^{i,n,-}\right) - \boldsymbol{\mathcal{F}}_{m-\frac{1}{2}}^{i,n,+} + \Theta_{m-\frac{1}{2}}^{i,n}\left(\widehat{\boldsymbol{F}}_{m-\frac{1}{2}}^{i,n,+} - \boldsymbol{\mathcal{F}}_{m-\frac{1}{2}}^{i,n,+}\right)\Big) + \Delta t^n \overline{\boldsymbol{B}}_m^{i,n}. \tag{1.36}$$

If one considers the case $\Theta_{m+\frac{1}{2}}^{i,n} = \Theta_{m-\frac{1}{2}}^{i,n} = 1$, then the monolithic scheme is equivalent to a DG scheme in its subcell formulation:

$$\overline{\boldsymbol{v}}_m^{i,n+1} = \overline{\boldsymbol{v}}_m^{i,n} - \frac{\Delta t^n}{|S_m^i|}\left(\widehat{\boldsymbol{F}}_{m+\frac{1}{2}}^{i,n} - \widehat{\boldsymbol{F}}_{m-\frac{1}{2}}^{i,n}\right) + \Delta t^n \overline{\boldsymbol{B}}_m^i.$$

Dropping the (i,n) superscripts for clarity, at the cell scale and at steady-state, the following identities are satisfied:

$$\eta_{i\pm\frac{1}{2}}^{+} = \eta_{i\pm\frac{1}{2}}^{-} = \eta^e, \qquad q_{i\pm\frac{1}{2}}^{+} = q_{i\pm\frac{1}{2}}^{-} = 0, \qquad b_{i\pm\frac{1}{2}}^{+} = b_{i\pm\frac{1}{2}}^{-}.$$

Then the high-order numerical flux on cells $\boldsymbol{\mathcal{F}}^{\mathrm{DG},\pm}$ verifies

$$\boldsymbol{\mathcal{F}}_{i\pm\frac{1}{2}}^{\mathrm{DG},\mp} = \boldsymbol{F}\left(\boldsymbol{v}_h(x_{i\pm\frac{1}{2}}), b_h(x_{i\pm\frac{1}{2}})\right),$$

such that

$$\widehat{\boldsymbol{F}}_{m\pm\frac{1}{2}}^{\mp} = \boldsymbol{F}\left(\boldsymbol{v}_h(\widetilde{x}_{m\pm\frac{1}{2}}), b_h(\widetilde{x}_{m\pm\frac{1}{2}})\right).$$

Consequently we get:

$$\widehat{\boldsymbol{F}}_{m+\frac{1}{2}}^{-} - \widehat{\boldsymbol{F}}_{m-\frac{1}{2}}^{+} = \int_{S_m} \partial_x \boldsymbol{F}\left(\boldsymbol{v}_h, b_h\right)\, \mathrm{d}x,$$

and we finally have

$$\overline{\boldsymbol{v}}_m^{i,n+1} = \overline{\boldsymbol{v}}_m^{i,n} - \frac{\Delta t^n}{|S_m^i|}\int_{S_m^i} \partial_x \boldsymbol{F}\left(\boldsymbol{v}_h, b_h\right)\, \mathrm{d}x + \frac{\Delta t^n}{|S_m^i|}\int_{S_m^i} \boldsymbol{B}(\boldsymbol{v}_h, \partial_x b_h)\, \mathrm{d}x = \overline{\boldsymbol{v}}_m^{i,n}.$$

If we consider the first-order case, i.e. $\Theta_{m+\frac{1}{2}}^{i,n} = \Theta_{m-\frac{1}{2}}^{i,n} = 0$, we get the following first-order scheme with subcell FV fluxes:

$$\overline{\boldsymbol{v}}_m^{i,n+1} = \overline{\boldsymbol{v}}_m^{i,n} - \frac{\Delta t^n}{|S_m^i|}\left(\boldsymbol{\mathcal{F}}_{m+\frac{1}{2}}^{i,n,-} - \boldsymbol{\mathcal{F}}_{m-\frac{1}{2}}^{i,n,+}\right) + \Delta t^n \overline{\boldsymbol{B}}_m^i.$$

At the subcell scale, we also have at steady state: $\overline{\eta}_m^{+} = \overline{\eta}_{m+1}^{-} = \eta^e$ and $\overline{q}_m^{+} = \overline{q}_{m+1}^{-} = 0$, leading to

$$\begin{aligned}
\boldsymbol{\mathcal{F}}\left(\overline{\boldsymbol{v}}_m^{+}, \overline{\boldsymbol{v}}_{m+1}^{-}, \overline{b}_m^{+}\right) &= \frac{1}{2}\left(\boldsymbol{F}\left(\overline{\boldsymbol{v}}_m^{+}, \overline{b}_m^{+}\right) + \boldsymbol{F}\left(\overline{\boldsymbol{v}}_{m+1}^{-}, \overline{b}_m^{+}\right)\right) - \frac{\sigma}{2}\left(\overline{\boldsymbol{v}}_{m+1}^{-} - \overline{\boldsymbol{v}}_m^{+}\right) \\
&= \frac{1}{2}\left(\boldsymbol{F}\left(\overline{\boldsymbol{v}}_m^{+}, \overline{b}_m^{+}\right) + \boldsymbol{F}\left(\overline{\boldsymbol{v}}_{m+1}^{-}, \overline{b}_m^{+}\right)\right) \\
&= \frac{1}{2}\begin{pmatrix} 0 \\ g\left((\eta^e)^2 - 2\eta^e \overline{b}_m^{+}\right)\end{pmatrix},
\end{aligned}$$

and then to:

$$\begin{aligned}
\boldsymbol{\mathcal{F}}_{m+\frac{1}{2}}^{-} &= \frac{1}{2}\begin{pmatrix} 0 \\ g\left((\eta^e)^2 - 2\eta^e \overline{b}_m^{+}\right)\end{pmatrix} + \begin{pmatrix} 0 \\ g\overline{\eta}_m^{+}\left(\overline{b}_m^{+} - 2\eta^e b_h(\widetilde{x}_{m+\frac{1}{2}})\right)\end{pmatrix} = \frac{1}{2}\begin{pmatrix} 0 \\ g\left((\eta^e)^2 - 2\eta^e b_h(\widetilde{x}_{m+\frac{1}{2}})\right)\end{pmatrix} \\
&= \boldsymbol{F}\left(\boldsymbol{v}_h(\widetilde{x}_{m+\frac{1}{2}}), b_h(\widetilde{x}_{m+\frac{1}{2}})\right),
\end{aligned}$$

and in a similar way

$$\boldsymbol{\mathcal{F}}_{m-\frac{1}{2}}^{+} = \boldsymbol{F}\left(\boldsymbol{v}_h(\widetilde{x}_{m-\frac{1}{2}}), b_h(\widetilde{x}_{m-\frac{1}{2}})\right).$$

It is then easy to see, as previously, that the fluxes are cancelling the source term exactly:

$$\overline{\boldsymbol{v}}_m^{i,n+1} = \overline{\boldsymbol{v}}_m^{i,n} - \frac{\Delta t^n}{|S_m^i|}\int_{S_m^i} \partial_x \boldsymbol{F}\left(\boldsymbol{v}_h, b_h\right)\,\mathrm{d}x + \frac{\Delta t^n}{|S_m^i|}\int_{S_m^i} \boldsymbol{B}(\boldsymbol{v}_h, \partial_x b_h)\,\mathrm{d}x = \overline{\boldsymbol{v}}_m^{i,n}.$$

If we now consider the more general case where $\left(\Theta^{i,n}_{m-\frac{1}{2}}, \Theta^{i,n}_{m+\frac{1}{2}}\right) \in [0,1]^2$, the following identities are still satisfied:

$$\begin{cases} \boldsymbol{\mathcal{F}}^{i,n,\mp}_{m\pm\frac{1}{2}} = \boldsymbol{F}\left(\boldsymbol{v}_h(\widetilde{x}_{m\pm\frac{1}{2}}), b_h(\widetilde{x}_{m\pm\frac{1}{2}})\right), \\ \widehat{\boldsymbol{F}}^{i,n,\mp}_{m\pm\frac{1}{2}} = \boldsymbol{F}\left(\boldsymbol{v}_h(\widetilde{x}_{m\pm\frac{1}{2}}), b_h(\widetilde{x}_{m\pm\frac{1}{2}})\right). \end{cases}$$

leading to $\widehat{\boldsymbol{F}}^{i,n,\mp}_{m\pm\frac{1}{2}} = \boldsymbol{\mathcal{F}}^{i,n,\mp}_{m\pm\frac{1}{2}}$ at steady-state. We finally get

$$\begin{aligned} \overline{\boldsymbol{v}}_m^{i,n+1} &= \overline{\boldsymbol{v}}_m^{i,n} - \frac{\Delta t^n}{|S_m^i|}\left(\boldsymbol{\mathcal{F}}^{i,n,-}_{m+\frac{1}{2}} - \boldsymbol{\mathcal{F}}^{i,n,+}_{m-\frac{1}{2}} + \Theta^{i,n}_{m+\frac{1}{2}}\left(\widehat{\boldsymbol{F}}^{i,n,-}_{m+\frac{1}{2}} - \boldsymbol{\mathcal{F}}^{i,n,-}_{m+\frac{1}{2}}\right) - \Theta^{i,n}_{m-\frac{1}{2}}\left(\widehat{\boldsymbol{F}}^{i,n,+}_{m-\frac{1}{2}} - \boldsymbol{\mathcal{F}}^{i,n,+}_{m-\frac{1}{2}}\right)\right) + \Delta t^n \overline{\boldsymbol{B}}_m^{i,n} \\ &= \overline{\boldsymbol{v}}_m^{i,n} - \frac{\Delta t^n}{|S_m^i|}\left(\boldsymbol{F}\left(\boldsymbol{v}_h(\widetilde{x}_{m+\frac{1}{2}}), b_h(\widetilde{x}_{m+\frac{1}{2}})\right) - \boldsymbol{F}\left(\boldsymbol{v}_h(\widetilde{x}_{m-\frac{1}{2}}), b_h(\widetilde{x}_{m-\frac{1}{2}})\right)\right) + \Delta t^n \overline{\boldsymbol{B}}_m^{i,n} \\ &= \overline{\boldsymbol{v}}_m^{i,n} - \frac{\Delta t^n}{|S_m^i|}\int_{S_m^i} \partial_x \boldsymbol{F}\left(\boldsymbol{v}_h, b_h\right)\,\mathrm{d}x + \frac{\Delta t^n}{|S_m^i|}\int_{S_m^i} \boldsymbol{B}(\boldsymbol{v}_h, \partial_x b_h)\,\mathrm{d}x \\ &= \overline{\boldsymbol{v}}_m^{i,n}. \end{aligned}$$

which concludes the proof. □

1.4 Numerical validations

In the following, we refer to « pure DG » when we are forcing our blending coefficient $\Theta = 1$ (fully equivalent to classical non-stabilized DG schemes). If not stated otherwise, we use in most of our numerical simulations the relaxed SubNAD and the PAD criteria.

1.4.1 Smooth sinusoidal solutions

Periodic solution

PARAMETERS

Computational domain: $\Omega = [0,1]$
Number of cells: $N_{\text{el}} = 50$ cells
Boundary conditions: periodic

Polynomial degree: $k = 1, 2, 3$
Time-marching order: $p = 3$
Final time: $T_{\max} = 0.3, 0.6$ sec

This first benchmark aims to assess the convergence rates of the present monolithic DG/FV scheme. To do so, we follow the methodoly introduced in [266] and [140], which permits to construct a familly of smooth analytical solutions. We then make use of the following initial data:

$$\eta_0 = \frac{u_0^2}{4g} \quad \text{and} \quad q_0 = \frac{u_0^3}{4g},$$

where the initial constant velocity is perturbed by a sinusoidal signal:

$$u_0(x) = 1 + 0.1\sin(2\pi x).$$

The Table 1.1 summarizes the global L^2-errors and convergence rates for various orders of approximation, computed on the surface elevation η at $t = 0.3$ sec. As anticipated, the observed convergence rates scale as $\mathcal{O}(k+1)$ for pure DG and SubNAD criteria. Similar behavior is observed for the horizontal discharge q.

Sinusoidal solution on a dry-bed

PARAMETERS

Computational domain: $\Omega = [0,1]$
Number of cells: $N_{\text{el}} = 50, 10$ cells
Boundary conditions: homogeneous Neumann

Polynomial degree: $k = 3, 8$
Time-marching order: $p = 3$
Final time: $T_{\max} = 0.09$ sec

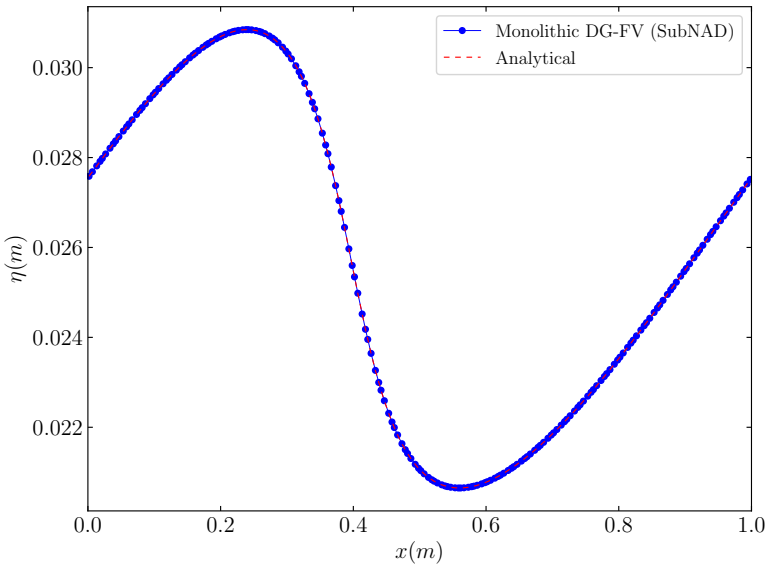


Figure 1.4: Periodic sinusoidal solution − Free surface elevation computed at $T_{\max} = 0.6$ for $k = 3$ and $N_{\text{el}} = 50$

k	1		2		3	
h	$E^{\eta}_{L^2}$	$q^{\eta}_{L^2}$	$E^{\eta}_{L^2}$	$q^{\eta}_{L^2}$	$E^{\eta}_{L^2}$	$q^{\eta}_{L^2}$
0.025	6.110E-6	2.03	1.491E-7	3.01	2.579E-9	3.98
0.012	1.477E-6	2.01	1.774E-8	3.00	1.561E-10	4.01
0.009	3.642E-7	2.01	5.094E-9	3.00	3.052E-11	4.00
0.006	1.64E-7	–	2.106E-9	–	9.609E-12	–

Table 1.1: Periodic sinusoidal solution − L^2-errors between numerical and analytical solutions and convergence rates for η at time $t = 0.3$ sec.

An other test-case obtained in a similar way is presented to test the robustness of our scheme, in particular the conservation of water-height positivity. To do so, we follow again the methodology as in the previous example, and then still initialize with the following:

$$\eta_0 = \frac{u_0^2}{4g} \quad \text{and} \quad q_0 = \frac{u_0^3}{4g},$$

where the initial velocity is

$$u_0(x) = \begin{cases} 2 & \text{if } x \leq \frac{1}{4}, \\ 1 + \sin(2\pi x) & \text{if } x \in \left[\frac{1}{4}, \frac{3}{4}\right], \\ 0 & \text{else} \end{cases}$$

The analytical velocity is then defined for all $t \in [0, t_c[$, where the critical time at which the characteristic lines cross is $t_c := \frac{1}{3\pi}$:

$$u(x,t) = \begin{cases} 2 & \text{if } x \leq \frac{1}{4} + 3t, \\ u_0(X) & \text{if } x \in \left[\frac{1}{4} + 3t, \frac{3}{4}\right], \\ 0 & \text{else} \end{cases}$$

where the characteristic lines read $x(X,t) = \frac{3}{4}u_0(X)t + X$. For practical applications, to assess the position of the characteristic line origin point X given x and t, we use an iterative root-finding process, as Newton's method, to solve the nonlinear problem $g(X) = 0$, where for given x and t, the function is $g(X) := \frac{3}{4}u_0(X)t + X - x$.

We present the results in two different frameworks: lower order with more cells ($k = 3$, $N_{\text{el}} = 50$) in Figure 1.5, and higher order with less cells ($k = 8$, $N_{\text{el}} = 10$) in Figure 1.6. While the *pure* DG scheme allows to

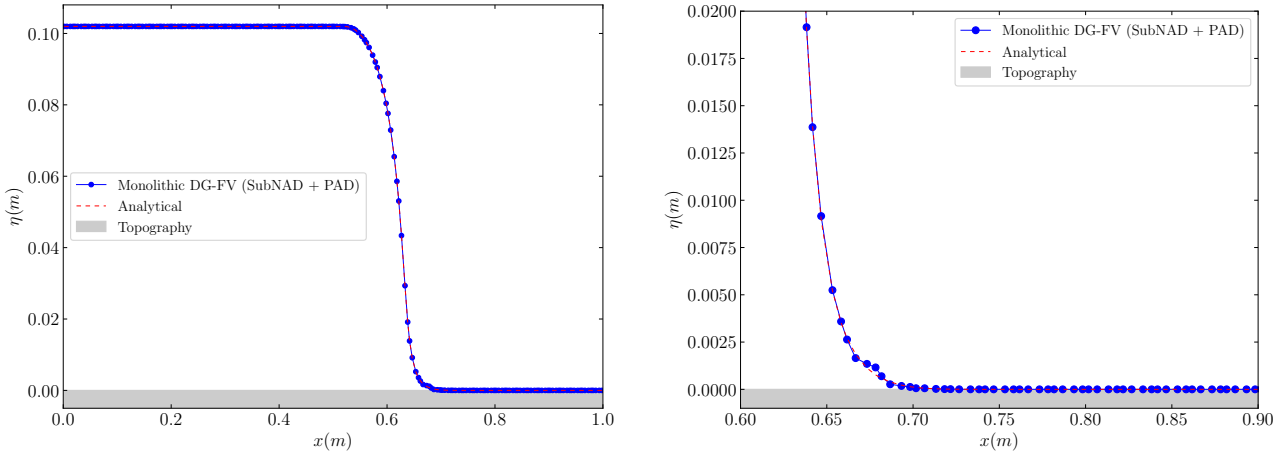


Figure 1.5: Sinusoidal solution on dry bed − Free surface elevation computed at $T_{\max} = 0.09$ sec for $k = 3$ and $n_{\text{el}} = 50$, with a zoom on dry-bed.

compute the solution without any robustness issue for small enough values of time, nonphysical oscillations may be generated for larger values of time, and so we can easily notice the efficiency of our method with only few cells for handling dry bed, especially when we go on higher orders.

1.4.2 Dam-break over flat bottoms

Dam-break test cases are interesting in shallow-water context because they provide a rigorous benchmark for validating the accuracy and robustness of numerical schemes in capturing rapidly varying flows and shock waves.

Dam-break on a wet-bed

PARAMETERS

Computational domain: $\Omega = [0, 1]$
Number of cells: $N_{\text{el}} = 50, 10$ cells
Boundary conditions: homogeneous Neumann

Polynomial degree: $k = 3, 9$
Time-marching order: $p = 3$
Final time: $T_{\max} = 0.075$ sec

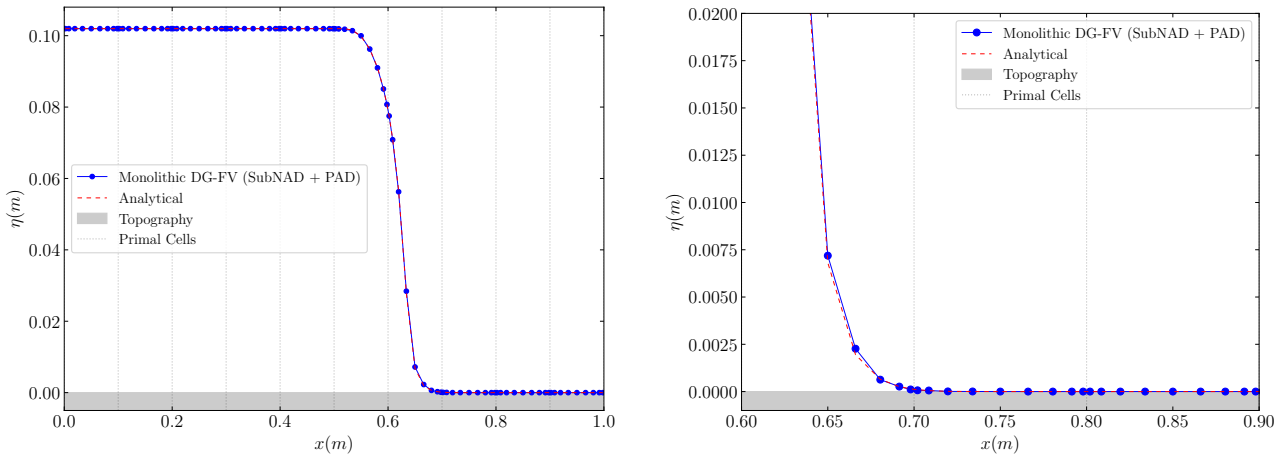


Figure 1.6: Sinusoidal solution on dry bed − Free surface elevation computed at $T_{\max} = 0.09\,s$ for $k = 8$ and $n_{\text{el}} = 10$ (left), with a zoom on the wet-dry transition (right).

The first set of initial conditions is defined as follows:

$$\eta_0(x) = \begin{cases} 1 & \text{if } x \leq 0.5, \\ 0.5 & \text{else,} \end{cases} \qquad q_0 = 0, \qquad b = 0.$$

We run the simulation on Figure 1.7 with fourth-order solution and 50 elements, with SubNAD criterion to get rid of non-physical oscillations. Figure 1.9 propose the same solution at a very higher order ($k = 9$) with only 10 cells. One can still note the efficiency of this method to retain the precise subcell resolution of DG schemes, allowing the use of very coarse meshes, along with being able to avoid the appearance of spurious oscillations or any unfortunate crash in the computation. This figure also displays how the present correction affects the solution only at the subcell level, allowing the resolution of the shock in only one mesh element.

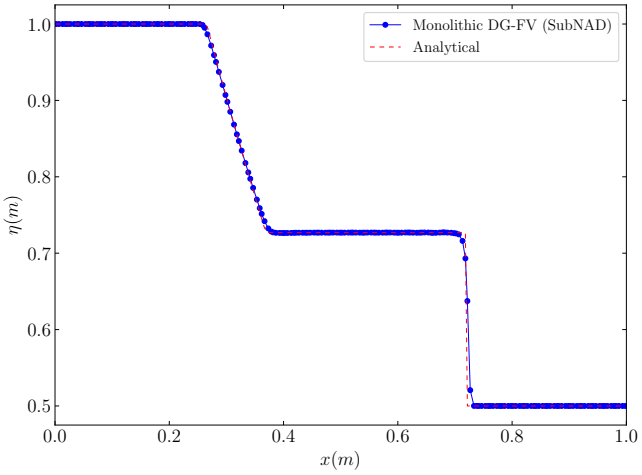


Figure 1.7: Dam-break on a wet bed − Free surface elevation computed at $T_{\max} = 0.075$ sec for $k = 3$ and $N_{\text{el}} = 50$.

Dam-break on a dry-bed

PARAMETERS

Computational domain: $\Omega = [0, 1]$
Number of cells: $N_{\text{el}} = 50$ cells
Boundary conditions: homogeneous Neumann

Polynomial degree: $k = 3$
Time-marching order: $p = 3$
Final time: $T_{\max} = 0.05$ sec

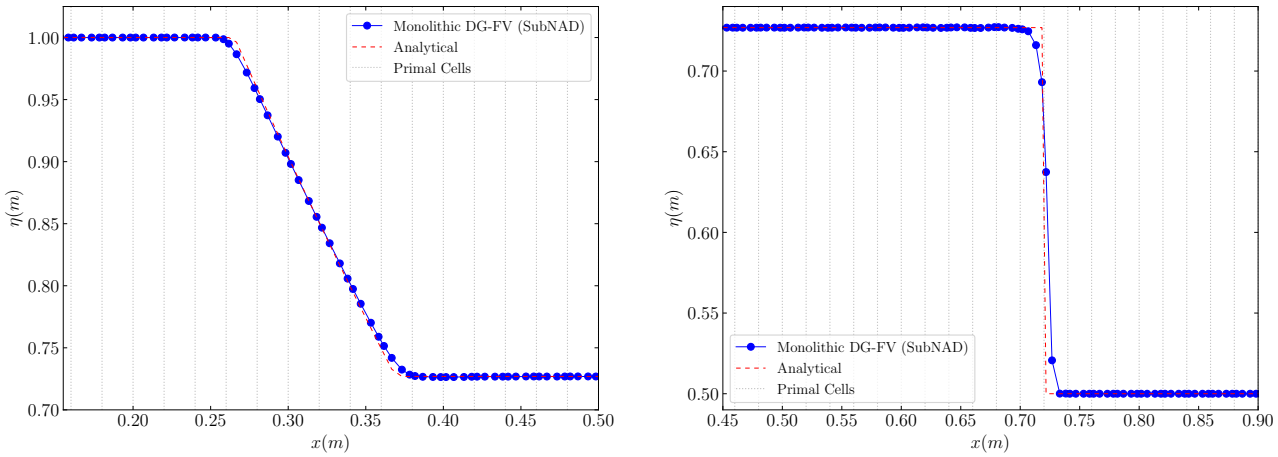


Figure 1.8: Dam-break − Free surface elevation computed at $T_{\max} = 0.075$ sec for $k = 3$ and $n_{\text{el}} = 50$, with a zoom on both rarefaction and shock waves.

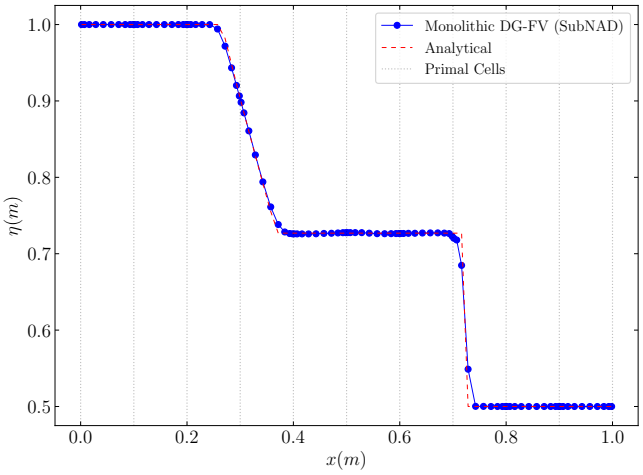


Figure 1.9: Dam-break on a wet bed − Free surface elevation computed at $T_{\max} = 0.075$ sec for $k = 9$ and $n_{\text{el}} = 10$.

We now modify the initial conditions as follows:

$$\eta_0(x) = \begin{cases} 1 & \text{if } x \leq 0.5, \\ 0 & \text{else,} \end{cases} \qquad q_0 = 0, \qquad b = 0,$$

in order to show the ability of the proposed method to handle the propagation of a classical wet/dry front. The results are presented in Figure 1.10 and show the ability of the method to handle dry bed, while the zoom on Figure 1.11 demonstrates that the method sharply capture both wet and dry fronts, as well as the transition between them.

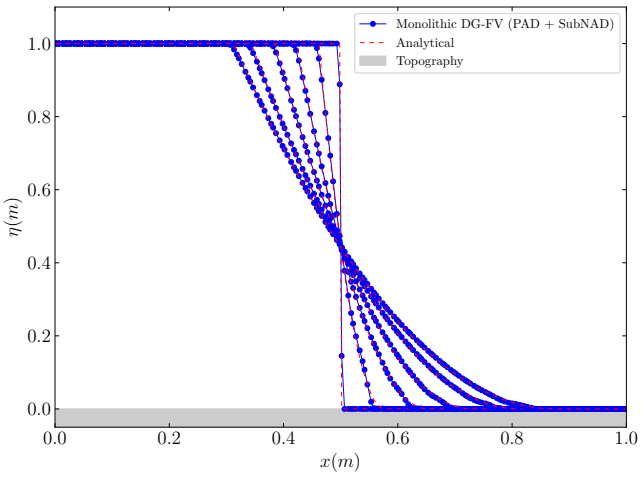


Figure 1.10: Dam-break on a dry bed − Free surface elevation computed in the time interval $[0.005, 0.05]$, for $k = 3$ and $N_{\text{el}} = 50$.

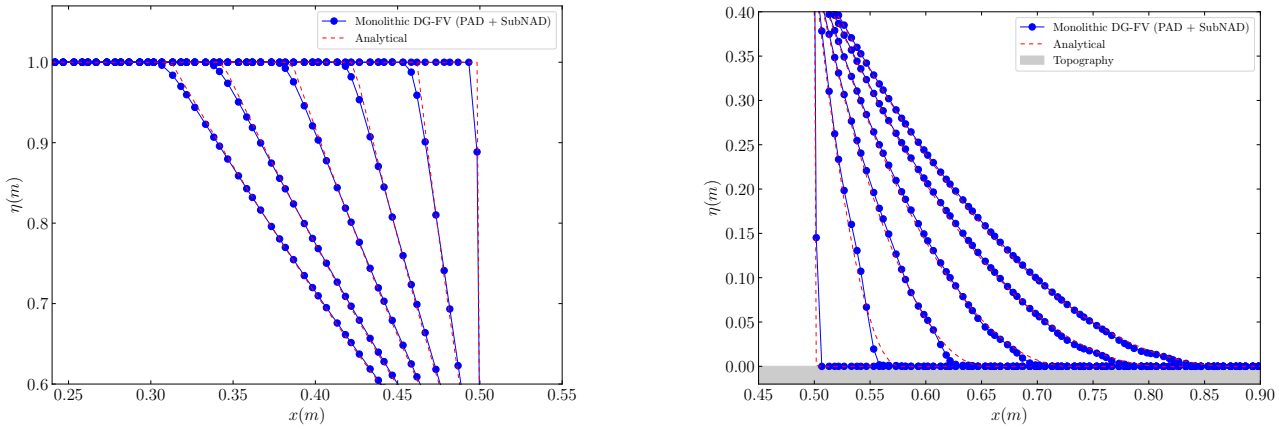


Figure 1.11: Dam-break on a dry bed − Zoom on top front and dry front, computed in the time interval $[0.005, 0.05]$, for $k = 3$ and $N_{\text{el}} = 50$.

1.4.3 Well-balancing with discontinuous bathymetry

PARAMETERS

Computational domain: $\Omega = [0, 1000]$
Number of cells: $N_{\text{el}} = 100$ cells
Boundary conditions: solid wall

Polynomial degree: $k = 3$
Time-marching order: $p = 1$
Final time: $T_{\max} = 500$ sec

We are now focusing on the preservation of the motionless steady states. We define here the following bathymetry profile:

$$b(x) = \begin{cases} A\left(\sin\left(\dfrac{(x - x_1)\cdot\pi}{750}\right)\right)^2 & \text{if } x_1 \leq x < x_2, \\ 5 & \text{if } x_2 \leq x < x_3, \\ 0 & \text{elsewhere,} \end{cases}$$

with $x_1 = 125$, $x_2 = 500$, $x_3 = 750$ and $A = 8.75$. This original parametrization shows very interesting properties:

1. The first section aims at testing the numerical method's ability to handle smooth bathymetry for the motionless steady states;
2. The constant part of the function, where $b(x) = 5$ for $x_2 \leq x < x_3$, adds a flat region to check if the method maintains hydrostatic balance over uniform segments and discontinuous bathymetry.

By definition this bathymetry includes two discontinuities, providing a comprehensive and difficult test for well-balancing methods. We will treat three different cases: fully immersed bathymetry (Figure 1.14), semi-immersed bathymetry with only one discontinuity outside the water (Figure 1.13), and bathymetry with both discontinuities exposed (Figure 1.14). These scenarios enable testing the method's effectiveness in handling different water coverage conditions, from completely submerged to partially and fully exposed discontinuities. For the first case, in Table 1.2 we can verify that the steady state is preserved up to double precision accuracy.

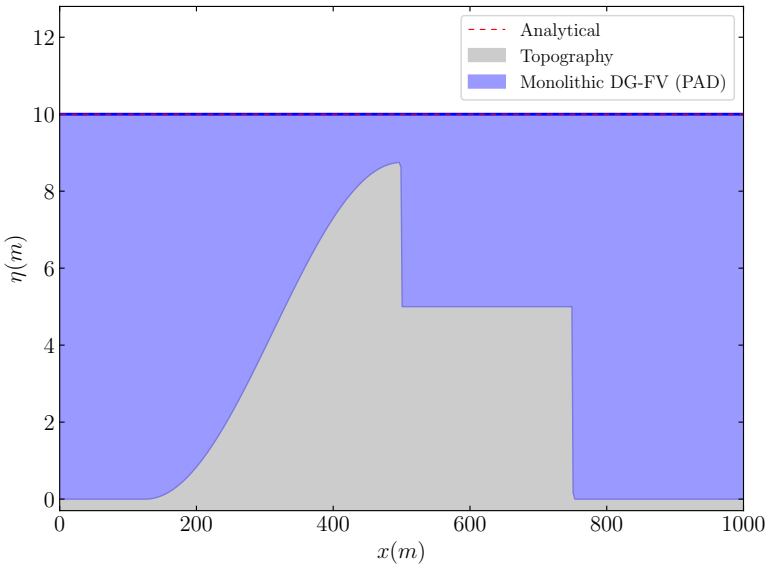


Figure 1.12: Preservation of a motionless steady state (first case) − Free surface elevation computed at $T_{\max} = 500$ sec, for $k = 3$ and $N_{\text{el}} = 100$.

1.4.4 Transcritical flow over a bump

PARAMETERS

Computational domain: $\Omega = [0, 25]$
Number of cells: $N_{\text{el}} = 100$ cells
Boundary conditions: fluvial upstream (L) & downstream (R)

Polynomial degree: $k = 3$
Time-marching order: $p = 3$
Final time: $T_{\max} = 500$ sec

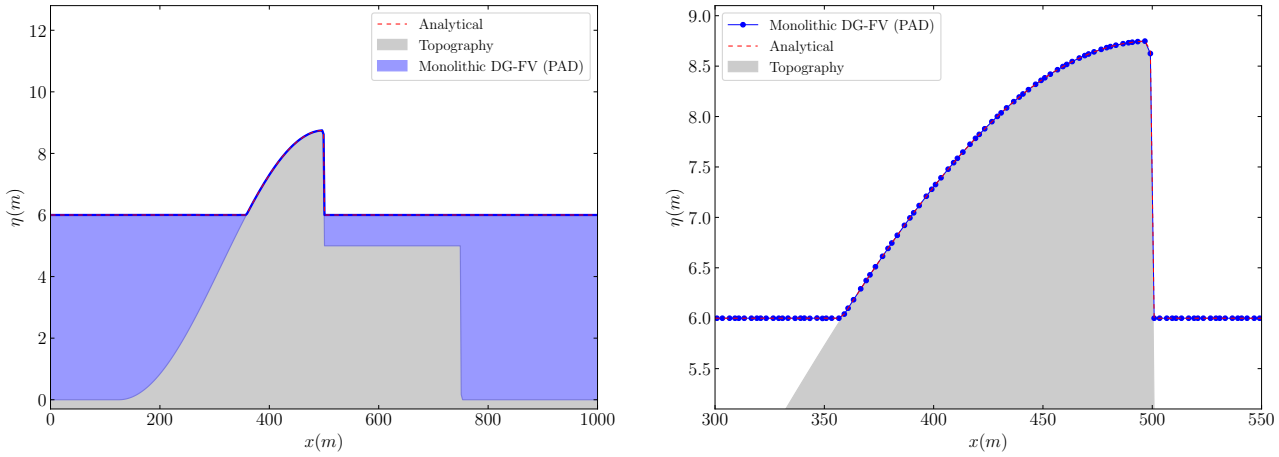


Figure 1.13: Preservation of a motionless steady state (second case) − Free surface elevation computed at $T_{\max} = 500$ sec, for $k = 3$ and $N_{\text{el}} = 100$, with a zoom on dry structure.

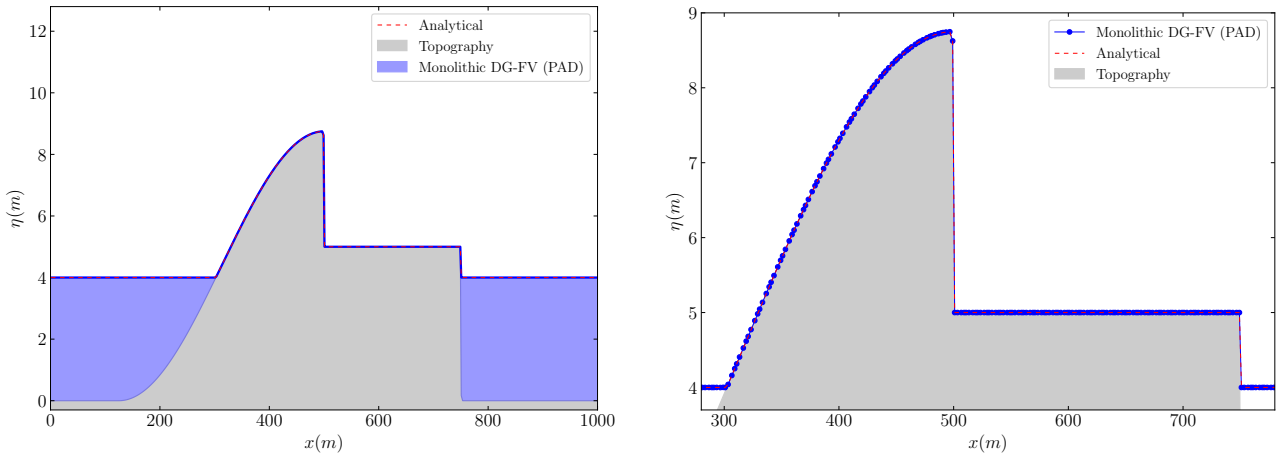


Figure 1.14: Preservation of a motionless steady state (third case) − Free surface elevation computed at $T_{\max} = 500$ sec, for $k = 3$ and $N_{\text{el}} = 100$, with a zoom on dry structure.

k	1	2	3
h	$E^{\eta}_{L^2}$	$E^{\eta}_{L^2}$	$E^{\eta}_{L^2}$
$\frac{1}{15}$	1.24E-15	1.09E-15	6.14E-16
$\frac{1}{30}$	5.01E-16	2.56E-16	8.89E-17
$\frac{1}{60}$	1.22E-16	5.44E-17	1.00E-17
$\frac{1}{120}$	6.64E-17	1.14E-17	1.36E-18

Table 1.2: Preservation of a motionless steady state (first case) − L^2-errors between numerical and exact steady state solutions for η at time $t = 50$ sec.

We now present the classical trans-critical flow without shock test-case, (see [127] for more details). We define the bathymetry profile as follows:

$$b(x) = \begin{cases} 0.2 - 0.05(x-10)^2 & \text{if } 8 < x < 12, \\ 0 & \text{elsewhere.} \end{cases} \tag{1.37}$$

In this test, the incoming flow is enforced to be fluvial upstream and becomes torrential at the top of the bump. The initial data is defined as:

$$\eta_0(x) = 0.66\ m \quad \text{and} \quad q_0(x) = 0\ m^3.\ \text{sec}^{-1},$$

and we prescribe the following boundary conditions:

$$\begin{cases} \text{upstream: } q = 1.53\ m^3.\ \text{sec}^{-1}, \\ \text{downstream: } H = 0.66\ m \text{ while the flow is subcritical.} \end{cases}$$

We show on Figure 1.15 both the free-surface elevation and the discharge obtained with the our scheme, at several moments during the transient part of the flow and when the steady state is reached.

1.4.5 Carrier & Greenspan's solutions

Transient solution

PARAMETERS

Computational domain: $\Omega = [-20, 6]$
Number of cells: $N_{\text{el}} = 50, 10$ cells
Boundary conditions: Dirichlet (L) & homogeneous Neumann (R)
Polynomial degree: $k = 3, 8$
Time-marching order: $p = 1$
Final time: $T_{\max} = 30$ sec

This test case, first introduced by Carrier in [54], describes a physical scenario in which the water level near the shoreline of a sloping beach is initially depressed and held motionless before being released at $t = 0$. This initial condition generates a transient wave that travels up the beach and eventually returns to an equilibrium state through a gradual convergence process. The test case provides a valuable benchmark for evaluating the robustness of the monolithic DG/FV scheme in computing long wave run-up scenarios. In [54], a hodograph transformation is employed to solve the NSW equations and derive an analytical solution. This transformation uses two dimensionless variables, denoted as σ^* and λ^*, which serve as space-like and time-like coordinates respectively, their expressions being:

$$\sigma^* = 4c^* \quad \text{and} \quad \lambda^* = 2\left(u^* + t^*\right).$$

Let l be the typical length scale here and α the beach slope. The scales used to obtain the nondimensionalized variables are:

$$x^* = x/l, \quad \eta^* = \eta/(\alpha l), \quad u^* = u/\sqrt{g\alpha l}, \quad t^* = t/\sqrt{l/\alpha g}, \tag{1.38}$$

and the non-dimensional phase speed is given by:

$$c^* = \sqrt{\eta^* - x^*}. \tag{1.39}$$

The initial solution is then specified by the following initial conditions:

$$\eta_0^*(\sigma^*) = e\left(1 - \frac{5}{2}\frac{a^3}{(a^2+\sigma^{*2})^{\frac{3}{2}}} + \frac{3}{2}\frac{a^5}{(a^2+\sigma^{*2})^{\frac{5}{2}}}\right), \qquad q_0^*(\sigma^*) = 0, \qquad x^* = -\frac{\sigma^{*2}}{16} + \eta_0^*,$$

where $a = \frac{3}{2}(1+0.9e)^{\frac{1}{2}}$ and e is a small parameter which characterizes the surface elevation profile. The analytical solution is then given by

$$\begin{cases} \eta^*(\sigma,\lambda) = -\dfrac{u^{*2}}{2} + e\,\Re\left\{1 - 2\dfrac{\frac{5}{4} - \mathrm{i}\lambda}{[(1-\mathrm{i}\lambda)^2+\sigma^2]^{\frac{3}{2}}} + \dfrac{3}{2}\dfrac{(1-\mathrm{i}\lambda)^2}{[(1-\mathrm{i}\lambda)^2+\sigma^2]^{\frac{5}{2}}}\right\}, \\ u^*(\sigma,\lambda) = \dfrac{8e}{a}\,\Im\left\{\dfrac{1}{[(1-\mathrm{i}\lambda)^2+\sigma^2]^{\frac{3}{2}}} - \dfrac{3}{4}\dfrac{1-\mathrm{i}\lambda}{[(1-\mathrm{i}\lambda)^2+\sigma^2]^{\frac{5}{2}}}\right\}, \\ t^* = \dfrac{1}{2}a\lambda - u^* \qquad \text{and} \qquad x^* = \eta^* - \dfrac{a^2\sigma^2}{16}, \end{cases}$$

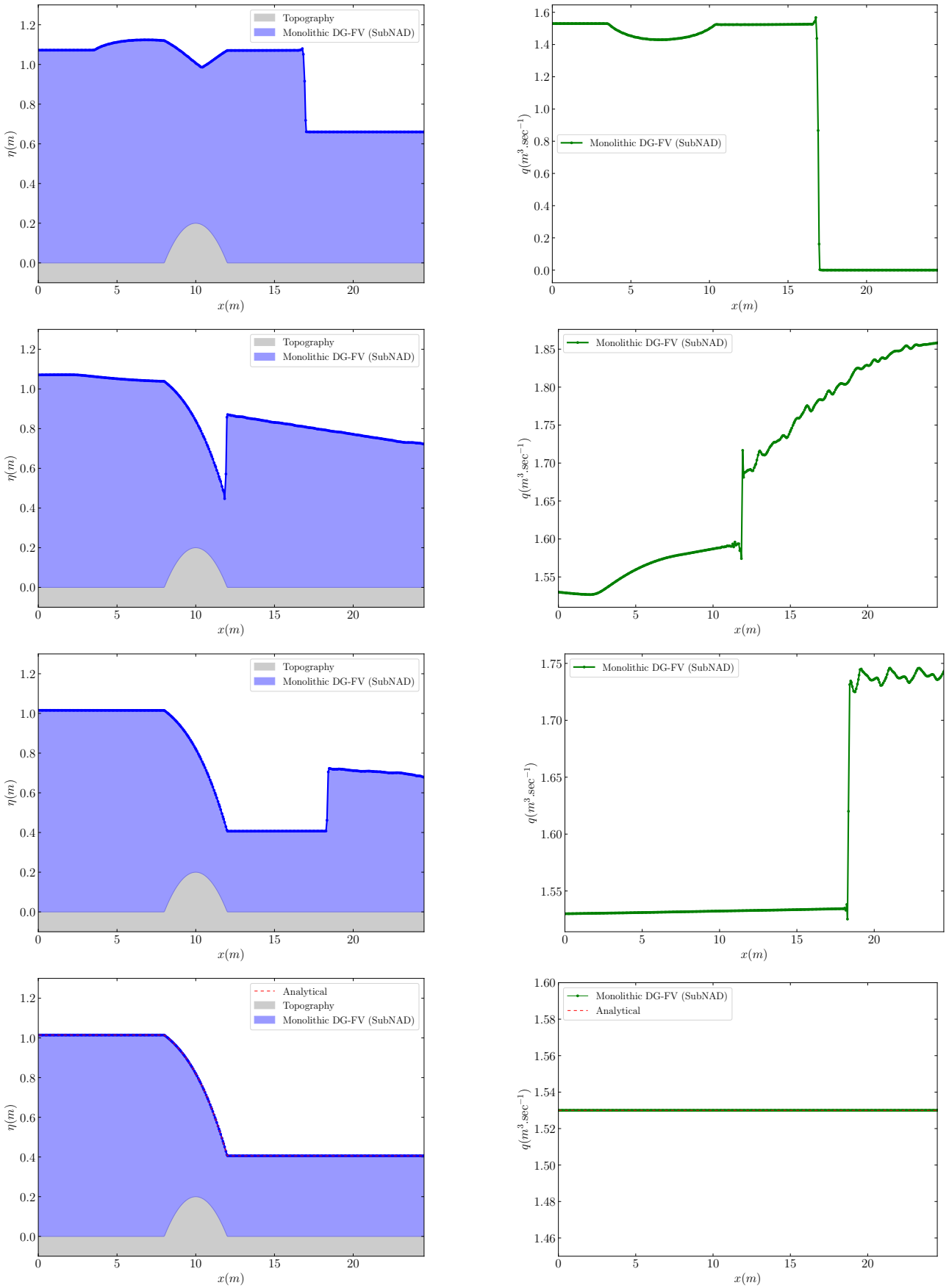


Figure 1.15: Transcritical flow over a bump without shock – Surface elevation (left) and discharge (right) computed at several times, for $k = 3$ and $N_{\text{el}} = 100$.

with $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denoting the real and imaginary parts respectively, where we have set $\sigma^* = a\sigma$ and $\lambda^* = a\lambda$. This set of equations may be solved by some iterative process, as mentioned earlier for smooth sinusoidal solution. In what follows, we set $e = 0.1$, $\alpha = 1/50$, the initial surface profile is provided in the dimensional case with the length scale $l = 20\ m$ and we define $\beta := \alpha e l$.

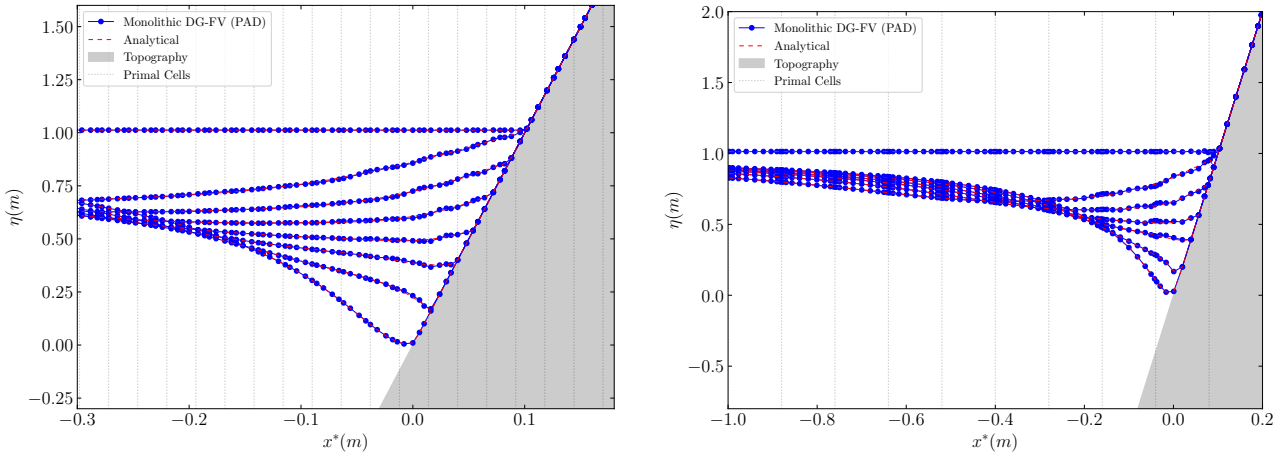


Figure 1.16: Carrier & Greenspan's transient solution − Surface elevation computed at several times between 0.0 sec and 100.0 sec − Comparison between $k = 3$, $N_{\text{el}} = 50$ and $k = 8$, $N_{\text{el}} = 10$ setups.

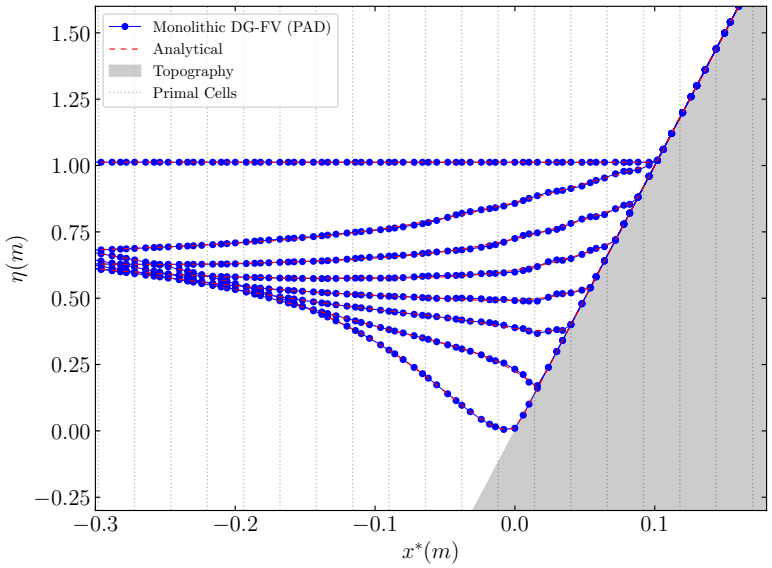


Figure 1.17: Carrier & Greenspan's transient solution − Surface elevation computed at several times between 0 and 100 sec for $k = 3$ and $N_{\text{el}} = 50$ (zoom on dry bed).

This benchmark is particularly interesting because it allows us to assess the ability of the method to handle dry bed for a long period of time, as the wave runs up and down the beach to finally stabilize in a lake-at-rest state. The results displayed in Figures 1.17 and 1.18 for $k = 3$, $N_{\text{el}} = 50$ and $k = 8$, $N_{\text{el}} = 10$ respectively, demonstrate once more the very good performance of our numerical sheme.

Periodic solution

PARAMETERS

Computational domain: $\Omega = [-20, 6]$
Number of cells: $N_{\text{el}} = 50, 10$ cells
Boundary conditions: Dirichlet (L) & homogeneous Neumann (R)
Polynomial degree: $k = 3, 8$
Time-marching order: $p = 1$
Final time: $T_{\max} = 300$ sec

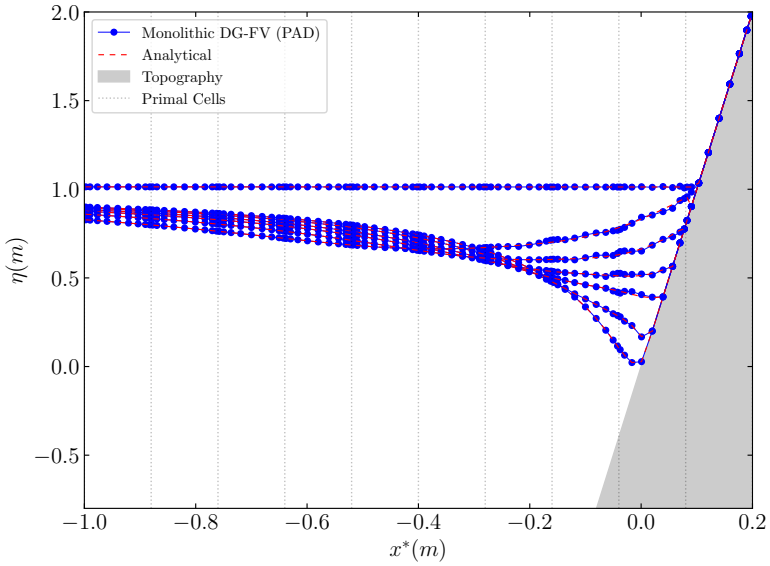


Figure 1.18: Carrier & Greenspan's transient solution − Surface elevation computed at several times between 0 and 100 sec for $k = 8$ and $N_{\text{el}} = 10$ (full domain).

This benchmark describes the motion of a monochromatic wave that is let run-up and run-down on a plane beach. The dimensionless amplitude of this periodic wave is denoted by A^* and its frequency by ω^*, traveling shoreward and being reflected out to sea generating a standing wave on a plane beach. Recalling the dimensionless quantities (3.40) and (3.41), the analytical solution is formulated as follows:

$$\begin{cases} u^* = -\dfrac{A^* J_1(\sigma^*)\sin(\lambda^*)}{\sigma^*}, \\ \eta^* = \dfrac{A^*}{4} J_0(\sigma^*)\cos(\lambda^*) - \dfrac{u^{*2}}{4}, \\ t^* = \dfrac{1}{2}\lambda^* - u^* \quad \text{and} \quad x^* = \eta^* - \dfrac{\sigma^{*2}}{16}, \end{cases}$$

where J_0 and J_1 stand for the Bessel functions of zero and first order. We consider the solution obtained for $A^* = 0.6$ and $\omega^* = 1$ (non-breaking wave), together with the length scale $l = 20\ m$ and a bottom slope $\alpha = \frac{1}{30}$. The value of this solution at $t = 0$ is supplied as initial condition, and similarly to the previous transient case, the analytical variations of the surface elevation at the left boundary is used as an offshore inlet boundary condition, generating the motion. We again refer the reader to [54] for a complete description. In particular, we compute the time evolution up to $T_{\max} = 1.5\,T$, where T is the time period of the periodic forcing.
Some snapshots are shown in both Figures 1.24,1.20 of the free-surface elevation plotted at several discrete time in the range $[20.25T, 20.50T]$, demonstrating the high accuracy of our numerical solution and the absence of any phase shift over long time periods.

1.4.6 Periodic oscillations in a parabolic basin

PARAMETERS

Computational domain: $\Omega = [0, 4]$
Number of cells: $N_{\text{el}} = 50, 30$ cells
Boundary conditions: solid wall

Polynomial degree: $k = 3, 6$
Time-marching order: $p = 3$
Final time: $T_{\max} = 10.0303$ sec

The exact solution described here is a simplification to 1D of an artificially 2D Thacker's solution developed in [78]. Indeed, for this solution, Thacker considered an infinite channel with a parabolic cross section but where the velocity has only one nonzero component (orthogonal to the axis of the channel). This case provides us with a relevant test in 1D for shallow-water model because it deals with a sloping bed as well as with wetting and drying. The bathymetry is a parabolic bowl given by:

$$b(x) = h_c \left(\frac{1}{a^2} \left(x - \frac{L}{2} \right)^2 - 1 \right),$$

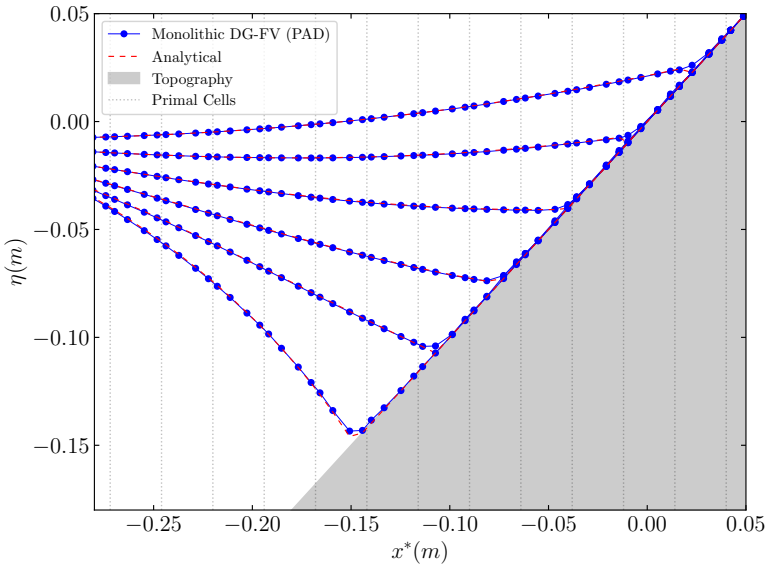


Figure 1.19: Carrier & Greenspan's periodic solution – Surface elevation computed in the time interval $[20.25T, 20.50T]$ for $k = 3$ and $N_{\text{el}} = 50$ (zoom on dry bed).

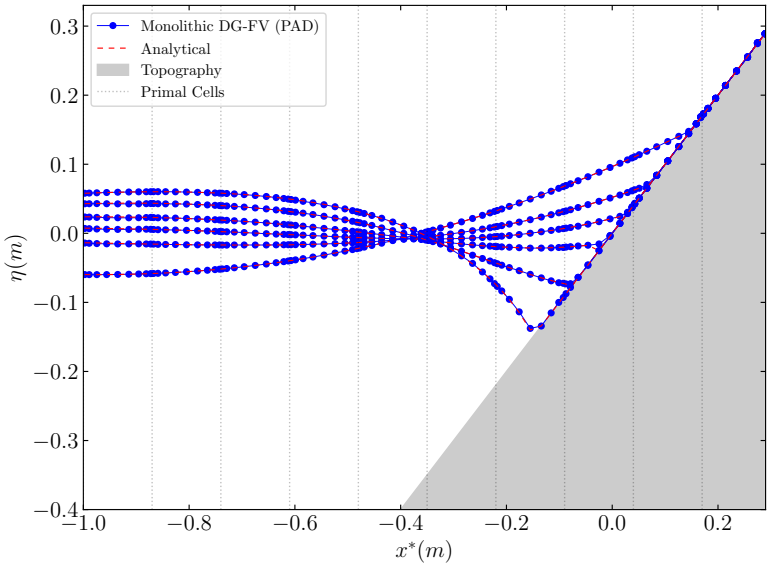


Figure 1.20: Carrier & Greenspan's periodic solution – Surface elevation computed in the time interval $[20.25T, 20.50T]$ for $k = 8$ and $N_{\text{el}} = 10$ (full domain).

and the initial condition on the water-height is

$$h_0(x) = \begin{cases} -h_c \left\{ \left(\frac{1}{a}\left(x - \frac{L}{2}\right) + \frac{B}{\sqrt{2gh_c}} \right)^2 - 1 \right\} & \text{for } x_1(0) \leq x \leq x_2(0), \\ 0 & \text{elsewhere,} \end{cases}$$

with $B = \frac{\sqrt{2gh_0}}{2a}$ and $u(x) = 0$ for the velocity. In the case where no friction is considered, Thacker's solution is a periodic solution and the free-surface remains planar in time. The analytic solution is given by:

$$\eta(t,x) = \begin{cases} b(x) - h_c \left\{ \left(\frac{1}{a}\left(x - \frac{L}{2}\right) + \frac{B}{\sqrt{2gh_0}} \cos\left(\frac{\sqrt{2gh_0}}{a} t \right) \right)^2 - 1 \right\} & \text{for } x_1(t) \leq x \leq x_2(t), \\ 0 & \text{otherwise,} \end{cases}$$

$$u(t,x) = \begin{cases} B \sin\left(\frac{\sqrt{2gh_0}}{a} t \right) & \text{for } x_1(t) \leq x \leq x_2(t), \\ 0 & \text{otherwise,} \end{cases}$$

where $x_\pm(t)$ are the locations of wet/dry interfaces at time t:

$$x_\pm(t) = -\frac{1}{2}\cos\left(\frac{\sqrt{2gh_0}}{a} t \right) \pm a + \frac{L}{2},$$

In our computation we define the constants as $a = 1$, $h_0 = 0.5$, $L = 4$ and $T_{\max} = 10.0303$ sec (5 periods). We show our results after 5 periods, i.e. in the time interval $[4T, 5T]$. In Table 1.23, we gather the global L^2-errors associated with the computation of η, for different polynomial orders, and increasing mesh refinements, computed at $t = 1$ sec. In such situations, the solution is only $\mathsf{H}^1(\Omega)$, so the regularity is not sufficient to obtain an order in $\mathcal{O}(k+1)$. As a consequence, it is no surprise that the observed rates of convergence are not the optimal rates generally associated, for a « sufficiently regular » solution, to high-order schemes. Measuring order of convergence in such test-cases is not a valid criterion to test accuracy of our high-order scheme. The order of convergence on a smooth solution being already computed in 1.1.

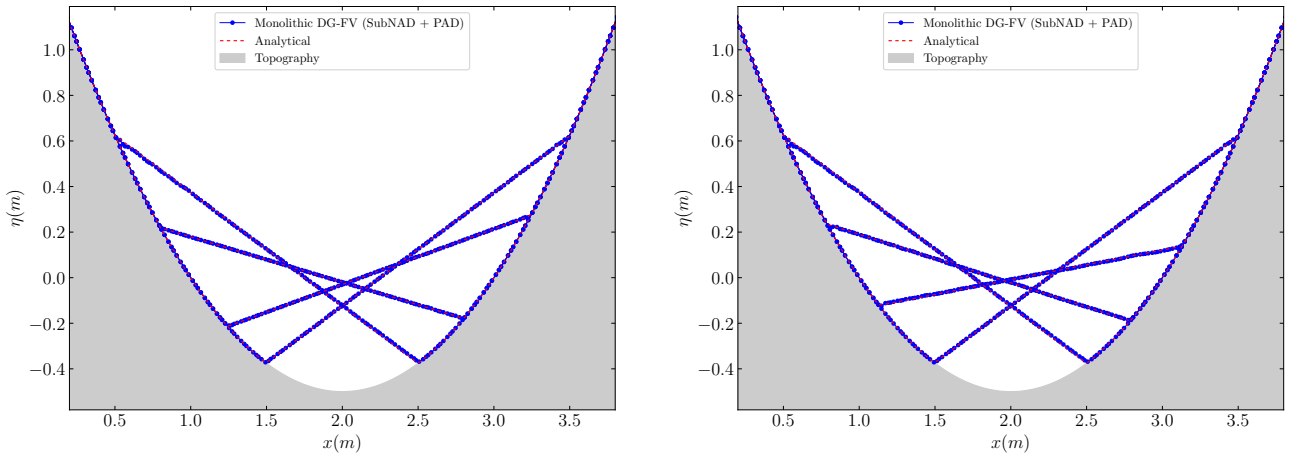


Figure 1.21: Periodic planar surface in a parabolic domain. Surface elevation computed for $k = 3$ and $N_{\text{el}} = 50$ over the time intervals $[4T, 4.5T]$ (left) and $[4.5T, 5T]$ (right).

This test-case is pretty similar to Carrier & Greenspan's periodic solution, but is more challenging because of the wetting and drying phenomena happening on both sides of the domain. The results are shown in Figure 1.21 for $k = 3$ and $N_{\text{el}} = 50$, and in Figure 1.22 for $k = 6$ and $N_{\text{el}} = 30$. We observe that the numerical solution is very close to the analytical one, even in the dry areas, which is a good indicator of the robustness and precision of our method.

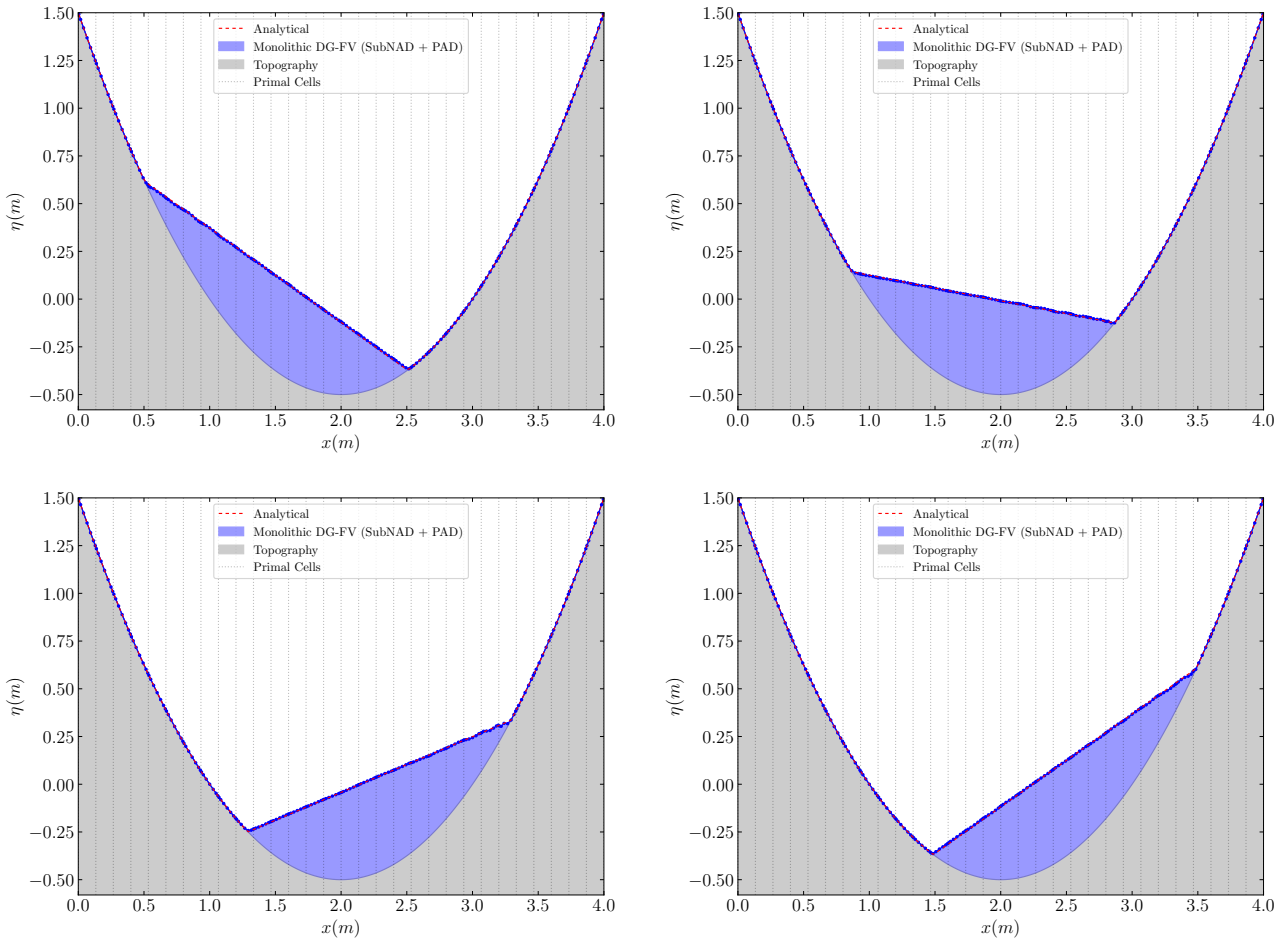


Figure 1.22: Periodic planar surface in a parabola – Surface elevation computed in the time interval $[4, 4.5T]$ for $k = 6$ and $N_{\text{el}} = 30$.

k	1		2		3	
h	$E^{\eta}_{L^2}$	$q^{\eta}_{L^2}$	$E^{\eta}_{L^2}$	$q^{\eta}_{L^2}$	$E^{\eta}_{L^2}$	$q^{\eta}_{L^2}$
0.2	1.194E-2	1.67	5.603E-3	1.41	4.872E-3	1.98
0.096	3.607E-3	1.64	2.035E-3	1.67	1.168E-3	1.53
0.065	1.827E-3	1.95	1.081E-3	1.76	9.392E-4	1.39
0.048	1.037E-3	–	6.091E-4	–	6.256E-4	–

Figure 1.23: Periodic planar surface in a parabola – L^2-errors between numerical and analytical solutions and convergence rates for η at time $t = 0.6$ sec.

1.4.7 Vacuum occurrence wave over a step

PARAMETERS

Computational domain: $\Omega = [0, 25]$
Number of cells: $N_{\text{el}} = 50, 100$ cells
Boundary conditions: solid wall

Polynomial degree: $k = 3, 4$
Time-marching order: $p = 3$
Final time: $T_{\max} = 20$ sec

This last test acts like a summary of the ability of our scheme, especially showing its robustness aspects. Indeed, we study here the ability of our method to compute vacuum. Moreover, the bathymetry is discontinuous, and defined like this:

$$b(x) = \begin{cases} 1 & \text{if } \frac{25}{3} < x < 12.5, \\ 0 & \text{elsewhere.} \end{cases}$$

This test was based on a test proposed by E. Toro with a non-trivial bathymetry introduced in [120]. The initial water elevation is $\eta_0(x) = 10\ m$ and the initial discharge is

$$q_0(x) = \begin{cases} -350 & \text{if } x < \frac{50}{3}, \\ 350 & \text{elsewhere.} \end{cases}$$

In the case of a flat bottom, the solution consists of two rarefaction waves, separated by a dry zone. However,

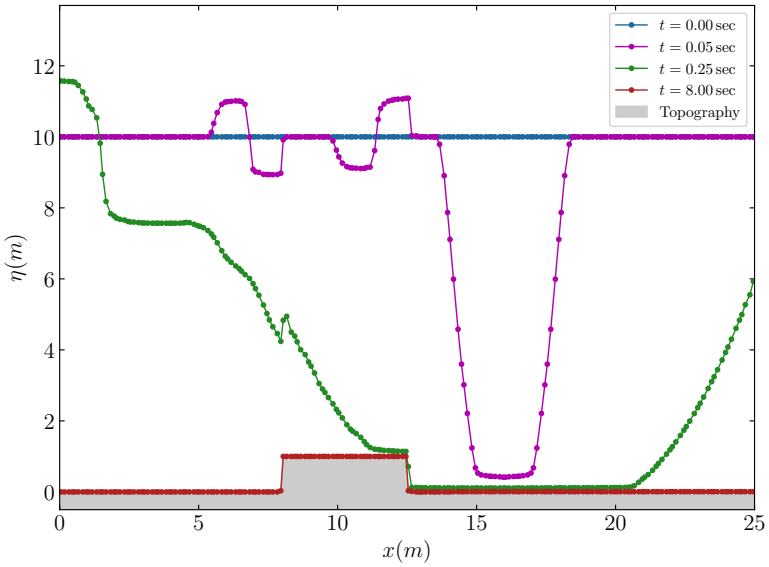


Figure 1.24: Vacuum occurrence wave over a step − Surface elevation computed during the time interval $[0.00, 8.00]$ for $k = 3$ and $N_{\text{el}} = 50$.

with a non-flat bathymetry, the scheme generates additional waves at topographic discontinuities. This test case presents multiple challenges, as it requires handling shocks, rarefaction waves, and maintaining a well-balanced dry state at the end of the simulation, all while dealing with a discontinuous bathymetry. As displayed in Figure 1.25, despite these complexities, the results demonstrate a highly accurate solution corresponding to the results in literature, sharply capturing both shocks and rarefaction waves, while ensuring that the water-height converges to zero at the final time with double-precision accuracy.

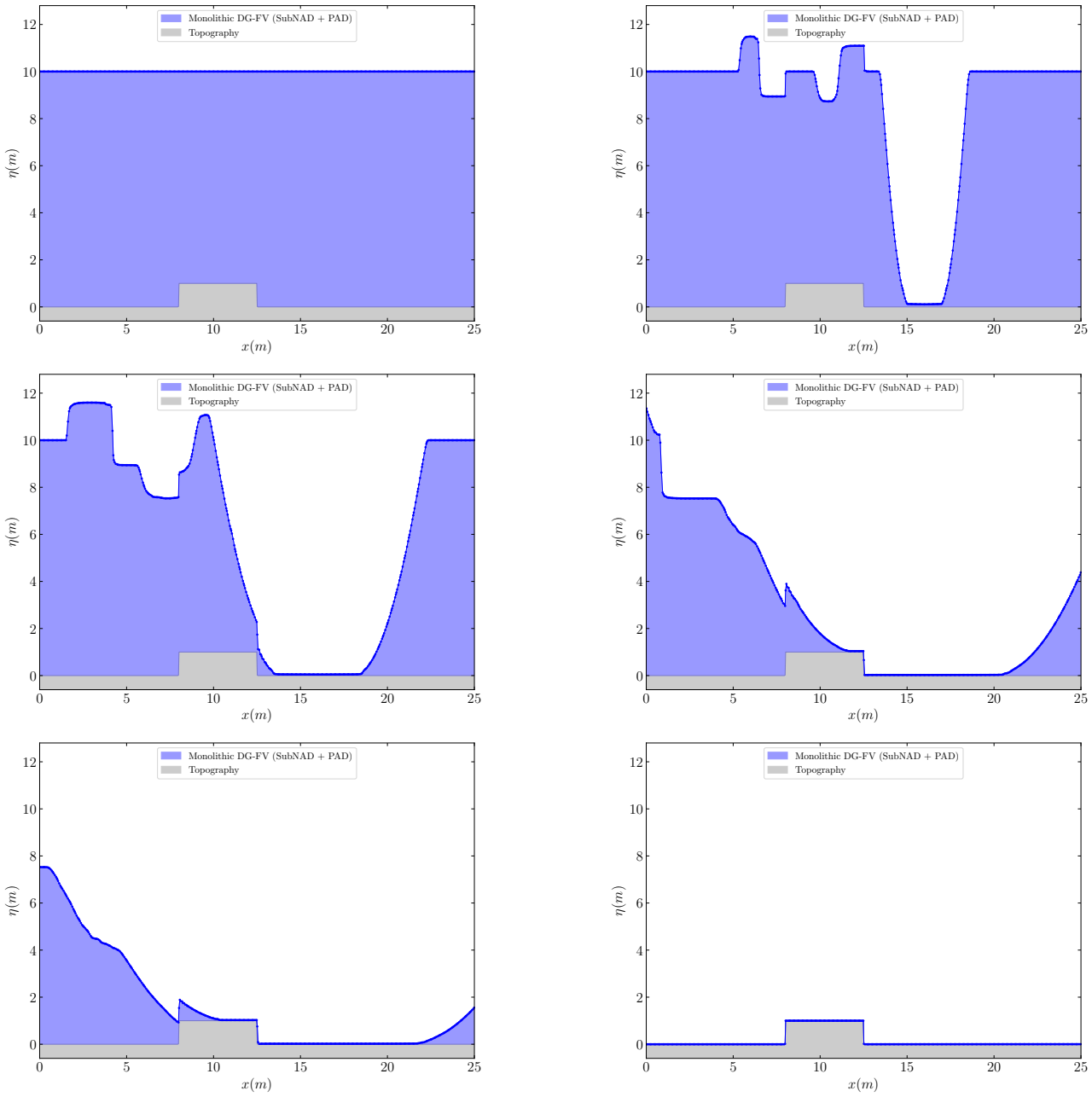


Figure 1.25: Vacuum occurrence wave over a step − Surface elevation computed from $t = 0$ sec to $t = T_{\max}$, for $k = 4$ and $N_{\text{el}} = 100$.

CHAPTER 2

Symmetric Interior Penalty DG/FV subcell formulation for 1D Serre-Green-Naghdi equations

The scheme presented here, together with the main development of Chapter 1, have been submitted [51].

This chapter investigates how the monolithic DG/FV subcell strategy can be coupled with an elliptic solver in order to approximate weakly dispersive shallow-water models. The target system is the one-dimensional Serre–Green–Naghdi model, whose dispersive terms introduce a nonlocal correction absent from the classical nonlinear shallow-water equations. This makes the problem structurally different from the purely hyperbolic setting of Chapter 1: the conservative shallow-water dynamics still requires a robust hyperbolic solver, but the dispersive contribution must be computed through the resolution of an additional elliptic problem.

The contribution of the chapter is to show that the monolithic DG/FV framework can be coupled in a consistent way with a high-order elliptic discretization. Following the approach of [292], the elliptic subproblem is approximated by a Symmetric Weighted Interior Penalty Discontinuous Galerkin (SWIP-DG) method, while the resulting dispersive correction is incorporated as an additional contribution in the monolithic DG/FV update. The resulting formulation provides a first hyperbolic-elliptic version of the subcell framework: the DG/FV component retains its robustness properties for the hyperbolic part, while the SWIP-DG component provides a stable and consistent approximation of the second-order operators involved in the dispersive correction. A more detailed investigation of this coupling, in particular in two space dimensions, is left for future work and will be the subject of further developments.

Contents

2.1 Some remarks about Serre-Green-Naghdi equations 54
2.1.1 Pre-balanced formulation 54
2.1.2 Symmetric discrete formulation 56
2.2 SWIP-DG discretization 57
2.2.1 Discrete bilinear form 57
2.2.2 Discrete Laplace and gradient operators 57
2.2.3 Semi-discrete DG–SWIP-DG formulation 58
2.2.4 Time discretization 59
2.3 Combined DG/FV–SWIP-DG fully-discrete formulations 59
2.4 Numerical results 61
2.4.1 Oscillating dam-break 61
2.4.2 Solitary-waves over a piecewise-linear bottom 62

Reaching this itch in my soul, it's like any good playing card
Must be why I'm thinking of Las Vegas, why it's more brighter than the sun is to me
≃ FROM: ***"Heaven or Las Vegas"***, Cocteau Twins – *Heaven or Las Vegas* (1990)

2.1 Some remarks about Serre-Green-Naghdi equations

Unless otherwise stated, we keep the notations introduced in the previous chapter. The reader may also refer to the General shallow-water model in one space dimension section for additional remarks on the Serre-Green-Naghdi equations. In what follows, we restrict ourselves to the one-dimensional setting, inspired by [292]. Recent work on the two-dimensional case, based on a coupled hyperbolic DG and Hybridizable DG approach, can be found in [211].

2.1.1 Pre-balanced formulation

Let $\zeta(t,x)$ again denote the free-surface elevation measured from the rest state, and let the bottom topography be given by $-H_0 + b(x)$. The total water depth is then defined by $H := H_0 + \zeta - b$.

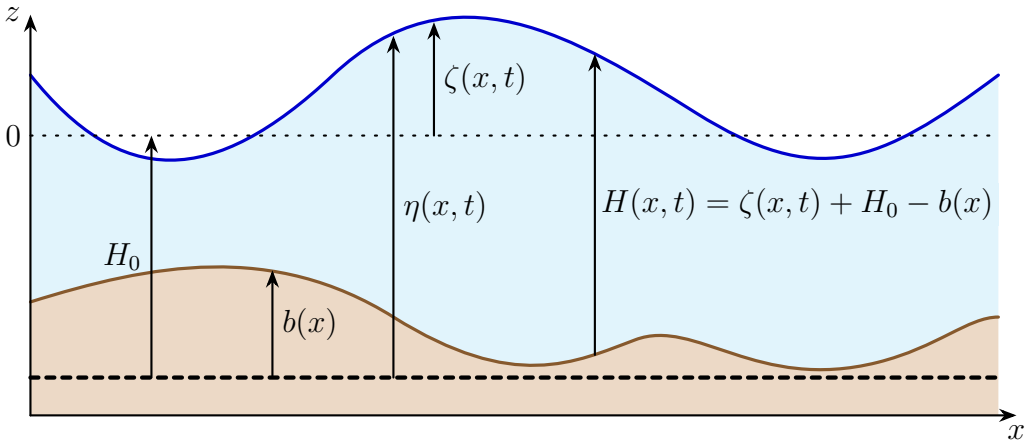


Figure 2.1: Main notations for Serre-Green-Naghdi equations. The free-surface elevation is denoted by η, ζ denoting the deviation from the rest state, the bathymetry is b and the water-height by H. The horizontal velocity is again denoted by u and the horizontal discharge by q.

Still using the discharge variable $q := Hu$ as the horizontal momentum variable, the classical Serre-Green-Naghdi (SGN) equations may be written as

$$\begin{cases} \partial_t \zeta + \partial_x q = 0, \\ (1 + \mathcal{T}[H,b])\,\partial_t u + u\,\partial_x u + \mathrm{g}\,\partial_x \zeta + \mathcal{Q}[H,b](u) = 0, \end{cases} \tag{2.1}$$

where the linear operator $\mathcal{T}[H,b]\cdot$ and the quadratic form $\mathcal{Q}[H,b](\cdot)$ are defined, for all sufficiently smooth scalar-valued functions w, by

$$\begin{aligned} \mathcal{T}[H,b]w &:= \mathcal{R}_1[H,b](\partial_x w) + \mathcal{R}_2[H,b](w\partial_x b), \\ \mathcal{Q}[H,b](w) &:= \mathcal{R}_1[H,b](\partial_x(w\partial_x w) - 2(\partial_x w)^2) + \mathcal{R}_2[H,b]((w\partial_x)^2 b), \end{aligned}$$

with

$$\mathcal{R}_1[H,b]w := -\frac{1}{3H}\partial_x(H^3 w) - \frac{H}{2} w\partial_x b, \qquad \mathcal{R}_2[H,b]w := \frac{1}{2H}\partial_x(H^2 w) + w\partial_x b.$$

In [85] has been introduced high-order fully discontinuous Galerkin (DG) formulations for the classical SGN equations, as well as for asymptotically improved SGN models [183], whose purpose is to enhance the dispersive properties of the classical SGN system. These formulations are based on the Symmetric Weighted Interior Penalty DG (SWIP-DG) method, in which the fully nonlinear weakly dispersive equations are recast as coupled nonlinear (pseudo) hyperbolic-elliptic problems. More precisely, they rely on the non-dispersive NSW equations, supplemented by additional algebraic source terms that account for the $\mathcal{O}(\mu^2)$ nonlinear dispersive corrections. These source terms are obtained by solving auxiliary linear second-order elliptic problems associated with the operator $1 + \mathcal{T}[H,b]$ appearing in (2.1).

More precisely, the SGN equations, written in pre-balanced form, read

$$\begin{cases} \partial_t \eta + \partial_x q = 0, \\ \partial_t q + \partial_x (uq) = -(1 + \boldsymbol{T}[H,b])^{-1} \mathbb{Q}[H,b](\eta, u), \end{cases} \tag{2.2}$$

where the linear operator $\boldsymbol{T}[H,b]\cdot$ and the nonlinear operator $\mathbb{Q}[H,b]$ are defined, for all sufficiently smooth scalar-valued functions w, by

$$\begin{aligned} \boldsymbol{T}[H,b]w &:= H\mathcal{T}[H,b]\tfrac{w}{H}, \\ \mathbb{Q}[H,b](\eta,u) &:= \mathrm{g}H\partial_x\eta + H\mathbb{Q}_1[H,b](u), \\ \mathbb{Q}_1[H,b](w) &:= -2\mathscr{R}_1[H,b]((\partial_x w)^2) + \mathscr{R}_2[H,b]((w\partial_x)^2 b). \end{aligned}$$

In the present work, the elliptic operator $1 + \boldsymbol{T}[H,b]$ will be discretized by means of the SWIP-DG method, while the dispersive source term $\mathbb{Q}[H,b](\eta,u)$ will be handled as an additional source contribution within the DG/FV subcell framework.

Remark 2.1.1 (*Well-posedness of SGN equations*). *Long-time existence for regular solutions holds with* $\boldsymbol{v}^0 \in \mathsf{H}^s(\mathbb{R}) \times \mathsf{H}^{s+1}(\mathbb{R})$*, see for instance [155]. When bathymetry variations should be accounted for, then the previous results still holds for* $b \in \mathscr{C}^\infty_{\mathrm{b}}(\mathbb{R})$.

Remark 2.1.2 (*Equivalent formulation of the SGN equations*). *The invertibility of* $(1 + \mathcal{T}[H,b])$ *is recalled in [292], where it is also shown that an equivalent formulation for the SGN equations:*

$$\partial_t \boldsymbol{v} + \partial_x \boldsymbol{F}(\boldsymbol{v}, b) = \boldsymbol{B}[b](\boldsymbol{v}) + \boldsymbol{D}[H](\boldsymbol{v}, \Lambda), \tag{2.3}$$

where the previous higher-order term is replaced by the following new one:

$$\boldsymbol{D}[H](\boldsymbol{v}, \Lambda) \coloneqq \Big(0,\ gH\partial_x\eta - H\Lambda\Big)^\top, \tag{2.4}$$

the auxiliary variable Λ *being obtained as the solution of a second-order linear differential elliptic problem*

$$\mathscr{L}[H,b]\Lambda = \mathbb{Q}[H,b](\boldsymbol{v}), \tag{2.5}$$

with

$$\mathscr{L}[H,b] \coloneqq H + \mathcal{S}_1[H,b]^\star H \mathcal{S}_1[H,b] + \mathcal{S}_2[b]^\star H \mathcal{S}_2[b]. \tag{2.6}$$

Such operator, supplemented with suitable boundary conditions when bounded domain are considered, see the remark below about boundary conditions for Λ*, allows some convenient numerical decoupling of differential operators approximations. Still anticipating on discrete formulations, it is also shown in the reference above that the following identity holds for all smooth enough scalar-valued functions* w*:*

$$\mathscr{L}[H,b]w = \partial_x(-\kappa(H)\partial_x w) - \beta(H,\partial_x b)\partial_x w + \partial_x(\beta(H,\partial_x b)w) + \delta(H,\partial_x b)w, \tag{2.7}$$

with the following coefficient functions defined as follows:

$$\kappa(H) := \frac{1}{3}H^3, \quad \beta(H,\partial_x b) := \frac{\sqrt{3}}{2}\kappa[H]^{\frac{1}{2}} H^{\frac{1}{2}}\partial_x b, \quad \delta(H,\partial_x b) := (H^{\frac{1}{2}}\partial_x b)^2 + H.$$

Remark 2.1.3 (*Asymptotic derivation*). *Considering flows with a vertical characteristic scale* H_0*, i.e. mean water-depth, with* λ *for the typical wavelength, together with the so-called dispersive parameter* $\mu \coloneqq H_0^2/\lambda^2$*, we recall that the SGN equations may be derived as some asymptotically consistent second-order approximation of the free-surface Euler equations, with respect to* μ *in the shallow-water or long-wave assumption:*

$$(\textit{shallow-water regime}) \qquad \mu \ll 1,$$

while the hyperbolic NSW equations [77] may be regarded as a first-order approximation, neglecting the terms of order μ *in the asymptotic expansion, see [181] for a complete derivation and proof of consistency. It is also worth noticing that no smallness assumption is made on the size of the surface perturbations for this derivation, and the corresponding regime is also said to be fully nonlinear:*

$$(\textit{fully nonlinear / large amplitude regime}) \qquad \varepsilon := \frac{a}{H_0} = \mathcal{O}(1),$$

where a *is the typical wave's amplitude. Numerous asymptotic models may be obtained from the water-wave equations with a free-surface boundary condition.*

2.1.2 Symmetric discrete formulation

Following [85], in order to build a symmetric discrete formulation associated with the operator $1+\boldsymbol{T}[H,b]$, one first observes that, provided the water depth H remains bounded away from zero, the following identity holds for every sufficiently smooth scalar-valued function v:

$$(1+\boldsymbol{T}[H,b])v = H(1+\mathcal{T}[H,b])\frac{v}{H}. \tag{2.8}$$

Moreover, for any sufficiently smooth scalar-valued function w, one has

$$H(1+\mathcal{T}[H,b])w = -\frac{1}{3}\partial_x(H^3\partial_x w) + \left(H+\frac{1}{2}\partial_x(H^2\partial_x b)+H(\partial_x b)^2\right)w. \tag{2.9}$$

As a consequence, the approximation of the right-hand side of (2.2) may be obtained as the solution to a non-homogeneous elliptic problem of diffusion-reaction type, whose time-dependent reaction-like coefficient is given by

$$H+\frac{1}{2}\partial_x(H^2\partial_x b)+H(\partial_x b)^2.$$

However, as already emphasized in [85], a main limitation of this formulation is that the sign of this coefficient depends on the sign of $\partial_x(H^2\partial_x b)$. Therefore, for general topographic configurations, one cannot guarantee a discrete coercivity property for the proposed discrete counterpart of the operator $H(1+\mathcal{T}[H,b])$.

In [292], a new DG formulation was introduced for the SGN equations, preserving the main structural properties of the continuous model. More precisely, for any $w\in\mathsf{H}^2(\mathbb{R})$, the following relation holds:

$$(1+\boldsymbol{T}[H,b])w = \left(H+(S_1[H,b])^\star HS_1[H,b]+(S_2[b])^\star HS_2[b]\right)\frac{w}{H}. \tag{2.10}$$

where the operators $(S_1[H,b])^\star$ and $(S_2[b])^\star$ denote the adjoint operators of $S_1[H,b]$ and $S_2[b]$, respectively, and are defined in General shallow-water model in one space dimension. Using (2.10), system (2.2) may be equivalently rewritten as

$$\begin{cases}\partial_t\eta+\partial_x q=0,\\ \partial_t q+\partial_x\left(uq+\frac{1}{2}g(\eta^2-2\eta b)\right)=-g\eta\partial_x b+gH\partial_x\eta-H\Lambda,\\ \left(H+(S_1[H,b])^\star HS_1[H,b]+(S_2[b])^\star HS_2[b]\right)\Lambda=\mathbb{Q}[H,b](\eta,u),\end{cases} \tag{2.11}$$

or, in a more compact form,

$$\begin{cases}\partial_t\boldsymbol{v}+\partial_x\boldsymbol{F}(\boldsymbol{v},b)=\boldsymbol{B}[b](\boldsymbol{v})+\boldsymbol{D}[H](\boldsymbol{v},\Lambda),\\ \left(H+(S_1[H,b])^\star HS_1[H,b]+(S_2[b])^\star HS_2[b]\right)\Lambda=\mathbb{Q}[H,b](\eta,u),\end{cases} \tag{2.12}$$

where $\boldsymbol{v}:=(\eta,q)$ gathers the flow variables, Λ is an auxiliary coupling variable between the hyperbolic equation $(2.12)_1$ and the elliptic equation $(2.12)_2$, and

$$\boldsymbol{F}(\boldsymbol{v},b):=\begin{pmatrix}q\\ uq+\frac{1}{2}g(\eta^2-2\eta b)\end{pmatrix},\quad \boldsymbol{B}[b](\boldsymbol{v}):=\begin{pmatrix}0\\ -g\eta\partial_x b\end{pmatrix},\quad \boldsymbol{D}[H](\boldsymbol{v},\Lambda):=\begin{pmatrix}0\\ gH\partial_x\eta-H\Lambda\end{pmatrix}.$$

In addition, observing that

$$\left(H+S_1[H,b]^\star HS_1[H,b]+S_2[b]^\star HS_2[b]\right)w=\partial_x(-\kappa[H]\partial_x w)-\beta[H,\partial_x b]\partial_x w+\partial_x(\beta[H,\partial_x b]w)+\delta[H,\partial_x b]w, \tag{2.13}$$

with

$$\kappa[H]:=\frac{1}{3}H^3,\quad \beta[H,\partial_x b]:=\frac{\sqrt{3}}{2}\kappa[H]^{\frac{1}{2}}H^{\frac{1}{2}}\partial_x b,\quad \delta[H,\partial_x b]:=(H^{\frac{1}{2}}\partial_x b)^2+H,$$

system (2.12) finally takes the form

$$\begin{cases}\partial_t\boldsymbol{v}+\partial_x\boldsymbol{F}(\boldsymbol{v},b)=\boldsymbol{B}[b](\boldsymbol{v})+\boldsymbol{D}[H](\boldsymbol{v},\Lambda),\\ \partial_x(-\kappa[H]\partial_x\Lambda)-\beta[H,\partial_x b]\partial_x\Lambda+\partial_x(\beta[H,\partial_x b]\Lambda)+\delta[H,\partial_x b]\Lambda=\mathbb{Q}[H,b](\eta,u).\end{cases} \tag{2.14}$$

Remark 2.1.4 (*Boundary conditions*). *In practice, the numerical solution of (2.14) is sought on a bounded spatial domain $\Omega\subset\mathbb{R}$. In order to close system (2.14), suitable boundary conditions must be prescribed not only for the primal variables $\boldsymbol{v}:=(\eta,q)$, but also for the auxiliary variable Λ, so that the associated discrete problem is well-defined on Ω. In this work, we only consider reflexive boundary conditions, designed to reproduce the presence of a solid wall. They are imposed as*

$$\partial_x\eta_{|\partial\Omega}=0,\quad q_{|\partial\Omega}=0,\quad \partial_x\Lambda_{|\partial\Omega}=0. \tag{2.15}$$

2.2 SWIP-DG discretization

We keep the notations introduced in Chapter 1 for the mesh $\mathcal{T}_h$ and its associated skeleton $\mathcal{F}_h$. In particular, we denote by $\mathcal{F}_h^0$ the set of interior vertices of the mesh, and by $\mathcal{F}_h^\partial$ the set of boundary vertices. For every vertex $\Gamma \in \mathcal{F}_h$, we denote by h_Γ the local mesh size associated with Γ, defined as the minimum of the diameters of the mesh elements sharing Γ. We also introduce a positive penalty parameter $\xi > 0$, which will be used to ensure the coercivity of the discrete bilinear form associated with the elliptic operator in (2.14).

2.2.1 Discrete bilinear form

Let $\kappa, \beta, \delta \in L^\infty(\Omega)$ be uniformly bounded coefficients and, for brevity, let us write $\kappa_{|\omega_i} := \kappa_{|\omega_i}$, $\beta_{\omega_i} := \beta_{|\omega_i}$ and $\delta_{\omega_i} := \delta_{|\omega_i}$ for every $\omega_i \in \mathcal{T}_h$. Following [82, 96], we introduce the jump and weighted average operators as follows. For any sufficiently regular function φ and any interior vertex $\Gamma \in \mathcal{F}_h^0$ shared by two distinct mesh elements ω_i and ω_j, that is, such that $\Gamma \subset \partial\omega_i \cap \partial\omega_j$, we set

$$[\![\varphi]\!]_\Gamma := \varphi_{|\omega_i} - \varphi_{|\omega_j}, \quad \{\!\{\varphi\}\!\}_{\lambda,\Gamma} := \lambda_2\varphi_{|\omega_i} + \lambda_1\varphi_{|\omega_j}, \quad \lambda_p := \frac{\kappa_{\omega_i}}{\kappa_{\omega_i}+\kappa_{\omega_j}}, \quad \forall p \in \{1,2\}. \tag{2.16}$$

In what follows, and whenever no ambiguity is possible, we may omit the subscript Γ in both $[\![\varphi]\!]$ and $\{\!\{\varphi\}\!\}$. In the particular case where $\kappa \equiv \mathsf{c}$ in Ω for some constant $\mathsf{c} > 0$, we simply take $\lambda_1 = \lambda_2 = \frac{1}{2}$ and omit the subscript λ as well. At boundary vertices, the definition of jumps and averages depends on the variable under consideration and on the boundary conditions that are prescribed. We refer to [84], for a discussion on weighted averages and harmonic means in the framework of heterogeneous diffusion problems.

For later use, we recall the bilinear form $\mathcal{A}_h(\kappa, \beta, \delta; \cdot, \cdot)$ on $\mathbb{P}^k(\mathcal{T}_h) \times \mathbb{P}^k(\mathcal{T}_h)$ introduced in [292], defined by

$$\begin{aligned}
\mathcal{A}_h(\kappa,\beta,\delta;v_h,w_h) :=& \sum_{\omega_i\in\mathcal{T}_h}\int_{\omega_i}\kappa\partial_x v_h\partial_x w_h\,\mathrm{d}x + \sum_{\Gamma\in\mathcal{F}_h}\xi\int_\Gamma \gamma_{\kappa,\Gamma}h_\Gamma^{-1}[\![v_h]\!][\![w_h]\!]\,\mathrm{d}s \\
&- \sum_{\Gamma\in\mathcal{F}_h}\left(\int_\Gamma \{\!\{\kappa\partial_x v_h\}\!\}_\lambda [\![w_h]\!]\,\mathrm{d}s + \int_\Gamma [\![v_h]\!]\{\!\{\kappa\partial_x w_h\}\!\}_\lambda\,\mathrm{d}s\right) \\
&- \sum_{\omega_i\in\mathcal{T}_h}\int_{\omega_i}\beta v_h\partial_x w_h\,\mathrm{d}x - \sum_{\omega_i\in\mathcal{T}_h}\int_{\omega_i}\partial_x v_h\beta w_h\,\mathrm{d}x \\
&+ \sum_{\Gamma\in\mathcal{F}_h}\left(\int_\Gamma \{\!\{\beta v_h\}\!\}_\lambda [\![w_h]\!]\,\mathrm{d}s + \int_\Gamma [\![v_h]\!]\{\!\{\beta w_h\}\!\}_\lambda\,\mathrm{d}s\right) \\
&+ \sum_{\omega_i\in\mathcal{T}_h}\int_{\omega_i}\delta v_h w_h\,\mathrm{d}x.
\end{aligned} \tag{2.17}$$

The κ-dependent penalty parameter $\gamma_{\kappa,\Gamma}$ is defined by

$$\gamma_{\kappa,\Gamma} := \begin{cases} \frac{2\kappa_{\omega_i}\kappa_{\omega_j}}{\kappa_{\omega_i}+\kappa_{\omega_j}}, & \text{if } \Gamma \in \mathcal{F}_h^0 \text{ and } \Gamma = \partial\omega_i \cap \partial\omega_j, \\ \kappa_{\omega_i}, & \text{if } \Gamma \in \mathcal{F}_h^\partial \text{ and } \Gamma = \partial\omega_i \cap \partial\Omega. \end{cases} \tag{2.18}$$

In (2.17), ξ stands for a user-chosen parameter, taken large enough to guarantee coercivity. Moreover, $\partial_x^h v_h$ denotes the partial derivative with respect to x, computed element by element on the mesh $\mathcal{T}_h$, namely, for any $v_h \in \mathbb{P}^k(\mathcal{T}_h)$,

$$(\partial_x^h v_h)_{|\omega_i} = \partial_x(v_{h|\omega_i}).$$

2.2.2 Discrete Laplace and gradient operators

As in [85], in order to discretize the linear and nonlinear operators appearing in our models, we introduce discrete analogues of the gradient and Laplace operators acting on broken polynomial functions. For any $v_h \in \mathbb{P}^k(\mathcal{T}_h)$, we define the global lifting of the jumps of v_h by

$$\mathcal{R}_h^k([\![v_h]\!]) := \sum_{\Gamma\in\mathcal{F}_h} r_\Gamma^k([\![v_h]\!]), \tag{2.19}$$

see, e.g., [84].

For every $\Gamma \in \mathcal{F}_h$, the local lifting operator $r_\Gamma^k([\![v_h]\!]) \in \mathbb{P}^k(\mathcal{T}_h)$ is defined as the unique solution to

$$\big(r_\Gamma^k([\![v_h]\!]), \psi_h\big)_\Omega = \big\langle [\![v_h]\!], \{\!\{\psi_h\}\!\}_{n_\Gamma} \big\rangle_\Gamma, \quad \forall \psi_h \in \mathbb{P}^k(\mathcal{T}_h), \tag{2.20}$$

where $\{\!\{\psi_h\}\!\}$ denotes the standard average operator defined as in (2.16), with weights $\omega_i = \omega_j = \frac{1}{2}$ at interior nodes, and extended to boundary nodes as described in the previous section.

Following [84], Section 2.3, we then define the discrete gradient operator

$$\nabla_h^k : \mathbb{P}^k(\mathcal{T}_h) \to \mathbb{P}^k(\mathcal{T}_h)$$

by setting, for all $v_h \in \mathbb{P}^k(\mathcal{T}_h)$,

$$\nabla_h^k v_h := \partial_x^h v_h - \mathcal{R}_h^k([\![v_h]\!]). \tag{2.21}$$

Remark 2.2.1 (*On the practical relevance of lifted gradients*). *This operator enjoys better consistency properties than the broken gradient ∂_x^h, since the second term accounts for the jumps of the discrete function. Indeed, evaluating spatial derivatives with the discrete lifted gradient ∇_h^k rather than the standard broken gradient ∂_x^h is particularly interesting for low-order approximations; for instance, in shallow-water models, the bottom topography is often represented by piecewise constant slopes ($k = 1$), meaning only the first two polynomial modes are non-zero. Differentiating this directly with ∂_x^h drastically reduces the available information to a single mode per cell, while completely missing the topographic variations concentrated at the cell interfaces. The local lifting operator $\mathcal{R}_h^k$ acts as a local L^2-projection that extends interface jumps into a volume-defined function on the adjacent cells. It effectively links boundary values to global test functions, enabling a consistent variational treatment of interface terms. Consequently, the lifted gradient ∇_h^k "sees" the interface jumps and recovers a much richer, consistent derivative. For this reason, computing all gradients, including topography and nonlinear fluxes, via this lifted operator systematically enhances the accuracy and stability of the formulation. We refer to [83], Theorem 2.2, for further details.*

Inspired by [168], Eq. (2.10), we also introduce the discrete Laplace operator

$$\Delta_h^k : \mathbb{P}^k(\mathcal{T}_h) \to \mathbb{P}^k(\mathcal{T}_h),$$

defined as follows: for every $v_h \in \mathbb{P}^k(\mathcal{T}_h)$, the function $\Delta_h^k(v_h)$ is the unique element satisfying

$$-\big(\Delta_h^k v_h, \psi_h\big)_\Omega = \mathcal{A}_h(v_h, \psi_h), \quad \forall \psi_h \in \mathbb{P}^k(\mathcal{T}_h), \tag{2.22}$$

where the bilinear form $\mathcal{A}_h(v_h, \psi_h)$ is given by (2.17) with the particular choices $\kappa := 1$, $\beta := 0$ and $\delta := 0$.

It can be shown that, for any $v \in \mathsf{H}_0^1(\Omega) \cap \mathsf{H}^{k+1}(\Omega)$, the following approximation estimates hold:

$$\inf_{v_h \in \mathbb{P}^k(\mathcal{T}_h)} \|\nabla v - \nabla_h^k v_h\| \lesssim h^k, \qquad \inf_{v_h \in \mathbb{P}^k(\mathcal{T}_h)} \|\Delta v - \Delta_h^k v_h\| \lesssim h^{k-1}, \tag{2.23}$$

where the notation $a \lesssim b$ means that $a \le \mathsf{c} b$ for some constant $\mathsf{c} > 0$ independent of the mesh size h. We also recall that the second estimate in (2.23) requires mesh quasi-uniformity; see [83, 168].

2.2.3 Semi-discrete DG–SWIP-DG formulation

The semi-discrete in space discontinuous Galerkin approximation of (2.14) reads:

Find $\boldsymbol{v}_h = (\eta_h, q_h) \in (\mathbb{P}^k(\mathcal{T}_h))^2$ and $\Lambda_h \in \mathbb{P}^k(\mathcal{T}_h)$ such that, for all $(\phi_h, \varphi_h) \in (\mathbb{P}^k(\mathcal{T}_h))^2$,

$$\begin{cases} \big(\partial_t \boldsymbol{v}_h, \varphi_h\big)_{\mathcal{T}_h} + \big(\mathcal{U}_h \circ \boldsymbol{v}_h, \varphi_h\big)_{\mathcal{T}_h} = 0 & \forall \varphi_h \in \mathbf{P}^k(\mathcal{T}_h), \\ \mathcal{A}_h(\kappa[H_h], \beta[H_h, \nabla b_h], \delta[H_h, \nabla b_h]; \Lambda_h, \phi_h) = \big(\mathbb{Q}_h[H_h, b_h](\eta_h, u_h), \phi_h\big)_{\mathcal{T}_h} & \forall \phi_h \in \mathbb{P}^k(\mathcal{T}_h), \end{cases} \tag{2.24}$$

where:

(i) The discrete nonlinear operator $\mathcal{U}_h$ is defined as

$$\begin{aligned} \big(\mathcal{U}_h \circ \boldsymbol{v}_h, \varphi_h\big)_{\mathcal{T}_h} := &- \big(\boldsymbol{F}(\boldsymbol{v}_h, b_h), \partial_x \varphi_h\big)_{\mathcal{T}_h} + \big\langle \boldsymbol{\mathcal{F}}_{\partial\omega_i}, \varphi_h \big\rangle_{\partial\mathcal{T}_h} \\ &- \big(\boldsymbol{B}[b_h](\boldsymbol{v}_h), \varphi_h\big)_{\mathcal{T}_h} - \big(\boldsymbol{D}[H_h](\boldsymbol{v}_h, \Lambda_h), \varphi_h\big)_{\mathcal{T}_h}, \end{aligned}$$

and the discrete dispersive correction reads:

$$\big(\boldsymbol{D}[H_h](\boldsymbol{v}_h, \Lambda_h), \varphi_h\big)_{\mathcal{T}_h} = \Big(0, \big(\mathrm{g} H_h \nabla_h^k \eta_h - H_h \Lambda_h, \varphi_h\big)_{\mathcal{T}_h}\Big)^\top,$$

where $\boldsymbol{\mathcal{F}}_{\partial\omega}$ is a suitable approximation of the normal face fluxes $\boldsymbol{F}(\boldsymbol{v}_h, b_h) \cdot \mathbf{n}_{\partial\omega}$, defined in Chapter 1.

(*ii*) The discrete operators $\kappa[H_h]$, $\beta[H_h,\nabla b_h]$, and $\delta[H_h,\nabla b_h]$ are recalled here:

$$\kappa[H_h] := \frac{\alpha}{3}H_h^3, \qquad \beta[H_h,\nabla b_h] := \frac{\sqrt{3}}{2}\kappa[H_h]^{\frac{1}{2}}H_h^{\frac{1}{2}}\partial_x b_h, \qquad \delta[H_h,\nabla b_h] := (H_h^{\frac{1}{2}}\partial_x b_h)^2 + H_h,$$

(*iii*) The discrete nonlinear operator $\mathbb{Q}_h[H_h,b_h]$ is defined by:

$$\mathbb{Q}_h[H_h,b_h](\eta_h,u_h) := \frac{1}{\alpha}gH_h\nabla_h^k\eta_h + H_h\mathbb{Q}_{1,h}[H_h,b_h](u_h).$$

where, for any $w_h \in \mathbb{P}^k(\mathcal{T}_h)$,

$$\begin{aligned}\mathbb{Q}_{1,h}[H_h,b_h](w_h) := {} & 2H_h\nabla_h^k(H_h + \frac{b_h}{2})(\nabla_h^k w_h)^2 + \frac{4}{3}H_h^2\nabla_h^k w_h\Delta_h^k w_h \\ & + H_h\Delta_h b_h(\nabla_h^k w_h)w_h + \left(\nabla_h^k\eta_h\Delta_h b_h + \frac{H_h}{2}\nabla^3 b_h\right)w_h^2.\end{aligned}$$

(*iv*) The discrete water-height is $H_h := \eta_h - b_h$ and velocity $u_h := \pi_h^k(q_h H_h^{-1})$.

Let us note that it has been proved in [292] that the proposed discrete bilinear form in (2.17) is symmetric, consistent with the elliptic operator $H(1+\mathcal{T}[H,b])$ and enjoys some discrete coercivity property, provided that the penalty coefficient ξ is large enough, so that the corresponding discrete problem is well-posed.

2.2.4 Time discretization

As usual, time integration is handled using explicit *Strong Stability Preserving Runge-Kutta* schemes (see Time discretization and time-marching algorithm), where the initial data $\boldsymbol{v}_h^0$ is defined as the L^2 projection of the initial datum. As our stabilization method both uses DG schemes on primal cells $\omega_i \in \mathcal{T}_h$ and FV schemes on subcells $S_m^i \in \mathcal{T}_{\omega_i}^k$, the time step Δt is computed adaptively using the following CFL condition on submean values of the velocity and water-height:

$$\Delta t^n := \frac{\min_{\omega_i\in\mathcal{T}_h}\left(\frac{h_{\omega_i}}{2k+1}, \min_{S_m^i\in\mathcal{T}_{\omega_i}^k}|S_m^i|\right)}{\sigma}$$

where σ is defined as

$$\sigma := \max_{\omega_i\in\mathcal{T}_h}\sigma_{\omega_i} = \max_{\omega_i\in\mathcal{T}_h}\left(\max_{m\in[\![1,k+1]\!]}\left[|\overline{u}_m^i| + \sqrt{g\overline{H}_m^i}\right]\right).$$

2.3 Combined DG/FV–SWIP-DG fully-discrete formulations

Following the lines of Chapter 1, one can show that the high-order DG method associated with a first-order time-marching method is equivalent to the following in-cell FV-like scheme written on the subcell mean values. We then get the following:

Find $(\overline{\boldsymbol{v}}_h^{n+1},\Lambda_h^n) \in \mathbf{P}^0(\widetilde{\mathcal{T}}_h)\times\mathbb{P}^k(\mathcal{T}_h)$ such that, for all $\omega_i\in\mathcal{T}_h$, all $\phi_h\in\mathbb{P}^k(\mathcal{T}_h)$ and all $m\in[\![1,k+1]\!]$,

$$\begin{cases}\overline{\boldsymbol{v}}_m^{i,n+1} = \overline{\boldsymbol{v}}_m^{i,n} - \frac{\Delta t^n}{|S_m^i|}\left(\widehat{\boldsymbol{F}}_{m+\frac{1}{2}}^{i,-} - \widehat{\boldsymbol{F}}_{m-\frac{1}{2}}^{i,+}\right) + \Delta t^n\left(\overline{\boldsymbol{B}}_m^i + \overline{\boldsymbol{D}}_m^i\right) & \forall\omega_i\in\mathcal{T}_h, \\ \overline{\boldsymbol{D}}_m^i := \overline{\pi}_i^k\circ\pi_i^k\boldsymbol{D}[H_h^n](\boldsymbol{v}_h^n,\Lambda_h^n)_{|\omega_i} & \forall\omega_i\in\mathcal{T}_h, \\ \mathscr{A}_h\big(\kappa[H_h^n],\beta[H_h^n,\nabla b_h],\delta[H_h^n,\nabla b_h];\Lambda_h^n,\phi_h\big) = \big(\mathbb{Q}_h[H_h^n,b_h](\eta_h^n,u_h^n),\phi_h\big)_{\mathcal{T}_h} & \forall\phi_h\in\mathbb{P}^k(\mathcal{T}_h).\end{cases} \tag{2.25}$$

where $\overline{\pi}_i^k\circ\pi_i^k\boldsymbol{D}[H_h^n](\boldsymbol{v}_h^n,\Lambda_h^n)_{|\omega_i}$ can be seen as the submean value of the dispersive correction $\boldsymbol{D}[H_h^n](\boldsymbol{v}_h^n,\Lambda_h^n)$:

$$\overline{\pi}_i^k\circ\pi_i^k\boldsymbol{D}[H_h^n](\boldsymbol{v}_h^n,\Lambda_h^n)_{|\omega_i} = \left(\frac{1}{|S_m^i|}\int_{S_m^i}\pi_i^k\boldsymbol{D}[H_h^n](\boldsymbol{v}_h^n,\Lambda_h^n)\,\mathrm{d}x\right)_{|\omega_i}.$$

Here, the reconstructed fluxes $\{\widehat{\boldsymbol{F}}_{m\pm\frac{1}{2}}^{i,\mp}\}_{m\in[\![0,k+1]\!]}$ are defined as in (1.25)–(1.26). The first equation in (2.25) is only a FV-like rewriting of the DG update on the subcell averages: it does not introduce any additional

stabilization. So, in order to enhance robustness near steep gradients, shocks, or nearly dry regions, we introduce the local monolithic DG/FV subcell stabilization operator

$$\mathcal{S}_h^{n,\star} : \mathbf{P}^k(\mathcal{T}_h) \ni \boldsymbol{v}_h \mapsto \mathcal{S}_h^{n,\star} \circ \boldsymbol{v}_h = (\overline{\pi}_h^k)^{-1} \circ \overline{\boldsymbol{v}}_h^\star \in \mathbf{P}^k(\mathcal{T}_h), \tag{2.26}$$

where $\overline{\boldsymbol{v}}_h^\star \in \mathbf{P}^0(\widetilde{\mathcal{T}}_h)$ denotes the stabilized subcell mean values obtained from the convex limiting procedure introduced in Chapter 1. In the present fully-discrete update, this operator is applied to the explicit DG candidate $\boldsymbol{v}_h = \boldsymbol{v}_h^n - \Delta t^n \left(\mathcal{U}_h \circ \boldsymbol{v}_h^n\right)$. Equivalently, the corrected subcell mean values are computed by replacing the reconstructed DG fluxes $\widehat{\boldsymbol{F}}$ by the blended fluxes $\widetilde{\boldsymbol{F}}$ in the FV-like subcell update.

The combined local subcell monolithic DG/FV–SWIP-DG fully-discrete formulation is then written as follows: find $(\overline{\boldsymbol{v}}_h^{n+1,\star}, \Lambda_h^n) \in \mathbf{P}^0(\widetilde{\mathcal{T}}_h) \times \mathbb{P}^k(\mathcal{T}_h)$ such that, for all $\omega_i \in \mathcal{T}_h$, all $\phi_h \in \mathbb{P}^k(\mathcal{T}_h)$ and all $m \in [\![1, k+1]\!]$,

$$\begin{cases} \overline{\boldsymbol{v}}_m^{i,n+1,\star} = \overline{\boldsymbol{v}}_m^{i,n} - \frac{\Delta t^n}{|S_m^i|}\left(\widetilde{\boldsymbol{F}}_{m+\frac{1}{2}}^{i,-} - \widetilde{\boldsymbol{F}}_{m-\frac{1}{2}}^{i,+}\right) + \Delta t^n \left(\overline{\boldsymbol{B}}_m^i + \overline{\boldsymbol{D}}_m^i\right) & \forall \omega_i \in \mathcal{T}_h, \\ \overline{\boldsymbol{D}}_m^i := \overline{\pi}_i^k \circ \pi_i^k \boldsymbol{D}[H_h^n](\boldsymbol{v}_h^n, \Lambda_h^n)_{|\omega_i} & \forall \omega_i \in \mathcal{T}_h, \\ \mathscr{A}_h\big(\kappa[H_h^n], \beta[H_h^n, \nabla b_h], \delta[H_h^n, \nabla b_h]; \Lambda_h^n, \phi_h\big) = \big(\mathbb{Q}_h[H_h^n, b_h](\eta_h^n, u_h^n), \phi_h\big)_{\mathcal{T}_h} & \forall \phi_h \in \mathbb{P}^k(\mathcal{T}_h). \end{cases} \tag{2.27}$$

The blended fluxes and the associated blending coefficients are defined in (1.27) and (1.2.5)-(1.3.4). Once the corrected subcell averages $\overline{\boldsymbol{v}}_h^{n+1,\star}$ have been computed, the updated DG global solution is reconstructed. Equivalently, the fully-discrete monolithic DG/FV–SWIP-DG scheme can be summarized in the compact form: find $(\boldsymbol{v}_h^{n+1}, \Lambda_h^n) \in \mathbf{P}^k(\mathcal{T}_h) \times \mathbb{P}^k(\mathcal{T}_h)$ such that, for all $\phi_h \in \mathbb{P}^k(\mathcal{T}_h)$,

$$\begin{cases} \boldsymbol{v}_h^{n+1} = \mathcal{S}_h^{n,\star} \circ \{\boldsymbol{v}_h^n - \Delta t^n \left(\mathcal{U}_h \circ \boldsymbol{v}_h^n\right)\} & \forall \varphi_h \in \mathbf{P}^k(\mathcal{T}_h), \\ \mathscr{A}_h\big(\kappa[H_h^n], \beta[H_h^n, \nabla b_h], \delta[H_h^n, \nabla b_h]; \Lambda_h^n, \phi_h\big) = \big(\mathbb{Q}_h[H_h^n, b_h](\eta_h^n, u_h^n), \phi_h\big)_{\mathcal{T}_h} & \forall \phi_h \in \mathbb{P}^k(\mathcal{T}_h). \end{cases} \tag{2.28}$$

where the action of $\mathcal{S}_h^{n,\star}$ is understood through the subcell update (2.27).

Remark 2.3.1 (*Interpretation of the stabilization operator*). *The operator $\mathcal{S}_h^{n,\star}$ is designed as a compact notation for the monolithic DG/FV update written on the subcell mean values. Starting from the explicit DG candidate $\boldsymbol{v}_h^n - \Delta t^n(\mathcal{U}_h \circ \boldsymbol{v}_h^n)$, the corresponding subcell update is corrected by replacing the reconstructed fluxes $\widehat{\boldsymbol{F}}$ with the blended fluxes $\widetilde{\boldsymbol{F}}$. The corrected subcell mean values $\overline{\boldsymbol{v}}_h^{n+1,\star}$ are then reconstructed into the DG space through $(\overline{\pi}_h^k)^{-1}$.*

Remark 2.3.2 (*Extension to SSP–RK stages*). *The formulation above is written for a first-order explicit time discretization. For an SSP–RK method, the same construction is applied at each Runge–Kutta stage. The state $\boldsymbol{v}_h^n$ is then replaced by the corresponding stage value, the SWIP-DG problem is solved with the stage data, and the stabilization operator $\mathcal{S}_h^{n,\star}$ is applied stage by stage through the same subcell flux blending procedure.*

Remark 2.3.3 (*Implementation*). *From a practical viewpoint, let us consider a "classical" DG numerical implementation in which, for any given mesh element ω_i, the dispersive contributions to the residuals in the momentum balance law $\boldsymbol{D\Psi}_h^i \in \mathbb{R}^{k+1}$ defined as*

$$(\boldsymbol{D\Psi}_h^i)_{j \in [\![1,k+1]\!]} = \int_{\omega_i} \boldsymbol{D}_h^i \psi_j^i \, \mathrm{d}x$$

are available. Then, we have the following identity:

$$\overline{\boldsymbol{D}}_h^i = \mathbb{P}_i \mathbb{M}_i^{-1} \boldsymbol{D\Psi}_h^i,$$

where $\mathbb{P}_i$ is the projection matrix introduced in Remark 1.1.3 and $\mathbb{M}_i$ is the mass-matrix associated with the element ω_i and the set of basis function (1.4).

Arbitrary order well-balancing property

The following well-balancing result follows directly from the well-balancing property of the pre-balanced shallow-water equations using DG/FV subcells method and the well-posedness of the elliptic sub-problem established earlier.

Proposition 2.3.4 (*Well-balancing property*). *The discrete formulation obtained in* (2.24) *together with a first-order Euler time-marching algorithm and DG/FV subcells numerical fluxes, preserves the motionless steady states, providing that the integrals of* (2.24) *are exactly computed for the motionless steady states. We then have:*

$$\forall n \in \mathbb{N}, \quad \forall \eta^e \in \mathbb{R}, \qquad (\eta_h^n = \eta^e \text{ and } q_h^n = 0) \Rightarrow \left(\eta_h^{n+1} = \eta^e \text{ and } q_h^{n+1} = 0\right).$$

Proof. Assuming $\boldsymbol{v}_h \equiv \boldsymbol{v}_h^e := (\eta^e, 0)$, we obtain homogeneous conditions for (2.24), meaning that

$$\mathbb{Q}_h[H_h, b_h](\eta_h, u_h) = 0,$$

along with homogeneous Neumann boundary conditions. Since (2.24) is well-posed (we refer to [292]), the unique solution is $\Lambda_h = 0$. Consequently, we have $\left(\boldsymbol{D}_h, \varphi_h\right)_{\mathcal{T}_h} = 0$. The proof then follows directly from the classical arguments developed for the (non-dispersive) shallow-water equations, together with the result from Proposition 1.3.6. Let us note that in particular, the dispersive contributions to the momentum balance equation vanishes with respect to the velocity, making the preservation of the steady-states directly related to the well-posedness of the discrete elliptic problems $\mathscr{A}_h(\kappa, \beta, \delta; \Lambda_h, \phi_h) = 0$. The interested reader can investigate [211]. □

Preserving water-height positivity

To ensure the stability of our scheme, meaning that the solution remains within the convex set $\mathscr{H}_b^+$, it is sufficient to use a numerical method that guarantees water-height positivity for the hyperbolic part of the equations. As demonstrated in the previous chapter, the monolithic DG/FV scheme effectively prevents simulation breakdowns caused by negative water-heights. This is achieved through an adaptive computation of the blending coefficient, which dynamically balances first-order FV fluxes and high-order reconstructed fluxes to maintain stability.

2.4 Numerical results

In this section, we present several numerical results obtained with the proposed monolithic DG/FV–SWIP-DG method for the SGN equations (2.3)-(2.4)-(2.5), in order to assess the accuracy and robustness of the scheme. We consider polynomial degrees $k = 2$, and use a third-order SSP-RK time-marching scheme (see Time discretization and time-marching algorithm) for all simulations. The penalty parameter in the SWIP-DG discretization is set to $\xi = 10$. Unless otherwise specified, solid wall boundary conditions are applied at the domain boundaries.

2.4.1 Oscillating dam-break

PARAMETERS

Computational domain: $\Omega = [-210, 210]$
Number of cells: $N_{\text{el}} = 800$ cells
Boundary conditions: solid wall

Polynomial degree: $k = 2$
Time-marching order: $p = 3$
Final time: $T_{\max} = 47,5$ sec

Inspired from [188], we investigate the dam-break problem for the SGN model (2.3)-(2.4)-(2.5), with the following initial data:

$$\eta_0(x) := \frac{H_L - H_R}{2}\left(1 - \tanh\left(\frac{x}{\chi}\right)\right), \quad q_0(x) = 0.,$$

and $H_L = 1.8$, $H_R = 1$ and $\chi = 0.4$, which is a regularized piecewise-constant initial data, with a sharp variation initially located at $x = 0$. We show on Figure 2.3 the surface profile observed at $t = 47.5$ sec, and we observe that the initial high-gradient area has divided into an oscillating shock-wave propagating rightward, and a rarefaction wave propagating leftward.

There is no closed-form for this transient oscillating solution, but as pointed out in [122], the analysis of the Witham system for the SGN equations and of Riemann invariants of the NSW equations leads to some approximations of the mean flow values dividing the rarefaction and the dispersive shock areas, and denoted by $(H^\star, u^\star)$:

$$H^\star = \left(\frac{\sqrt{H_L} + \sqrt{H_R}}{2}\right)^2, \qquad u^\star = \left(\sqrt{gH^\star} - \sqrt{gH_R}\right).$$

see Figure 2.3. We observe that these approximate values are correctly estimated by the model. Note also that an approximation of the leading soliton's amplitude is also available, denoted by a^+, and also accurately reproduced, see Figure 2.3.

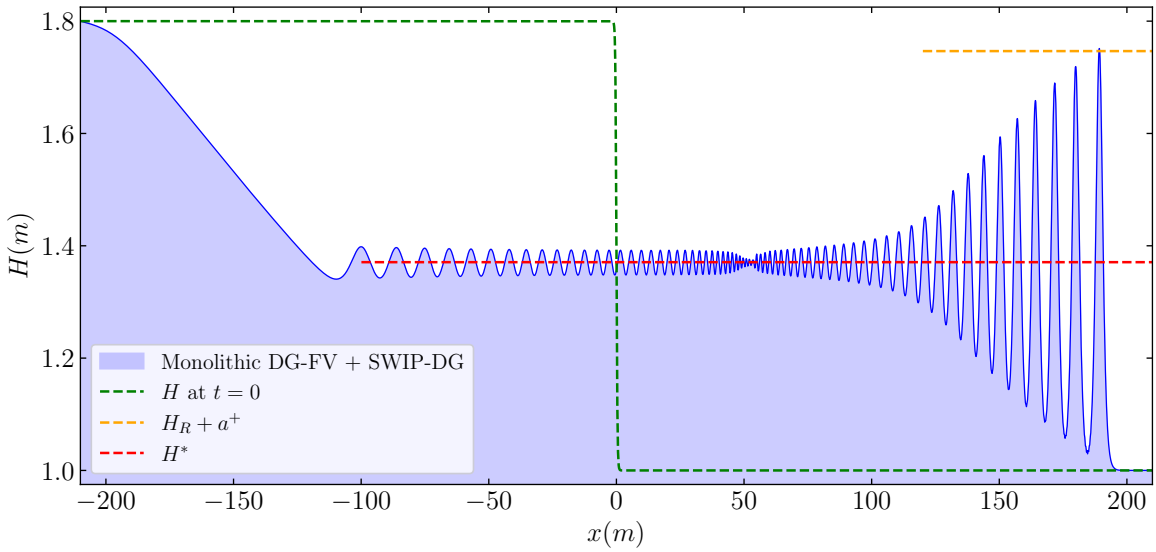


Figure 2.2: Oscillating dam-break − Surface elevation computed for $k = 2$ and $N_{\text{el}} = 800$ at time $T_{\max} = 47.5$ sec (animation).

2.4.2 Solitary-waves over a piecewise-linear bottom

PARAMETERS

Computational domain: $\Omega = [-24, 12]$
Number of cells: $N_{\text{el}} = 150$ cells
Boundary conditions: solid wall

Polynomial degree: $k = 2$
Time-marching order: $p = 3$
Final time: $T_{\max} = 7,5$ sec

In this last test-case, we consider the propagation of a solitary wave, which is a surface perturbation propagating without any temporal evolution in shape or size when viewed in the reference frame moving with the group velocity of the wave. The SGN equations (2.3) do have solitary wave solutions defined as follows:

$$\begin{cases} \eta(t,x) = H_0 + \epsilon H_0 \text{sech}^2\left(\kappa(x-ct)\right), \\ q(t,x) = c\left(\eta(t,x) - H_0\right), \end{cases} \tag{2.29}$$

where ϵ is the amplitude of the wave and $\kappa := \sqrt{\frac{3\epsilon}{4H_0^2(1+\epsilon)}}$ and $c := \sqrt{gH_0(1+\epsilon)}$. For this particular test-case, the computational domain is a $36\,m$ channel with a flat bottom for $x \leq 0$ followed by a linear profile with a slope $1:30$, and a solitary wave propagates from the left boundary, with a water level at rest $H_0 = 0.25\,m$. Measurements of the free-surface are available at various locations along the channel, obtained from wave-gauges from the laboratory study performed at LEGI (Grenoble, France, see [135]). We consider 4 series of experiments, involving increasing wave amplitudes, from $\epsilon = 0.096$ to $\epsilon = 0.534$, which stand for highly nonlinear evolutions. For each experiment, we compare the numerical results with time-series of the free-surface elevation at wave gauges locations (we do not recall the exact location of the gauges, which may be found, for instance, in Table 2 of [37]). We observe some excellent agreements, especially regarding the maximum surface elevation during the shoaling process, and considering the low number of mesh elements. The wave asymmetrical profile is also accurately reproduced even in the most nonlinear cases. We emphasize that neither stabilization process is needed (as the solution is ensured to remain smooth), nor additional filtering treatment to control possible aliasing effects due to the numerical approximations of highly nonlinear processes.

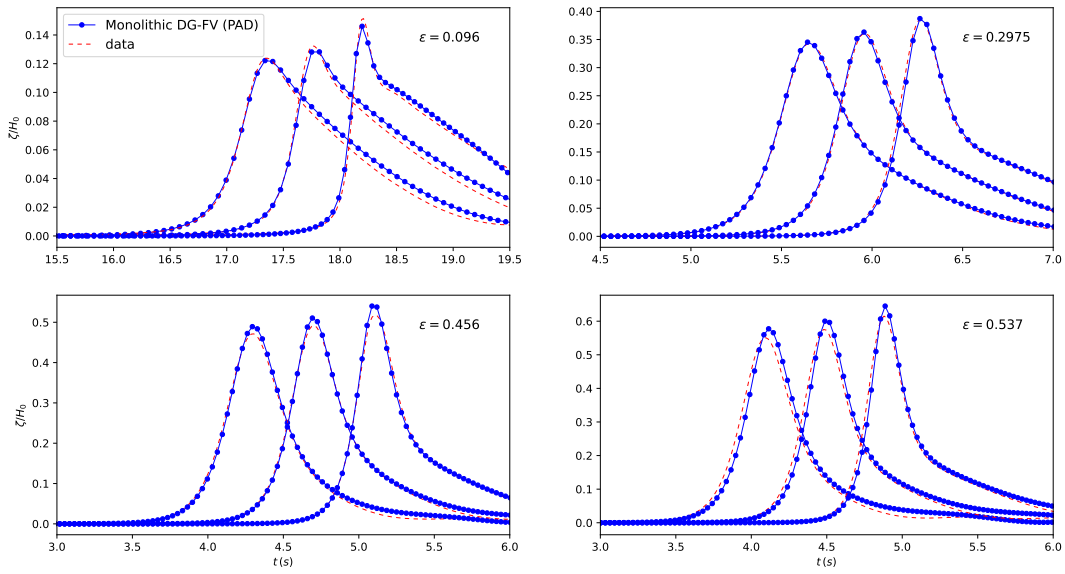


Figure 2.3: Solitary-waves over a piecewise-linear bottom – Time-series of the free-surface elevation at wave gauges locations, for the 4 experiments with increasing wave amplitudes.

CHAPTER 3

Local monolithic DG/FV subcell scheme for 2D nonlinear shallow-water equations on unstructured grids

The main results presented in this chapter have been submitted [52].

This chapter develops the two-dimensional version of the monolithic DG/FV subcell convex property preserving strategy for the nonlinear shallow-water equations on unstructured meshes. The extension is not a purely formal generalization of the one-dimensional construction: in two space dimensions, the subcell partitions, the definition of local states, the evaluation of interface fluxes and the treatment of geometrical information must all be adapted to general unstructured elements. The objective is to preserve the main properties obtained in Chapter 1, while making the method applicable to genuinely multidimensional configurations.

The resulting scheme keeps the central idea of the monolithic construction: the high-order DG update is interpreted locally as a finite-volume balance on a subcell partition and is blended with a robust low-order update through interface-based convex coefficients. These coefficients are designed to enforce the relevant admissibility constraints, including positivity of the water-height and nonlinear stability indicators, while allowing the high-order DG flux to be recovered in smooth regions. A specific effort is also devoted to the well-balanced treatment of topography and friction on unstructured meshes, so that the two-dimensional formulation retains the equilibrium-preserving properties of the one-dimensional method. The numerical results show that the scheme is able to compute demanding shallow-water benchmarks on complex meshes, with sharp shock resolution, robust control of nonphysical oscillations and physically admissible states.

Contents

3.1 Continuous and discrete settings 66
3.1.1 Nonlinear shallow-water model in two space dimension 66
3.1.2 Discrete framework 67
3.2 Discrete formulations and flux reconstruction 71
3.2.1 Classical DG formulation 72
3.2.2 Equivalency between DG and a subcell FV-like scheme 72
3.3 Local subcell monolithic DG/FV scheme 76
3.3.1 High and low-order topography-dependent numerical fluxes 76
3.3.2 Blending high and low-order fluxes 78
3.3.3 Source term discretization 78
3.4 Calibrating the blended fluxes 79
3.4.1 Blended and first-order Riemann intermediate states 80
3.4.2 Blending coefficient smoothening 80
3.4.3 Water-height positivity (PAD criterion) 81
3.4.4 Relaxed subcell numerical admissibility detection (SubNAD criterion) 82
3.4.5 Well-balancing for motionless steady-states 83
3.5 Numerical simulations 85
3.5.1 Steady vortex with smooth topography 85
3.5.2 Well-balancing 86
3.5.3 Dam-breaks in a channel 90
3.5.4 Circular dam-breaks 95
3.5.5 Carrier-Greenspan solutions 95
3.5.6 Wave-island interactions 101
3.5.7 Tsunami on an island with conical obstacles 108

Τα λεφτα μου όλα δίνω γιά ένα tango
κι ένα άγγιγμά σου κάτω από το τραπέζι
FROM: ***"Gia Ena Tango"***, Háris Alexíou – *Ena Fili Tou Kosmou* (1997)

3.1 Continuous and discrete settings

Here we introduce the continuous and discrete settings for the two-dimensional nonlinear shallow-water equations (NSW) with topography and friction effects.

3.1.1 Nonlinear shallow-water model in two space dimension

Considering $b \in \mathscr{C}^\infty(\Omega)$ the topography variation, denoting again the free-surface elevation by η, the water discharge by $\boldsymbol{q} = (q_x, q_y)^\top$ and $\boldsymbol{u} = (u_x, u_y)^\top$ as the depth-averaged water velocity (see Figure 3.1), the 2D pre-balanced NSW written as follows

$$\begin{cases} \partial_t \eta + \nabla_{\boldsymbol{x}} \cdot \boldsymbol{q} = S_1[\tau](\boldsymbol{v}), \\ \partial_t \boldsymbol{q} + \nabla_{\boldsymbol{x}} \cdot \left(\boldsymbol{u} \otimes \boldsymbol{q} + \frac{\mathrm{g}\eta}{2}(\eta - 2b)\mathbb{I}_2 \right) = \boldsymbol{S}_2[\tau](\boldsymbol{v}), \end{cases}$$

or in a compact form

$$\partial_t \boldsymbol{v} + \nabla_{\boldsymbol{x}} \cdot \mathbb{F}(\boldsymbol{v}, b) = \boldsymbol{S}[\tau](\boldsymbol{v}), \tag{3.1}$$

supplemented with some initial-data $\boldsymbol{v}(\cdot, 0) \coloneqq \boldsymbol{v}^0$ and where:

- $\mathbb{R}^2 \times \mathbb{R}_+ \ni (\boldsymbol{x}, t) \mapsto \boldsymbol{v}(\boldsymbol{x}, t) \coloneqq (\eta, q_x, q_y)(\boldsymbol{x}, t) \in \mathscr{H}_b^+$ gathers the chosen flow description variables, with
$$\mathscr{H}_b^+ = \{(\eta, q_x, q_y) \in \mathbb{R}^3 \mid H \coloneqq \eta - b \geq 0\},$$
- $\mathscr{H}_b^+ \times \mathbb{R} \ni (\boldsymbol{v}, b) \mapsto \mathbb{F}(\boldsymbol{v}, b) \coloneqq \left(\boldsymbol{q},\ \boldsymbol{u} \otimes \boldsymbol{q} + \frac{\mathrm{g}\eta}{2}(\eta - 2b)\mathbb{I}_2 \right)^\top \in \mathscr{M}_{3\times 2}(\mathbb{R})$ is the (nonlinear) flux tensor,
- $\mathscr{H}_b^+ \times \mathbb{R} \ni (\boldsymbol{v}, \tau) \mapsto \boldsymbol{S}[\tau](\boldsymbol{v}) \coloneqq (S_1[\tau](\boldsymbol{v}), \boldsymbol{S}_2[\tau](\boldsymbol{v}))^\top \in \mathbb{R}^3$ denotes a generic algebraic source term depending on external data τ. Relevant choices for the present work can include, as in the 1D case:
  - a bathymetry source term $\boldsymbol{B}[b](\boldsymbol{v}) \coloneqq (0,\ -\mathrm{g}\eta\nabla_{\boldsymbol{x}} b)^\top$,
  - a pressure-related source term $\boldsymbol{P}[\underline{\mathrm{p}}](\boldsymbol{v}) \coloneqq (0, -\frac{1}{\rho} H \nabla_{\boldsymbol{x}} \underline{\mathrm{p}})^\top$, with $\underline{\mathrm{p}}$ the pressure at the free-surface, ρ the fluid density,
  - a nonlinear Manning-type friction term $\boldsymbol{R}[\mathrm{n_f}](\boldsymbol{v}) \coloneqq (0, -\mathrm{n_f}^2 \frac{\boldsymbol{q}\|\boldsymbol{q}\|}{(\eta - b)^\gamma})^\top$ or a simpler linear one $\boldsymbol{R}^{\mathrm{lin}}[\mathrm{n_f}](\boldsymbol{v}) \coloneqq (0, -\mathrm{n_f}\boldsymbol{q})^\top$.

  These effects may be considered independently or combined.

Remark 3.1.1 (*Friction source term*). *This friction source term will not be considered in the mathematical framework presented in this chapter, but has been handled numerically the same way as the topography one, and a simple cut-off has been added numerically to avoid division by zero when $H^\gamma = (\eta - b)^\gamma$ become small, as we do not investigate further onto those asymptotic behaviors.*

Remark 3.1.2 (*Boundary conditions*). *In practical computations, the NSW system is solved on a bounded spatial domain $\Omega \subset \mathbb{R}^2$ and must be supplemented with appropriate boundary conditions. Depending on the physical configuration, one may consider, for instance,* Dirichlet *boundary conditions, where the values of $(\eta, \boldsymbol{q})$ are prescribed on $\partial\Omega$, or* non-homogeneous Neumann *conditions, where a normal flux or a water-level gradient is imposed. Another standard choice is given by* solid-wall *boundary conditions, which express impermeability through the relation $\boldsymbol{q} \cdot \mathbf{n} = 0$ on $\partial\Omega$. For open boundaries, one may also distinguish* inflow *conditions, where the discharge or velocity profile is prescribed at the entrance, and* outflow *conditions, typically designed as free-exit or radiation-type conditions in order to reduce artificial reflections. These classical boundary conditions can be implemented in various ways within finite-volume or finite-element frameworks for example through ghost cells, penalty terms, or suitable modifications of the numerical fluxes at the boundary.*

Remark 3.1.3 (*Well-posedness*). *For well-posedness results as well as for practical computations, it is standard to assume that the bathymetry b belongs to $\mathscr{C}_{\mathrm{b}}^\infty(\mathbb{R}^d)$. This ensures in particular that $\nabla_{\boldsymbol{x}} b$ does not introduce singular source terms in (3.1). In practice, even piecewise $\mathscr{C}^1$ bathymetries are commonly used in finite-volume or DG discretisations, provided that the numerical scheme consistently handles discontinuities of $\nabla_{\boldsymbol{x}} b$. For a smooth bathymetry and strictly positive water depth $H = \eta - b$, the NSW equations form a Friedrichs-symmetrizable hyperbolic system. Given initial data $\boldsymbol{v}^0 \in (\mathsf{H}^s(\mathbb{R}^d))^{d+1}$ with $s > 1 + d/2$, the associated Cauchy problem is locally well-posed in*

$$\mathscr{C}([0, T), \mathsf{H}^s(\mathbb{R}^d)) \cap \mathscr{C}^1([0, T), \mathsf{H}^{s-1}(\mathbb{R}^d))$$

for some $T > 0$. This result applies both in one and two space dimensions ($d = 1, 2$), so that the one-dimensional setting discussed previously falls within the same analytical framework, see for instance [167]

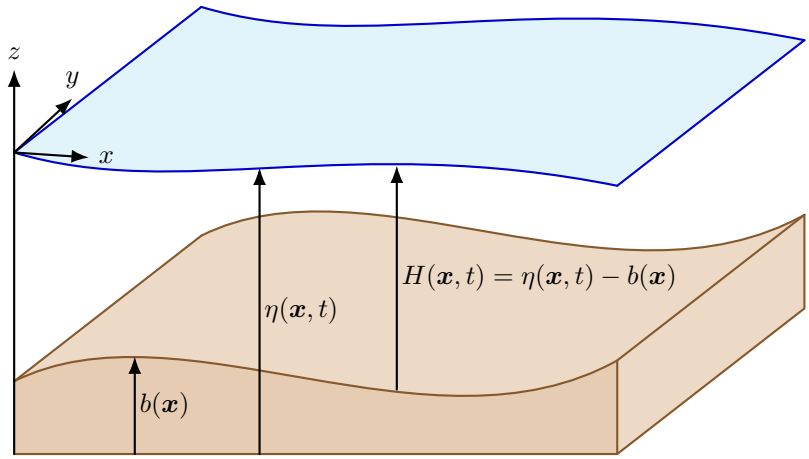


Figure 3.1: Schematic representation of the 2D shallow-water flow configuration, illustrating the free-surface elevation $\eta(\boldsymbol{x},t)$, the bathymetry $b(\boldsymbol{x})$, and the water-height $H(\boldsymbol{x},t)$.

3.1.2 Discrete framework

We try to keep the notations as consistent as possible with those of Chapter 1, but some adaptations are necessary to account for the two-dimensional setting and the use of unstructured meshes.

Domain partitioning and related polynomial approximation

We denote by $\Omega \subset \mathbb{R}^2$ the computational domain, which is assumed to be a bounded open set with Lipschitz boundary $\partial\Omega$. We denote by $\mathcal{T}_h$ a triangulation of Ω into $N_{\rm el}$ open cells ω_c, $c \in [\![1, N_{\rm el}]\!]$, such that

$$\overline{\Omega} = \bigcup_{\omega_c \in \mathcal{T}_h} \overline{\omega_c}, \quad \text{and} \quad \mathring{\omega_c} \cap \mathring{\omega_v} = \emptyset, \ \text{ for all } c \neq v.$$

Given $\omega_c \in \mathcal{T}_h$, we denote by h_c its diameter, and define $h := \max_{c\in[\![1,N_{\rm el}]\!]} h_c$ as the diameter of the largest cell in the mesh. Similarly, $\boldsymbol{x}_{\rm b}^c := (x_{\rm b}^c, y_{\rm b}^c)$ denotes the barycenter of ω_c, and h_x^c (resp. h_y^c) stands for the maximum x- (resp. y-) amplitude of ω_c.

The set of all mesh faces is denoted by $\mathcal{F}_h$ and is partitioned as $\mathcal{F}_h = \mathcal{F}_h^0 \cup \mathcal{F}_h^\partial$, where $\mathcal{F}_h^0$ collects the internal faces (i.e. there exist $\omega_c, \omega_v \in \mathcal{T}_h$ with $c \neq v$ such that $\Gamma \subset \partial\omega_c \cap \partial\omega_v$), and $\mathcal{F}_h^\partial$ the boundary faces (i.e. there exists $\omega_c \in \mathcal{T}_h$ such that $\Gamma \subset \partial\omega_c \cap \partial\Omega$). The length of a face $\Gamma \in \mathcal{F}_h$ is denoted by ℓ_Γ, and its barycenter by $\boldsymbol{x}_{\rm b}^\Gamma$. We denote by h_Γ the minimum of the diameters of the elements to which Γ belongs.

For each cell $\omega_c \in \mathcal{T}_h$, we denote by $\mathcal{F}_c := \{\Gamma \in \mathcal{F}_h \mid \Gamma \subset \partial\omega_c\}$ the set of its faces. During our numerical investigations, we consider simplicial meshes, i.e. $\#\mathcal{F}_c = 3$ for all $\omega_c \in \mathcal{T}_h$, but the method can of course be extended to more general elements. If $\Gamma \in \mathcal{F}_c$, we define $\mathbf{n}_\Gamma$ as the unit normal to Γ pointing outward from ω_c. For internal faces $\Gamma \in \mathcal{F}_h^0$, we fix an arbitrary but consistent orientation and denote the associated normal by $\mathbf{n}_\Gamma := \mathbf{n}_{c\Gamma}$. For boundary faces $\Gamma \in \mathcal{F}_h^\partial$, we take $\mathbf{n}_\Gamma := \mathbf{n}_{c\Gamma}$. If there is no ambiguity, we will often denote $\mathbf{n}_{cv(k)}$ as the normal vector associated to the face $\Gamma_{cv(k)}$ of the element ω_c shared with its neighbor $\omega_{v(k)}$, $k \in [\![1, \#\mathcal{F}_c]\!]$.

For any integer $k \geq 0$, let us define the broken-polynomials space:

$$\mathbb{P}^k(\mathcal{T}_h) := \{v \in L^2(\Omega) \mid \forall \omega_c \in \mathcal{T}_h,\ v_{|\omega_c} \in \mathbb{P}^k(\omega_c)\},$$

where $\mathbb{P}^k(\omega_c)$ denotes the space of polynomials in ω_c of total degree at most k, with $\dim \mathbb{P}^k(\omega_c) = \frac{(k+1)(k+2)}{2}$. In the following, we use the shortcut $N_k := \dim \mathbb{P}^k(\omega_c)$, for all $\omega_c \in \mathcal{T}_h$. Piecewise polynomial functions belonging to some broken-polynomial space $\mathbb{P}^k(\mathcal{T}_h)$ are denoted with a subscript h in the following, and for any $v_h \in \mathbb{P}^k(\mathcal{T}_h)$, and for all $\omega_c \in \mathcal{T}_h$, we use the shortcut $v_h^c := v_{h_{|\omega_c}}$. We also adopt the shorthand notation $\mathbf{P}^k(U) := (\mathbb{P}^k(U))^3$ for any relevant domain U (e.g., $U = \mathcal{T}_h$, $\mathcal{F}_h$, ω_c etc.). For any $\omega_c \in \mathcal{T}_h$, a basis for $\mathbb{P}^k(\omega_c)$ is denoted by

$$\boldsymbol{\Psi}_{\omega_c} := \{\psi_j^c\}_{j\in[\![1,N_k]\!]}, \tag{3.2}$$

and for $v_h^c \in \mathbb{P}^k(\omega_c)$, $\underline{v}^c := \left\{\underline{v}_j^c \mid j \in [\![1, N_k]\!]\right\}$ stand for the *local* degrees of freedom associated with ω_c, introduced as follows:

$$v_h^c(\boldsymbol{x}, t) := \sum_{m=1}^{N_k} \underline{v}_m^c(t)\psi_m^c(\boldsymbol{x}).$$

We also denote by $\underline{v} := \{\underline{v}^c \mid c \in [\![1, N_{\text{el}}]\!]\}$ the *global* degrees of freedom of v_h. For $\omega_i \in \mathcal{T}_h$, we denote by π_c^k the L^2-orthogonal projector onto $\mathbb{P}^k(\omega_c)$ and π_h^k the global L^2-orthogonal projector onto $\mathbb{P}^k(\mathcal{T}_h)$, gathering all the local L^2 projectors π_c^k.

The broken gradient operator is defined as $\nabla_h^k \colon \mathbb{P}^k(\mathcal{T}_h) \to \mathbb{P}^{k-1}(\mathcal{T}_h)$, with, for all $v_h \in \mathbb{P}^k(\mathcal{T}_h)$,

$$(\nabla_h^k v_h)_{|\omega_c} := \nabla_{\boldsymbol{x}} \left(v_{h|\omega_c}\right), \qquad \forall \omega_c \in \mathcal{T}_h.$$

Thus, ∇_h^k acts elementwise: it is the classical gradient applied to the polynomial restriction of v_h on each cell, without imposing any continuity across interfaces. In the following, when no ambiguity is possible, we will simply write $\nabla_{\boldsymbol{x}} v_h$ instead of $\nabla_h^k v_h$, keeping in mind that this notation refers to the broken gradient when applied to discontinuous or piecewise polynomial functions. It may be worth mentioning that the broken gradient operator is not the only possible choice for the discrete gradient in this context, and other definitions can be considered, such as the one based on the lifting operator from (2.21), with the interesting properties discussed in Remark 2.2.1.

As usual, the discrete (broken) scalar products for $(v, w) \in L^2(\mathcal{T}_h)^2$ and $(\mu, \nu) \in L^2(\partial\mathcal{T}_h)^2$ may be written as follows:

$$\begin{aligned}
(v, w)_{\mathcal{T}_h} &:= \sum_{\omega_c \in \mathcal{T}_h} (v, w)_{\omega_c} = \sum_{\omega_c \in \mathcal{T}_h} \int_{\omega_c} v(\boldsymbol{x}, t)\, w(\boldsymbol{x}, t)\, \mathrm{d}\boldsymbol{x}, \\
\left\langle \mu, \nu \right\rangle_{\partial\mathcal{T}_h} &:= \sum_{\omega_c \in \mathcal{T}_h} \left\langle \mu, \nu \right\rangle_{\partial\omega_c} = \sum_{\omega_c \in \mathcal{T}_h} \int_{\partial\omega_c} \mu(\boldsymbol{x}, t)\, \nu(\boldsymbol{x}, t)\, \mathrm{d}s, \\
\left\langle \mu, \nu \right\rangle_{\mathcal{T}_h} &:= \sum_{\omega_c \in \mathcal{T}_h} \left\langle \mu, \nu \right\rangle_{\mathcal{T}_c} = \sum_{\omega_c \in \mathcal{T}_h} \sum_{\Gamma \in \mathcal{T}_c} \int_{\Gamma} \mu(\boldsymbol{x}, t)\, \nu(\boldsymbol{x}, t)\, \mathrm{d}s.
\end{aligned}$$

The extension to vector-valued functions $\boldsymbol{v}, \boldsymbol{w}$ and $\boldsymbol{\mu}, \boldsymbol{\nu}$ follows the same principle, replacing pointwise products by Euclidean dot products.

Seeking for approximate solutions of (3.1), the next sections are devoted to the computation of $\boldsymbol{v}_h := (\eta_h, \boldsymbol{q}_h)$, such that for any time value, $\boldsymbol{v}_h(\cdot, t) \in \mathbf{P}^k(\mathcal{T}_h)$. For any given $\boldsymbol{v}_h \in \mathbf{P}^k(\mathcal{T}_h)$, and an interface $\Gamma_{cv(k)}$, let us define $\boldsymbol{v}_k^-$ and $\boldsymbol{v}_k^+$ respectively the restrictions of $\boldsymbol{v}_{h_{|\omega_c}}$ and $\boldsymbol{v}_{h_{|\omega_{v(k)}}}$ to $\Gamma_{cv(k)}$ (ie. the interior and exterior traces with respect to ω_c).

In the context of NSW equations, the approximated vector solution $\boldsymbol{v}_h^c$ is expressed as the following vector of polynomials:

$$\boldsymbol{v}_h^c(\boldsymbol{x}, t) = \sum_{m=1}^{N_k} \underline{\boldsymbol{v}}_m^c(t)\psi_m^c(\boldsymbol{x}), \tag{3.3}$$

with $\boldsymbol{v}_h^c = (\eta_h^c, (q_x)_h^c, (q_y)_h^c)$. Similarly, we consider a polynomial expansion of the bathymetry parameterization b as $b_h^c(\boldsymbol{x}) = \sum_{m=1}^{N_k} \underline{b}_m^c \phi_m^c(\boldsymbol{x})$.

Sub-mesh and related discrete operators

Let us divide the initial mesh into subcells: for any given mesh element $\omega_c \in \mathcal{T}_h$, we introduce a sub-partition $\mathcal{T}_{\omega_c}^k$ into N_s open disjoint subcells:

$$\overline{\omega}_c = \bigcup_{m \in [\![1, N_s]\!]} \overline{S}_m^c,$$

where the subcell S_m^c has the volume $|S_m^c|$, gathered in the diagonal volume matrix $\mathbb{D}_c \in \mathscr{M}_{N_s}(\mathbb{R})$ s.t. $\mathbb{D}_c := \mathrm{diag}_{m \in [\![1, N_s]\!]}(|S_m^c|)$.

Remark 3.1.4 (*Subcell partitioning*). *For the numerical applications considered in this work, we employed two types of subdivisions: a hybrid quadrilateral/triangle configuration and a fully triangular one. This means that the subcells can be either quadrangular or triangular. However, any subdivision that satisfies the required geometric and numerical properties can be used, allowing for very general subcell shapes.*

We also define Γ^c_{mp} being the face between subcell S^c_m and its neighbor S_p, its length is denoted by ℓ_{mp} while N^m_f is the number of subcell S^c_m faces. The set containing such faces is denoted by $\mathcal{F}^c_m$ s.t. $\#\mathcal{F}^c_m = N^m_f$. Also, we define N^c_f such as the number of subcell faces inside ω_c (meaning not belonging to $\partial\omega_c$).
Let us also define the set $\mathcal{V}^c_m$, that contains the face neighboring subcells of S^c_m (s.t. $\#\mathcal{V}^c_m = N^m_f$), and $\breve{\mathcal{V}}^c_m$ stands for the set containing only the face neighboring subcells of S^c_m inside cell ω_c. Let us mention that a subneighbor S_p can be either inside cell ω_c i.e. $S_p \in \mathcal{T}^k_{\omega_c}$ (then $S_p = S^c_p$), or in one of its neighboring cells ω_v i.e. $S_p \in \mathcal{T}^k_{\omega_v}$ (then $S_p = S^v_p$), see Figure .

We will denote by $\mathbf{n}_{mp}$ the normal vector associated to the face Γ^c_{mp} of the element S^c_m shared with its neighbor $S^v_p \in \mathcal{V}^c_m$. These normals are oriented through the sign function ε^c_{mp} defined as follows:

$$\varepsilon^c_{mp} = \begin{cases} 1 & \text{if } S_m \text{ and } S_p \text{ are neighbors and } m < p, \\ -1 & \text{if } S_m \text{ and } S_p \text{ are neighbors and } m > p, \\ 0 & \text{if } S_p \notin \mathcal{V}^c_m. \end{cases}$$

We also introduce $\mathcal{P}^c_m$ the set of vertices $\boldsymbol{x}_p$ of subcell S^c_m, as well as $\mathcal{V}_p$ the set of subcells that share $\boldsymbol{x}_p$ as a vertex. Let us then define the set $\mathcal{N}(S^c_m)$ such that

$$\mathcal{N}(S^c_m) := \bigcup_{\boldsymbol{x}_p \in \mathcal{P}^c_m} \mathcal{V}_p$$

being constituted by subcell S^c_m as well as its face and node neighboring subcells S^v_p, either they belong to the same cell or not. Finally, the set $\mathcal{V}_{mp}$ contains all subcells sharing at least one node vertex the face Γ_{mp}.

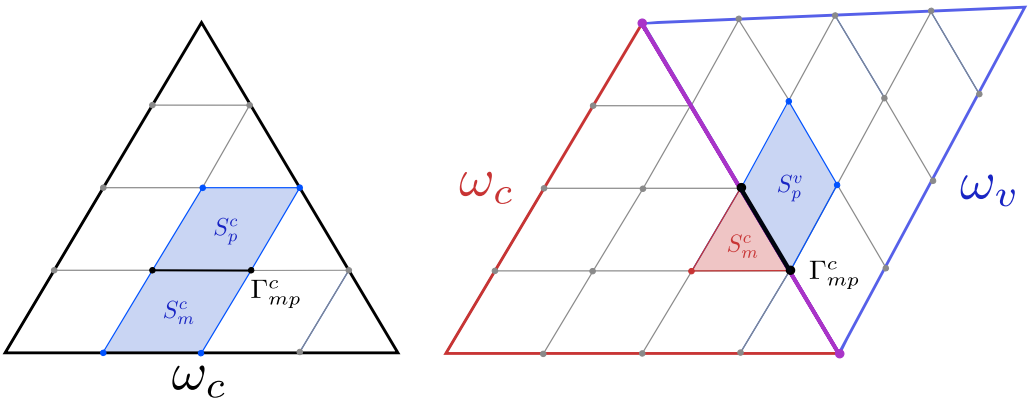


Figure 3.2: Two cases: subneighbor S_p inside cell ω_c (left), and subneighbor S_p inside neighbor cell ω_v (right).

For any given $\omega_c \in \mathcal{T}_h$, let us introduce the set of piecewise-constant functions on the sub-mesh:

$$\mathbb{P}^0(\mathcal{T}^k_{\omega_c}) := \{v \in L^\infty(\omega_c) \mid \forall S^c_m \in \mathcal{T}^k_{\omega_c},\ v_{|S^c_m} \in \mathbb{P}^0(S^c_m)\},$$

together with the subcell-indicator functions

$$\{\mathbb{1}^c_m \mid m \in [\![1, N_s]\!]\},$$

such that $\mathbb{1}^c_m(\boldsymbol{x}) := 1$ if $\boldsymbol{x} \in S^c_m$, and 0 otherwise, allowing to define the mean-value projector $\overline{\pi}^k_c$ onto $\mathbb{P}^0(\mathcal{T}^k_{\omega_c})$ such that

$$\mathbb{P}^k(\omega_c) \ni v^c_h \mapsto \overline{\pi}^k_c v^c_h := \sum_{m=1}^{N_s} \overline{v}^c_m \mathbb{1}^c_m =: \overline{v}^c_h \in \mathbb{P}^0(\mathcal{T}^k_{\omega_c}),$$

where

$$\{\overline{v}^c_m \mid m \in [\![1, N_s]\!]\} =: \overline{v}^c,$$

is the collection of piecewise-constant components defined as the mean-values of v_h^c on the subcells belonging to the subdivision $\mathcal{T}_{\omega_c}^k$, and called *sub-mean values*. Note that, in what follows and when no confusion is possible, $\overline{v}^c$ may be seen as a vector of $\mathbb{R}^{N_s}$, a set of N_s mean-values or identified with the associated piecewise-constant function $\overline{v}_h^c \in \mathbb{P}^0(\mathcal{T}_{\omega_c}^k)$. In order to reach a global description of subcells and related operators, we gather the local subcell partitions and define the associated global approximation space as follows:

$$\widetilde{\mathcal{T}_h} := \bigcup_{\omega_c \in \mathcal{T}_h} \mathcal{T}_{\omega_c}^k, \qquad \mathbb{P}^0(\widetilde{\mathcal{T}_h}) := \bigtimes_{\omega_c \in \mathcal{T}_h} \mathbb{P}^0(\mathcal{T}_{\omega_c}^k).$$

The global projector onto $\mathbb{P}^0(\widetilde{\mathcal{T}_h})$, obtained by gathering the local ones $\overline{\pi}_c^k$, is denoted by $\overline{\pi}_h^k$. Note that $\overline{\pi}_h^k$ can be extended to larger functional spaces (typically $\mathsf{H}^s(\mathbb{R})$, $s > 1 + d/2$ for the initial data, and $\mathscr{C}_{\mathrm{b}}^\infty(\mathbb{R})$ for the bathymetry) and denoted $\widetilde{\pi}_h^k$. In what follows, piecewise-constant functions belonging to $\mathbb{P}^0(\widetilde{\mathcal{T}_h})$ are denoted with an overline and a subscript h, and for any $\overline{v}_h \in \mathbb{P}^0(\widetilde{\mathcal{T}_h})$, and for all $S_m^c \in \widetilde{\mathcal{T}_h}$, we obviously have the identity $\overline{v}_m^c = \overline{v}_{h|S_m^c}$. We may also use the notations $\mathbf{P}^0(\mathcal{T}_{\omega_c}^k) := (\mathbb{P}^0(\mathcal{T}_{\omega_c}^k))^3$ and $\mathbf{P}^0(\widetilde{\mathcal{T}_h}) := (\mathbb{P}^0(\widetilde{\mathcal{T}_h}))^3$ for vectorial quantities. For a more compact representation of the principal geometric and topological subcell notations introduced above, we can refer the reader to Table A.1 in Appendix A.

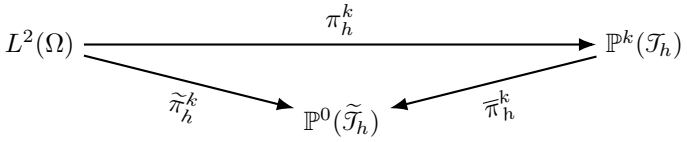


Figure 3.3: Relation between the projectors π_h^k, $\overline{\pi}_h^k$ and its extension $\widetilde{\pi}_h^k$.

Remark 3.1.5 (*Projection and reconstruction operators*). *For the coupling between the classical DG schemes and the subcell FV-like method, the projection operator $\overline{\pi}_c^k : \mathbb{P}^k(\omega_c) \to \mathbb{P}^0(\mathcal{T}_{\omega_c}^k)$ must be injective. This naturally requires the number of polynomial degrees of freedom N_k to be less than or equal to the number of subcells per cell N_s ($N_k \leq N_s$). This injectivity ensures that the polynomial moments can be uniquely recovered from the subcell mean values. The associated local transformation matrix $\mathbb{P}_c \in \mathcal{M}_{N_s \times N_k}(\mathbb{R})$ is defined as*

$$(\mathbb{P}_c)_{(m,p) \in [\![1,N_s]\!] \times [\![1,N_k]\!]} = \frac{1}{|S_m^c|} \int_{S_m^c} \psi_p^c(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}.$$

For our subdivision to be admissible, $\mathbb{P}_c$ must have full column rank. This guarantees the existence of a reconstruction operator $\mathbb{R}_c$ acting as a left inverse, such that $\mathbb{R}_c \mathbb{P}_c = \mathbb{I}_{N_k}$. These operators are then used to switch between the polynomial representation of the solution and its subcell mean values:

$$\overline{\boldsymbol{v}}^c = \mathbb{P}_c \, \underline{\boldsymbol{v}}^c \quad \textit{and} \quad \underline{\boldsymbol{v}}^c = \mathbb{R}_c \, \overline{\boldsymbol{v}}^c.$$

When $N_s > N_k$, the reconstruction problem is overdetermined. To properly recover the polynomial moments $\underline{\boldsymbol{v}}^c$ from the subcell averages $\overline{\boldsymbol{v}}^c$, we rely on a least-squares procedure, defining the reconstruction operator as the pseudo-inverse:

$$\mathbb{R}_c := \left(\mathbb{P}_c^\top \mathbb{P}_c \right)^{-1} \mathbb{P}_c^\top.$$

In the light of [2, 267], a subdivision is then admissible if and only if $\mathbb{P}_c^\top \mathbb{P}_c$ is indeed invertible. Of course the reconstruction operator $\mathbb{R}_c$ simply reduces to $\mathbb{P}_c^{-1}$ when $N_s = N_k$. Note that these local problems can be easily vectorized on modern computational ressources.

Remark 3.1.6 (*On the choice of the subdivision and the number of subcells*). *The procedure outlined below is not limited to the case $N_s = N_k$, and it is entirely possible to exhibit the corresponding numerical fluxes even when the number of subcells does not match the dimension of the polynomial space $\mathbb{P}^k$. In scenarios where there are more subcells than DoF, one can either apply a least squares procedure to the set of submean values obtained through the FV-like method to return to a $(k+1)$-th order DG scheme, or directly transition to a higher-order DG scheme if the number of subcells corresponds to the dimension of a higher degree polynomial space.*
Conversely, if fewer subcells than DoF are used, the same techniques can be employed to transition from the subcell mean values set to a lower DG scheme. For further details concerning the cell subdivision, we refer to [265, 267].

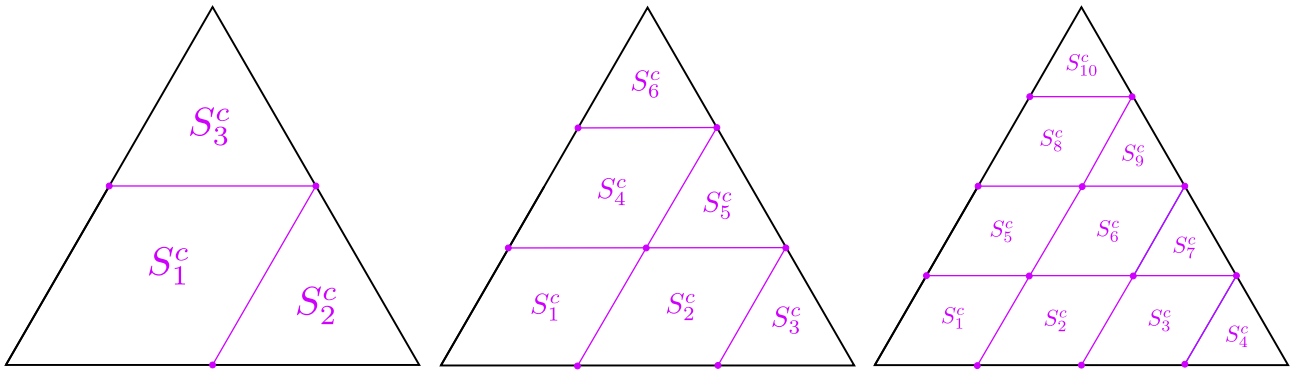


Figure 3.4: Partition of a mesh element ω_c into $N_s = N_k$ subcells for $\mathbb{P}^1$ (left), $\mathbb{P}^2$ (center) and $\mathbb{P}^3$ (right) cases.

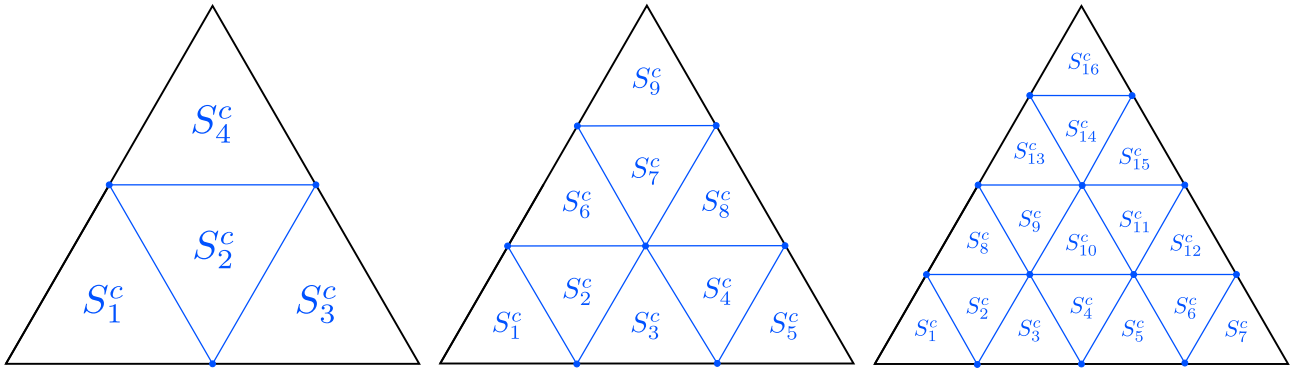


Figure 3.5: Partition of a mesh element ω_c into $N_s > N_k$ subcells for $\mathbb{P}^1$ (left), $\mathbb{P}^2$ (center) and $\mathbb{P}^3$ (right) cases.

Time discretization and time-marching algorithm

Given a final time $T > 0$, we consider a general partition of the time interval $[0, T]$ defined by $0 =: t^0 < t^1 < \cdots < t^n < \cdots < t^N := T$. The time step is denoted by $\Delta t^n \in \mathbb{R}_+^*$ and may vary with time, such that for any $n \geq 0$, we have $t^{n+1} = t^n + \Delta t^n$. Time integration is still performed here using explicit *Strong Stability Preserving Runge-Kutta* (SSP-RK) schemes [126] defined in Time discretization and time-marching algorithm.

As the monolithic scheme couples DG methods on primal cells $\omega_c \in \mathcal{T}_h$ with FV methods on subcells $S_m^c \in \mathcal{T}_{\omega_c}^k$, the time step Δt^n is chosen adaptively according to the following CFL condition:

$$\Delta t^n = \frac{\min\limits_{\omega_c \in \mathcal{T}_h,\, S_m^c \in \mathcal{T}_{\omega_c}^k} |S_m^c|}{\sigma \max\limits_{\omega_c \in \mathcal{T}_h} \left(\sum\limits_{k=1}^{\#\mathcal{F}_c} \ell_{cv(k)} \right)}, \qquad \sigma = \max_{\omega_c \in \mathcal{T}_h} \left[\max_{S_m^c \in \mathcal{T}_{\omega_c}^k,\, S_p^v \in \mathcal{V}_m^c} \left(|\overline{\boldsymbol{u}}_m^c \cdot \mathbf{n}_{mp}| + \sqrt{\mathrm{g} \overline{H}_m^c} \right) \right], \tag{3.4}$$

3.2 Discrete formulations and flux reconstruction

In this section, we focus on the discrete formulations associated with (3.1). In the following, and when no confusion is possible, $\omega_{v(k)}$, with $k \in [\![1, \#\mathcal{F}_c]\!]$, refers to a generic face-neighbor of ω_c and we introduce the following shortcut for the normal component of the physical flux across the (oriented) face $\Gamma_{cv(k)}$:

$$\mathbb{F}_{cv(k)}^* := \mathbb{F}(\boldsymbol{v}_h, b_h)_{|\Gamma_{cv(k)}} \cdot \mathbf{n}_{cv(k)}.$$

To approximate $\mathbb{F}_{cv(k)}^*$ on the k-th interface $\Gamma_{cv(k)}$, we may use any consistent numerical flux $\mathbb{F}_{cv(k)}^{*,\mathrm{DG}}$, like the global Lax-Friedrichs numerical flux for instance. We will introduce later a hierarchy of numerical fluxes designed to approximate the normal fluxes, operating at both the cell and subcell levels and taking into account the topography source term handling, in the same fashion as Chapter 1.

3.2.1 Classical DG formulation

Relying on the discrete framework introduced in the previous section, a semi-discrete global DG formulation of the NSW equations with topography source term may reads as: find $\boldsymbol{v}_h := (\eta_h, \boldsymbol{q}_h) \in \mathbf{P}^k(\mathcal{T}_h)$ such that, for all $\varphi_h \in \mathbb{P}^k(\mathcal{T}_h)$,

$$\big(\partial_t \boldsymbol{v}_h, \varphi_h\big)_{\mathcal{T}_h} - \big(\mathbb{F}(\boldsymbol{v}_h, b_h), \nabla_{\boldsymbol{x}}\varphi_h\big)_{\mathcal{T}_h} + \big\langle \mathbb{F}_\Gamma^{*,\mathrm{DG}}, \varphi_h\big\rangle_{\partial\mathcal{T}_h} = \big(\boldsymbol{B}[b_h](\boldsymbol{v}_h), \varphi_h\big)_{\mathcal{T}_h}, \tag{3.5}$$

which locally traduces as: for all $\omega_c \in \mathcal{T}_h$, find $\boldsymbol{v}_h^c := (\eta_h^c, \boldsymbol{q}_h^c) \in \mathbf{P}^k(\omega_c)$ such that, for all $\varphi_h^c \in \mathbb{P}^k(\omega_c)$,

$$\int_{\omega_c} \partial_t \boldsymbol{v}_h^c \varphi_h^c \,\mathrm{d}\boldsymbol{x} - \int_{\omega_c} \mathbb{F}(\boldsymbol{v}_h^c, b_h^c)\cdot\nabla_{\boldsymbol{x}}\varphi_h^c \,\mathrm{d}\boldsymbol{x} + \sum_{\Gamma\in\mathcal{F}_c}\int_\Gamma \mathbb{F}_\Gamma^{*,\mathrm{DG}}\varphi_h^c \,\mathrm{d}s = \int_{\omega_c} \boldsymbol{B}[b_h^c](\boldsymbol{v}_h^c)\varphi_h^c \,\mathrm{d}\boldsymbol{x}. \tag{3.6}$$

Remark 3.2.1 (*Matrix-vector form of the DG formulation*). *Replacing $\boldsymbol{v}_h^c$ by its polynomial expansion, φ_h^c by basis functions ψ_p^c, we obtain the following semi-discrete formulation for the polynomial moments $\underline{\boldsymbol{v}}^c$:*

$$\sum_{m=1}^{N_k} \frac{\mathrm{d}\underline{\boldsymbol{v}}_m^c}{\mathrm{d}t}\int_{\omega_c}\psi_m^c\psi_p^c \,\mathrm{d}\boldsymbol{x} - \int_{\omega_c}\mathbb{F}(\boldsymbol{v}_h^c, b_h^c)\cdot\nabla_{\boldsymbol{x}}\psi_p^c \,\mathrm{d}\boldsymbol{x} + \sum_{\Gamma\in\mathcal{F}_c}\int_\Gamma \mathbb{F}_\Gamma^{*,\mathrm{DG}}\psi_p^c \,\mathrm{d}s = \int_{\omega_c}\boldsymbol{B}[b_h^c](\boldsymbol{v}_h^c)\psi_p^c \,\mathrm{d}\boldsymbol{x}. \tag{3.7}$$

Note that this may now be also written in a compact matrix-vector form as follows

$$\mathbb{M}_c \frac{\mathrm{d}\underline{\boldsymbol{v}}^c}{\mathrm{d}t} = \boldsymbol{\Xi}_c + \boldsymbol{S}_c, \tag{3.8}$$

- $\mathbb{R}^{N_k} \ni \underline{\boldsymbol{v}}^c = (\underline{\boldsymbol{v}}_m^c)_{m\in[\![1,N_k]\!]}$ *is the solution vector filled with the polynomial moments,*
- $\mathbb{R}^{N_k} \ni \boldsymbol{\Xi}_c$ *is the vector of DG residual which writes as*

$$(\boldsymbol{\Xi}_p^c)_{p\in[\![1,N_k]\!]} := \sum_{\Gamma\in\mathcal{F}_c}\int_\Gamma \mathbb{F}_\Gamma^{*,\mathrm{DG}}\psi_p^c \,\mathrm{d}s - \int_{\omega_c}\mathbb{F}(\boldsymbol{v}_h^c, b_h^c)\cdot\nabla_{\boldsymbol{x}}\psi_p^c \,\mathrm{d}\boldsymbol{x},$$

- $\mathscr{M}_{N_k}(\mathbb{R}) \ni \mathbb{M}_c$ *gathers the integrals of basis functions, generally known as the mass matrix*

$$(\mathbb{M}_{mp}^c)_{(m,p)\in[\![1,N_k]\!]^2} := \int_{\omega_c}\psi_m^c\psi_p^c \,\mathrm{d}\boldsymbol{x},$$

- $\mathbb{R}^{N_k} \ni \boldsymbol{S}_c$ *is the source term vector defined as*

$$(\boldsymbol{S}_p^c)_{p\in[\![1,N_k]\!]} := \int_{\omega_c}\boldsymbol{B}[b_h^c](\boldsymbol{v}_h^c)\psi_p^c \,\mathrm{d}\boldsymbol{x}.$$

3.2.2 Equivalency between DG and a subcell FV-like scheme

There is a well-known equivalence between the DG formulation and a subcell FV-like scheme, described in [265, 267], which can be established by using the submean values of the local DG polynomials as the primary unknowns.

Reconstruction through fluxes

In this section, our goal is to express the reconstructed fluxes solely in terms of the interior flux $\mathbb{F}(\boldsymbol{v}_h^c, b_h^c)$, along with a correction term that accounts for the jumps of the normal component approximations on cells interfaces.

Lemma 3.2.2 (*Equivalence between semi-discrete DG formulation and subcell formulation*). *The global semi-discrete DG formulation (4.19) is locally equivalent to the following formulation relying on submean values:*

$$\frac{\mathrm{d}\overline{\boldsymbol{v}}_m^c}{\mathrm{d}t} = -\frac{1}{|S_m^c|}\left(\sum_{S_p^v\in\mathcal{V}_m^c}\int_{\Gamma_{mp}^c}\mathbb{F}_h^c\cdot\mathbf{n}_{mp}\,\mathrm{d}s - \sum_{k=1}^{\#\mathcal{F}_c}\int_{\Gamma_{ij(k)}^c}\left(\mathbb{F}_h^c\cdot\mathbf{n}_{cv(k)} - \mathbb{F}_{cv(k)}^{*,\mathrm{DG}}\right)\widetilde{\phi}_m^c \,\mathrm{d}s\right) + \int_{S_m^c}\boldsymbol{B}_h^c \,\mathrm{d}\boldsymbol{x}$$

for all $m \in [\![1, N_s]\!]$, with the sub-resolution spanning set of functions $\widetilde{\boldsymbol{\Phi}}_{\omega_c} := \{\widetilde{\phi}_m^c\}_{m\in[\![1,N_s]\!]}$ with $\widetilde{\phi}_m^c(\boldsymbol{x}) = \pi_c^k \mathbb{1}_m^c(\boldsymbol{x})$, and where

$$\mathbb{F}_h := \pi_h^k\mathbb{F}(\boldsymbol{v}_h, b_h) \quad \textit{and} \quad \boldsymbol{B}_h := \pi_h^k\boldsymbol{B}[b_h^c](\boldsymbol{v}_h^c),$$

the L^2-projections of the flux function and source term onto $\mathbb{P}^k(\mathcal{T}_h)$

Proof. Replacing the flux and source terms by their L^2-projections into (4.20), we get:

$$\int_{\omega_c} \partial_t \boldsymbol{v}_h^c \varphi_h^c \,\mathrm{d}\boldsymbol{x} - \int_{\omega_c} \mathbb{F}_h^c \cdot \nabla_{\boldsymbol{x}} \varphi_h^c \,\mathrm{d}\boldsymbol{x} + \sum_{k=1}^{\#\mathcal{F}_c} \int_{\Gamma_{cv(k)}} \mathbb{F}_{cv(k)}^{*,\mathrm{DG}} \varphi_h^c \,\mathrm{d}s = \int_{\omega_c} \boldsymbol{B}_h^c \varphi_h^c \,\mathrm{d}\boldsymbol{x}, \quad \forall \varphi_h^c \in \mathbb{P}^k(\omega_c).$$

Integrating by parts gives:

$$\int_{\omega_c} \partial_t \boldsymbol{v}_h^c \varphi_h^c \,\mathrm{d}\boldsymbol{x} - \sum_{k=1}^{\#\mathcal{F}_c} \int_{\Gamma_{cv(k)}} \left(\mathbb{F}_h^c \cdot \mathbf{n}_{cv(k)} - \mathbb{F}_{cv(k)}^{*,\mathrm{DG}} \right) \varphi_h^c \,\mathrm{d}s + \int_{\omega_c} \varphi_h^c \nabla_{\boldsymbol{x}} \cdot \mathbb{F}_h^c \,\mathrm{d}\boldsymbol{x} = \int_{\omega_c} \boldsymbol{B}_h^c \varphi_h^c \,\mathrm{d}\boldsymbol{x}, \tag{3.9}$$

which is generally called the *strong DG form*. Since $\partial_t \boldsymbol{v}_h^c$, $(\nabla_{\boldsymbol{x}} \cdot \mathbb{F}_h^c)$, $\boldsymbol{B}_h^c \in \mathbf{P}^k(\omega_c)$, substituting φ_h^c with $\widetilde{\phi}_m^c$ into (4.27) gives, for all $m \in [\![1, N_s]\!]$:

$$\int_{S_m^c} \partial_t \boldsymbol{v}_h^c \,\mathrm{d}\boldsymbol{x} = \sum_{k=1}^{\#\mathcal{F}_c} \int_{\Gamma_{cv(k)}} \left(\mathbb{F}_h^c \cdot \mathbf{n}_{cv(k)} - \mathbb{F}_{cv(k)}^{*,\mathrm{DG}} \right) \widetilde{\phi}_m^c \,\mathrm{d}s - \int_{S_m^c} \nabla_{\boldsymbol{x}} \cdot \mathbb{F}_h^c \,\mathrm{d}\boldsymbol{x} + \int_{S_m^c} \boldsymbol{B}_h^c \,\mathrm{d}\boldsymbol{x}.$$

We can then rewrite this with submean values, to get

$$\frac{\mathrm{d}\overline{\boldsymbol{v}}_m^c}{\mathrm{d}t} = -\frac{1}{|S_m^c|} \left(\sum_{S_p^v \in \mathcal{V}_m^c} \int_{\Gamma_{mp}^c} \mathbb{F}_h^c \cdot \mathbf{n}_{mp} \,\mathrm{d}s - \sum_{k=1}^{\#\mathcal{F}_c} \int_{\Gamma_{cv(k)}} \left(\mathbb{F}_h^c \cdot \mathbf{n}_{cv(k)} - \mathbb{F}_{cv(k)}^{*,\mathrm{DG}} \right) \widetilde{\phi}_m^c \,\mathrm{d}s \right) + \overline{\boldsymbol{B}}_m^c,$$

for all $m \in [\![1, N_s]\!]$, where $\overline{\boldsymbol{B}}_m^c$ and $\overline{\boldsymbol{v}}_m^c$ are respectively the submean values of $\boldsymbol{B}_h$ and $\boldsymbol{v}_h$ on subcell S_m^c, ie. $\overline{\boldsymbol{B}}_m^c := \frac{1}{|S_m^c|} \int_{S_m^c} \boldsymbol{B}_h^c \,\mathrm{d}\boldsymbol{x}$ and $\overline{\boldsymbol{v}}_m^c := \frac{1}{|S_m^c|} \int_{S_m^c} \boldsymbol{v}_h^c \,\mathrm{d}\boldsymbol{x}$. □

Lemma 3.2.3 (*Reconstructed flux representation*). *The reconstructed flux $\widehat{\mathbb{F}}_c \in \mathbb{R}^{N_f^c}$ such that*

$$\frac{\mathrm{d}\overline{\boldsymbol{v}}_m^c}{\mathrm{d}t} = -\frac{1}{|S_m^c|} \left(\sum_{S_p^v \in \mathcal{V}_m^c} \int_{\Gamma_{mp}^c} \widehat{\mathbb{F}}_n \,\mathrm{d}s \right) + \overline{\boldsymbol{B}}_m^c, \qquad \forall m \in [\![1, N_s]\!],$$

with

$$\sum_{S_p^v \in \mathcal{V}_m^c} \int_{\Gamma_{mp}^c} \widehat{\mathbb{F}}_n \,\mathrm{d}s = \sum_{S_p^v \in \mathcal{V}_m^c} \int_{\Gamma_{mp}^c} \mathbb{F}_h^c \cdot \mathbf{n}_{mp} \,\mathrm{d}s - \sum_{k=1}^{\#\mathcal{F}_c} \int_{\Gamma_{cv(k)}} \left(\mathbb{F}_h^c \cdot \mathbf{n}_{cv(k)} - \mathbb{F}_{cv(k)}^{*,\mathrm{DG}} \right) \widetilde{\phi}_m^c \,\mathrm{d}s,$$

writes as

$$\widehat{\mathbb{F}}_c = \mathbb{F}_c - \mathbb{A}_c^\top \mathscr{L}_c^{-1} \partial \mathbb{F}_c, \tag{3.10}$$

where:

- $\mathbb{R}^{N_f^c} \ni \mathbb{F}_c$ *collects the traces of the normal component of the interior flux on the subinterfaces s.t. for all* $S_m^c \in \mathcal{J}_{\omega_c}^k$,

$$(\ell_{mp} \mathbb{F}_{mp}^c)_{p \in [\![1, N_f^c]\!]} := \int_{\Gamma_{mp}^c} \mathbb{F}_h^c \cdot \mathbf{n}_{mp} \,\mathrm{d}s,$$

- $\mathscr{M}_{N_k \times N_f^c}(\mathbb{R}) \ni \mathbb{A}_c$ *stands for the adjacency matrix of cell* ω_c, *where*

$$\left(\mathbb{A}_{mp}^c\right)_{(m,p) \in [\![1, N_k]\!] \times [\![1, N_f^c]\!]} = \varepsilon_{mp}^c,$$

- $\mathbb{R}^{N_s} \ni \partial \mathbb{F}_c$ *contains the following boundary contribution*

$$(\partial \mathbb{F}_m^c)_{m \in [\![1, N_s]\!]} = \int_{\partial \omega_c} \left(\mathbb{F}_h^c \cdot \mathbf{n}_{\partial \omega_c} - \mathbb{F}_{\partial \omega_c}^{*,\mathrm{DG}} \right) \widehat{\phi}_m^c(s) \,\mathrm{d}s,$$

  where $\widehat{\phi}_m^c$ *is defined as follows:*

$$\widehat{\phi}_m^c(s) := \begin{cases} \widetilde{\phi}_m^c(s) & \text{if } s \in \partial \omega_c \setminus \partial S_m^c, \\ \widetilde{\phi}_m^c(s) - 1 & \text{if } s \in \partial \omega_c \cap \partial S_m^c. \end{cases}$$

– $\mathscr{M}_{N_f^c \times N_s}(\mathbb{R}) \ni \mathscr{L}_c^{-1}$ *is the inverse of the Laplacian matrix of the interior subgrid graph* $\mathbb{L}_c := \mathbb{A}_c^t \mathbb{A}_c$ *on the orthogonal of its kernel. For any* $\lambda \neq 0$*, this generalized inverse writes:*

$$\mathscr{L}_c^{-1} = (\mathbb{L}_c + \lambda \Pi)^{-1} - \frac{1}{\lambda}\Pi \tag{3.11}$$

with $\Pi = \frac{1}{N_s}(\mathbf{1} \otimes \mathbf{1}) \in \mathscr{M}_{N_k}(\mathbb{R})$.

Proof. In order to develop previous formulation, we need to impose the equality between reconstructed flux $\widehat{\mathbb{F}}_n$ and DG numerical flux $\mathbb{F}_n^{*,\mathrm{DG}}$ on the boundary $\partial\omega_c$, i.e. $\widehat{\mathbb{F}}_{n|\partial\omega} = \mathbb{F}_n^{*,\mathrm{DG}}$. Noticing that numerical flux term

$$\sum_{S_p^v \in \mathscr{V}_m^c} \int_{\Gamma_{mp}^c} \widehat{\mathbb{F}}_n \, \mathrm{d}s = \sum_{S_p^v \in \mathscr{V}_m^c} \int_{\Gamma_{mp}^c} \mathbb{F}_h^c \cdot \mathbf{n}_{mp} \, \mathrm{d}s - \sum_{k=1}^{\#\mathscr{F}_c} \int_{\Gamma_{cv(k)}^c} \left(\mathbb{F}_h^c \cdot \mathbf{n}_{cv(k)} - \mathbb{F}_{cv(k)}^* \right) \widetilde{\phi}_m^c \, \mathrm{d}s,$$

can be equivalently written as

$$\int_{\partial S_m^c} \widehat{\mathbb{F}}_n \, \mathrm{d}s = \int_{\partial S_m^c} \mathbb{F}_h^c \cdot \mathbf{n} \, \mathrm{d}s - \int_{\partial\omega_c} \left(\mathbb{F}_h^c \cdot \mathbf{n}_{\partial\omega_c} - \mathbb{F}_{\partial\omega_c}^{*,\mathrm{DG}} \right) \widetilde{\phi}_m^c \, \mathrm{d}s, \tag{3.12}$$

we directly get

$$\int_{\partial S_m^c \setminus \partial\omega_c} \widehat{\mathbb{F}}_n \, \mathrm{d}s = \int_{\partial S_m^c \setminus \partial\omega_c} \mathbb{F}_h^c \cdot \mathbf{n} \, \mathrm{d}s - \int_{\partial\omega_c} \left(\mathbb{F}_h^c \cdot \mathbf{n}_{\partial\omega_c} - \mathbb{F}_{\partial\omega_c}^{*,\mathrm{DG}} \right) \widehat{\phi}_m^c \, \mathrm{d}s, \tag{3.13}$$

where $\widehat{\phi}_m^c$ is defined above. We now use a face integrated value of the high-order DG reconstructed flux. Indeed, for a face Γ_{mp}^c, let $\widehat{\mathbb{F}}_{mp}$ be defined as:

$$\int_{\Gamma_{mp}^c} \widehat{\mathbb{F}}_n \, \mathrm{d}s = \varepsilon_{mp}^c \widehat{\mathbb{F}}_{mp},$$

and, similarly, let $\mathbb{F}_{mp}$ be the face integrated value of the polynomial interior flux

$$\int_{\Gamma_{mp}^c} \mathbb{F}_h^c \cdot \mathbf{n} \, \mathrm{d}s = \varepsilon_{mp}^c \mathbb{F}_{mp}.$$

Now, considering the vectors $\mathbb{F}_c \in \mathbb{R}^{N_f^c}$ and $\widehat{\mathbb{F}}_c \in \mathbb{R}^{N_f^c}$ defined earlier gets us to the following:

$$\mathbb{A}_c \widehat{\mathbb{F}}_c = \mathbb{A}_c \mathbb{F}_c - \partial \mathbb{F}_c.$$

Remarking that $\mathbb{A}_c$ is not necessarily a square matrix and $\ker \mathbb{A}_c \neq \mathbf{0}$, we make use of the graph Laplacian technique employed in [2, 267], allowing us to solve such system and express explicitly the reconstructed flux $\widehat{\mathbb{F}}_c$ through the interior flux and a boundary correction term. We get the following inverse of $\mathbb{L}_c$ on the orthogonal of its kernel:

$$\mathscr{L}_c^{-1} = (\mathbb{L}_c + \lambda \Pi)^{-1} - \frac{1}{\lambda}\Pi,$$

for any $\lambda \neq 0$, with $\Pi = \frac{1}{N_s}(1 \otimes 1) \in \mathscr{M}_{N_k}(\mathbb{R})$. The reconstructed flux then writes as:

$$\widehat{\mathbb{F}}_c = \mathbb{F}_c - \mathbb{A}_c^\top \mathscr{L}_c^{-1} \partial \mathbb{F}_c. \tag{3.14}$$

We are then able to recast the local semi-discrete DG formulation (4.20) into a subcell FV-like scheme, as follows:

Theorem 3.2.4 (*Equivalence between DG and FV-subcell formulations*). *For the Saint-Venant equations, considering any* $\varphi_h^c \in \mathbb{P}^k(\omega_c)$*, the classical unlimited Discontinuous Galerkin scheme in cell* ω_c

$$\int_{\omega_c} \partial_t \boldsymbol{v}_h^c \varphi_h^c \, \mathrm{d}\boldsymbol{x} - \int_{\omega_c} \mathbb{F}(\boldsymbol{v}_h^c, b_h^c) \cdot \nabla_{\boldsymbol{x}} \varphi_h^c \, \mathrm{d}\boldsymbol{x} + \sum_{k=1}^{\#\mathscr{F}_c} \int_{\Gamma_{cv(k)}} \mathbb{F}_{cv(k)}^{*,\mathrm{DG}} \varphi_h^c \, \mathrm{d}s = \int_{\omega_c} \boldsymbol{B}(\boldsymbol{v}_h^c, \nabla_{\boldsymbol{x}} b_h^c) \varphi_h^c \, \mathrm{d}\boldsymbol{x}, \tag{3.15}$$

can be recast into N_s *FV-like subcell schemes as*

$$\frac{\mathrm{d}\overline{\boldsymbol{v}}_m^c}{\mathrm{d}t} = -\frac{1}{|S_m^c|} \sum_{S_p^v \in \mathscr{V}_m^c} \ell_{mp} \widehat{\mathbb{F}}_{mp} + \overline{\mathbf{B}}_m^c, \qquad \forall m \in [\![1, N_s]\!], \tag{3.16}$$

where the reconstructed flux is defined in (4.25).

Proof. The result is immediate from Lemmas 3.2.2 and 3.2.3. □

Reconstruction through residuals

While in the previous section we obtained the reconstructed fluxes by expressing them through the interior flux $\mathbb{F}(\boldsymbol{v}_h^c, b_h^c)$ and a boundary correction term, we can also retain them by considering classical DG residuals. This version has the advantage of being usually easier to implement, since we compute everything in terms of matrix-vector products.

Theorem 3.2.5 (*Equivalence between DG and FV-subcell formulations*). *Considering NSW equations, the classical unlimited Discontinuous Galerkin scheme expressed in compact form through residuals*

$$\frac{\mathrm{d}\underline{\boldsymbol{v}}^c}{\mathrm{d}t} = \mathbb{M}_c^{-1}(\boldsymbol{\Xi}_c + \boldsymbol{S}_c), \tag{3.17}$$

can be recast into N_s FV-like subcell schemes as

$$\frac{\mathrm{d}\overline{\boldsymbol{v}}^c}{\mathrm{d}t} = -\mathbb{D}_c^{-1}\left(\mathbb{A}_c\widehat{\mathbb{F}}_c + \partial\mathbb{F}_c\right) + \overline{\boldsymbol{S}}_c \tag{3.18}$$

where $\overline{\boldsymbol{S}}_c := \mathbb{P}_c\mathbb{M}_c^{-1}\boldsymbol{S}_c$ is a vector containing the submean values of source term projection, i.e. $(\overline{\boldsymbol{S}}_c)_m = \overline{\boldsymbol{B}}_m^c$. The reconstructed flux $\widehat{\mathbb{F}}_c$, that contains all the interior faces reconstructed fluxes weighted by the face length, i.e. $\ell_{mp}\widehat{\mathbb{F}}_{mp}$ writes as

$$\widehat{\mathbb{F}}_c = -\mathbb{A}_c^\top\mathscr{L}_c^{-1}\left(\mathbb{D}_c\mathbb{P}_c\mathbb{M}_c^{-1}\boldsymbol{\Xi}_c + \partial\mathbb{F}_c\right), \tag{3.19}$$

where $\partial\mathbb{F}_c \in \mathbb{R}^{N_s}$ carries the cell boundary contribution as the integral of DG fluxes over the cell boundary, i.e.

$$(\partial\mathbb{F}_m^c)_{m\in[\![1,N_s]\!]} = \int_{\partial\omega_c\cap\partial S_m^c} \mathbb{F}_n^{*,\mathrm{DG}}\,\mathrm{d}\boldsymbol{x}.$$

Proof. Using the polynomial expansion of $\boldsymbol{v}_h^c$ and the definition of submean value, we get

$$\frac{\mathrm{d}\overline{\boldsymbol{v}}_m^c}{\mathrm{d}t} = \frac{1}{|S_m^c|}\sum_{q=1}^{N_k}\frac{\mathrm{d}\underline{\boldsymbol{v}}_q^c}{\mathrm{d}t}\int_{S_m^c}\psi_q^c\,\mathrm{d}\boldsymbol{x},$$

that can be rewritten in a matrix-vector form as

$$\frac{\mathrm{d}\overline{\boldsymbol{v}}^c}{\mathrm{d}t} = \mathbb{P}_c\frac{\mathrm{d}\underline{\boldsymbol{v}}^c}{\mathrm{d}t}. \tag{3.20}$$

By means of DG scheme definition given in (3.8), it follows that

$$\frac{\mathrm{d}\overline{\boldsymbol{V}}_c}{\mathrm{d}t} = \mathbb{P}_c\mathbb{M}_c^{-1}(\boldsymbol{\Xi}_c + \boldsymbol{S}_c), \tag{3.21}$$

and in order to express it as a FV-like scheme, the same way as previously, we consider the following DG reconstructed flux $\widehat{\mathbb{F}}_n$:

$$\frac{\mathrm{d}\overline{\boldsymbol{v}}_m^c}{\mathrm{d}t} = -\frac{1}{|S_m^c|}\int_{\partial S_m^c}\widehat{\mathbb{F}}_n\,\mathrm{d}\boldsymbol{x} + (\mathbb{P}_c\mathbb{M}_c^{-1}\boldsymbol{S}_c)_m = -\frac{1}{|S_m^c|}\sum_{S_p^v\in\mathscr{V}_m^c}\int_{\Gamma_{mp}^c}\widehat{\mathbb{F}}_n\,\mathrm{d}\boldsymbol{x} + (\mathbb{P}_c\mathbb{M}_c^{-1}\boldsymbol{S}_c)_m, \tag{3.22}$$

and we impose the same equality condition $\widehat{\mathbb{F}}_{n|\partial\omega} = \mathbb{F}_n^{*,\mathrm{DG}}$ on boundary of cell ω_c. Our previous formulation then can be rewritten as

$$\frac{\mathrm{d}\overline{\boldsymbol{v}}_m^c}{\mathrm{d}t} = -\frac{1}{|S_m^c|}\left(\sum_{S_p^v\in\breve{\mathscr{V}}_m^c}\int_{\Gamma_{mp}^c}\widehat{\mathbb{F}}_n\,\mathrm{d}\boldsymbol{x} + \int_{\partial\omega_c\cap\partial S_m^c}\mathbb{F}_n^{*,\mathrm{DG}}\,\mathrm{d}\boldsymbol{x}\right) + (\mathbb{P}_c\mathbb{M}_c^{-1}\boldsymbol{S}_c)_m, \tag{3.23}$$

As in reconstruction through fluxes, each subface get assigned an orientation, and we define reconstructed flux $\widehat{\mathbb{F}}_{mp}$ at interface Γ_{mp}^c as follows:

$$\int_{\Gamma_{mp}^c}\widehat{\mathbb{F}}_n\,\mathrm{d}\boldsymbol{x} = \varepsilon_{mp}^c\widehat{\mathbb{F}}_{mp}. \tag{3.24}$$

The governing equations (3.23) then yield to the following system

$$-\mathbb{A}_c\widehat{\mathbb{F}}_c = \mathbb{D}_c\frac{\mathrm{d}\overline{\boldsymbol{V}}_c}{\mathrm{d}t} + \partial\mathbb{F}_c, \tag{3.25}$$

where all the quantities above have been already defined, and $\partial\mathbb{F}_c$ is the vector containing the following boundary integral:

$$(\partial\mathbb{F}_c)_m = \int_{\partial\omega_c\cap\partial S_m^c} \mathbb{F}_n^* \,\mathrm{d}\boldsymbol{x}.$$

Substituting (3.21) into (3.25), we get

$$-\mathbb{A}_c\widehat{\mathbb{F}}_c = \mathbb{D}_c\mathbb{P}_c\mathbb{M}_c^{-1}\boldsymbol{\Xi}_c + \partial\mathbb{F}_c. \tag{3.26}$$

To solve this system, employing the same Laplacian graph technique as before, we finally obtain the residual definition of the reconstructed flux

$$\widehat{\mathbb{F}}_c = -\mathbb{A}_c^\top\mathscr{L}_c^{-1}\left(\mathbb{D}_c\mathbb{P}_c\mathbb{M}_c^{-1}\boldsymbol{\Xi}_c + \partial\mathbb{F}_c\right),$$

where $\mathscr{L}_c^{-1}$ is defined in (4.26). □

Remark 3.2.6 (*Implementation aspects*). *The only terms depending on time are $\boldsymbol{\Xi}_c$, the residual which is directly available in any DG code, and the boundary contribution $\partial\mathbb{F}_c$. All the other terms can be precomputed, and even can be reduced to an only matrix for all ω_c, by means of a mapping to a referential triangle, according that all the mesh cells have the same structure. Indeed, choosing the same set of basis functions across all elements, as Legendre functions for example, will lead to an unique definition for $\mathbb{A}_c$, $\mathbb{P}_c$ and $\mathscr{L}_c^{-1}$. However, in practice, for implementation purpose we rely on physical basis functions defined as*

$$\Psi_c^{(i,j)}(\boldsymbol{x}) := \left(\frac{x - x_{\mathrm{b}}^c}{h_x^c}\right)^i\left(\frac{y - y_{\mathrm{b}}^c}{h_y^c}\right)^j,$$

allowing us to handle elements of arbitrary shape with good performance, and also ensuring that the basis is invariant under translations of the element ω_c. Let us remark that whenever needed, improving basis quality can be straightforward by simply orthonormalizing it through a Gram–Schmidt algorithm.

3.3 Local subcell monolithic DG/FV scheme

We are now able to define the local monolithic DG/FV scheme, by simply replacing the high-order DG fluxes $\mathbb{F}_{cv(k)}^{*,\mathrm{DG}}$ by some blended fluxes yet to be defined, that will allow us to combine the high-order DG scheme with a low-order FV scheme, while preserving the well-balanced property and ensuring consistency with the original PDEs.

3.3.1 High and low-order topography-dependent numerical fluxes

In the following, we rely on several numerical fluxes defined at both mesh and submesh interfaces. Due to the presence of non-flat bathymetry, in contrast with [265], the design of these fluxes must be carefully modified to ensure consistency, stability, preserve the expected high-order accuracy and, in particular, maintain the *well-balancing* property.

Inspired by the *hydrostatic reconstruction* for FV methods [12], extended to pre-balanced and DG formulations in [107,200], a key feature of our approach is to finely tune this reconstruction to be strategically applied both on high-order polynomial traces or low-order approximations *both at cells and subcells levels*, using properly defined intermediate reconstructed states, in the same fashion as [51] for the 1D case (see also Chapter 1). This involves non-conservative corrections to the numerical fluxes, which are designed to vanish at motionless steady states. As shown in previous works, these corrections do not affect the accuracy nor the consistency of the solution, and the conservation error is of the order of the scheme. Gathered with the previously introduced *reconstructed* fluxes used to relate high-order DG scheme to local lowest-order FV scheme, the construction of the local subcell monolithic DG/FV method is achieved through a combination of various numerical approximate interface fluxes. For clarity and ease of reference, the definitions of these numerical fluxes are gathered and summarized below:

– *Global Lax-Friedrichs numerical flux and bathymetry variations.* The chosen interface flux has to account for the bathymetry variations from cell to cell or subcell to subcell, and is defined as follows:

$$(\mathscr{H}_b^+)^2 \times \mathbb{R}^4 \ni (\boldsymbol{v}^-, \boldsymbol{v}^+, b^-, b^+, \mathbf{n}) \mapsto \mathbb{F}^*(\boldsymbol{v}^-, \boldsymbol{v}^+, b^-, b^+, \mathbf{n}) := \frac{\mathbb{F}(\boldsymbol{v}^-, b^-) + \mathbb{F}(\boldsymbol{v}^+, b^+)}{2} \cdot \mathbf{n} - \frac{\sigma}{2}(\boldsymbol{v}^+ - \boldsymbol{v}^-), \tag{3.27}$$

– *Nonconservative high-order DG fluxes $\mathbb{F}^{*,\mathrm{DG}}_{cv}$ and cell-scale hydrostatic reconstruction.* Some local reconstructed interface values for the interface topography are considered, relying on the left and right traces of high-order approximations:

$$\widetilde{b}_k = \max\left(b_k^-, b_k^+\right), \qquad \check{b}_k = \widetilde{b}_k - \max\left(0, \widetilde{b}_k - \eta_k^-\right),$$

together with reconstructed interface values for the water-height and total elevation as follows:

$$\check{H}_k^{\pm} = \max\left(0, \eta_k^{\pm} - \widetilde{b}_k\right) \quad \text{and} \quad \check{\eta}_k^{\pm} = \check{H}_k^{\pm} + \widetilde{b}_k,$$

and interior/exterior solution values:

$$\boldsymbol{v}_k^{\pm} = \left(\check{\eta}_k^{\pm}, \frac{\check{H}_k^{\pm}}{H_k^{\pm}} \boldsymbol{q}_k^{\pm}\right)^{\top}.$$

Using these reconstructed values, we define the following (non-conservative) high-order interface flux through the interface between ω_c and $\omega_{v(k)}$, $k \in [\![1, \#\mathcal{F}_c]\!]$, as follows:

$$\mathbb{F}^{*,\mathrm{DG}}_{cv(k)} = \mathbb{F}^*\left(\check{\boldsymbol{v}}_k^-, \check{\boldsymbol{v}}_k^+, \check{b}_k, \check{b}_k, \mathbf{n}_{cv(k)}\right) + \check{\mathbb{F}}^{\mathrm{DG}}_{cv(k)}, \tag{3.28}$$

where $\check{\mathbb{F}}^{\mathrm{DG}}_{cv(k)}$ is a correction term needed to ensure flux balancing at motionless steady states, defined as follows, inspired from [285]:

$$\check{\mathbb{F}}^{\mathrm{DG}}_{cv(k)} = \begin{pmatrix} 0 & 0 \\ \mathrm{g}\check{\eta}_k^-(\check{b}_k - b_k^-) & 0 \\ 0 & \mathrm{g}\check{\eta}_k^-(\check{b}_k - b_k^-) \end{pmatrix} \cdot \mathbf{n}_{cv(k)}.$$

– *First-order subcell FV fluxes $\mathbb{F}^{*,\mathrm{FV}}_{mp}$ and subcell-scale hydrostatic reconstruction.* For a neighbor subcell S_p^v sharing a subface Γ^c_{mp} with S_m^c, we will denote by $\overline{\boldsymbol{v}}_m^c$ and $\overline{\boldsymbol{v}}_p^v$ respectively the interior and exterior mean-values with respect to Γ^c_{mp}. We then now proceed the same way as usual by new defining state reconstructions as follows:

$$\begin{aligned} \widetilde{\overline{b}}_{mp} &= \max\left(\overline{b}_m^c, \overline{b}_p^v\right), \qquad \check{\overline{b}}_{mp} = \widetilde{\overline{b}}_{mp} - \max\left(0, \widetilde{\overline{b}}_{mp} - \overline{\eta}_m^c\right), \\ \check{\overline{H}}_m^c &= \max\left(0, \overline{\eta}_m^c - \widetilde{\overline{b}}_{mp}\right), \qquad \check{\overline{H}}_p^v = \max\left(0, \overline{\eta}_p^v - \widetilde{\overline{b}}_{mp}\right), \end{aligned}$$

and the reconstructed states for total elevation are then defined as

$$\check{\overline{\eta}}_m^c = \check{\overline{H}}_m^c + \check{\overline{b}}_{mp}, \qquad \check{\overline{\eta}}_p^v = \check{\overline{H}}_p^v + \check{\overline{b}}_{mp},$$

leading to new exterior and interior values:

$$\check{\overline{\boldsymbol{v}}}_m^c = \left(\check{\overline{\eta}}_m^c, \frac{\check{\overline{H}}_m^c}{\overline{H}_m^c} \overline{\boldsymbol{q}}_m^c\right)^{\top}, \qquad \check{\overline{\boldsymbol{v}}}_p^v = \left(\check{\overline{\eta}}_p^v, \frac{\check{\overline{H}}_p^v}{\overline{H}_p^v} \overline{\boldsymbol{q}}_p^v\right)^{\top}.$$

The corrected FV numerical flux writes then as

$$\mathbb{F}^{*,\mathrm{FV}}_{mp} = \mathbb{F}^*\left(\check{\overline{\boldsymbol{v}}}_m^c, \check{\overline{\boldsymbol{v}}}_p^v, \check{\overline{b}}_{mp}, \check{\overline{b}}_{mp}, \mathbf{n}_{mp}\right) + \check{\mathbb{F}}^{\mathrm{FV}}_{mp}, \tag{3.29}$$

where $\check{\mathbb{F}}^{\mathrm{FV}}_{mp}$ is the correction term assuring flux balancing at motionless steady states

$$\check{\mathbb{F}}^{\mathrm{FV}}_{mp} = \begin{pmatrix} 0 & 0 \\ \mathrm{g}\check{\overline{\eta}}_m^c\left(\check{\overline{b}}_{mp} - b_{mp}\right) & 0 \\ 0 & \mathrm{g}\check{\overline{\eta}}_m^c\left(\check{\overline{b}}_{mp} - b_{mp}\right) \end{pmatrix},$$

with b_{mp} denoting the mean-value of the bathymetry on cell ω_c at the face Γ^c_{mp}, i.e.

$$b_{mp} := \frac{1}{|S^c_m|}\int_{\Gamma^c_{mp}} b^c_h(s)\,\mathrm{d}s.$$

– *High-order subcell DG reconstructed fluxes.* These fluxes are obtained in (3.14), within the equivalency given in Theorem 3.2.4 and aim at locally mimicking the behavior of an "in-cell" FV scheme, while preserving the global high-order accuracy of the DG scheme in the encompassing mesh element:

$$\widehat{\mathbb{F}}_c = \mathbb{F}_c - \mathbb{A}_c^\top \mathscr{L}_c^{-1}\partial\mathbb{F}_c,$$

all the quantities in the above equation being defined in the previous section.

For the computation of the velocity from polynomial or subcell values, we use the regularization described in Remark 1.1.4; its scalar formula is applied componentwise to $\boldsymbol{q}/H$ in two dimensions.

3.3.2 Blending high and low-order fluxes

The earlier reformulations of the DG scheme into a subcell FV-like approach, achieved through the introduction of reconstructed fluxes and the first-order FV well-balanced corrected flux, now provide the foundation for constructing our monolithic FV subcell scheme, adapting the ideas of [265]. These schemes are convex combinations of high-order fluxes with robust low-order fluxes. Indeed, for every face Γ^c_{mp} of each subcell S^c_m, we compute the high-order DG reconstructed flux $\widehat{\mathbb{F}}_{mp}$ and the first-order robust FV numerical flux $\mathbb{F}^{*,\mathrm{FV}}_{mp}$. Then, those two quantities are gathered as follows:

$$\widetilde{\mathbb{F}}_{mp} = \mathbb{F}^{*,\mathrm{FV}}_{mp} + \Theta_{mp}\Delta\mathbb{F}_{mp}, \tag{3.30}$$

where $\Theta_{mp} \in [0,1]$ and $\Delta\mathbb{F}_{mp} := \widehat{\mathbb{F}}_{mp} - \mathbb{F}^{*,\mathrm{FV}}_{mp}$. In the following, we refer to Θ_{mp} as *blending coefficients* and to $\widetilde{\mathbb{F}}_{mp}$ as *blended fluxes*.

Obviously, setting the blending coefficient to zero leads to a first-order FV numerical flux, while the value 1 gives the high-order DG reconstructed flux, leading to a local approximation equivalent to the usual DG formulation. For consistency reason, we define a unique blending coefficient for every subedge, meaning for every $S^v_p \in \mathscr{V}^c_m$, $\Theta_{mp} = \Theta_{pm}$. The local subcell monolithic DG/FV convex property preserving scheme for NSW then writes as

$$\frac{\mathrm{d}\overline{\boldsymbol{v}}^c_m}{\mathrm{d}t} = -\frac{1}{|S^c_m|}\sum_{S^v_p\in\mathscr{V}^c_m} \ell_{mp}\widetilde{\mathbb{F}}_{mp} + \overline{\boldsymbol{B}}^c_m. \tag{3.31}$$

Using a forward Euler time integration, the subcell monolithic DG/FV convex property preserving scheme reads as follows:

$$\overline{\boldsymbol{v}}^{c,n+1}_m = \overline{\boldsymbol{v}}^{c,n}_m - \frac{\Delta t^n}{|S^c_m|}\sum_{S^v_p\in\mathscr{V}^c_m} \ell_{mp}\widetilde{\mathbb{F}}_{mp} + \Delta t^n\overline{\boldsymbol{B}}^{c,n}_m, \tag{3.32}$$

where $\widetilde{\mathbb{F}}_{mp} = \mathbb{F}^{*,\mathrm{FV}}_{mp} + \Theta_{mp}\left(\widehat{\mathbb{F}}_{mp} - \mathbb{F}^{*,\mathrm{FV}}_{mp}\right)$, with the reconstructed high-order flux defined in (3.14) and the FV flux treated in (3.29). In practice, higher-order strong stability preserving Runge-Kutta methods (SSP-RK) [126] are employed for time integration, as they can be seen as convex combination of Forward-Euler schemes.

3.3.3 Source term discretization

As far as the source term discretization is concerned, we emphasize the issue of juggling between modal apprximations and piecewise-constant subcell values, in order to achieve a welcome well-balancing property for steady-states. Considering the topography source term $\boldsymbol{B}[b](\boldsymbol{v}) = (0, -\mathrm{g}\eta\nabla_{\boldsymbol{x}}b)^\top$, the proposed discretization should follow the following flowchart to ensure the consistency and balancing property with the discretization of the fluxes:

1. on each cell, subcell averages of b and η are computed. Using the projection matrix $\mathbb{P}_c$, b_h and η_h are reconstructed,
2. the source term is then projected onto $\mathbb{P}^k$ to obtain $\boldsymbol{B}_h := \pi^k_h\boldsymbol{B}[b^c_h](\boldsymbol{v}^c_h)$,

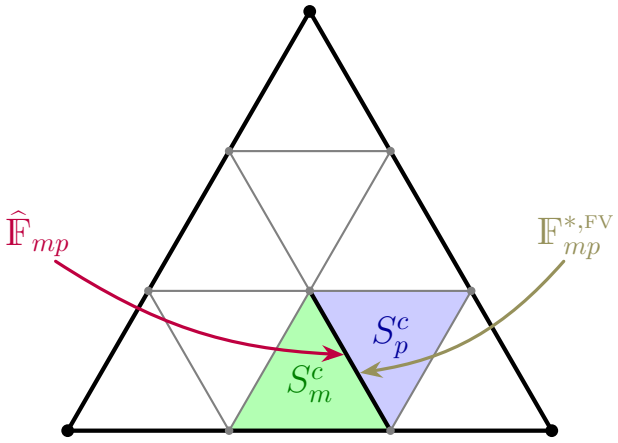


Figure 3.6: Representation of the two fluxes computed on each subinterface between two neighboring subcells S^c_m and S^v_p. The blending coefficient Θ_{mp} then allows to combine those two fluxes in a convex way to get the final flux $\widetilde{\mathbb{F}}_{mp}$ used in the monolithic scheme.

3. the mean integral is computed and we get $\overline{\boldsymbol{B}}^c_m := \frac{1}{|S^c_m|}\int_{S^c_m} \boldsymbol{B}_h \,\mathrm{d}\boldsymbol{x}$.

From a practical viewpoint, and conveniently starting from a "classical" DG local approximation of the source term $\int_{\omega_c} \boldsymbol{B}_h\varphi_h \,\mathrm{d}x$, this may be formulated in a compact way using previously defined operators-matrix products:

$$\overline{\boldsymbol{B}}^c_m = \mathbb{P}_c\mathbb{M}_c^{-1}\left(\int_{\omega_c} \boldsymbol{B}_h\varphi_h \,\mathrm{d}x\right).$$

The reader may also find additional informations on this source term discretization process within subcell monolithic schemes in [51].

Remark 3.3.1 (*Alternative discretizations*). *Other discretizations are of course possible for the topography source term, like for example using the blending coefficients Θ in a mimetic way with the monolithic blended fluxes, representing the source term $\overline{\boldsymbol{B}}^c_m$ as a convex combination between a $\overline{\boldsymbol{B}}^{c,\mathrm{FV}}_m$ and $\overline{\boldsymbol{B}}^{c,\mathrm{DG}}_m$ as so:*

$$\overline{\boldsymbol{B}}^c_m = \overline{\boldsymbol{B}}^{c,\mathrm{FV}}_m + \Theta^c_m\left(\overline{\boldsymbol{B}}^{c,\mathrm{DG}}_m - \overline{\boldsymbol{B}}^{c,\mathrm{FV}}_m\right),$$

with $\Theta^c_m := \frac{1}{\#\mathcal{V}^c_m}\sum_{S^v_p\in\mathcal{V}^c_m}\Theta_{mp}$, and $\overline{\boldsymbol{B}}^{c,\mathrm{DG}}_m = \frac{1}{|S^c_m|}\int_{S^c_m}\boldsymbol{B}_h \,\mathrm{d}\boldsymbol{x}$. For the FV source term, we define it as

$$\overline{\boldsymbol{B}}^{c,\mathrm{FV}}_m = \frac{1}{|S^c_m|}\int_{S^c_m}\boldsymbol{B}(\overline{\boldsymbol{v}}^c_m,\nabla_{\boldsymbol{x}}b^c_h)\,\mathrm{d}\boldsymbol{x} = -\frac{\mathrm{g}\overline{\eta}^c_m}{|S^c_m|}\int_{S^c_m}\begin{pmatrix}0\\ \nabla_{\boldsymbol{x}}b^c_h\end{pmatrix}\mathrm{d}\boldsymbol{x} = -\frac{\mathrm{g}\overline{\eta}^c_m}{|S^c_m|}\int_{\partial S^c_m}\begin{pmatrix}0\\ b^c_h\cdot\mathbf{n}\end{pmatrix}\mathrm{d}s.$$

Let us add that we also discretized it as in [109], i.e.

$$\overline{\boldsymbol{B}}^{c,\mathrm{FV}}_m = \frac{1}{|S^c_m|}\sum_{S^v_p\in\mathcal{V}^c_m}\ell_{mp}\begin{pmatrix}0\\ \frac{\mathrm{g}}{2}\left(\overline{\eta}^c_m - \breve{\overline{\eta}}^c_m\right)\left(\overline{b}^c_m - \breve{\overline{b}}^c_m\right)\mathbf{n}_{mp}\end{pmatrix}.$$

Numerical investigations do not show noticeable variations on the studied test-cases.

3.4 Calibrating the blended fluxes

We are left with the definition of some blending coefficients, which aim at ensuring that the required properties on the resulting approximate solution are strictly enforced, while preserving the highest possible level of accuracy and convergence rate of the scheme. To achieve this, the idea introduced in [265] is to locally reformulate the monolithic scheme (3.32) as a Godunov-like scheme at subcells interfaces, and we follow these lines, but accounting for the particularity of the NSW equations.
In what follows, the novel discrete form and the subsequent analysis are developed relying on semi-discrete formulations, together with an extension to fully-discrete case with the first-order Euler time-marching algorithm (various higher-order Runge-Kutta time-marching methods may be used to achieve better convergence in time, and these multistage methods are convex combinations of first-order forward Euler steps).

3.4.1 Blended and first-order Riemann intermediate states

Let us recall that since $\sum_{S_p^v \in \mathcal{V}_m^c} \ell_{mp} \mathbf{n}_{mp} = \mathbf{0}$, we can add this to our discrete scheme, and also add then substract a certain quantity, in order to rewrite the new solution as a convex combination of quantities defined at the previous time step:

$$\begin{aligned}\overline{\boldsymbol{v}}_m^{c,n+1} &= \overline{\boldsymbol{v}}_m^{c,n} - \frac{\Delta t^n}{|S_m^c|} \sum_{S_p^v \in \mathcal{V}_m^c} \ell_{mp} \widetilde{\mathbb{F}}_{mp} + \Delta t^n \overline{\boldsymbol{B}}_m^{c,n} + \frac{\Delta t^n}{|S_m^c|} \mathbb{F}\left(\overline{\boldsymbol{v}}_m^{c,n}, \overline{b}_m^c\right) \cdot \sum_{S_p^v \in \mathcal{V}_m^c} \ell_{mp} \mathbf{n}_{mp} \pm \frac{\sigma \Delta t^n}{|S_m^c|} \sum_{S_p^v \in \mathcal{V}_m^c} \ell_{mp} \overline{\boldsymbol{v}}_m^{c,n} \\ &= \left(1 - \frac{\sigma \Delta t^n}{|S_m^c|} \sum_{S_p^v \in \mathcal{V}_m^c} \ell_{mp}\right) \overline{\boldsymbol{v}}_m^{c,n} + \frac{\sigma \Delta t^n}{|S_m^c|} \sum_{S_p^v \in \mathcal{V}_m^c} \ell_{mp} \left(\overline{\boldsymbol{v}}_m^{c,n} - \frac{\widetilde{\mathbb{F}}_{mp} - \mathbb{F}\left(\overline{\boldsymbol{v}}_m^{c,n}, \overline{b}_m^c\right) \cdot \mathbf{n}_{mp}}{\sigma}\right) + \Delta t^n \overline{\boldsymbol{B}}_m^{c,n}.\end{aligned}$$

Defining the *interior blended Riemann intermediate states* $\widetilde{\boldsymbol{v}}_{mp}^* := \overline{\boldsymbol{v}}_m^{c,n} - \frac{\widetilde{\mathbb{F}}_{mp} - \mathbb{F}\left(\overline{\boldsymbol{v}}_m^{c,n}, \overline{b}_m^c\right) \cdot \mathbf{n}_{mp}}{\sigma}$, the previous expression can be recast into the following form

$$\overline{\boldsymbol{v}}_m^{c,n+1} = \left(1 - \frac{\sigma \Delta t^n}{|S_m^c|} \sum_{S_p^v \in \mathcal{V}_m^c} \ell_{mp}\right) \overline{\boldsymbol{v}}_m^{c,n} + \frac{\sigma \Delta t^n}{|S_m^c|} \sum_{S_p^v \in \mathcal{V}_m^c} \ell_{mp} \widetilde{\boldsymbol{v}}_{mp}^* + \Delta t^n \overline{\boldsymbol{B}}_m^{c,n}.$$

The sum parts of this formulation of $\overline{\boldsymbol{v}}_m^{c,n+1}$ is a convex combination of solution-states, provided that the time step is computed under the CFL condition (3.4). By means of LF flux definition, we may rewrite the interior blended Riemann intermediate state as

$$\widetilde{\boldsymbol{v}}_{mp}^* = \frac{\overline{\boldsymbol{v}}_m^{c,n} + \overline{\boldsymbol{v}}_p^{v,n}}{2} - \frac{\left(\mathbb{F}\left(\overline{\boldsymbol{v}}_m^{c,n}, \overline{b}_m^c\right) - \mathbb{F}\left(\overline{\boldsymbol{v}}_p^{v,n}, \overline{b}_p^v\right)\right) \cdot \mathbf{n}_{mp}}{2\sigma} - \Theta_{mp} \frac{\Delta \mathbb{F}_{mp}}{\sigma} = \boldsymbol{v}_{mp}^* - \Theta_{mp} \frac{\Delta \mathbb{F}_{mp}}{\sigma},$$

where $\boldsymbol{v}_{mp}^*$ is the interior first-order FV Riemann intermediate state. Since the proposed numerical construction relies on non-conservative numerical fluxes (3.29), we need interior and exterior trace-values for the numerical fluxes, and we have the following definitions of blended Riemann intermediate states:

$$\widetilde{\boldsymbol{v}}_{mp}^* := \boldsymbol{v}_{mp}^* - \Theta_{mp} \left(\frac{\widehat{\mathbb{F}}_{mp} - \mathbb{F}_{mp}^{*,\mathrm{FV}}}{\sigma}\right), \qquad \widetilde{\boldsymbol{v}}_{pm}^* := \boldsymbol{v}_{pm}^* + \Theta_{mp} \left(\frac{\widehat{\mathbb{F}}_{pm} - \mathbb{F}_{pm}^{*,\mathrm{FV}}}{\sigma}\right),$$

where $\boldsymbol{v}_{mp}^{*,\pm}$ are the FV intermediate states defined as

$$\boldsymbol{v}_{mp}^* := \overline{\boldsymbol{v}}_m^c - \frac{\mathbb{F}_{mp}^{*,\mathrm{FV}} - \mathbb{F}\left(\overline{\boldsymbol{v}}_m^{c,n}, \overline{b}_m^c\right) \cdot \mathbf{n}_{mp}}{\sigma}, \qquad \boldsymbol{v}_{pm}^* := \overline{\boldsymbol{v}}_m^c - \frac{\mathbb{F}_{pm}^{*,\mathrm{FV}} - \mathbb{F}\left(\overline{\boldsymbol{v}}_p^{v,n}, \overline{b}_p^v\right) \cdot \mathbf{n}_{pm}}{\sigma}.$$

Consequently, let us consider the following set of admissible submean values (the same way as in the 1D case given in Definition 1.3.3):

$$\overline{\mathscr{H}}_b^+ = \left\{\overline{\boldsymbol{v}}_h \in \mathbf{P}^0(\widetilde{\mathcal{T}_h}) \mid \overline{H}_h := \overline{\eta}_h - \overline{b}_h \geq 0\right\},$$

that is derived from the continuous set $\mathscr{H}_b^+$. Then, if the numerical initial-data does lie in $\mathscr{H}_b^+$, we show in the following that it is possible to define local blending coefficients Θ_{mp} which allows to reach optimal convergence order in smooth areas, while ensuring that $\overline{\boldsymbol{v}}_m^c$ remains in $\mathscr{H}_b^+$ during the time-evolution (provided, as usual, that the first-order underlying FV scheme satisfies the positivity property), preventing in the meantime from spurious oscillations in the vicinity of classical shocks, and which is entirely compatible with the preservation of motionless steady-states. The various constraints on the blending coefficients will be expressed as inequalities. Consequently, to simultaneously enforce multiple properties, one simply needs to consider the minimum of the corresponding conditions, in the same fashion as the 1D case.

3.4.2 Blending coefficient smoothening

Before describing the detailed construction of such blending coefficients, let us emphasize that it has been previously noted in [265, 267] that, for some particularly nonlinear evolutions, in the two-dimensional case, a sharp switching between a first-order scheme and a fully high-order scheme can lead to unphysical oscillatory behaviors. For instance, some subcells may be assigned first-order FV fluxes on certain faces (hence $\Theta_{mp} = 0$) and fully high-order reconstructed fluxes (hence $\Theta_{mp} = 1$) on others. To mitigate such sharp variations, a *blending coefficient smoother* may be introduced, which aims at smoothing the blending coefficients across the mesh.

Mean values blending smoother

Each subcell $S_m^c \in \mathcal{T}_{\omega_c}^k$ is assigned a blending coefficient Θ_m^c, defined as the average of the blending coefficients of its faces, as follows:

$$\Theta_m^c := \frac{1}{\#\mathcal{V}_m^c} \sum_{S_p^v \in \mathcal{V}_m^c} \Theta_{mp}.$$

Subsequently, the blending coefficient for each subcell face may be adjusted by reducing it to the average of the blending coefficients of all subcells sharing a node with the face Γ_{mp}. This is expressed as:

$$\widetilde{\Theta}_{mp} := \min\left(\Theta_{mp}, \frac{1}{\#\mathcal{V}_{mp}} \sum_{S_q^v \in \mathcal{V}_{mp}} \Theta_q^v\right). \tag{3.33}$$

and is the one used to produce the numerical results of section 3.5.

Minimum values blending smoother

This second smoothing technique is a variation of the first, where the averaging operation is replaced by a minimization. Consequently, each subcell $S_m^c \in \mathcal{T}_{\omega_c}^k$ is assigned a blending coefficient Θ_m^c, defined as the minimum of the blending coefficients of its faces:

$$\Theta_m^c := \min_{S_p^v \in \mathcal{V}_m^c} \Theta_{mp}.$$

Therefore, the blending coefficient for each subcell face may be further refined by taking the minimum between its initial value and the smallest blending coefficient among all subcells sharing a node with face Γ_{mp}, as follows:

$$\widetilde{\Theta}_{mp} := \min\left(\Theta_{mp}, \min_{S_q^v \in \mathcal{V}_{mp}} \Theta_q^v\right).$$

Those two smoothening techniques have proved to improve the numerical results while preserving all the different properties presented in the next sections. In most of the numerical experiments, we will use the mean values blending smoother, as it is offering a less diffusive solution.

3.4.3 Water-height positivity (PAD criterion)

To develop a robust scheme for shallow-water problems, it is essential to ensure the positivity of the scheme, meaning that the numerical solution must not violate the non-negativity of the water-height. This requirement is notoriously challenging for high-order methods. Initially, it is necessary to demonstrate that the first-order finite volume (FV) Riemann intermediate states preserve the positivity of the water-height. Specifically, for a subcell $S_m^c \in \mathcal{T}_{\omega_c}^k$ and its neighboring subcell $S_p^v \in \mathcal{V}_m^c$, we must show that $\eta_{mp}^{*,-} \geq \overline{b}_m^c$ and $\eta_{mp}^{*,+} \geq \overline{b}_p^v$.

Proposition 3.4.1 (*Positivity of the first-order FV Riemann intermediate states*). *The water-height positivity is ensured on first-order FV Riemann intermediate states for the subcell FV fluxes.*

Proof. See Appendix B. □

We then can easily find a blending coefficient $\Theta_{mp}^{\mathscr{H}^+}$ that ensures preservation of water-height positivity.

Proposition 3.4.2 (*Positivity preservation*). *Let us consider the local-subcell monolithic update in two space dimensions. Assume that the subcell representation $\overline{\boldsymbol{v}}_h^n \in \mathbf{P}^0(\widetilde{\mathcal{T}_h})$ belongs to $\overline{\mathscr{H}}_b^+$, and that the time step satisfies the CFL condition (3.4). Then, for all $\omega_c \in \mathcal{T}_h$ and for all $S_m^c \in \mathcal{T}_{\omega_c}$, the updated local solution $\overline{\boldsymbol{v}}_h^{c,n+1} \in \mathbf{P}^0(\mathcal{T}_{\omega_c}^k)$ is admissible, provided that the blending coefficient satisfies on each subface Γ_{mp} the following constraint:*

$$\Theta_{mp} \leq \Theta_{mp}^{\mathscr{H}^+} := \min\left(1, \Theta_{mp}^{\mathscr{H}^+,-}, \Theta_{mp}^{\mathscr{H}^+,+}\right),$$

where the two quantities $\Theta_{mp}^{\mathscr{H}^+,\pm}$ are imposed only when the corresponding constraint is active:

$$\Theta_{mp}^{\mathscr{H}^+,-} = \frac{\sigma\left(\eta_{mp}^{*,-} - \overline{b}_m^c\right)}{\Delta\mathbb{F}_{mp}} \quad \text{if } \Delta\mathbb{F}_{mp} > 0, \qquad \Theta_{mp}^{\mathscr{H}^+,+} = \frac{\sigma\left(\overline{b}_p^v - \eta_{mp}^{*,+}\right)}{\Delta\mathbb{F}_{pm}} \quad \text{if } \Delta\mathbb{F}_{pm} < 0,$$

and are otherwise set equal to 1.

Remark 3.4.3 (*Initial-data projection*). *We emphasize that a neat computation of the discrete initial-data is required, to ensure the initial positivity of the water-height both at the cells and subcells scales. To achieve this, we first compute the subcell mean-values of the variables and then apply the projection matrix $\mathbb{P}_c$ to recover the cell polynomial representation, ensuring that the approximate initial-data belongs to $\overline{\mathscr{H}}_b^+$ (or, more generally, to any convex set of admissible states of interest). Indeed, the approximate initial data $\boldsymbol{v}_h^0$ must be admissible in order to generate a sequence of admissible states under the CFL condition and the chosen blending coefficients. Starting from a physical or idealized (analytical) initial-data $\boldsymbol{v}^0 \coloneqq (\eta^0, q_x^0, q_y^0)$ and a bathymetry parameterization b such that $\eta^0 - b \geq 0$, we observe that $\widetilde{\pi}_h^k(\eta^0 - b) \geq 0$. Therefore, the subcell representation $\overline{\boldsymbol{v}}_h^0 \coloneqq \widetilde{\pi}_h^k \boldsymbol{v}^0$ is admissible. In the same spirit, we also have that $\boldsymbol{v}_h^0 \coloneqq (\overline{\pi}_h^k)^{-1}\overline{\boldsymbol{v}}_h^0$ remains admissible. From this last relation, we see that the approximate initial surface elevation must be defined consistently with the previous definition of the approximate bathymetry:*

$$\eta_h^0 \coloneqq (\overline{\pi}_h^k)^{-1}\widetilde{\pi}_h^k \eta^0 - b_h. \tag{3.34}$$

3.4.4 Relaxed subcell numerical admissibility detection (SubNAD criterion)

In order to tackle the issue of spurious non-physical oscillations near discontinuities, we want to mimic a local maximum principle (LMP) at the discrete level on the submean values. We therefore impose the submean value on total elevation $\overline{\eta}_m^{c,n+1}$ to be bounded by the submean values at the previous time step (or previous RK step in time integration) in a given subcell set. Also, as theoretically there is no local maximum for the conserved variables in the case of nonlinear system, we add the interior FV Riemann intermediate state to the local bounds, and get

$$\alpha_m^c \coloneqq \min_{S_p^v \in \mathcal{N}(S_m^c)} \left(\overline{\eta}_p^{v,n}, \eta_{mp}^{*,-}\right) \leq \overline{\eta}_m^{c,n+1} \leq \max_{S_p^v \in \mathcal{N}(S_m^c)} \left(\overline{\eta}_p^{v,n}, \eta_{mp}^{*,-}\right) =: \beta_m^c, \tag{3.35}$$

Remark 3.4.4 (*LMP on Riemann invariants*). *Let us note that we also worked on applying the SubNAD and relaxation principles on Riemann invariants $I^{\pm} = \frac{\boldsymbol{q}\cdot\mathbf{n}}{H} \pm 2\sqrt{gH}$ (under the hypothesis that $b \equiv 0$) and its linearization*

$$\mathrm{d}I^{\pm} = \frac{\mathrm{d}\boldsymbol{q}\cdot\mathbf{n}}{H} - \frac{\boldsymbol{q}\cdot\mathbf{n}}{H^2} \pm \sqrt{\frac{\mathrm{g}}{H}}\mathrm{d}H,$$

as it allows to stay closer to the physics of the equation. We then defined the local subcell linearized Riemann invariants

$$\overline{\mathfrak{I}}_m^{c,\pm,n} = \frac{\overline{\boldsymbol{q}}_m^{c,n}}{\overline{H}_m^{c,n}} - \overline{H}_m^{c,n}\overline{\Lambda}_m^{c,\pm,n}, \qquad \overline{\Lambda}_m^{c,\pm,n} \coloneqq \overline{\left[\frac{\boldsymbol{q}_h^{c,n}\cdot\mathbf{n}}{(H_h^{c,n})^2} \mp \sqrt{\frac{\mathrm{g}}{H_h^{c,n}}}\right]}_m^{c,n}.$$

That gets us to rewrite this linearized Riemann invariants as the following convex combination:

$$\begin{aligned}
\overline{\mathfrak{I}}_m^{c,\pm,n+1} &= \left(1 - \frac{\sigma\Delta t^n}{|S_m^c|}\sum_{S_p^v \in \mathcal{V}_m^c} \ell_{mp}\right)\overline{\mathfrak{I}}_m^{c,\pm,n} + \frac{\sigma\Delta t^n}{|S_m^c|}\sum_{S_p^v \in \mathcal{V}_m^c} \ell_{mp}\left(\overline{\mathfrak{I}}_m^{c,\pm,n} - \frac{\widetilde{\mathbb{F}}_{mp}^{\boldsymbol{q}} - \overline{H}_m^{c,n}\overline{\Lambda}_m^{c,\pm,n}\widetilde{\mathbb{F}}_{mp}^{\eta}}{\sigma\overline{H}_m^{c,n}}\right) \\
&= \left(1 - \frac{\sigma\Delta t^n}{|S_m^c|}\sum_{S_p^v \in \mathcal{V}_m^c} \ell_{mp}\right)\overline{\mathfrak{I}}_m^{c,\pm,n} + \frac{\sigma\Delta t^n}{|S_m^c|}\sum_{S_p^v \in \mathcal{V}_m^c} \ell_{mp}\widetilde{\mathfrak{I}}_{mp}^{*,-}.
\end{aligned}$$

with $\widetilde{\mathfrak{I}}_{mp}^{*,-} = \overline{\mathfrak{I}}_m^{c,\pm,n} - \frac{\widetilde{\mathbb{F}}_{mp}^{\boldsymbol{q}} - \overline{H}_m^{c,n}\overline{\Lambda}_m^{c,\pm,n}\widetilde{\mathbb{F}}_{mp}^{\eta}}{\sigma\overline{H}_m^{c,n}}$ the interior blended Riemann states. We will not go further in this direction, as the numerical results showed no significant improvement.

Proposition 3.4.5 (*Subcell numerical admissibility detection*). *To ensure condition (3.35) is satisfied for any subcell, it is sufficient to verify that $\widetilde{\eta}_{mp}^{*,-} \in [\alpha_m^c, \beta_m^c]$ and $\widetilde{\eta}_{mp}^{*,+} \in [\alpha_p^v, \beta_p^v]$. Since $\eta_{mp}^{*,\pm}$ satisfies both conditions, it suffices to impose the following constraint on Θ_{mp}:*

$$\Theta_{mp}^{\mathrm{nad}} \leq \min\left(1, \left|\frac{\sigma}{\Delta\mathbb{F}_{mp}}\right| \begin{cases} \min\left(\beta_p^v - \eta_{mp}^{*,+},\ \eta_{mp}^{*,-} - \alpha_m^c\right) & \text{if } \Delta\mathbb{F}_{mp} > 0, \\ \min\left(\beta_m^c - \eta_{mp}^{*,-},\ \eta_{mp}^{*,+} - \alpha_p^v\right) & \text{if } \Delta\mathbb{F}_{mp} < 0. \end{cases}\right).$$

Let us say again that the local maximum principle (3.35) is based on subcell mean values. Since this constraint does not account for the full polynomial set of values, it is well-known that relaxing it is necessary to preserve the scheme's accuracy in the presence of smooth extrema.

Relaxation through a smooth extrema detector

As in [265, 267], we also introduce some relaxation through a smooth extrema detector in order to preserve the scheme's accuracy in the presence of smooth extrema. This detector is based on the smoothness indicator for finite elements, as described in [178]. The main idea is as follows: the numerical solution is considered to exhibit a smooth extrema profile if the linearized version of the numerical solution's spatial derivatives represents a monotonous profile. To achieve this, we introduce the following subcell linear reconstructions:

$$\mathfrak{E}_x^m(\boldsymbol{x}) = \overline{\partial_x \eta_h^c}^m + \overline{\nabla_{\boldsymbol{x}}(\partial_x \eta_h^c)}^m \cdot (\boldsymbol{x} - \boldsymbol{x}_m^c), \tag{3.36}$$

$$\mathfrak{E}_y^m(\boldsymbol{x}) = \overline{\partial_y \eta_h^c}^m + \overline{\nabla_{\boldsymbol{x}}(\partial_y \eta_h^c)}^m \cdot (\boldsymbol{x} - \boldsymbol{x}_m^c). \tag{3.37}$$

In the above, $\boldsymbol{x}_m^c$ represents the barycenter of the subcell S_m^c, and $\overline{\partial_{x\setminus y}\eta_h^c}^m$ and $\overline{\nabla_{\boldsymbol{x}}(\partial_{x\setminus y}\eta_h^c)}^m$ denote the subcell averages of the successive partial derivatives of η_h^c. The smoothness indicator functions $\mathfrak{E}_x^m$ and $\mathfrak{E}_y^m$, being linear, attain their extrema at the vertices $\boldsymbol{x}_q \in \mathcal{P}_m^c$. These extrema act as vertex-based limiters. To confirm that the numerical solution η_h presents a smooth profile within subcell S_m^c, we require that the linearized spatial derivative functions satisfy the following constraints:

$$\mathfrak{E}_{x,q}^{\min} \leq \mathfrak{E}_x^m(\boldsymbol{x}_q) \leq \mathfrak{E}_{x,q}^{\max} \quad \text{and} \quad \mathfrak{E}_{y,q}^{\min} \leq \mathfrak{E}_y^m(\boldsymbol{x}_q) \leq \mathfrak{E}_{y,q}^{\max}, \tag{3.38}$$

where $\mathfrak{E}_{x\setminus y,q}^{\min}$ and $\mathfrak{E}_{x\setminus y,q}^{\max}$ are defined as:

$$\mathfrak{E}_{x\setminus y,q}^{\min} = \min_{S_m^c \in \mathcal{V}_q} \overline{\partial_{x\setminus y}\eta_h^c}^m \quad \text{and} \quad \mathfrak{E}_{x\setminus y,q}^{\max} = \max_{S_m^c \in \mathcal{V}_q} \overline{\partial_{x\setminus y}\eta_h^c}^m.$$

In practice, if condition (3.38) holds for all vertices $x_q \in \mathcal{P}_m^c$, the numerical solution η_h can be regarded as smooth within subcell S_m^c. Additionally, if the solution is smooth in both subcells S_m^c and S_p^v, the blending coefficient constraint through the local maximum condition (3.35) can be relaxed (ie. $\Theta_{mp} = 1$). This approach ensures the preservation of smooth extrema with high-order accuracy, particularly for smooth problems.

Remark 3.4.6 (*Link between the detector and the order of the scheme*). *Smooth extrema detection is not required for second-order approximations. For third-order local subcell monolithic DG/FV schemes, however, the smoothness detector must operate at the cell level. Indeed, in this case, the second derivatives $\nabla_{\boldsymbol{x}}(\partial_{x\setminus y}\eta_h^c)$ are constant and do not require additional subcell-level computations.*

3.4.5 Well-balancing for motionless steady-states

In this section, we show that the previous constrained choices of blended coefficients, when combined together with local and global hydrostatic-like reconstructions, do allow to preserve a well-balanced property.

Remark 3.4.7 (*Steady-state relation*). *Let us note that under the motionless steady-state assumption $\eta_h = \eta^e > 0$ and $\boldsymbol{q}_h^n = \boldsymbol{0}$, the following relation holds:*

$$\nabla_{\boldsymbol{x}} \cdot \mathbb{F}(\boldsymbol{v}_c, b_c) = \boldsymbol{B}[b_c](\boldsymbol{v}_c), \qquad \forall \omega_c \in \mathcal{T}_h.$$

Moreover, as we have

$$\mathbb{F}(\boldsymbol{v}_h, b_h) = \begin{pmatrix} 0 & 0 \\ \frac{1}{2}g\left(\eta_h^2 - 2\eta_h b_h\right) & 0 \\ 0 & \frac{1}{2}g\left(\eta_h^2 - 2\eta_h b_h\right) \end{pmatrix},$$

we emphasize that $\mathbb{F}(\boldsymbol{v}_h, b_h)$ belongs to $\mathscr{M}_{2\times 3}\left(\mathbb{P}^{2k}(\mathcal{T}_h)\right)$. Therefore at steady state,

$$\mathbb{F}_h := \pi_h^k \mathbb{F}(\boldsymbol{v}_h, b_h) = \mathbb{F}(\boldsymbol{v}_h, b_h).$$

Proposition 3.4.8 (*Well-balancing property*). *The discrete formulation (3.32) preserves the motionless steady states, provided that the integrals in both the DG and FV subcell formulations are exactly computed at steady-states. We then have:*

$$\forall n \in \mathbb{N}, \quad \forall \eta^e \in \mathbb{R}, \qquad (\eta_h^n = \eta^e \text{ and } \boldsymbol{q}_h^n = \boldsymbol{0}) \Longrightarrow \left(\eta_h^{n+1} = \eta^e \text{ and } \boldsymbol{q}_h^{n+1} = \boldsymbol{0}\right).$$

Proof. Let us recall that the discrete formulation of the monolithic DG/FV subcell scheme is given by:

$$\overline{\boldsymbol{v}}_m^{c,n+1} = \overline{\boldsymbol{v}}_m^{c,n} - \frac{\Delta t^n}{|S_m^c|} \sum_{S_p^v \in \mathscr{V}_m^c} \ell_{mp} \widetilde{\mathbb{F}}_{mp} + \Delta t^n \overline{\boldsymbol{B}}_m^{c,n}.$$

At motionless steady state, we want to show that the numerical fluxes are cancelling the source term,

$$\frac{1}{|S_m^c|} \sum_{S_p^v \in \mathscr{V}_m^c} \ell_{mp} \widetilde{\mathbb{F}}_{mp} = \overline{\boldsymbol{B}}_m^{c,n} \qquad \text{s.t.} \qquad \overline{\boldsymbol{v}}_m^{c,n+1} = \overline{\boldsymbol{v}}_m^{c,n}. \tag{3.39}$$

Since the blended flux $\widetilde{\mathbb{F}}_{mp}$ is a convex combination of the first-order FV flux $\mathbb{F}_{mp}^{*,\mathrm{FV}}$ and the high-order reconstructed flux $\widehat{\mathbb{F}}_{mp}$, it is enough to prove that the source term is exactly cancelled by both fluxes. Let us first consider the case where the blending Θ is set to 1, meaning we consider only the high-order contribution. We have

$$\widehat{\mathbb{F}}_c = \begin{cases} \mathbb{F}_c - \mathbb{A}_c^\top \mathscr{L}_c^{-1} \partial \mathbb{F}_c & \text{in } \omega_c \setminus \overline{\partial\omega_c}, \\ \mathbb{F}_{\partial\omega_c}^{*,\mathrm{DG}} & \text{on } \partial\omega_c, \end{cases}$$

where

$$(\partial \mathbb{F}_c)_m = \int_{\partial\omega_c} \left(\mathbb{F}_h^c \cdot \mathbf{n}_{\partial\omega_c} - \mathbb{F}_{\partial\omega_c}^{*,\mathrm{DG}} \right) \widetilde{\phi}_m^c(s) \, \mathrm{d}s.$$

Let us show that, under steady-states assumption, for each interface $\Gamma_{cv(k)} \subset \partial\omega_c$, $k \in [\![1, \#\mathscr{F}_c]\!]$, we have $\mathbb{F}_{cv(k)}^{*,\mathrm{DG}} = \mathbb{F}_h^c \cdot \mathbf{n}_{cv(k)}$. This can simply be obtained by remarking that

$$\mathbb{F}_h^c \cdot \mathbf{n}_{cv(k)} = \mathbb{F}(\boldsymbol{v}_k^-, b_k^-) \cdot \mathbf{n}_{cv(k)} = \mathbb{F}\left(\begin{pmatrix} \eta^e \\ \mathbf{0} \end{pmatrix}, b_k^- \right) \cdot \mathbf{n}_{cv(k)} = \begin{pmatrix} 0 & 0 \\ \frac{1}{2} g \left((\eta^e)^2 - 2\eta^e b_k^- \right) & 0 \\ 0 & \frac{1}{2} g \left((\eta^e)^2 - 2\eta^e b_k^- \right) \end{pmatrix} \cdot \mathbf{n}_{cv(k)}$$

and also

$$\begin{aligned} \mathbb{F}_{cv(k)}^{*,\mathrm{DG}} &= \frac{\mathbb{F}(\check{\boldsymbol{v}}_k^-, \check{b}_k) + \mathbb{F}(\check{\boldsymbol{v}}_k^+, \check{b}_k)}{2} \cdot \mathbf{n}_{cv(k)} - \frac{\sigma}{2} \left(\check{\boldsymbol{v}}_k^+ - \check{\boldsymbol{v}}_k^- \right) + \check{\mathbb{F}}_{cv(k)} = \frac{\mathbb{F}(\check{\boldsymbol{v}}_k^-, \check{b}_k) + \mathbb{F}(\check{\boldsymbol{v}}_k^+, \check{b}_k)}{2} \cdot \mathbf{n}_{cv(k)} + \check{\mathbb{F}}_{cv(k)} \\ &= \left\{ \begin{pmatrix} 0 & 0 \\ \frac{1}{2} g \left((\eta^e)^2 - 2\eta^e \check{b}_k \right) & 0 \\ 0 & \frac{1}{2} g \left((\eta^e)^2 - 2\eta^e \check{b}_k \right) \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \mathrm{g}\eta^e (\check{b}_k - b_k^-) & 0 \\ 0 & \mathrm{g}\eta^e (\check{b}_k - b_k^-) \end{pmatrix} \right\} \cdot \mathbf{n}_{cv(k)} \\ &= \begin{pmatrix} 0 & 0 \\ \frac{1}{2} g \left((\eta^e)^2 - 2\eta^e b_k^- \right) & 0 \\ 0 & \frac{1}{2} g \left((\eta^e)^2 - 2\eta^e b_k^- \right) \end{pmatrix} \cdot \mathbf{n}_{cv(k)}. \end{aligned}$$

This means that we have

$$\widehat{\mathbb{F}}_c = \begin{cases} \mathbb{F}_c & \text{in } \omega_c \setminus \overline{\partial\omega_c}, \\ \mathbb{F}_{\partial\omega_c}^{*,\mathrm{DG}} & \text{on } \partial\omega_c, \end{cases}$$

and therefore, by Remark 3.4.7, we have $\widehat{\mathbb{F}}_c = \mathbb{F}(\boldsymbol{v}_h^c, b_h^c)$ at steady states and we can write:

$$\sum_{S_p^v \in \mathscr{V}_m^c} \ell_{mp} \widehat{\mathbb{F}}_{mp} = \int_{S_m^c} \nabla_{\boldsymbol{x}} \cdot \mathbb{F}_h^c \, \mathrm{d}\boldsymbol{x} = \int_{S_m^c} \nabla_{\boldsymbol{x}} \cdot \mathbb{F}(\boldsymbol{v}_h^c, b_h^c) \, \mathrm{d}\boldsymbol{x}.$$

We get then

$$\frac{1}{|S_m^c|} \sum_{S_p^v \in \mathscr{V}_m^c} \ell_{mp} \widehat{\mathbb{F}}_{mp} - \overline{\boldsymbol{B}}_m^{c,n} = \frac{1}{|S_m^c|} \left(\int_{S_m^c} \nabla_{\boldsymbol{x}} \cdot \mathbb{F}(\boldsymbol{v}_h^c, b_h^c) \, \mathrm{d}\boldsymbol{x} - \int_{S_m^c} \boldsymbol{B}(\boldsymbol{v}_h^c, \nabla_{\boldsymbol{x}} b_h^c) \, \mathrm{d}\boldsymbol{x} \right),$$

and since we have $\nabla_{\boldsymbol{x}} \cdot \mathbb{F}(\boldsymbol{v}_h^c, b_h^c) = \boldsymbol{B}(\boldsymbol{v}_h^c, \nabla_{\boldsymbol{x}} b_h^c)$ at steady states,

$$\frac{1}{|S_m^c|} \sum_{S_p^v \in \mathscr{V}_m^c} \ell_{mp} \widehat{\mathbb{F}}_{mp} - \overline{\boldsymbol{B}}_m^{c,n} = 0 \qquad \text{s.t.} \qquad \overline{\boldsymbol{v}}_m^{c,n+1} = \overline{\boldsymbol{v}}_m^{c,n}.$$

Let us now consider $\Theta = 0$, meaning that we consider only the first-order FV flux contributions, we show that

$$\frac{1}{|S_m^c|}\sum_{S_p^v\in\mathscr{V}_m^c}\ell_{mp}\mathbb{F}_{mp}^{*,\mathrm{FV}} = \overline{\boldsymbol{B}}_m^{c,n} \qquad \text{s.t.} \qquad \overline{\boldsymbol{v}}_m^{c,n+1} = \overline{\boldsymbol{v}}_m^{c,n}.$$

Let us note that

$$\begin{aligned}
\mathbb{F}_{mp}^{*,\mathrm{FV}} &= \frac{\mathbb{F}(\breve{\overline{\boldsymbol{v}}}_m^c, \breve{\overline{b}}_{mp}) + \mathbb{F}(\breve{\overline{\boldsymbol{v}}}_p^v, \breve{\overline{b}}_{mp})}{2}\cdot\mathbf{n}_{mp} - \frac{\sigma}{2}\left(\breve{\overline{\boldsymbol{v}}}_p^v - \breve{\overline{\boldsymbol{v}}}_m^c\right) + \breve{\mathbb{F}}_{mp}^{\mathrm{FV}} = \frac{\mathbb{F}(\breve{\overline{\boldsymbol{v}}}_m^c, \breve{\overline{b}}_{mp}) + \mathbb{F}(\breve{\overline{\boldsymbol{v}}}_p^v, \breve{\overline{b}}_{mp})}{2}\cdot\mathbf{n}_{mp} + \breve{\mathbb{F}}_{mp}^{\mathrm{FV}} \\
&= \left\{\begin{pmatrix} 0 & 0 \\ \frac{1}{2}g\left((\eta^e)^2 - 2\eta^e\breve{\overline{b}}_{mp}\right) & 0 \\ 0 & \frac{1}{2}g\left((\eta^e)^2 - 2\eta^e\breve{\overline{b}}_{mp}\right)\end{pmatrix} + \begin{pmatrix} 0 & 0 \\ g\eta^e\left(\breve{\overline{b}}_{mp} - b_{mp}\right) & 0 \\ 0 & g\eta^e\left(\breve{\overline{b}}_{mp} - b_{mp}\right)\end{pmatrix}\right\}\cdot\mathbf{n}_{mp} \\
&= \begin{pmatrix} 0 & 0 \\ \frac{1}{2}g\left((\eta^e)^2 - 2\eta^e b_{mp}\right) & 0 \\ 0 & \frac{1}{2}g\left((\eta^e)^2 - 2\eta^e b_{mp}\right)\end{pmatrix}\cdot\mathbf{n}_{mp}.
\end{aligned}$$

Since $b_{mp} := \frac{1}{|S_m^c|}\int_{\Gamma_{mp}^c} b_h^c(s)\,\mathrm{d}s$, it indeed implies equality between the numerical flux $\mathbb{F}_{mp}^{*,\mathrm{FV}}$ and the flux function $\mathbb{F}_h^c\cdot\mathbf{n}_{mp}$ at steady-states. For any interface Γ_{mp} and in the general case $\Theta_{mp}\in[0,1]$ we have for all $S_m^c\in\mathscr{T}_{\omega_c}^k$, for all $S_p^v\in\mathscr{V}_m^c$, $\widehat{\mathbb{F}}_{mp} = \mathbb{F}_{mp}^{*,\mathrm{FV}} = \mathbb{F}_h^c\cdot\mathbf{n}_{mp}$ at steady-states, and we easily get the following:

$$\begin{aligned}
\frac{1}{|S_m^c|}\sum_{S_p^v\in\mathscr{V}_m^c}\ell_{mp}\widetilde{\mathbb{F}}_{mp} &= \frac{1}{|S_m^c|}\sum_{S_p^v\in\mathscr{V}_m^c}\ell_{mp}\mathbb{F}_{mp}^{*,\mathrm{FV}} + \frac{1}{|S_m^c|}\sum_{S_p^v\in\mathscr{V}_m^c}\ell_{mp}\Theta_{mp}\left(\widehat{\mathbb{F}}_{mp} - \mathbb{F}_{mp}^{*,\mathrm{FV}}\right) \\
&= \frac{1}{|S_m^c|}\sum_{S_p^v\in\mathscr{V}_m^c}\ell_{mp}\mathbb{F}_{mp}^{*,\mathrm{FV}} = \overline{\boldsymbol{B}}_m^{c,n},
\end{aligned}$$

which shows the exact cancellation of the source term by the blended numerical fluxes. □

3.5 Numerical simulations

In this section, we show that the previous discrete formulation is operational and that the claimed properties are ensured in various existing scenarios, demonstrating that the expected and previously proven properties are effectively observable in a reliable manner. If not stated otherwise, the numerical simulations are produced using the monolithic DG/FV scheme with relaxed SubNAD and PAD criteria, with the blending smoother (3.33). Both subdivisions are shown (i.e. fully triangular and quadrilateral/triangular one), as neither of them affects the quality of numerical results. When presenting pseudo-1D results (i.e. domain slices and density profiles), we either display the subcell mean values or the cell-averaged values (i.e., the mean of their corresponding subcell values) to ensure that the plots are readable. For some benchmarks, animations are also provided for illustration purposes. Please note that they do not always correspond exactly to the static configuration displayed in the figures, but they are intended to give a clearer view of the dynamics; when available, the corresponding link is given in the caption.

3.5.1 Steady vortex with smooth topography

PARAMETERS

Computational domain: $\Omega = [-5,5]^2$
Number of cells: $N_{\mathrm{el}} = 200,\ 800,\ 3200,\ 12800$ cells
Boundary conditions: Dirichlet
Polynomial degree: $k = 1,2,3$
Time-marching order: $p = 3$
Final time: $T_{\max} = 1$ sec

This first test-case aims at assessing the convergence properties of the method in some optimal set-up, with the occurrence of a smooth bathymetry: we consider a steady vortex configuration detailed in [218]. The bathymetry is $b(\boldsymbol{x}) = 0.2e^{0.5(1-r^2)}$ ($r^2 = x^2+y^2$) and the corresponding exact solution, illustrated in Figure 3.7, $\boldsymbol{v} = (\eta, \boldsymbol{q})^\top$, is given by:

$$\begin{aligned}
H(\boldsymbol{x},t) &= 1 - \frac{1}{4g}e^{2(1-r^2)} - b(\boldsymbol{x}), \qquad \eta(\boldsymbol{x},t) = H(\boldsymbol{x},t) + b(\boldsymbol{x}), \\
q_x(\boldsymbol{x},t) &= H(\boldsymbol{x},t)\left(ye^{1-r^2}\right), \qquad q_y(\boldsymbol{x},t) = H(\boldsymbol{x},t)\left(-xe^{1-r^2}\right),
\end{aligned}$$

which allows to quantitatively compare the approximate and expected solutions at a given discrete-time. Even if not fully displayed here for the sake of conciseness, we report some excellent agreements for both conservative variables, and we emphasize that an "optimal" convergence-order of $\mathcal{O}(k+1)$ is observed in every studied computational scenario. As an example, we show color-valued free-surface elevation on Figure 3.7 for a particular simulation (similar results are observable with the discharge) and we report the corresponding error-values in Table 3.9, with practical details of the computation provided in the grey summary-box above. We emphasize that such an optimal behavior is achieved even using the relaxed-SubNAD criterion, ensuring that this particular numerical trick does not modulate the leading error terms, even when considering higher-order approximations.

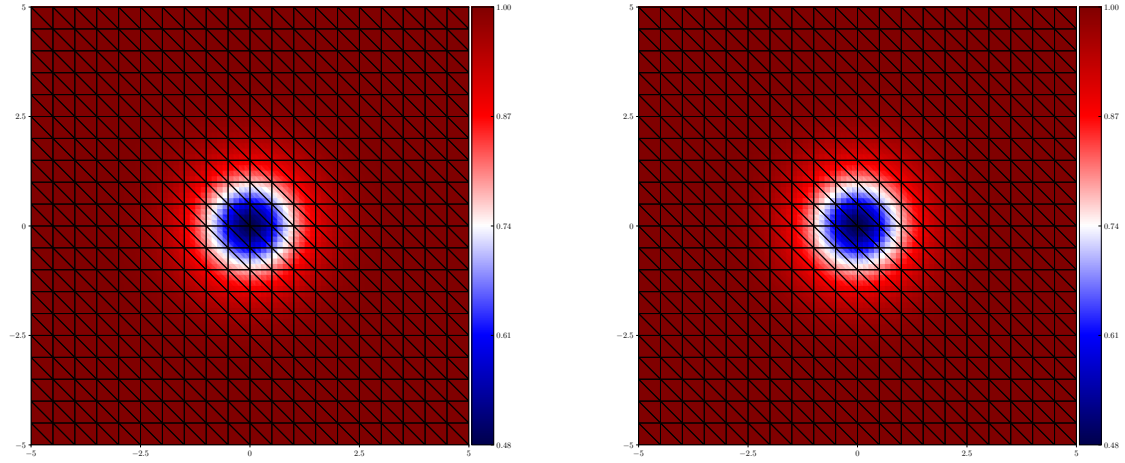


Figure 3.7: Steady vortex – Exact (left) and $\mathbb{P}^3$ numerical (right) height at final time $t = 1$ sec on 800 cells.

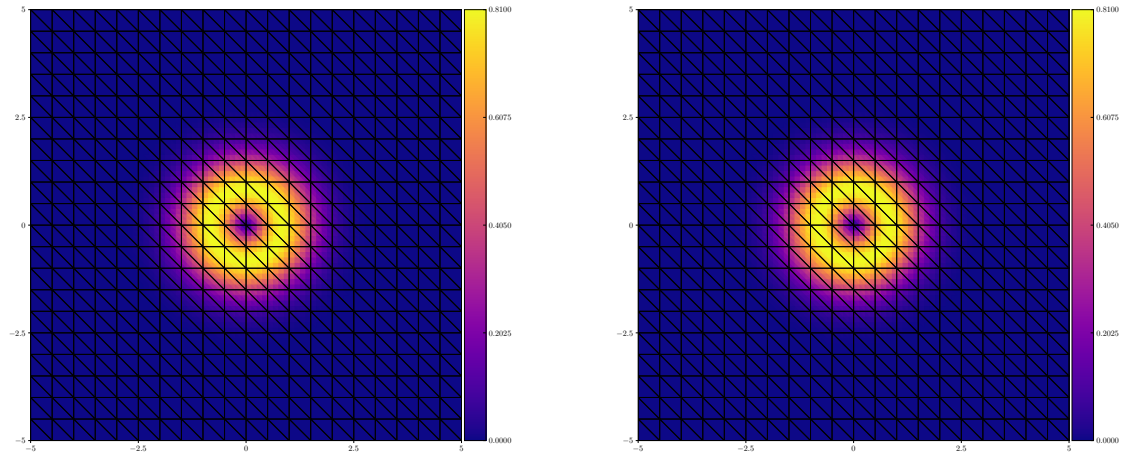


Figure 3.8: Steady vortex – Exact (left) and $\mathbb{P}^3$ numerical (right) discharge norm at final time $t = 1$ sec on 800 cells.

3.5.2 Well-balancing

PARAMETERS

Computational domain: $\Omega = [0, 2] \times [0, 1]$
Number of cells: $N_{\text{el}} = 506, 2064$ cells
Boundary conditions: solid walls

Polynomial degree: $k = 4$
Time-marching order: $p = 3$
Final time: $T_{\max} = 20$ sec

We perform numerical experiments to assess the well-balancedness of the scheme and we consider a non-uniform topography defined as follows:

$$b(\boldsymbol{x}) = \begin{cases} 0.5e^{-100((x-1.2)^2+150(y-0.7)^2)} & \text{if } x > 0.68, \\ -0.5e^{-100((x-0.45)^2+150(y-0.4)^2)} & \text{otherwise.} \end{cases}$$

k	1		2		3	
h	$E^{\eta}_{L^2}$	$q^{\eta}_{L^2}$	$E^{\eta}_{L^2}$	$q^{\eta}_{L^2}$	$E^{\eta}_{L^2}$	$q^{\eta}_{L^2}$
1	9.445E-2	2.35	1.529E-2	2.91	4.580E-3	4.19
$\frac{1}{2}$	1.854E-2	2.16	2.039E-3	3.03	2.505E-4	4.10
$\frac{1}{4}$	4.158E-3	2.07	2.491E-4	2.97	1.465E-5	4.00
$\frac{1}{8}$	9.923E-4	–	3.187E-5	–	9.165E-7	–

Figure 3.9: Steady vortex – L^2-errors between numerical and analytical solutions and convergence rates for η at time $t = 1$ sec.

Two different set-ups are investigated on unstructured and coarse meshes, in order to assess the capability of our numerical method to preserve motionless balance within a high-order setting: fully-immersed and partly-immersed bathymetry, see Figures 3.10–3.12 for a visual description.

Fully immersed

At the initial time $t = 0$, we first prescribe $\eta_0 = 1$ and $\boldsymbol{q}_0 = 0$. We show on Figure 3.10 the corresponding flat free-surface profile, and the associated blending coefficient, which remains stuck to 1 as expected.

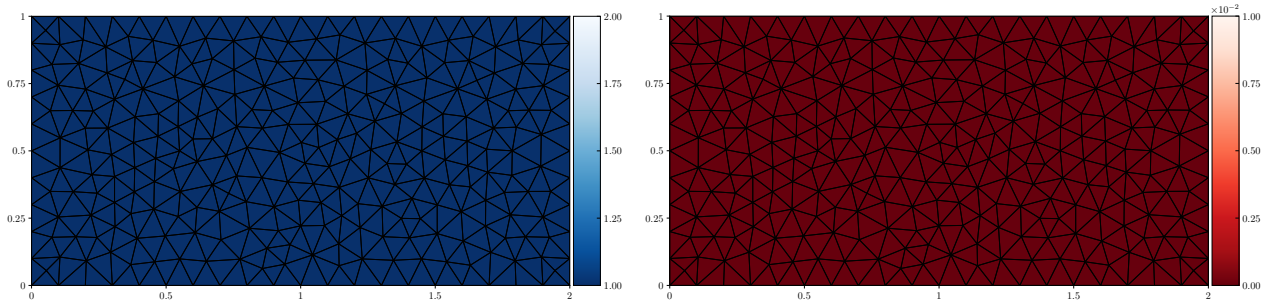


Figure 3.10: Well-balancing – $\mathbb{P}^4$ solution at final time $t = 20$ sec – 2D free-surface elevation (left) and discharge norm (right).

Partly immersed

Going further into the practical investigations, we consider a finer mesh of 2064 cells, and initially prescribe a localized dry area by setting $\eta_0 = \max(b, 0.2)$ and $\boldsymbol{q}_0 = 0$. Previous machine-accuracy preservation may be also obtained provided that a matching mesh, perfectly compatible with the wet/dry contact line, is used. With general (non-matching) meshes, tiny oscillations may be classically observed near the contact-line for the discharge and for higher-order approximations, as expected and exhibited in Figures 3.12–3.13. We emphasize that such an issue is inherent to the nature of wet-dry interfaces located *inside* mesh elements, as the same local polynomial approximations have to alternatively approximate, into the same mesh element, free-surface and bathymetry. If such a situation generally generates numerical instabilities for globally high-order methods, it is remarkable to observe that the subcell resolution capability of the new proposed approach, although not perfect, allows to largely mitigate such mesh-geometry disagreement oscillations, as expected from operational and robust methods. These neither propagate nor amplify with respect to time, and from our extensive investigations, they do not affect the global accuracy. Thanks to the PAD, the simulation remains globally stable and the blending coefficients are automatically adapted to properly handle the presence of the dry area, as shown in Figure 3.12 within a vertical slice of the computational domain through the bump, while Figure 3.13 displays the bathymetry bump/hole locations, together with free-surface elevation, discharge norm, and a map of blending coefficients mean-values per subcell, in order to clearly observe the PAD activation and effective corrections. Importantly, this shows that the correction mechanism is only switched-on in the vicinity of dry areas, as expected, which highlights the robustness and efficiency of the proposed framework in globally optimizing the preservation of the full accuracy of the higher-order underlying DG scheme. Note that

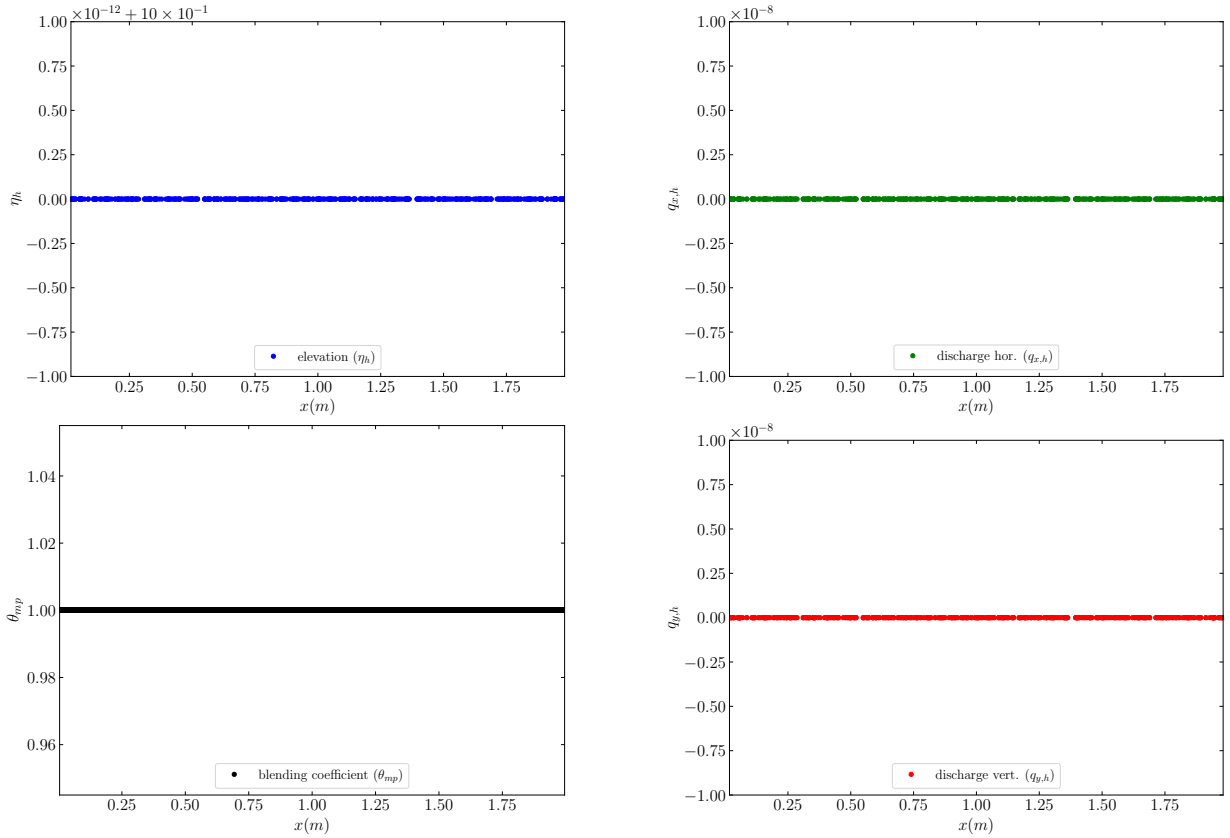


Figure 3.11: Well-balancing – At $t = 20$ sec, from left to right and top to bottom: free-surface elevation, horizontal discharge, blending coefficient and vertical discharge density profiles.

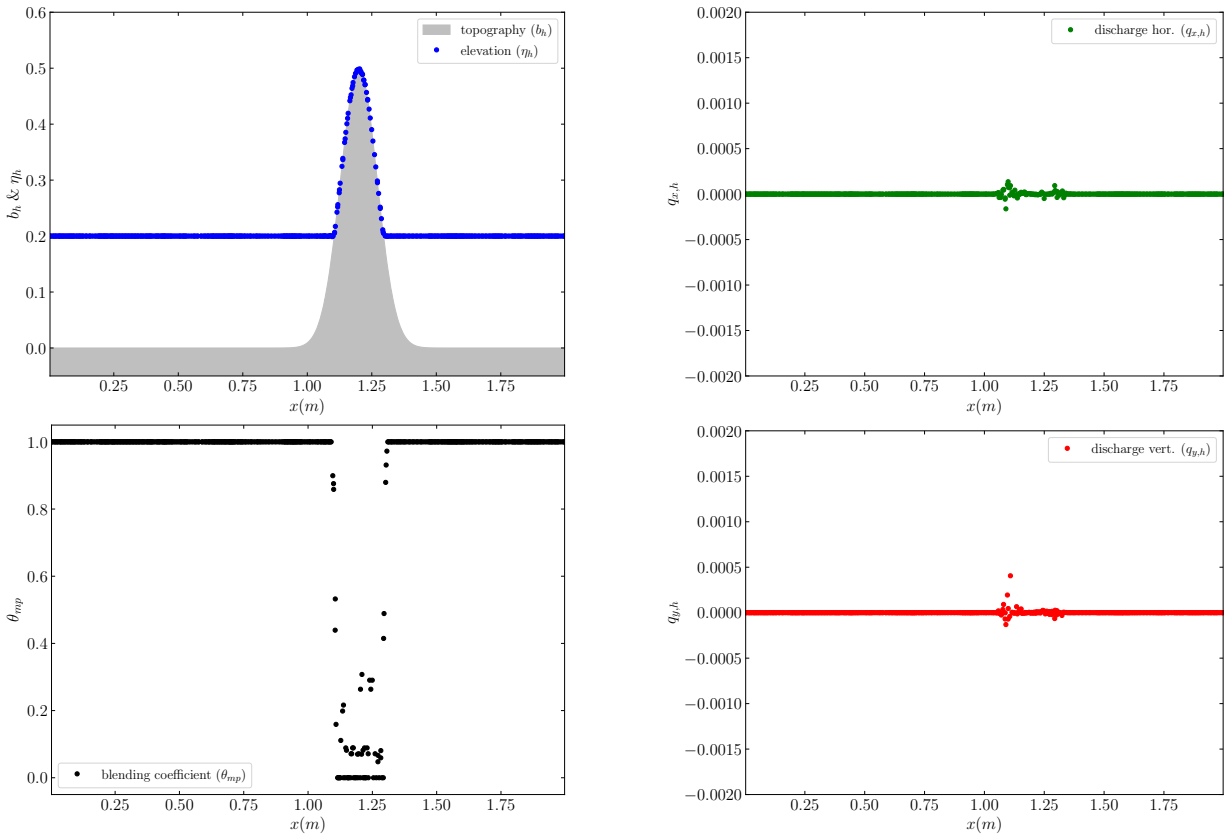


Figure 3.12: Well-balancing with dry area – $\mathbb{P}^4$ solution on a domain slice ($y = 0.7$), located on the dry area.

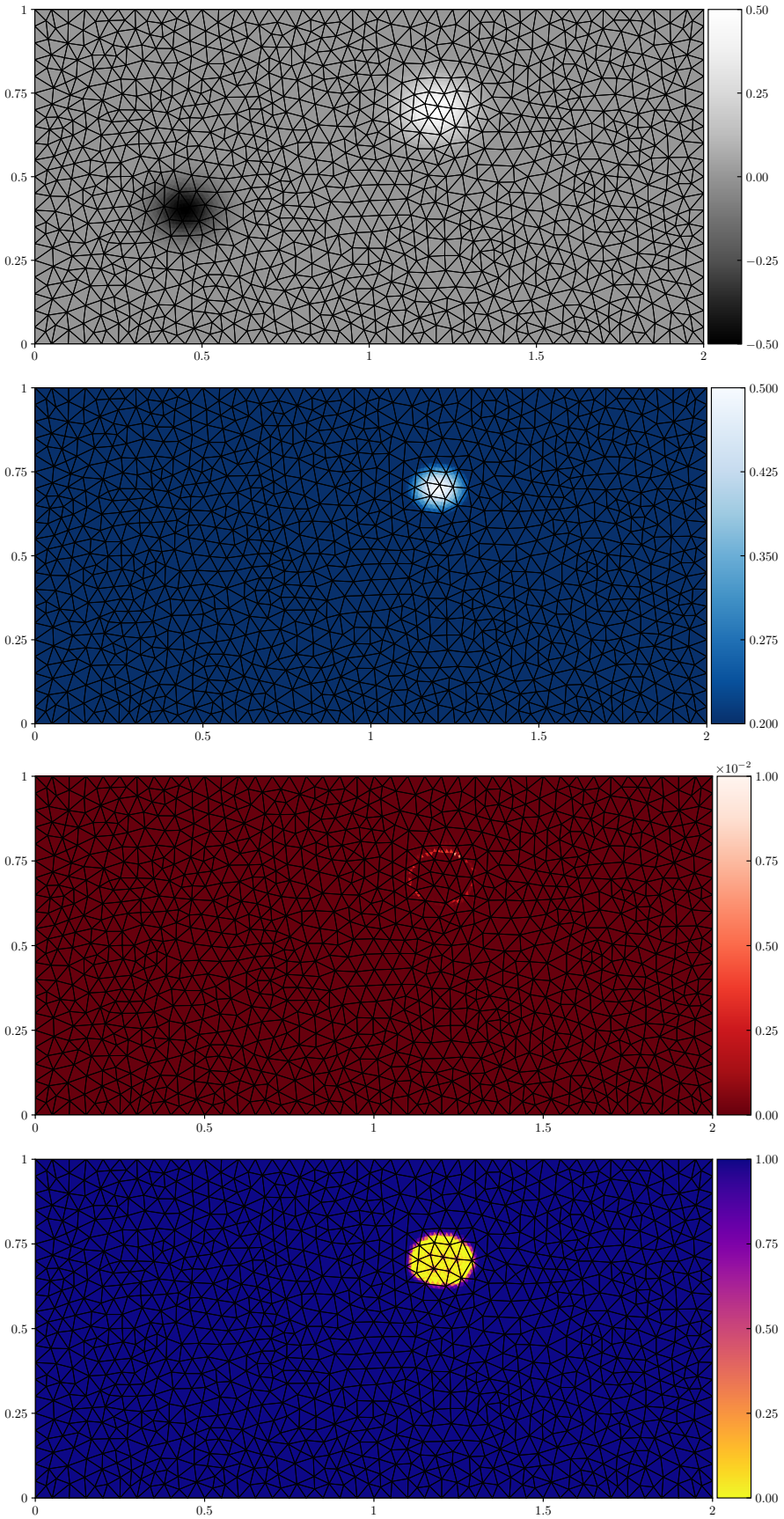


Figure 3.13: Well-balancing with dry area – At $t = 20$ sec, from top to bottom: topography, free-surface elevation, discharge norm, and map of blending coefficient means per subcell.

in both cases, we also run the computations for (very) long times of several physical hours without noticing any significant error increase due to local flux/bathymetry balancing failure.

3.5.3 Dam-breaks in a channel

In this third test-case, we consider extruded 1D dam-break problems in a channel, inducing a directional propagation (along the x-direction), but using general unstructured meshes without anisotropy, providing a (classical) benchmark to evaluate the ability of the numerical methods to handle strong discontinuities or singularities, together with the preservation of the water-height positivity in the limit cases. We set $b(\boldsymbol{x}) = 0$ and investigate both the wet/wet and wet/dry situations, described in the following.

On a wet-bed

PARAMETERS

Computational domain: $\Omega = [0, 1000] \times [0, 200]$
Number of cells: $N_{\text{el}} = 350$ cells
Boundary conditions: solid walls

Polynomial degree: $k = 4$
Time-marching order: $p = 3$
Final time: $T_{\max} = 32$ sec

We consider the initial-data

$$\eta(\boldsymbol{x}, 0) = \begin{cases} 10 & \text{if } x < 500, \\ 7 & \text{otherwise,} \end{cases} \qquad \boldsymbol{q}(\boldsymbol{x}, 0) = \boldsymbol{0},$$

and use a very coarse mesh of only 350 (unstructured) cells, not aligned with the propagation direction, together with an approximate piecewise-polynomial solution of degree $k = 4$. We emphasize that considering such a mesh-size, the sharp variations of the Riemann solution may be entirely located inside one mesh-element (in the lateral direction), as shown in Figures 3.15–3.16 and that even within such a demanding computational configuration, the numerical method still allows to sharply capture the propagating discontinuity and the rarefaction wave. This may be clearly observed in Figure 3.15 with a comparison between exact and numerical solutions on the whole domain, together with the related map of blending coefficients on the subcells, highlighting the efficiency of these *a priori* local corrections, only activated very locally in the vicinity of the solution's singularities, which is an optimal behavior generally observed only with *a posteriori* correction, showing again the practical efficiency of the relaxed subNAD.

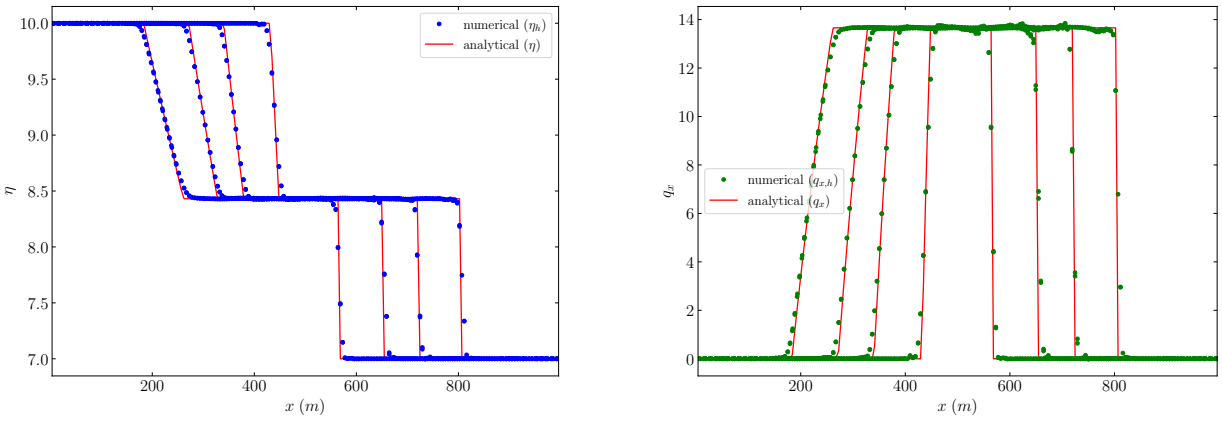


Figure 3.14: Dam break on a wet bed – $\mathbb{P}^4$ free-surface elevation and horizontal discharge between $t = 0$ and $t = 32$ sec on the slice $y = 100$ (animation).

Indeed, the resulting numerical solution is in excellent agreement with the analytical solution, as observed on Fig.3.16 where a sliced solution is plotted, with the corresponding subcells values and compared with the exact solution. We also briefly show on Figure 3.17 that (non-stiff) friction source terms may also be straightforwardly handled in the same way than the bathymetry source term, by further considering an additional Manning friction term, with $\gamma = \frac{10}{7}$ and several values of the Manning coefficient n_f, on the same mesh. Observing a consistent behavior, we choose not to further (quantitatively) investigate approximations of friction source terms and leave it for future works as extensions of [30, 108] for asymptotic parabolic regimes.

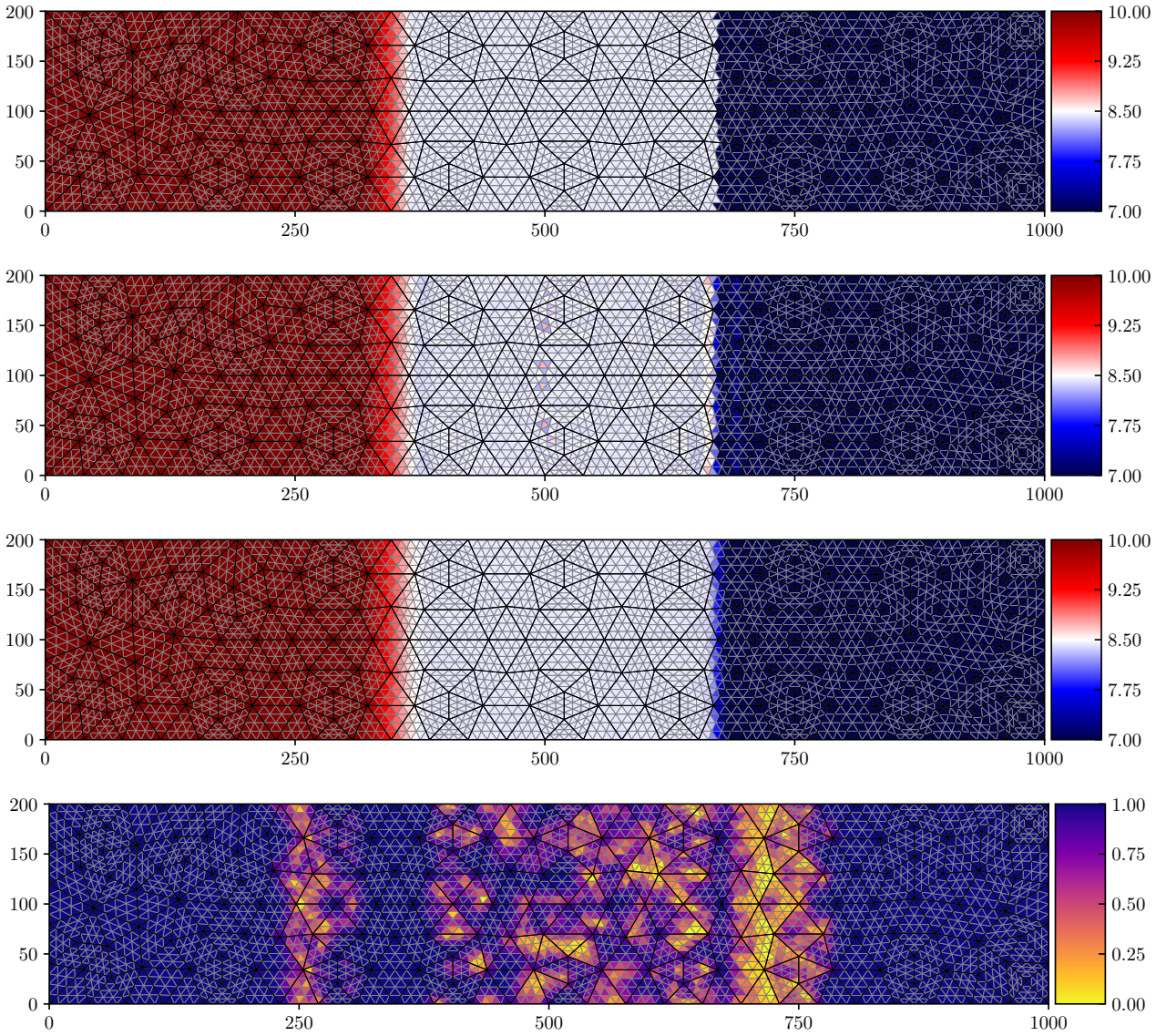


Figure 3.15: Dam break on a wet bed – At $t = 18$ sec, from top to bottom: exact elevation, unlimited DG scheme, monolithic DG/FV scheme and map of blending coefficient means per subcell (animation).

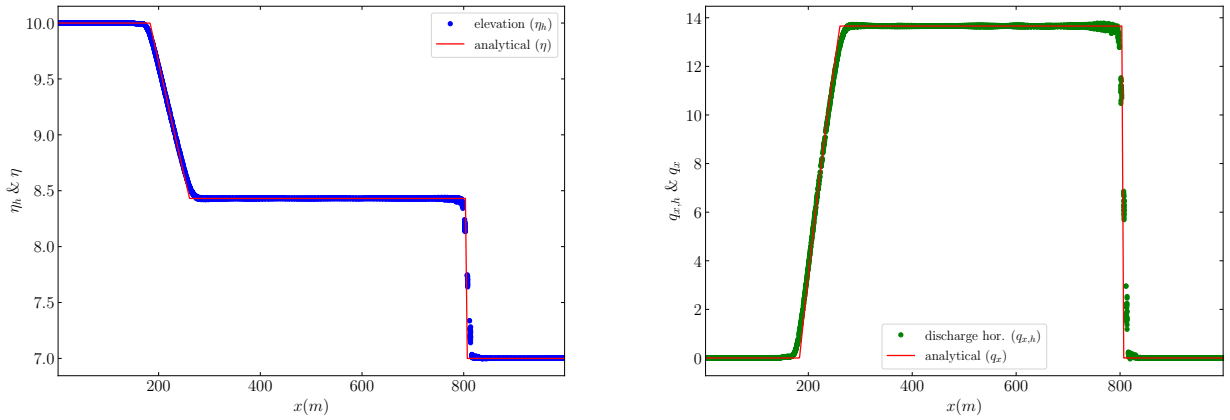


Figure 3.16: Dam break on a wet bed – $\mathbb{P}^4$ free-surface elevation and horizontal discharge density profiles at $t = 32$ sec (animation).

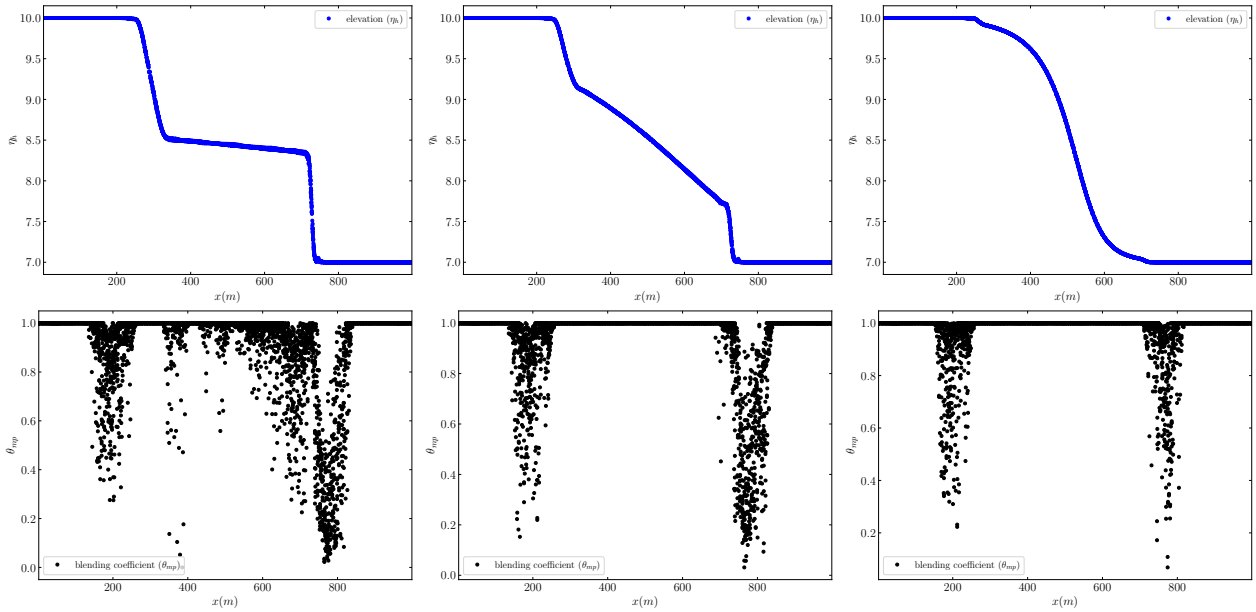


Figure 3.17: Dam break on a wet bed with friction – $\mathbb{P}^4$ free-surface elevation and blending density profiles at $t = 24$ sec for $n_f = 0.5$ (left), 2 (center) and 10 (right).

On a dry-bed

PARAMETERS

Computational domain: $\Omega = [0, 1000] \times [0, 200]$
Number of cells: $N_{\text{el}} = 350$ cells (unstruct.), 240 cells (struct.)
Boundary conditions: solid walls

Polynomial degree: $k = 3$
Time-marching order: $p = 3$
Final time: $T_{\max} = 20, 60$ sec

Now, we set the initial-data as follows:

$$\eta(\boldsymbol{x}, 0) = \begin{cases} 10 & \text{if } x < 500, \\ 0 & \text{otherwise,} \end{cases} \qquad \boldsymbol{q}(\boldsymbol{x}, 0) = \boldsymbol{0},$$

resulting in a dam-break on a dry-bed, challenging the robustness of our method, and we show that the local subcell monolithic DG/FV scheme is able to accurately capture the flow's dynamic, while maintaining accuracy and robustness.

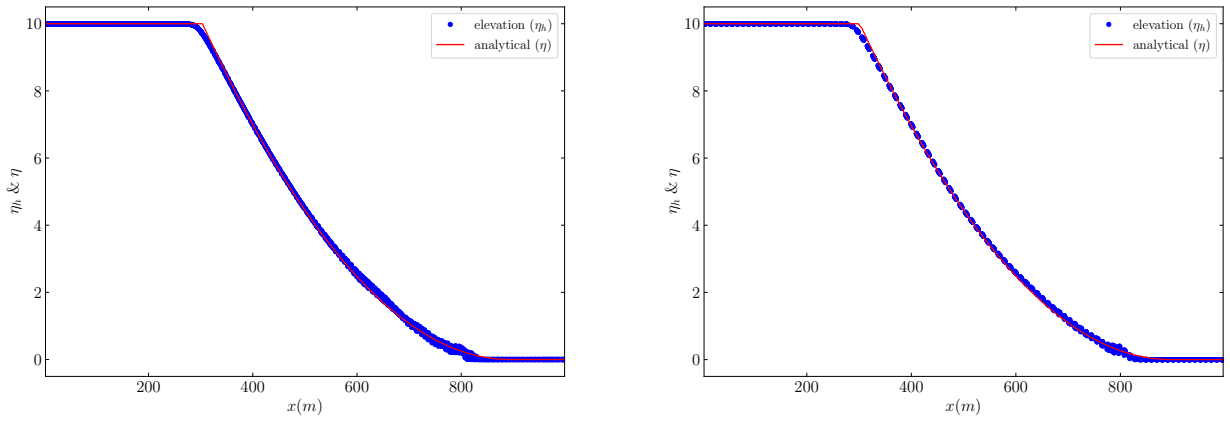


Figure 3.18: Dam break on a dry bed – Density profile of $\mathbb{P}^3$ solution on unstructured (left) and structured (right) meshes.

Again, we consider our coarse unstructured mesh of 350 cells, together with an approximate piecewise-polynomial solution of degree $k = 3$. We emphasize that considering such a large mesh-size, the solution's singularities (wet-dry transitions) are again generally located inside mesh elements, exhibiting sub-mesh resolution capabilities and

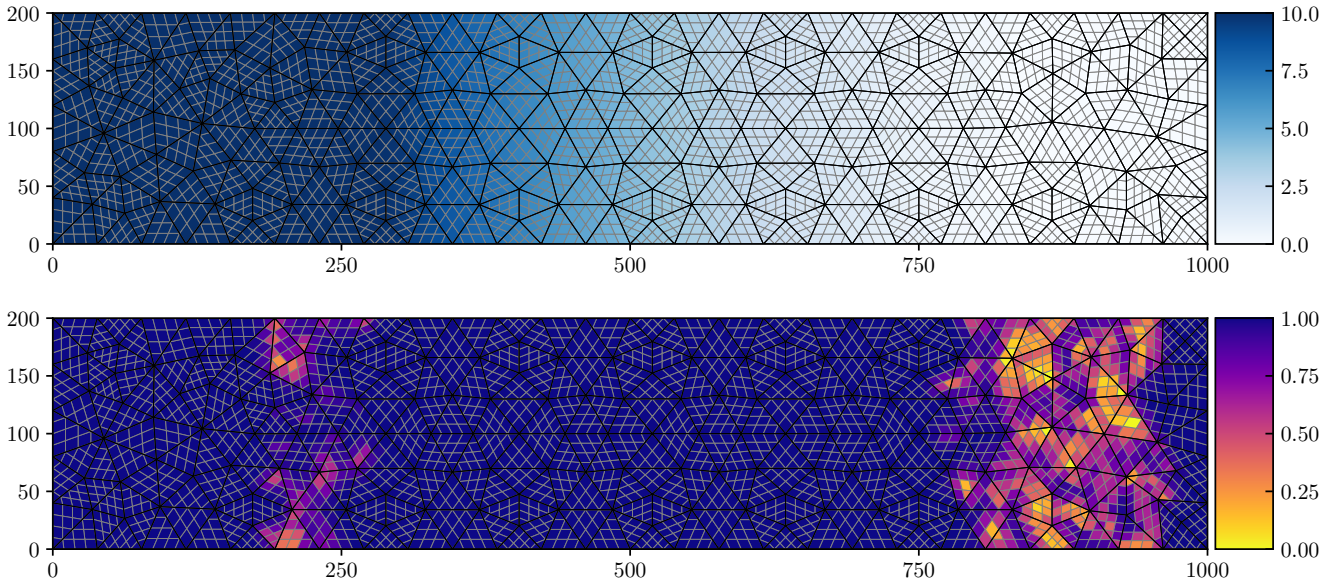


Figure 3.19: Dam break on a dry bed – At $t = 20$ sec, $\mathbb{P}^3$ free-surface elevation and blending map on unstructured mesh.

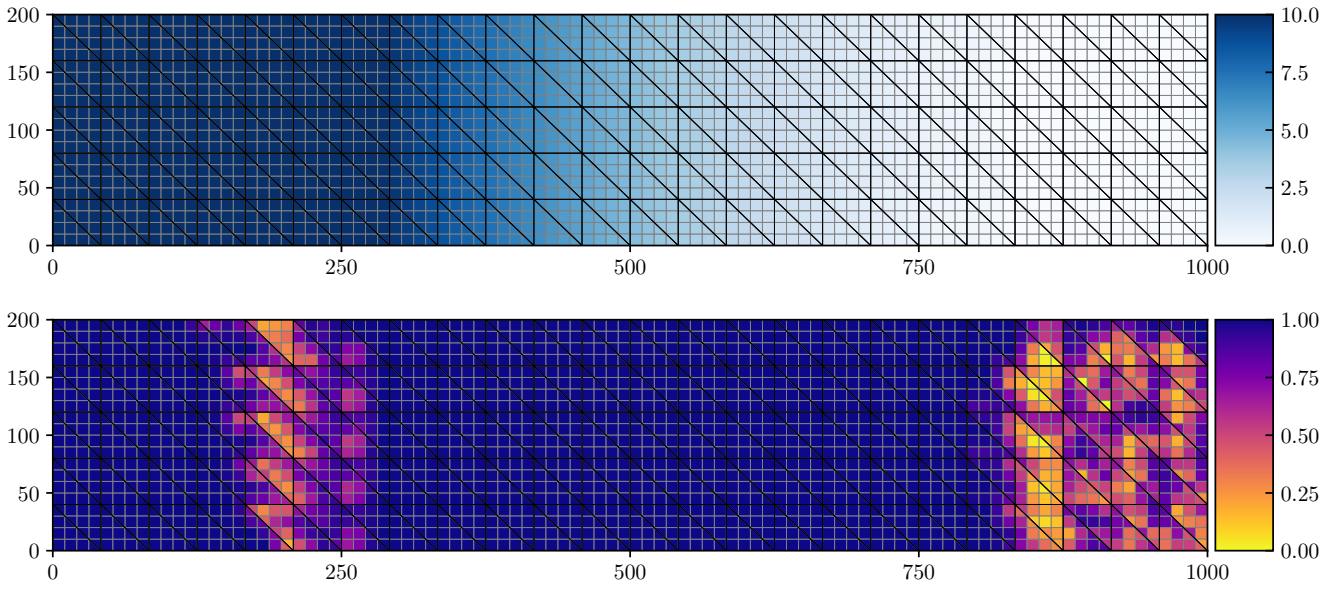


Figure 3.20: Dam break on a dry bed – At $t = 20$ sec, $\mathbb{P}^3$ free-surface elevation and blending map on structured mesh.

real efficiency and true robustness for higher-order approximations. We show in Figures 3.22 the approximate solution obtained at $t = 20$ sec. on the whole domain, together with the related map of blending coefficients on the subcells, assessing the practical efficiency of the PAD. Note that a comparison with the analytical solution is also shown for a sliced solution in the propagation direction in Figure 3.18, showing an excellent agreement, emphasizing again the chosen large mesh-size. To investigate the effect of mesh topology, we compare the approximate results obtained on both structured (240 elements) and unstructured meshes, as shown in Figure 3.19. We observe only marginal differences, which confirms that the chosen corrective mechanisms are activated in similar areas. This demonstrates the robustness and flexibility of the scheme with respect to the mesh choice. A comparison between the analytical solution and numerical results for both mesh types at $t = 20\,s$ is also presented in Figure 3.18. Some additional results are also shown (Figures 3.21 and 3.22) with the linear friction source term.

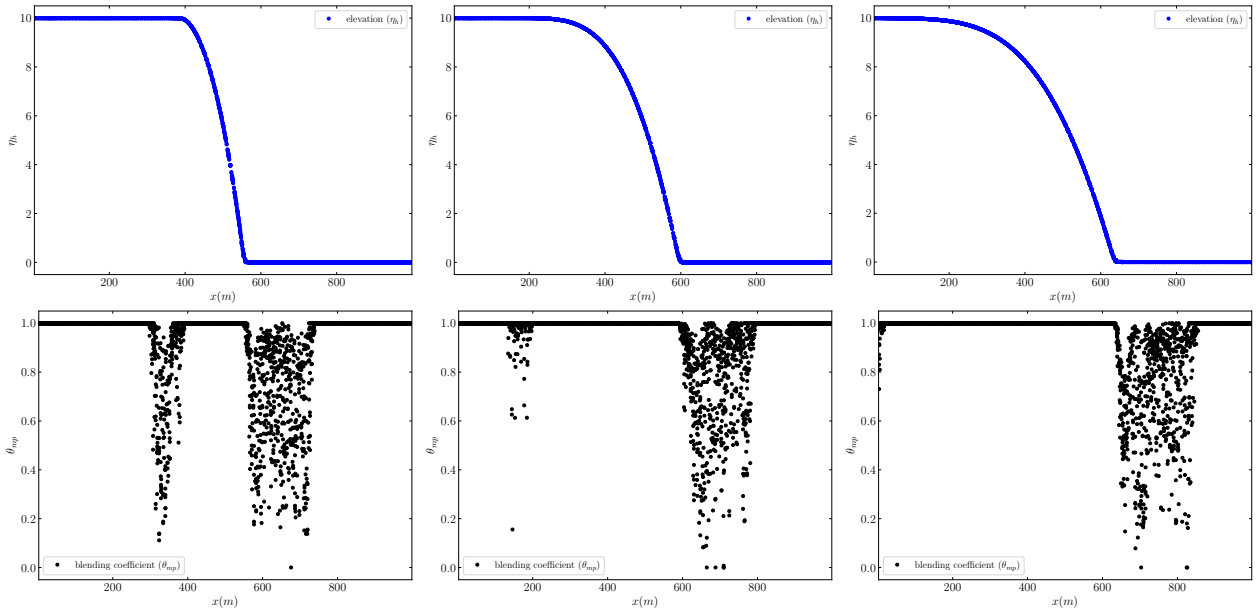


Figure 3.21: Dam break on a dry bed with friction – Snapshots of $\mathbb{P}^3$ free-surface elevation and blending density profiles for $t \in [10, 60]$ sec for $k_f = 0.5$.

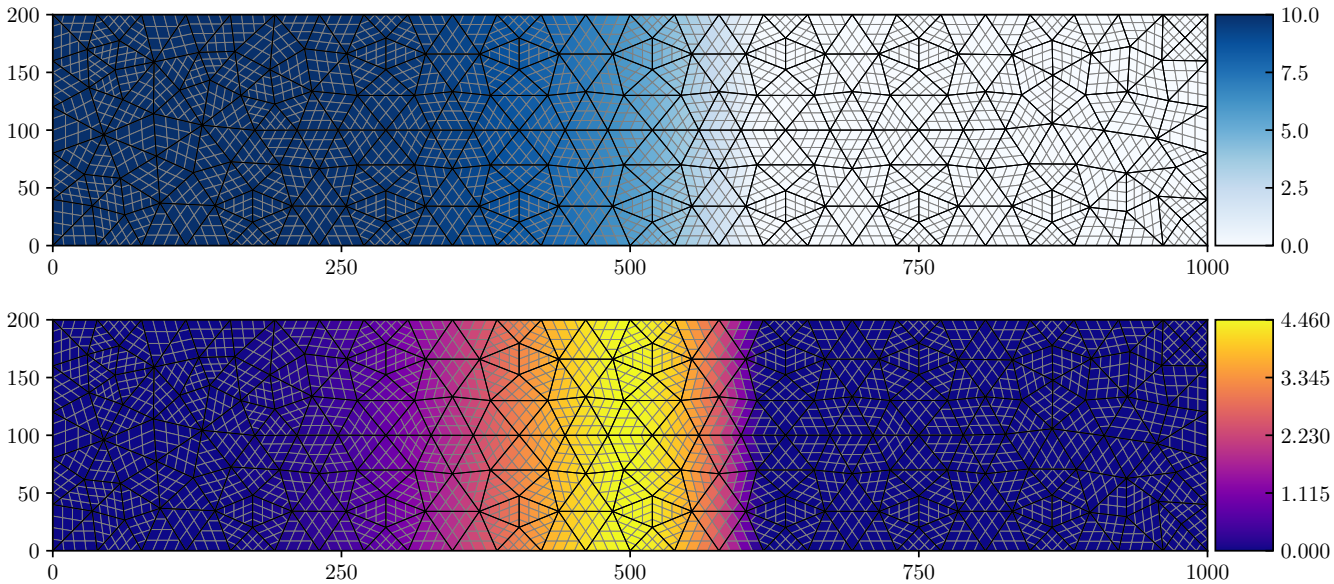


Figure 3.22: Dam break on a dry bed with friction – At $t = 40$ sec, $\mathbb{P}^3$ free-surface elevation (top) and discharge norm (bottom).

3.5.4 Circular dam-breaks

PARAMETERS

Computational domain: $\Omega = \{\boldsymbol{x} \in \mathbb{R}^2 \mid x^2 + y^2 \leq 10^2\}$
Number of cells: $N_{\text{el}} = 870$ cells
Boundary conditions: homogeneous Neumann
Polynomial degree: $k = 4$
Time-marching order: $p = 3$
Final time: $T_{\max} = 1$ sec

In this fourth test-case, we study the dam-break with a radial symmetry, in a circular computational domain, in order to produce a solution with true two-dimensional features and symmetry (again, the topography is set to $b(\boldsymbol{x}) = 0$). Such a circular dam-break problem on a dry bed is a classic yet highly challenging benchmark for evaluating the robustness and accuracy of numerical schemes. Its complexity stems from several critical factors which are combined: (*i*) the transition to a dry bed combined to the radial symmetry preservation (maintaining the circularity of the solution on non-radial grids to assess the scheme's anisotropy), (*ii*) wave focusing and singularities: as the depression wave propagates toward the center, the convergence of flow leads to a geometric focalization. This concentration of energy tests the scheme's ability to handle high gradients and sharp transitions without generating spurious oscillations associated to Gibbs phenomenon, (*iii*) shock-capturing for high-order methods, as the initial discontinuity represents a severe test for high-order methods. In the following, we show that the proposed scheme clearly resolves the resulting complex and combined 2D wave patterns.

On a wet-bed

The initial condition is set as follows:

$$\eta(\boldsymbol{x}, 0) = \begin{cases} 1.5 & \text{if } \sqrt{x^2 + y^2} < 4, \\ 0.5 & \text{otherwise.} \end{cases}, \qquad \boldsymbol{q}(\boldsymbol{x}, 0) = \boldsymbol{0}.$$

Free-surface elevation and discharge norm respectively computed at some intermediate (left) and final (right) times are shown in Figure 3.23, while some one-dimensional sliced profiles obtained along the horizontal and vertical axes are reported in Figure 3.24. Compared against a first-order FV baseline on a very fine grid, our numerical solution preserves excellent radial symmetry throughout the simulation, even on such a coarse unstructured mesh. The circular front remains perfectly isotropic, with no visible grid-deformation effects or 'squaring' of the wave as observed with cartesian meshes. Also, the wave focusing at the center is handled without any spurious oscillations: the peak water-height at the moment of convergence is in close agreement with reference solutions, demonstrating the effectiveness of the proposed method and again, the sub-grid resolution capabilities. Considering the anisotropic unstructured mesh, we observe that the numerical results demonstrate an accurate symmetry in every direction of space, as further confirmed by the diagonal profiles in Figure 3.25.

On a dry-bed

Now, let consider the case of an initially dry area:

$$\eta(\boldsymbol{x}, 0) = \begin{cases} 1.5 & \text{if } \sqrt{x^2 + y^2} < 4, \\ 0 & \text{otherwise,} \end{cases}, \qquad \boldsymbol{q}(\boldsymbol{x}, 0) = \boldsymbol{0}.$$

We show on Figure 3.26 the free-surface elevation at final time on the computational domain (left), together with the corresponding colormap of the blending coefficients, automatically adapted to ensure stability. Again, we observe some very satisfying results, as confirmed on Figure 3.27 where the free-surface elevation (left) and related blending coefficients (right) at $t = T_{\max}$ on slice $y = 0$ are displayed. The scheme demonstrates strict positivity-preserving properties at the wet/dry interface. No non-physical negative water-depths were observed, and the wetting-drying front is captured sharply without the need for excessive *a priori* corrections.

3.5.5 Carrier-Greenspan solutions

Transient solution

PARAMETERS

Computational domain: $\Omega = [-20, 6] \times [0, 4]$
Number of cells: $N_{\text{el}} = 1092$ cells
Boundary conditions: prescribed Dirichlet (left), homogeneous Neumann (right)
Polynomial degree: $k = 2$
Time-marching order: $p = 1$
Final time: $T_{\max} = 100$ sec

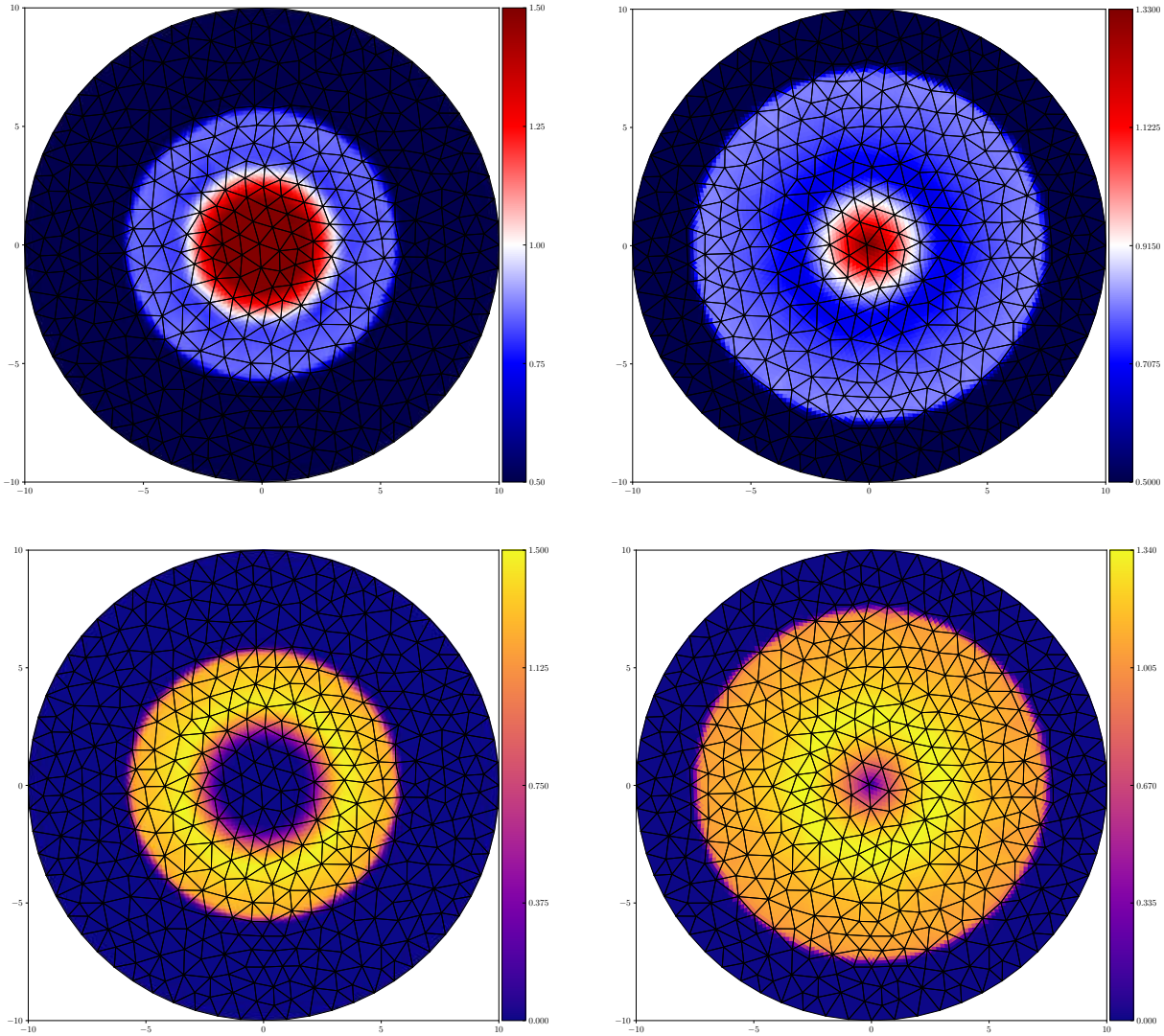


Figure 3.23: Circular dam break on a wet bed – $\mathbb{P}^4$ free-surface elevation (top) and discharge norm (bottom) at $t = \frac{T_{\max}}{2}$ (left) and at $t = T_{\max}$ (right) (animation).

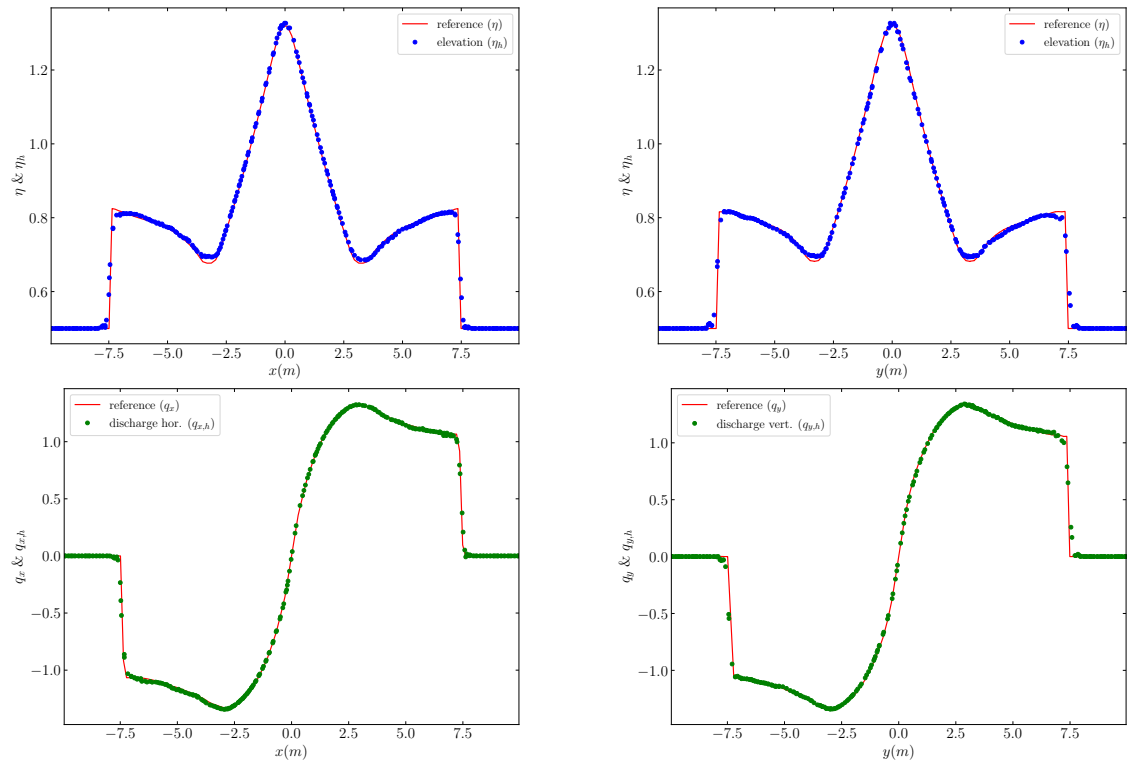


Figure 3.24: Circular dam-break on a wet-bed – $\mathbb{P}^4$ free-surface elevation at $t = T_{\max}$ on slice $y = 0$ (resp. $x = 0$) and horizontal (resp. vertical) discharge (animation).

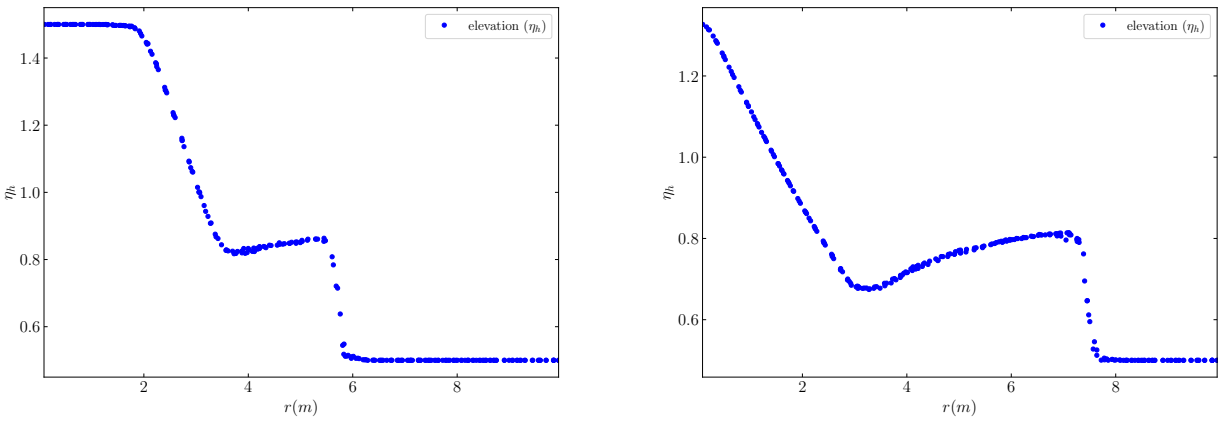


Figure 3.25: Circular dam-break on a wet-bed – $\mathbb{P}^4$ free-surface elevation at $t = \frac{T_{\max}}{2}$ (left) and at $t = T_{\max}$ (right) on $\{(x, y) \in \Omega \mid y = \frac{\pi x}{3}$ and $x, y \geq 0\}$ (animation).

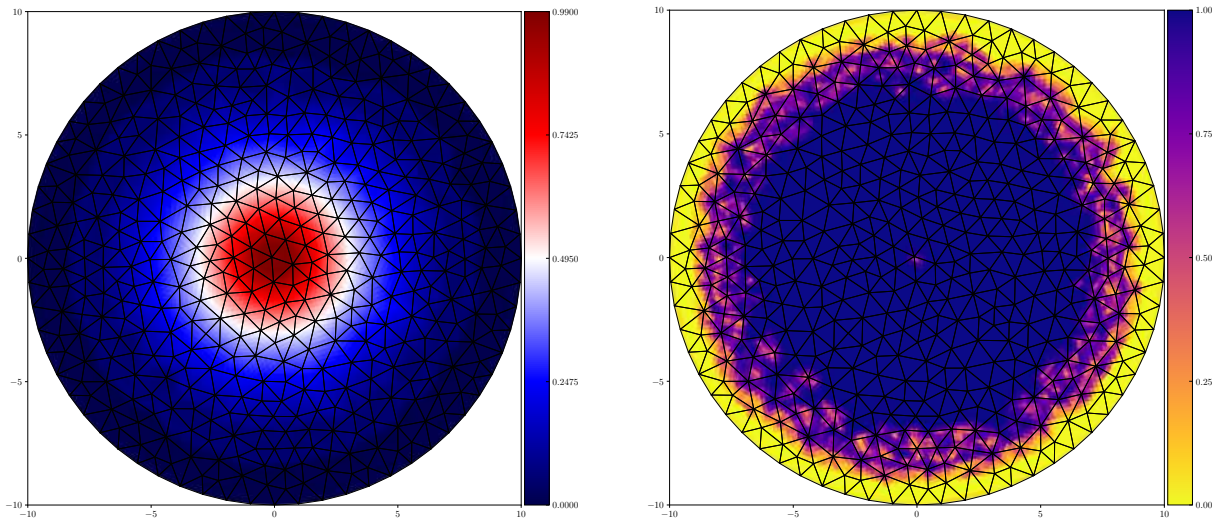


Figure 3.26: Circular dam break on a dry bed – $\mathbb{P}^4$ free-surface elevation (left) and blending coefficients (right) at $t = T_{\max}$ (right).

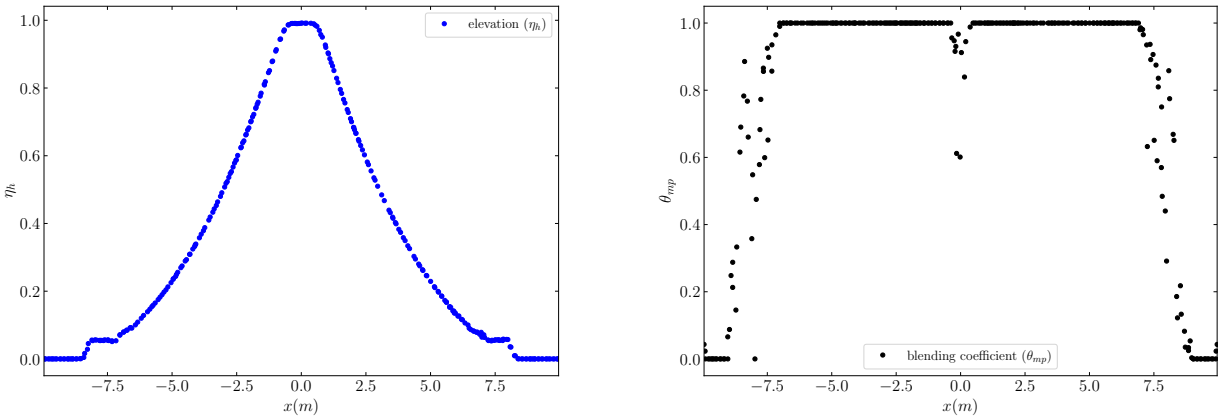


Figure 3.27: Circular dam break on a dry bed – $\mathbb{P}^4$ free-surface elevation (left) and blending coefficients (right) at $t = T_{\max}$ on slice $y = 0$.

In order to combine moving wet-dry contact lines with varying bathymetry, we consider the test-case originally introduced by Carrier in [54], modeling evolutions of the surface-elevation near the shoreline of a sloping beach, from an initially depressed and held stationary configuration, before being released at $t = 0$. This setup generates a transient wave that moves up the beach and eventually returns to equilibrium through a slow convergence process. It offers an excellent benchmark to assess the robustness of the monolithic DG/FV scheme when dealing with long-wave run-up problems.
In [54], an analytical solution to the NSW equations is obtained using a hodograph transformation, introducing two dimensionless variables, σ^* and λ^*, which act respectively as space-like and time-like coordinates:

$$\sigma^* = 4c^*, \quad \lambda^* = 2\left(u^* + t^*\right).$$

Let l denote the characteristic length and α the beach slope. The dimensionless variables are defined using the following scaling:

$$x^* = x/l, \quad \eta^* = \eta/(\alpha l), \quad u^* = u/\sqrt{g\alpha l}, \quad t^* = t/\sqrt{l/(\alpha g)}, \tag{3.40}$$

and the non-dimensional phase speed is given by

$$c^* = \sqrt{\eta^* - x^*}. \tag{3.41}$$

The initial condition for the solution is specified as:

$$\eta_0^*(\sigma^*) = e\left(1 - \frac{5}{2}\frac{a^3}{(a^2+\sigma^{*2})^{\frac{3}{2}}} + \frac{3}{2}\frac{a^5}{(a^2+\sigma^{*2})^{\frac{5}{2}}}\right), \qquad \boldsymbol{q}_0^*(\sigma^*, \cdot) = \mathbf{0}, \qquad x^* = -\frac{\sigma^{*2}}{16} + \eta_0^*,$$

where $a = \frac{3}{2}(1 + 0.9e)^{1/2}$, and e is a small parameter characterizing the initial surface depression.
The analytical solution is given by:

$$\begin{cases} \eta^*(\sigma,\lambda) = -\dfrac{u^{*2}}{2} + e\,\Re\left[1 - 2\dfrac{\frac{5}{4} - i\lambda}{[(1-\mathrm{i}\lambda)^2+\sigma^2]^{3/2}} + \dfrac{3}{2}\dfrac{(1-\mathrm{i}\lambda)^2}{[(1-\mathrm{i}\lambda)^2+\sigma^2]^{5/2}}\right], \\ u^*(\sigma,\lambda) = \dfrac{8e}{a}\,\Im\left[\dfrac{1}{[(1-\mathrm{i}\lambda)^2+\sigma^2]^{3/2}} - \dfrac{3}{4}\dfrac{1-\mathrm{i}\lambda}{[(1-\mathrm{i}\lambda)^2+\sigma^2]^{5/2}}\right], \\ t^* = \dfrac{1}{2}a\lambda - u^* \qquad \text{and} \qquad x^* = \eta^* - \dfrac{a^2\sigma^2}{16}, \end{cases}$$

where the scaled variables $\sigma^* = a\sigma$ and $\lambda^* = a\lambda$ are used. In the following, we choose $e = 0.1$, $\alpha = 1/50$, the initial free-surface profile is expressed in dimensional form, with $l = 20$ m and the analytical variation of the free-surface elevation at the offshore boundary is prescribed as the inlet boundary condition to drive the motion. The evolution of the free-surface during the transient phase is displayed in Figure 3.28. At $t = 10$ sec, a direct comparison between the analytical and numerical elevations is shown in Figure 3.29, while the horizontal discharge profiles and the final elevation profile are illustrated in Figure 3.30. The numerical results for the Carrier-Greenspan test case demonstrate the significant advantages of the high-order local subcell monolithic DG/FV scheme. Even when operating on such a coarse unstructured mesh, the scheme accurately captures the nonlinear oscillation and the precise movement of the shoreline. The high-order sub-grid representation within each element compensates for the limited spatial resolution of the global mesh, preserving the analytical wave profile with minimal dissipation. This performance highlights the ability of the method to resolve complex run-up and draw-down dynamics on irregular geometries without the numerical damping typically associated with low-order approximations on similar under-resolved grids.

Periodic solution

PARAMETERS

Computational domain: $\Omega = [-20, 6] \times [0, 4]$
Number of cells: $N_{\text{el}} = 1092$ cells
Boundary conditions: prescribed Dirichlet (left), homogeneous Neumann (right)
Polynomial degree: $k = 2$
Time-marching order: $p = 1$
Final time: $T_{\max} = 300$ sec

This benchmark models the run-up and run-down of a monochromatic wave propagating over a plane beach. The dimensionless amplitude of the periodic wave is denoted by A^* and its frequency by ω^*. As the wave moves

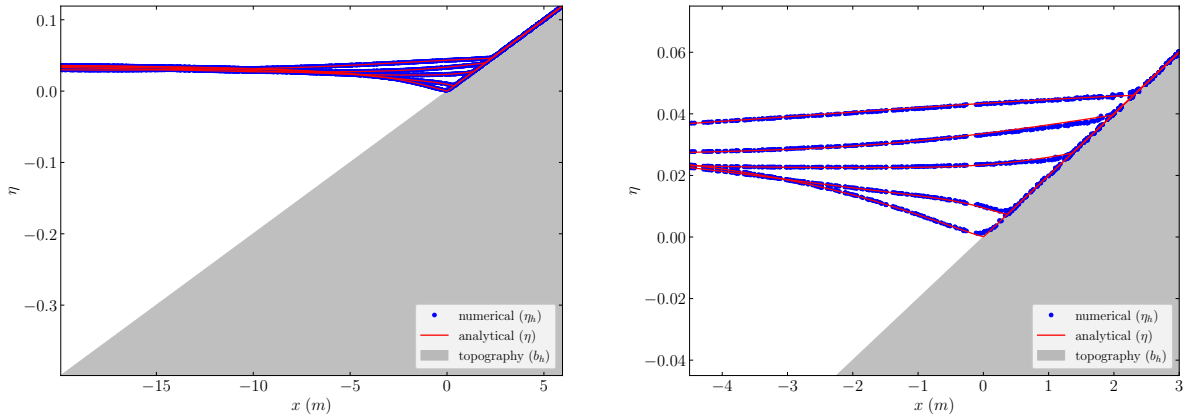


Figure 3.28: Carrier & Greenspan transient solution − Snapshots of $\mathbb{P}^2$ elevation density profile on full domain (left) and zoom on dry bed (right) for $t \in [1, 21]$ sec.

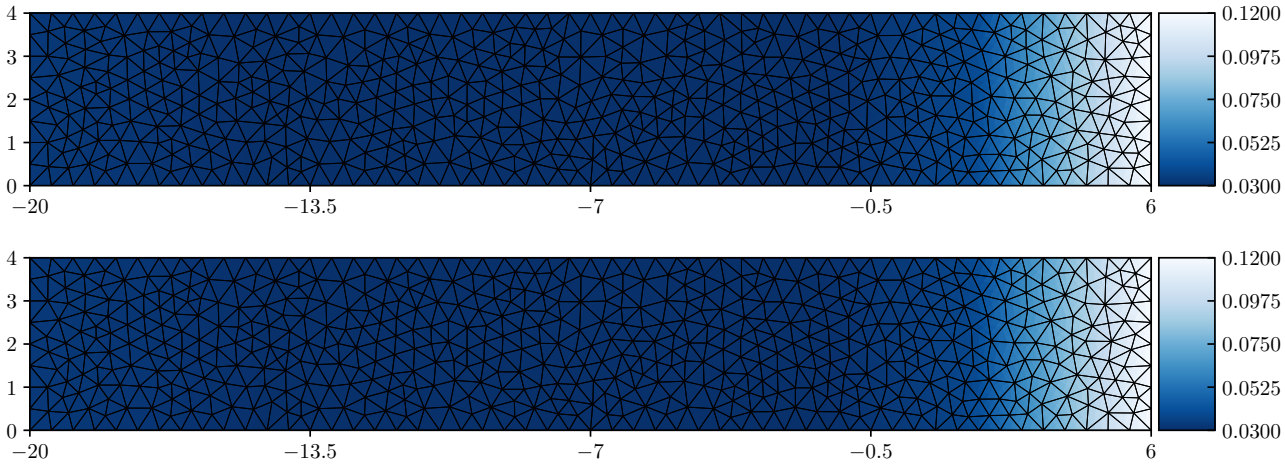


Figure 3.29: Carrier & Greenspan transient solution − At $t = 10$ sec, exact elevation (top) and $\mathbb{P}^2$ numerical elevation (bottom).

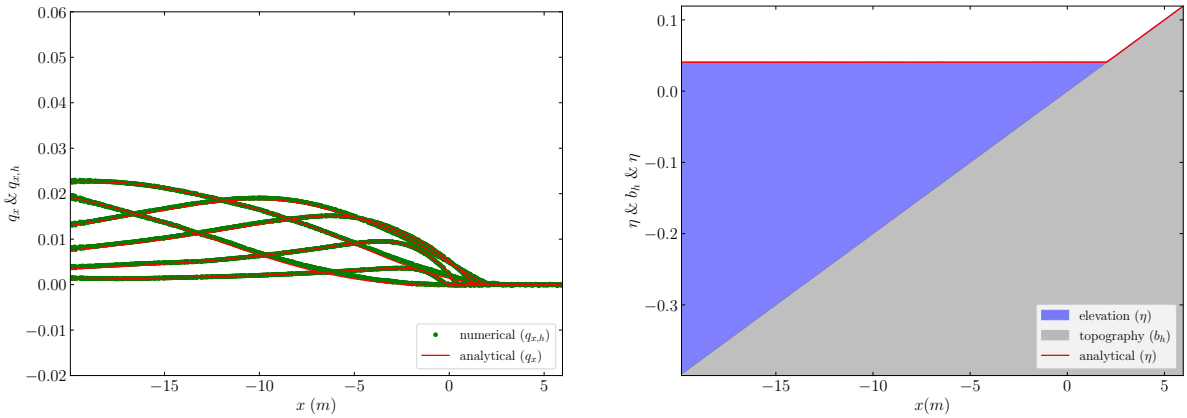


Figure 3.30: Carrier & Greenspan transient solution − Snapshots of $\mathbb{P}^2$ horizontal discharge density profile for $t \in [1, 21]$ sec (left) and elevation density profile at final time (right).

toward the shore and reflects back into the sea, it generates a standing wave-pattern over the sloping bottom. Using the dimensionless variables introduced in (3.40) and (3.41), the analytical solution is given by:

$$\begin{cases} u^* = -\dfrac{A^* J_1(\sigma^*)\sin(\lambda^*)}{\sigma^*}, \\ \eta^* = \dfrac{A^*}{4} J_0(\sigma^*)\cos(\lambda^*) - \dfrac{u^{*2}}{4}, \\ t^* = \dfrac{1}{2}\lambda^* - u^* \quad \text{and} \quad x^* = \eta^* - \dfrac{\sigma^{*2}}{16}, \end{cases}$$

where J_0 and J_1 denote the Bessel functions of the first kind of order zero and one, respectively. We set $A^* = 0.6$ and $\omega^* = 1$, corresponding to a non-breaking wave, with a reference length $l = 20$ m and a beach slope of $\alpha = 1/30$. The solution at $t = 0$ is used as the initial-data, and, similarly to the previous transient case, the analytic solution at the offshore boundary is prescribed as the inlet boundary condition, driving the motion. For a detailed description, we refer again to [54]. The simulation is carried out up to $T_{\max} = 1.5\,T$, where T is the period of the external forcing. The spatio-temporal evolution of the free-surface is illustrated in Figure 3.31.

The quality of the results is once again clear, showing an excellent agreement with the expected wave run-up

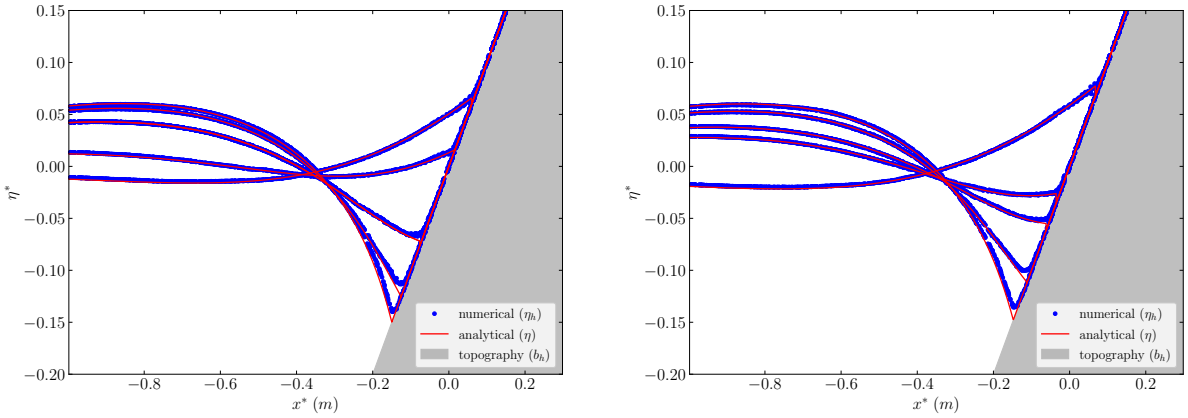


Figure 3.31: Carrier & Greenspan periodic solution − Snapshots of $\mathbb{P}^2$ elevation density profile for $t \in [80, 86]$ sec (left) and for $t \in [154, 160]$ sec (right) (animation).

and run-down patterns. The local subcell monolithic DG/FV scheme approximation effectively resolves the nonlinear transformation of the wave profile as it approaches the shore, maintaining a sharp and oscillation-free representation of the moving boundary. Even throughout multiple oscillation cycles, the numerical solution shows no sign of amplitude decay or phase shift, confirming the low-dissipative nature of the high-order scheme. This highlights the robustness of the wet/dry front treatment in capturing the periodic shoreline migration over the sloping beach. Notably, these results are achieved using a very coarse and non-matching mesh, reflecting the demanding conditions typically encountered in operational engineering contexts. The ability of the scheme to maintain high fidelity under such under-resolved constraints further underscores its suitability for practical large-scale applications. As a first step towards such applications, we investigate some more demanding test-cases in the remainder of this section.

3.5.6 Wave-island interactions

Single-wave on a conical island

PARAMETERS

Computational domain: $\Omega = [0, 30] \times [0, 25]$
Number of cells: $N_{\text{el}} = 3790$ cells
Boundary conditions: solid walls (top/bottom), Dirichlet (left/right)
Polynomial degree: $k = 2$
Time-marching order: $p = 3$
Final time: $T_{\max} = 12$ sec

In this subsection, we investigate a test case directly inspired by the experimental setup of Liu et al. [196]. It consists of a solitary wave propagating and running up over an idealized conical island within a rectangular basin. The wave possesses sufficient initial amplitude and energy to produce significant run-up on both the front and lee sides of the island. Consequently, this test case evaluates the robustness of the numerical method under challenging and physically realistic conditions.

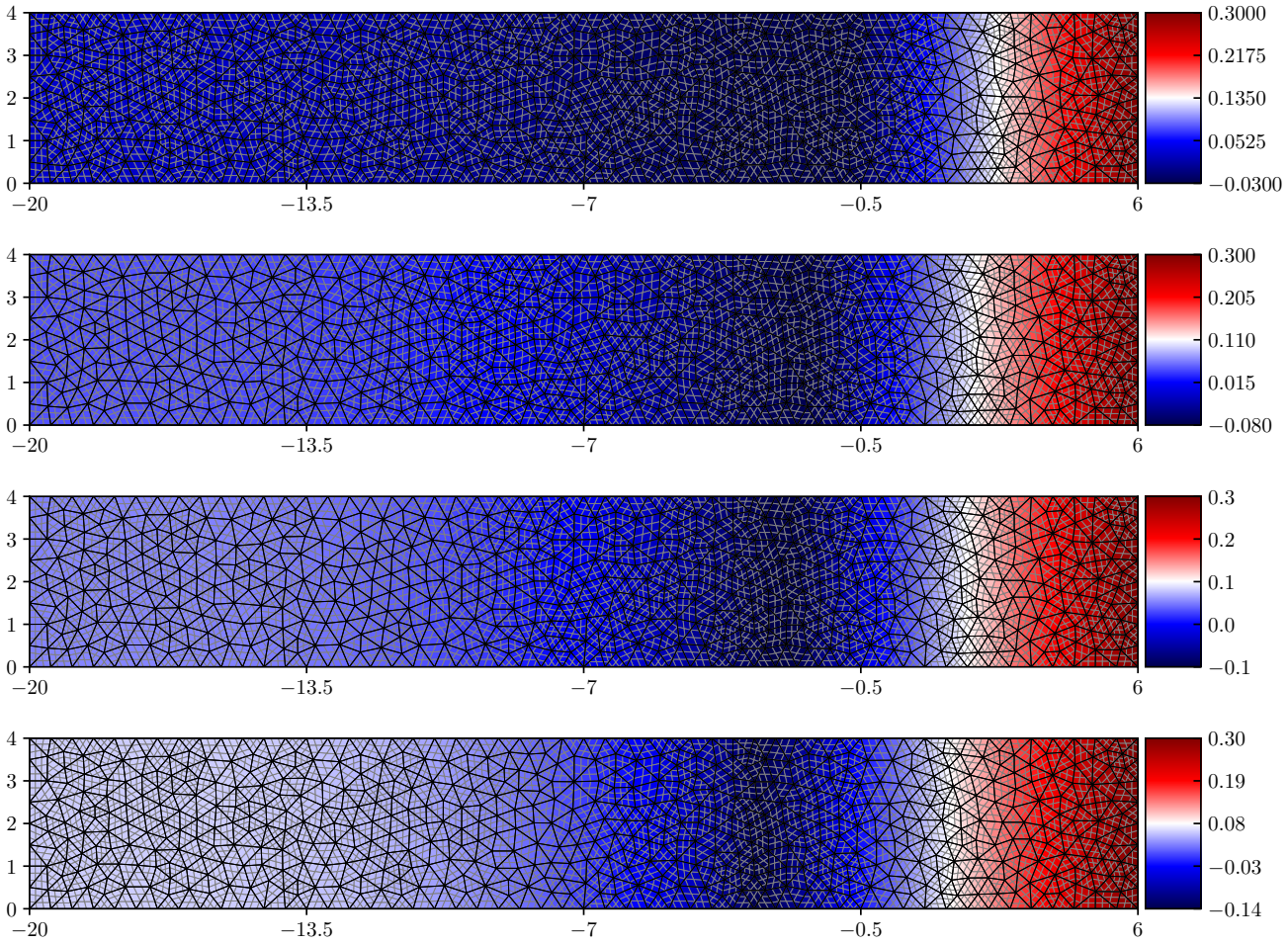


Figure 3.32: Carrier & Greenspan periodic solution – Snapshots of $\mathbb{P}^2$ free-surface elevation for $t \in [154, 160]$ sec (animation).

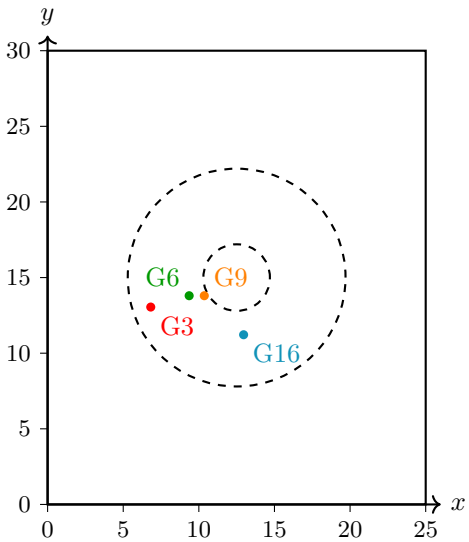


Figure 3.33: Single-wave on a conical island – Experimental setup : computational domain with island geometry and gauge locations.

Experimental data are available from a series of gauges distributed around the island. In the laboratory experiments, the solitary wave shoals and breaks on the front slope, with the peak impact recorded at gauge 9 ($t \approx 5.6$ s). The wave energy is subsequently partitioned and refracted around the island, leading to a secondary run-up on the rear side at gauge 16 ($t \approx 9.7$ s), while a circular reflected front propagates away from the obstacle. Subsequent wave splitting and backwash processes generate secondary oscillations, particularly in the sheltered region behind the island. These phenomena have been numerically reproduced in numerous studies and serve as a standard benchmark for model validation.
Following the original experimental description, the island bathymetry $b(r)$ is modeled as:

$$b(r) = \begin{cases} \max\left(0.625,\, 0.9 - \frac{r}{4}\right), & r < 3.6, \\ 0, & \text{otherwise}, \end{cases} \tag{3.42}$$

where r denotes the radial distance from the center of the island. The quiescent water depth is set to $h_0 = 0.32$. The incident wave is generated via a time-dependent boundary condition:

$$\begin{cases} h = h_0 + \alpha h_0 \operatorname{sech}^2\left(\dfrac{\sqrt{g h_0}}{L}\,\chi t\right), \\ \chi = \sqrt{\dfrac{3\alpha}{4\beta}(1+\alpha)}, \quad \beta = \left(\dfrac{h_0}{L}\right)^2, \quad L = 15, \quad \alpha = 0.3, \end{cases} \tag{3.43}$$

although there is no consensus concerning this standard solitary wave forcing. A schematic of the computational setup and gauge locations is provided in Figure 3.33, while the comparison between numerical results and experimental data is presented in Figure 3.34. In the present study, we perform simulations using polynomial orders

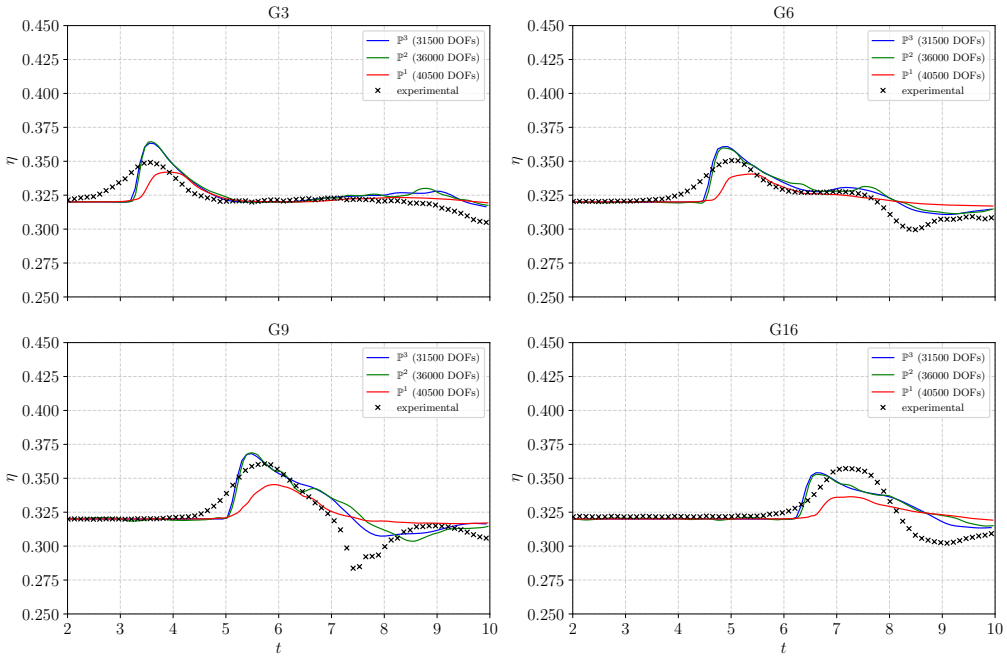


Figure 3.34: Single-wave on a conical island – Numerical/experimental comparisons of the time-series of the surface-elevation at the different gauges for $t \in [2, 10]$ sec.

$\mathbb{P}^1$ (13,500 cells), $\mathbb{P}^2$ (6,000 cells), and $\mathbb{P}^3$ (3,150 cells). The resulting free-surface elevations are compared with experimental time-series at selected gauges, as shown in Figure 3.34. The good agreement between numerical results and experimental data confirms the ability of the solver to capture the primary features of the wave–island interaction. This includes the breaking-induced run-up on the front slope and the delayed response at the rear gauges, demonstrating that the contact-line velocity around the island is accurately estimated. While higher-order schemes notably improve both phase and amplitude accuracy, some discrepancies persist in the backwash oscillations. These differences are classically attributed to dispersive effects, which are not accounted for in the hyperbolic Nonlinear Shallow Water (NSW) equations.
We emphasize that to maintain a consistent computational resolution, the mesh resolution is adjusted such that

the total number of degrees of freedom remains within the same order of magnitude as the polynomial degree increases. Despite a slight reduction in the actual number of DOFs for higher orders, the $\mathbb{P}^3$ scheme yields the most accurate results. We observe a particularly significant improvement when moving from $\mathbb{P}^1$ to $\mathbb{P}^2$, while the gain between $\mathbb{P}^2$ and $\mathbb{P}^3$ is more incremental, but entails a substantial reduction in the number of mesh elements.

It is worth noting that these results are obtained on meshes significantly coarser than those typically reported in the literature, where simulations often involve one to two orders of magnitude more elements. Even without local mesh refinement around the island or in the vicinity of the gauges, the present method successfully reproduces the primary wave characteristics observed in the experiments. This highlights the ability of high-order schemes to provide accurate and robust solutions on relatively modest computational grids, while suggesting that further improvements could be achieved with local refinement.

In particular, an artificial viscosity coefficient could be introduced, as is common practice in industrial codes. Furthermore, adopting a dispersive model would likely better capture the complex wave interactions, especially in the sheltered region behind the island. However, the objective here was primarily to evaluate the scheme in a standard, "ready-to-use" form. Finally, it should be noted that, to the authors' knowledge, a definitive consensus on the optimal setup for this benchmark has yet to be established—and this study does not claim to provide one.

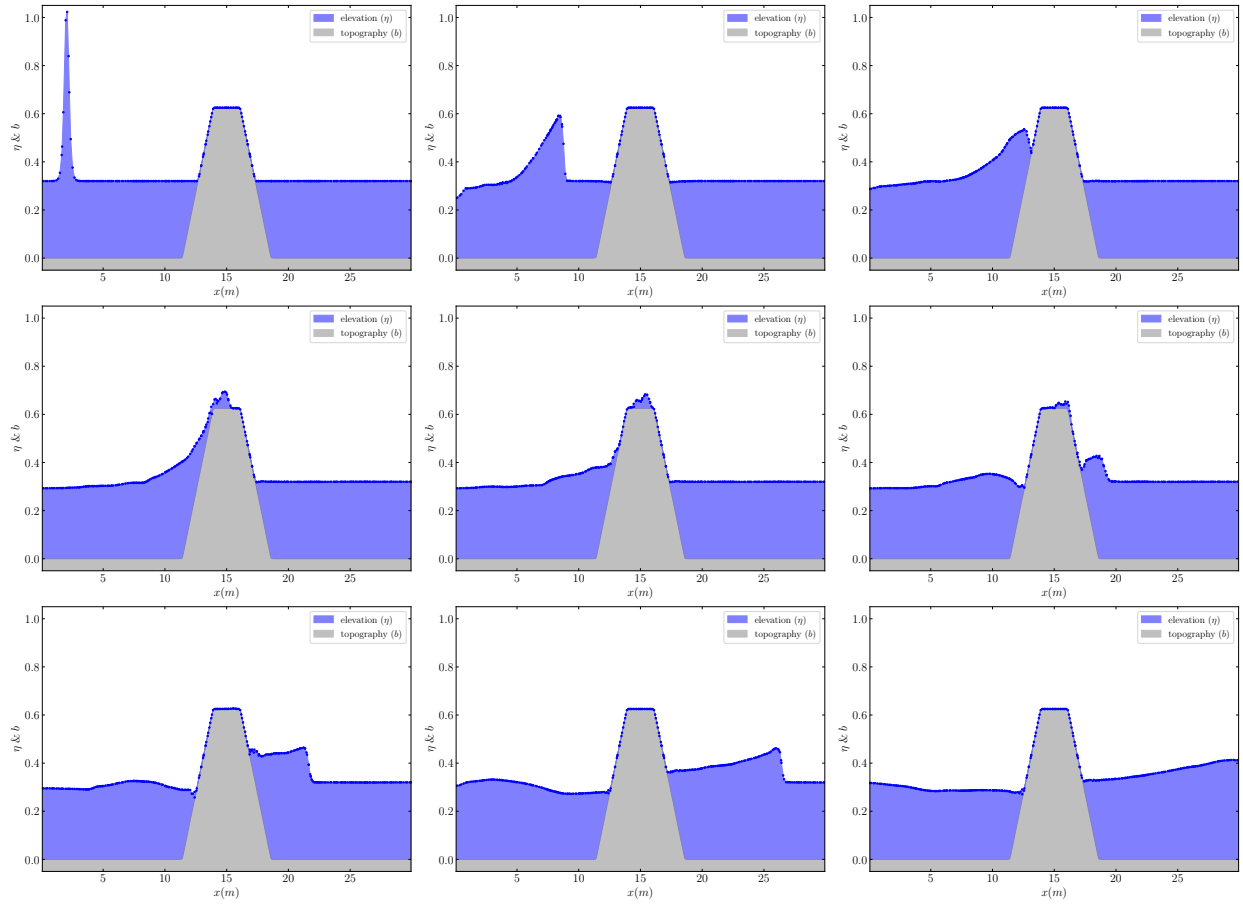


Figure 3.35: Tidal-wave overtopping the conical island − Snapshots of $\mathbb{P}^2$ free-surface elevation on slice $y = 12.5$ for $t \in [0, 12]$ sec.

Tidal-wave overtopping the conical island

PARAMETERS

Computational domain: $\Omega = [0, 30] \times [0, 25]$
Number of cells: $n_{\text{el}} = 3790$ cells
Boundary conditions: solid walls (top/bottom), Dirichlet (left/right)

Polynomial degree: $k = 2$
Time-marching order: $p = 3$
Final time: $T_{\max} = 12$ sec

We now quantitatively investigate the effect of a higher wave amplitude by reproducing the configuration recently proposed in [132]. Unlike the previous benchmark, this case features a traveling wave with sufficient initial amplitude and energy to fully overtop the island, thereby demonstrating the robustness of the method under even more dynamic and challenging conditions. We briefly recall the specific parameters for this setup, which involves slight modifications from the previous case, with the bottom topography defined as:

$$b(\boldsymbol{x}) := \min\left(H_{\text{top}}, \max\left(H_{\text{cone}} - \frac{r(\boldsymbol{x})}{s_{\text{cone}}}, 0\right)\right), \qquad r(\boldsymbol{x}) := \sqrt{(x-15)^2 + (y-12.5)^2}.$$

Here, $H_{\text{top}} = 0.625$, $H_{\text{cone}} = 0.9$, and $s_{\text{cone}} = 4$ (all dimensions in meters). The initial-data is chosen to model a solitary wave approaching the island from the left. The initial water-height and horizontal velocity are defined as:

$$H(\boldsymbol{x}, 0) := \max\left(H_0 + \frac{A}{\cosh^2\left(\sqrt{\frac{3A}{4H_0^3}}(x - x_s)\right)} - b(\boldsymbol{x}), 0\right),$$
$$u_x(\boldsymbol{x}, 0) := \frac{A}{\cosh^2\left(\sqrt{\frac{3A}{4H_0^3}}(x - x_s)\right)}\sqrt{\frac{\mathrm{g}}{H_0}},$$

with the parameters $h_0 = 0.32$, $A = 2.5h_0$, and $x_s = 2.04$. The initial momentum is subsequently defined as $\boldsymbol{q}(\boldsymbol{x}, 0) = (u_x(\boldsymbol{x}, 0)H(\boldsymbol{x}, 0), 0)^\top$. The time evolution of the free-surface elevation along the horizontal cross-section at $y = 12.5$ is presented in Figure 3.39. A complementary view of the two-dimensional free-surface elevation and the corresponding blending maps at various stages of the simulation is provided in Figure 3.36. For this final test case, we emphasize that only 3,700 mesh elements are employed. This clearly demonstrates the ability of the proposed method to ensure both accuracy and robustness on very coarse meshes, highlighting its suitability for operational purposes, where bathymetry and computational grids are often constrained by low-resolution experimental measurements.

Single-wave on a Gaussian-shaped rock

PARAMETERS

Computational domain: $\Omega = [5, 25] \times [0, 30]$
Number of cells: $N_{\text{el}} = 584$ cells
Boundary conditions: solid walls
Polynomial degree: $k = 6$
Time-marching order: $p = 3$
Final time: $T_{\max} = 5$ sec

We now turn to another particularly challenging two-dimensional test case, involving the propagation of a solitary wave over a fixed solid obstacle, with a Gaussian-shaped topography ad a very high-order polynomial degree. The computational domain contains the following Gaussian-shaped topography representing the rock:

$$b(x, y) = 5e^{-\left(0.4(x-15)^2 + 0.2(y-15)^2\right)}.$$

At initial time, the solitary wave is defined by:

$$\eta_0(x, y) = H_0 + A\,\mathrm{sech}\left(\gamma(2y - y_1)\right) \qquad \text{and} \qquad \boldsymbol{q}_0(x, y) = \begin{pmatrix} 0 \\ 0.3\sqrt{\mathrm{g}\left(\eta_0(x, y) - H_0\right)} \end{pmatrix},$$

where $\gamma = \sqrt{\dfrac{A}{600H_0}}$ and $y_1 = 180$. The simulation is performed with $A = 2\,\mathrm{m}$ and $H_0 = 2\,\mathrm{m}$.
Again one can clearly observe the satisfying interaction between the solitary wave and the rock. As the wave approaches the obstacle, its front steepens and eventually collides with the rock, generating a cylindrical, outward-propagating shock. The resulting wave interactions around the rock are complex and dynamic. These results further highlight the robustness of the corrected scheme, even in this highly demanding test configuration and for a (very) high-order polynomial degree.

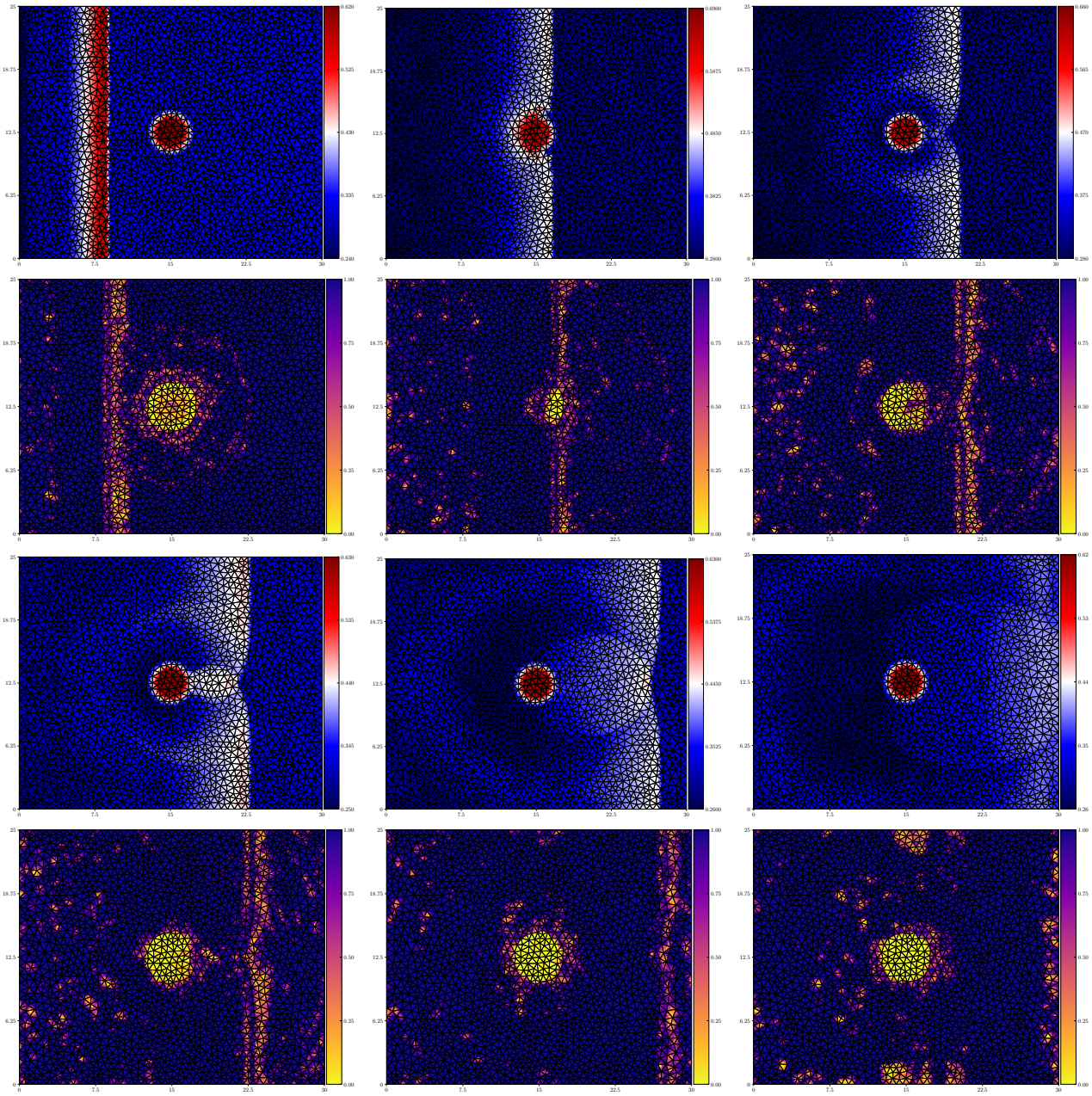


Figure 3.36: Tidal-wave overtopping the conical island – Snapshots of $\mathbb{P}^2$ free-surface elevation (1^{st} and 3^{rd} line) and their corresponding blending map (2^{nd} and 4^{th} line) for $t \in [1, 12]$ sec (animation).

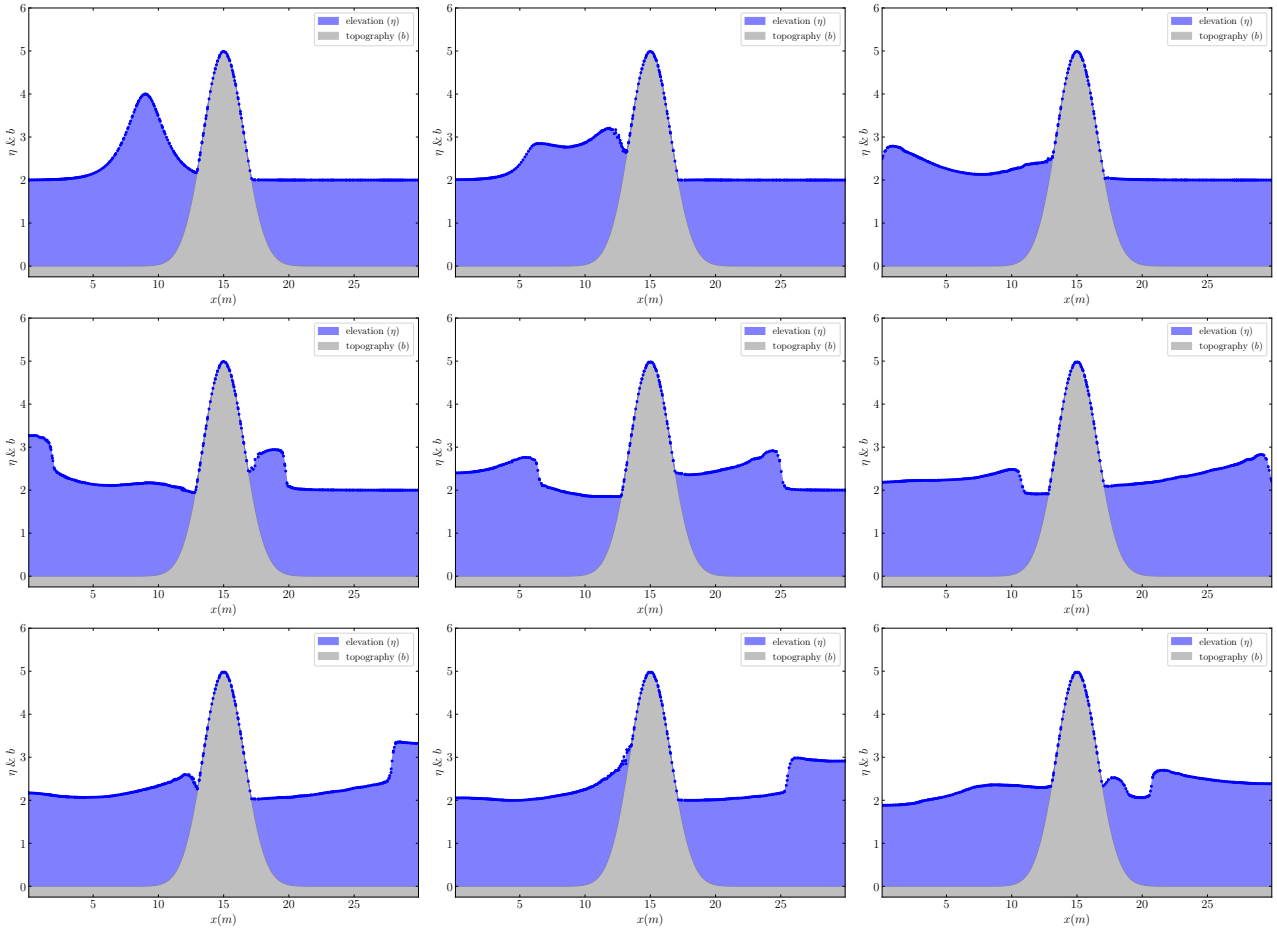


Figure 3.37: Single wave-rock interaction − Snapshots of $\mathbb{P}^6$ free-surface elevation on slice $x = 15$ for $t \in [0, 5]$ sec (animation).

3.5.7 Tsunami on an island with conical obstacles

PARAMETERS

Computational domain: $\Omega = [-5, 7] \times [-2, 2]$ — *Polynomial degree*: $k = 2$
Number of cells: $N_{\text{el}} = 7096$ cells — *Time-marching order*: $p = 3$
Boundary conditions: inflow (left), solid walls (top/bottom), Dirichlet (right) — *Final time*: $T_{\max} = 2$ sec

We conclude with a test case simulating a tsunami interacting with multiple obstacles. Such simulations are frequently carried out [69] as they provide a solid foundation for tackling realistic coastal engineering problems. In this case, a shock wave impacts three conical obstacles. The bathymetry is given by

$$b(\boldsymbol{x}) := \sum_{i=1}^{3} b_i(\boldsymbol{x}) + \begin{cases} 1 + 0.2x, & \text{for } x < 0, \\ 1, & \text{for } 0 \le x \le 3, \\ 1 + 0.4(x-3), & \text{for } x > 3, \end{cases}$$

where each $b_i(\boldsymbol{x})$ is defined as $b_i(\boldsymbol{x}) := c(\boldsymbol{x}, \boldsymbol{x}_i, R_i, A_i)$, with c being a cone-shaped function defined by:

$$c(\boldsymbol{x}, \boldsymbol{x}_c, R, A) := \begin{cases} \frac{A}{R}\left(R - \sqrt{(x-x_c)^2 + (y-y_c)^2}\right), & \text{if } \sqrt{(x-x_c)^2 + (y-y_c)^2} < R, \\ 0, & \text{otherwise.} \end{cases}$$

In this setup, we take $R_i := 0.5$ and $A_i := 3$ for all i, and place the cones at the locations $(x_1, y_1) := (1, -1)^\top$, $(x_2, y_2) := (1, 1)$, and $(x_3, y_3) := (2, 0)$. The initial condition is defined component-wise as follows:

$$H(\boldsymbol{x}, 0) := \begin{cases} 1.5 - b(\boldsymbol{x}), & \text{if } x < -3.5 \\ 0, & \text{otherwise} \end{cases}, \qquad u(\boldsymbol{x}, 0) := \begin{cases} 4, & \text{if } x < -3.5 \\ 0, & \text{otherwise} \end{cases}, \qquad v(\boldsymbol{x}, 0) := 0.$$

To reflect a more realistic tsunami scenario, a time-dependent inflow condition is used to model a sequence of waves striking the obstacles. It is set as follows on the left boundary:

$$\boldsymbol{q}(x = -5, y, t) := (10(1 + \cos(2\pi t))e^{-2t}, 0)^\top.$$

Figures 3.39–3.40 illustrate again the robustness of the method for this demanding configuration, where the incoming waves interact with several obstacles and generate complex reflected patterns on dry areas.

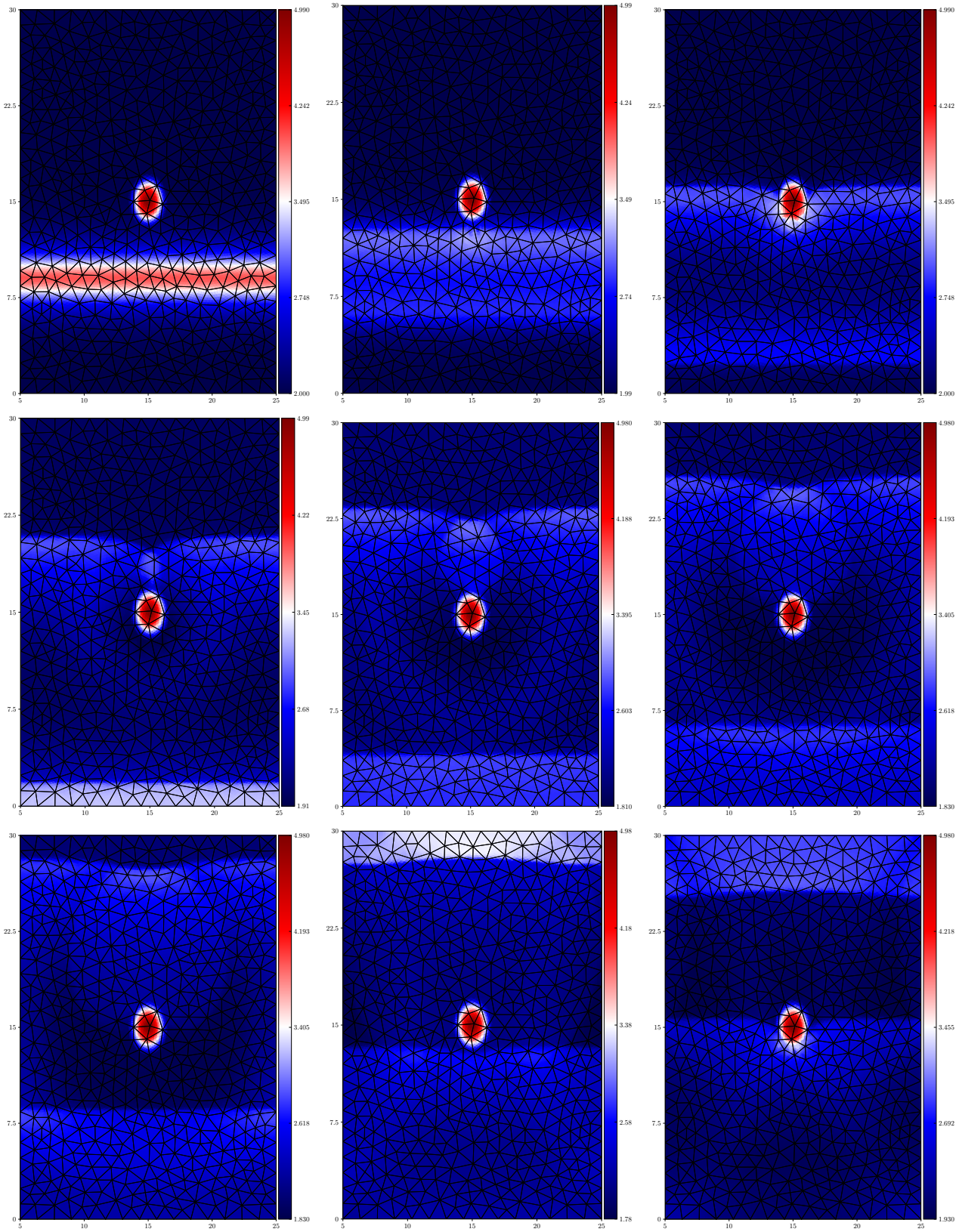


Figure 3.38: Single wave-rock interaction − Snapshots of $\mathbb{P}^6$ free-surface elevation for $t \in [0, 5]$ sec (animation).

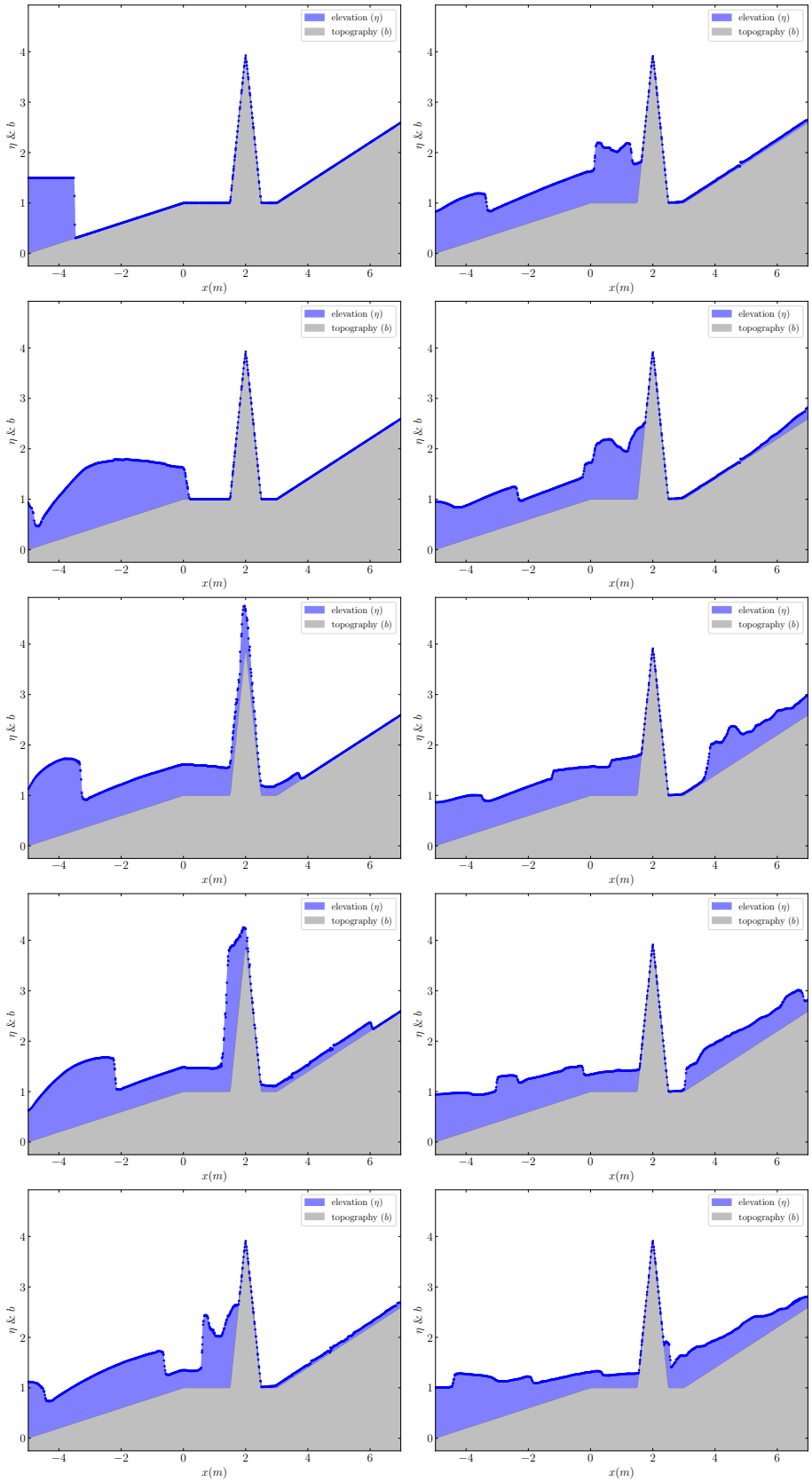


Figure 3.39: Tsunami with conical obstacles – Columnwise, from top to bottom: snapshots of $\mathbb{P}^2$ free-surface elevation on slice $y = 12.5$ for $t \in [0, 2]$ sec (animation).

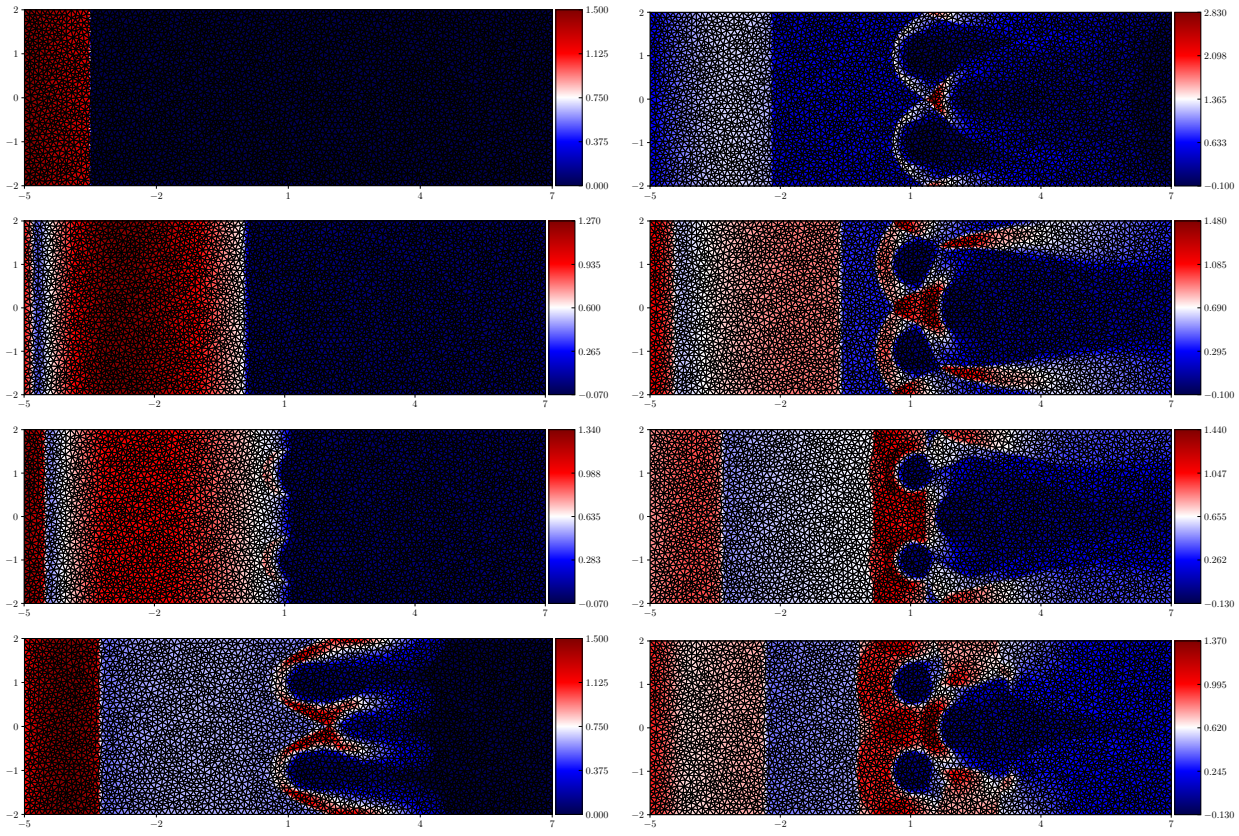


Figure 3.40: Tsunami with conical obstacles – Snapshots of $\mathbb{P}^2$ water-height for $t \in [0, 2]$ sec, columnwise, from left to right (animation).

CHAPTER 4

Subcell Arbitrary–Lagrangian Eulerian scheme for shallow-water flows on unstructured moving meshes

The ALE scheme presented here is expected to form part of a forthcoming publication on coupled surge–heave–pitch wave-structure interactions.

This chapter constructs an Arbitrary Lagrangian–Eulerian version of the two-dimensional monolithic DG/FV subcell method introduced in Chapter 3. The objective is to formulate the method on moving unstructured meshes while preserving the intrinsic properties of the Eulerian scheme: high-order accuracy in smooth regimes, local subcell robustness, convex admissibility preservation and compatibility with well-balanced source discretizations.

The construction starts from a time-dependent ALE mapping between a fixed reference configuration and the moving physical domain. This leads to the grid velocity, the ALE differential identities and the geometric conservation law, which are then incorporated into the DG and DG/FV subcell formulations. The scheme is rewritten on time-dependent triangulations, using invariant basis functions, moving control volumes and ALE numerical fluxes. A key point is that the Eulerian subcell strategy extends naturally to the ALE setting: once the physical fluxes are replaced by ALE fluxes and the subcell volumes and faces are updated consistently, the same monolithic convex blending mechanism can be used on the moving mesh. The chapter also introduces the computation of the grid velocity through elliptic problems solved with continuous Lagrange finite elements. This provides one of the very first direct high-order subcell ALE framework on unstructured moving meshes and forms a necessary step towards the simulation of free-floating structures.

Contents

4.1 Continuous and discrete settings 114
4.1.1 The 2D pre-balanced nonlinear shallow-water equations 114
4.1.2 Discrete framework 114
4.2 Arbitrary Lagrangian–Eulerian description 117
4.2.1 Grid velocity and geometric conservation law 117
4.2.2 Adapted basis functions 118
4.2.3 ALE integral formulation 121
4.3 Discrete formulations and ALE extension 121
4.3.1 DG formulation and DG-ALE formulation 122
4.3.2 Monolithic DG/FV-ALE subcell formulation 122
4.4 Computation of the ALE grid velocity 124
4.4.1 From boundary motion to mesh motion 124
4.4.2 On the choice of the mesh velocity 125
4.4.3 Finite element formulations for the grid velocity 126
4.4.4 Reduction to explicit affine mapping for pseudo-1D configurations 129
4.5 Local subcell monolithic DG/FV–ALE scheme 132
4.6 Conservation properties of the ALE formulation 134
4.6.1 Semi and fully-discrete geometric conservation laws 135
4.6.2 Well-balancing property 139
4.7 Numerical illustrations 140
4.7.1 Well-balancing on a moving unstructured mesh 140
4.7.2 Periodic wave generation by an oscillating piston 142
4.7.3 Wave run-up over a sloping beach generated by an impulsively stopped piston 143
4.7.4 Wave generation by an oscillating cylindrical wavemaker 147

Eu tô perdida num mar de ondas suas
E remo sem destino esse barco de papel
FROM: ***"Barquinho de papel"***, Anavitória – *Trilhas* (2022)

4.1 Continuous and discrete settings

In this section, we introduce the continuous and discrete settings for the Arbitrary Lagrangian–Eulerian (ALE) formulation of the two-dimensional nonlinear shallow-water (NSW) equations. The continuous setting is based on a time-dependent mapping between a fixed reference configuration and the moving physical domain, which leads to the ALE differential identities and the geometric conservation law. The discrete setting is then adapted to the ALE framework, with all discrete objects defined on time-dependent meshes.

4.1.1 The 2D pre-balanced nonlinear shallow-water equations

Given a parameterization $b \in \mathscr{C}_c^\infty(\Omega)$ of the bathymetry variation, denoting the free-surface elevation by η, the water discharge by $\boldsymbol{q} = (q_x, q_y)^\top$ and $\boldsymbol{u} = (u_x, u_y)^\top$ as the depth-averaged water velocity, we still consider the 2D pre-balanced NSW from Chapter 3:

$$\partial_t \boldsymbol{v} + \nabla_{\boldsymbol{x}} \cdot \mathbb{F}(\boldsymbol{v}, b) = \boldsymbol{B}[b](\boldsymbol{v}),$$

supplemented with some initial-data $\boldsymbol{v}(\cdot, 0) \coloneqq \boldsymbol{v}^0$ and where all the physical quantities are defined in Nonlinear shallow-water model in two space dimension.

4.1.2 Discrete framework

Here, the discrete setting developed in Chapter 3 is adapted to the Arbitrary Lagrangian-Eulerian (ALE) framework, and all discrete objects are defined on time-dependent triangulations. For completeness and to keep the presentation rigorous, we redefine the main mesh, polynomial, and subcell notations with their explicit dependence on time. Readers already familiar with the notation introduced in Chapter 3 may therefore skip most of the technical definitions below and only keep in mind that the same objects are now considered on a moving mesh.

The domain partitioning, the polynomial approximation, and the sub-mesh construction are designed to be consistent with the moving mesh geometry, and to allow for a direct extension of the monolithic DG/FV method to the ALE context. Let us remark that this extension does not require any modification of the topological structure of the mesh: in our implementation, the same connectivity is preserved during the mesh motion. This allows us to keep the same data structure and to focus on the geometric and numerical issues related to mesh motion, without dealing with remeshing or interpolation between different meshes.

Domain time-dependent partitioning and related polynomial approximation

We consider a computational domain $\Omega \subset \mathbb{R}^2$, assumed to be a bounded open set with Lipschitz boundary $\partial\Omega(t)$. At each time t, the domain is equipped with a moving triangulation $\mathcal{T}_h(t)$ made of N_{el} open cells $\omega_c(t)$, $c \in [\![1, N_{\mathrm{el}}]\!]$, satisfying

$$\overline{\Omega} = \bigcup_{\omega_c(t) \in \mathcal{T}_h(t)} \overline{\omega_c(t)}, \quad \text{and} \quad \mathring{\omega}_c(t) \cap \mathring{\omega}_v(t) = \emptyset, \ \text{ for all } c \neq v.$$

For a given element $\omega_c(t) \in \mathcal{T}_h(t)$, its diameter is denoted by $h_c(t)$, while the global mesh size is defined by $h(t) \coloneqq \max_{c \in [\![1, N_{\mathrm{el}}]\!]} h_c(t)$. We also denote by $\boldsymbol{x}_{\mathrm{b}}^c(t) \coloneqq (x_{\mathrm{b}}^c(t), y_{\mathrm{b}}^c(t))$ the barycenter of $\omega_c(t)$, and by $h_x^c(t)$ and $h_y^c(t)$ the maximum extensions of $\omega_c(t)$ in the x- and y-directions, respectively.

The set of all mesh faces is denoted by $\mathcal{F}_h(t)$ and is decomposed into internal and boundary faces as $\mathcal{F}_h(t) = \mathcal{F}_h^0(t) \cup \mathcal{F}_h^\partial$. The subset $\mathcal{F}_h^0(t)$ contains the faces shared by two distinct cells, namely the faces Γ such that there exist $\omega_c(t), \omega_v(t) \in \mathcal{T}_h(t)$, with $c \neq v$, satisfying $\Gamma \subset \partial\omega_c(t) \cap \partial\omega_v(t)$. The subset $\mathcal{F}_h^\partial$ contains the boundary faces, i.e. the faces Γ such that there exists $\omega_c(t) \in \mathcal{T}_h(t)$ with $\Gamma \subset \partial\omega_c(t) \cap \partial\Omega$. The length of a face $\Gamma \in \mathcal{F}_h(t)$ is denoted by ℓ_Γ, and its barycenter by $\boldsymbol{x}_{\mathrm{b}}^\Gamma(t)$. We also denote by h_Γ the minimum of the diameters of the elements sharing the face Γ. Finally, $\mathcal{V}_h(t)$ stands for the set of vertices of the triangulation $\mathcal{T}_h(t)$, with cardinality $N_v \coloneqq \#\mathcal{V}_h(t)$. Each vertex is denoted by $\boldsymbol{x}_p(t) \coloneqq (x_p(t), y_p(t))$, with $p \in [\![1, N_v]\!]$.

For each cell $\omega_c(t) \in \mathcal{T}_h(t)$, we denote by $\mathcal{F}_c(t) \coloneqq \{\Gamma \in \mathcal{F}_h(t) \mid \Gamma \subset \partial\omega_c(t)\}$ the collection of its faces. In the numerical experiments considered in this work, the meshes are simplicial, so that $\#\mathcal{F}_c(t) = 3$ for all $\omega_c(t) \in \mathcal{T}_h(t)$. Nevertheless, the construction is not restricted to simplices and may be extended to more general cells. If $\Gamma \in \mathcal{F}_c(t)$, the unit normal to Γ pointing outward from $\omega_c(t)$ is denoted by $\mathbf{n}_\Gamma$. For an internal face $\Gamma \in \mathcal{F}_h^0(t)$, an arbitrary but fixed orientation is chosen, and the associated normal is denoted by $\mathbf{n}_\Gamma \coloneqq \mathbf{n}_{c\Gamma}$. For

a boundary face $\Gamma \in \mathcal{F}_h^\partial$, we set $\mathbf{n}_\Gamma := \mathbf{n}_{c\Gamma}$. When no ambiguity is possible, we will also write $\mathbf{n}_{cv(k)}$ for the normal vector associated with the face $\Gamma_{cv(k)}$ of the element $\omega_c(t)$ shared with its neighboring element $\omega_{v(k)}$, $k \in [\![1, \#\mathcal{F}_c(t)]\!]$.

For any integer $k \geq 0$, let us define the broken-polynomials space:

$$\mathbb{P}^k(\mathcal{T}_h(t)) := \{v \in L^2(\Omega) \mid \forall \omega_c(t) \in \mathcal{T}_h(t),\ v_{|\omega_c(t)} \in \mathbb{P}^k(\omega_c(t))\},$$

where $\mathbb{P}^k(\omega_c(t))$ denotes the space of polynomials in $\omega_c(t)$ of total degree at most k, with $\dim \mathbb{P}^k(\omega_c(t)) = \frac{(k+1)(k+2)}{2}$. In the following, we use the shortcut $N_k := \dim \mathbb{P}^k(\omega_c(t))$, for all $\omega_c(t) \in \mathcal{T}_h(t)$. Functions belonging to the broken-polynomial space $\mathbb{P}^k(\mathcal{T}_h(t))$ will be denoted with a subscript h. For such a function $v_h \in \mathbb{P}^k(\mathcal{T}_h(t))$, and for all cells $\omega_c(t) \in \mathcal{T}_h(t)$, we use the local notation $v_h^c := v_{h_{|\omega_c(t)}}$. We also use the shorthand notation $\mathbf{P}^k(U) := (\mathbb{P}^k(U))^3$ for any relevant domain U (e.g., $U = \mathcal{T}_h(t)$, $\mathcal{F}_h(t)$, $\omega_c(t)$ etc.). For any element $\omega_c(t) \in \mathcal{T}_h(t)$, a basis of $\mathbb{P}^k(\omega_c(t))$ is denoted by

$$\mathbf{\Psi}_{\omega_c(t)} := \{\psi_j^c\}_{j \in [\![1, N_k]\!]}, \tag{4.1}$$

and, for $v_h^c \in \mathbb{P}^k(\omega_c(t))$, the collection $\underline{v}^c := \left\{\underline{v}_j^c \mid j \in [\![1, N_k]\!]\right\}$ denotes the *local* degrees of freedom associated with $\omega_c(t)$, through the expansion

$$v_h^c(\boldsymbol{x}, t) := \sum_{m=1}^{N_k} \underline{v}_m^c(t)\psi_m^c(\boldsymbol{x}).$$

The *global* degrees of freedom of v_h are then gathered in the collection $\underline{v} := \{\underline{v}^c \mid c \in [\![1, N_{\mathrm{el}}]\!]\}$. For $\omega_c \in \mathcal{T}_h(t)$, we denote by π_c^k the L^2-orthogonal projector onto $\mathbb{P}^k(\omega_c(t))$ and by π_h^k the global L^2-orthogonal projector onto $\mathbb{P}^k(\mathcal{T}_h(t))$, obtained by assembling all the local projectors π_c^k. The broken gradient operator $\nabla_h^k \colon \mathbb{P}^k(\mathcal{T}_h(t)) \to \mathbb{P}^{k-1}(\mathcal{T}_h(t))$ is defined elementwise by, for all $v_h \in \mathbb{P}^k(\mathcal{T}_h(t))$,

$$(\nabla_h^k v_h)_{|\omega_c(t)} := \nabla_{\boldsymbol{x}}(v_{h|\omega_c(t)}), \qquad \forall \omega_c(t) \in \mathcal{T}_h(t).$$

We also introduce the following broken scalar products for $(v, w) \in L^2(\mathcal{T}_h)^2$ and $(\mu, \nu) \in L^2(\partial\mathcal{T}_h)^2$:

$$\begin{aligned}
(v, w)_{\mathcal{T}_h(t)} &:= \sum_{\omega_c(t) \in \mathcal{T}_h(t)} (v, w)_{\omega_c(t)} = \sum_{\omega_c(t) \in \mathcal{T}_h(t)} \int_{\omega_c(t)} v(\boldsymbol{x}, t)\, w(\boldsymbol{x}, t)\, \mathrm{d}\boldsymbol{x}, \\
\left\langle \mu, \nu \right\rangle_{\partial\mathcal{T}_h(t)} &:= \sum_{\omega_c(t) \in \mathcal{T}_h(t)} \left\langle \mu, \nu \right\rangle_{\partial\omega_c(t)} = \sum_{\omega_c(t) \in \mathcal{T}_h(t)} \int_{\partial\omega_c(t)} \mu(\boldsymbol{x}, t)\, \nu(\boldsymbol{x}, t)\, \mathrm{d}s, \\
\left\langle \mu, \nu \right\rangle_{\mathcal{T}_h(t)} &:= \sum_{\omega_c(t) \in \mathcal{T}_h(t)} \left\langle \mu, \nu \right\rangle_{\mathcal{F}_c(t)} = \sum_{\omega_c(t) \in \mathcal{T}_h(t)} \sum_{\Gamma \in \mathcal{F}_c(t)} \int_\Gamma \mu(\boldsymbol{x}, t)\, \nu(\boldsymbol{x}, t)\, \mathrm{d}s.
\end{aligned}$$

The same notation is used for vector-valued functions $\boldsymbol{v}, \boldsymbol{w}$ and $\boldsymbol{\mu}, \boldsymbol{\nu}$, with scalar products replaced by the Euclidean dot product.

We now seek approximate solutions of (3.1) in the form $\boldsymbol{v}_h := (\eta_h, \boldsymbol{q}_h)$, with $\boldsymbol{v}_h(\cdot, t) \in \mathbf{P}^k(\mathcal{T}_h(t))$ for any time t. Given such a function $\boldsymbol{v}_h \in \mathbf{P}^k(\mathcal{T}_h(t))$ and an interface $\Gamma_{cv(k)}$, we denote by $\boldsymbol{v}_k^-$ and $\boldsymbol{v}_k^+$ the restrictions of $\boldsymbol{v}_{h_{|\omega_c(t)}}$ and $\boldsymbol{v}_{h_{|\omega_{v(k)}}}$ to $\Gamma_{cv(k)}$, respectively, i.e. the interior and exterior traces with respect to $\omega_c(t)$. On each element, the approximate vector solution $\boldsymbol{v}_h^c$ is expanded as

$$\boldsymbol{v}_h^c(\boldsymbol{x}, t) = \sum_{m=1}^{N_k} \underline{\boldsymbol{v}}_m^c(t)\psi_m^c(\boldsymbol{x}), \tag{4.2}$$

with $\boldsymbol{v}_h^c = (\eta_h^c, (q_x)_h^c, (q_y)_h^c)$. Similarly, the bathymetry parameterization is approximated locally by the polynomial expansion $b_h^c(\boldsymbol{x}) = \sum_{m=1}^{N_k} \underline{b}_m^c \phi_m^c(\boldsymbol{x})$.

Sub-mesh and related discrete operators

We now introduce the subcell structure associated with the moving mesh. For any given mesh element $\omega_c(t) \in \mathcal{T}_h(t)$, we define a sub-partition $\mathcal{T}_{\omega_c}^k(t)$ made of N_s open disjoint subcells:

$$\overline{\omega}_c(t) = \bigcup_{m \in [\![1, N_s]\!]} \overline{S}_m^c(t),$$

where the subcell $S_m^c(t)$ has volume $|S_m^c(t)|$. These volumes are collected in the diagonal matrix $\mathbb{D}_c \in \mathscr{M}_{N_s}(\mathbb{R})$ defined by $\mathbb{D}_c(t) := \mathrm{diag}_{m\in[\![1,N_s]\!]}(|S_m^c(t)|)$.

We denote by Γ_{mp}^c the face shared by the subcell $S_m^c(t)$ and a neighboring subcell S_p. Its length is denoted by ℓ_{mp}, and N_f^m denotes the number of faces of the subcell $S_m^c(t)$. The set of such faces is denoted by $\mathscr{F}_m^c$, so that $\#\mathscr{F}_m^c = N_f^m$. We also introduce N_f^c as the number of subcell faces lying inside $\omega_c(t)$, namely those not belonging to $\partial\omega_c(t)$.
The set of face-neighboring subcells of $S_m^c(t)$ is denoted by $\mathscr{V}_m^c(t)$, with $\#\mathscr{V}_m^c(t) = N_f^m$, while $\breve{\mathscr{V}}_m^c$ denotes the subset containing only the face-neighboring subcells of $S_m^c(t)$ located inside the same cell $\omega_c(t)$. A neighboring subcell S_p may therefore either belong to the same element $\omega_c(t)$, in which case $S_p \in \mathcal{T}_{\omega_c}^k(t)$ and $S_p = S_p^c$, or to a neighboring element $\omega_v(t)$, in which case $S_p \in \mathcal{T}_{\omega_v}^k(t)$ and $S_p = S_p^v(t)$.

The normal vector associated with the face Γ_{mp}^c of $S_m^c(t)$ shared with a neighboring subcell $S_p^v(t) \in \mathscr{V}_m^c(t)$ is denoted by $\mathbf{n}_{mp}$. Its orientation is encoded through the sign function ε_{mp}^c, defined by

$$\varepsilon_{mp}^c = \begin{cases} 1 & \text{if } S_m \text{ and } S_p \text{ are neighbors and } m < p, \\ -1 & \text{if } S_m \text{ and } S_p \text{ are neighbors and } m > p, \\ 0 & \text{if } S_p \notin \mathscr{V}_m^c. \end{cases}$$

We denote by $\mathscr{P}_m^c(t)$ the set of vertices $\boldsymbol{x}_p(t)$ of the subcell $S_m^c(t)$, and by $\mathscr{V}_p(t)$ the set of subcells sharing the vertex $\boldsymbol{x}_p(t)$. We then define

$$\mathscr{N}(S_m^c(t)) := \bigcup_{\boldsymbol{x}_p(t)\in\mathscr{P}_m^c(t)} \mathscr{V}_p(t)$$

which gathers the subcell $S_m^c(t)$ together with its face- and vertex-neighboring subcells $S_p^v(t)$, whether they belong to the same mesh cell or to a neighboring one. Finally, the set $\mathscr{V}_{mp}(t)$ contains all subcells sharing at least one vertex with the face Γ_{mp}.

For any given cell $\omega_c(t) \in \mathcal{T}_h(t)$, we introduce the space of piecewise-constant functions on the local sub-mesh:

$$\mathbb{P}^0(\mathcal{T}_{\omega_c}^k((t)) := \{v \in L^\infty(\omega_c(t)) \mid \forall S_m^c(t) \in \mathcal{T}_{\omega_c}^k((t),\ v_{|S_m^c(t)} \in \mathbb{P}^0(S_m^c(t))\},$$

together with the subcell-indicator functions

$$\{\mathbb{1}_m^c \mid m \in [\![1, N_s]\!]\},$$

defined by $\mathbb{1}_m^c(\boldsymbol{x}) := 1$ if $\boldsymbol{x} \in S_m^c(t)$, and 0 otherwise. These indicator functions allow us to define the local mean-value projector $\overline{\pi}_c^k$ onto $\mathbb{P}^0(\mathcal{T}_{\omega_c}^k(t))$ by

$$\mathbb{P}^k(\omega_c(t)) \ni v_h^c \mapsto \overline{\pi}_c^k v_h^c := \sum_{m=1}^{N_s} \overline{v}_m^c \mathbb{1}_m^c =: \overline{v}_h^c \in \mathbb{P}^0(\mathcal{T}_{\omega_c}^k(t)),$$

where

$$\{\overline{v}_m^c \mid m \in [\![1, N_s]\!]\} =: \overline{v}^c,$$

is the collection of piecewise-constant components given by the mean values of v_h^c over the subcells of $\mathcal{T}_{\omega_c}^k(t)$. These values are referred to as *sub-mean values*. When no confusion is possible, $\overline{v}^c$ may equivalently denote a vector of $\mathbb{R}^{N_s}$, a collection of N_s scalar mean values, or the associated piecewise-constant function $\overline{v}_h^c \in \mathbb{P}^0(\mathcal{T}_{\omega_c}^k(t))$. To obtain a global subcell description, we gather all local sub-partitions and introduce

$$\widetilde{\mathcal{T}_h}(t) := \bigcup_{\omega_c\in\mathcal{T}_h} \mathcal{T}_{\omega_c}^k(t), \qquad \mathbb{P}^0(\widetilde{\mathcal{T}_h}(t)) := \bigtimes_{\omega_c\in\mathcal{T}_h(t)} \mathbb{P}^0(\mathcal{T}_{\omega_c}^k(t)).$$

The global projector onto $\mathbb{P}^0(\widetilde{\mathcal{T}_h}(t))$, obtained by collecting the local projectors $\overline{\pi}_c^k$, is denoted by $\overline{\pi}_h^k$. This projector may also be extended to larger functional spaces, typically $H^s(\mathbb{R})$, $s > 1 + d/2$ for the initial data, and $\mathscr{C}_{\mathrm{b}}^\infty(\mathbb{R})$ for the bathymetry, and is then denoted by $\widetilde{\pi}_h^k$. In what follows, piecewise-constant functions in $\mathbb{P}^0(\widetilde{\mathcal{T}_h}(t))$ are denoted with an overline and a subscript h. Thus, for any $\overline{v}_h \in \mathbb{P}^0(\widetilde{\mathcal{T}_h}(t))$ and any subcell $S_m^c \in \widetilde{\mathcal{T}_h}(t)$, we have the identity $\overline{v}_m^c = \overline{v}_{h|S_m^c}$. For vector-valued quantities, we use the product space notations $\mathbf{P}^0(\mathcal{T}_{\omega_c(t)}^k) := (\mathbb{P}^0(\mathcal{T}_{\omega_c(t)}^k))^3$ and $\mathbf{P}^0(\widetilde{\mathcal{T}_h}(t)) := (\mathbb{P}^0(\widetilde{\mathcal{T}_h}(t)))^3$.

Time discretization and time-marching algorithm.

Given a final time $T > 0$, we consider a general partition of the time interval $[0, T]$ defined by $0 =: t^0 < t^1 < \cdots < t^n < \cdots < t^N := T$. The time step is denoted by $\Delta t^n \in \mathbb{R}_+^*$ and may vary with time, such that for any $n \geq 0$, we have $t^{n+1} = t^n + \Delta t^n$. Time integration is performed using explicit *Strong Stability Preserving Runge-Kutta* (SSP-RK) schemes from Time discretization and time-marching algorithm.

As in Chapter 1, we define $\boldsymbol{v}_h^n := \boldsymbol{v}_h(\cdot, t^n)$, and let $\overline{\boldsymbol{v}}_h^n := \overline{\pi}_h^k \boldsymbol{v}_h^n =: (\overline{\eta}_h^n, \overline{\boldsymbol{q}}_h^n)$ denote the associated piecewise-constant function over the sub-mesh. For each cell $\omega_c(t)$, we denote

$$\overline{\boldsymbol{v}}^{c,n} := \{\overline{\boldsymbol{v}}_m^{c,n} \mid m \in [\![1, N_s]\!]\},$$

as the set of sub-mean values at time t^n, where each $\overline{\boldsymbol{v}}_m^{c,n}$ corresponds to the value of $\overline{\boldsymbol{v}}_h^n$ restricted to S_m^c. When the context is clear, we may omit the superscripts c and n to simplify notation. Also, for the sake of clarity, time-dependent quantities will sometimes be denoted using a superscript n, rather than explicitly writing (t^n) or $(\cdot, t^n)$.

Since the scheme evolves both polynomial degrees of freedom on the primal cells and mean values on the subcells, the time step must be constrained by the subcell geometry and by the largest local wave speed. We therefore choose Δt^n adaptively according to

$$\Delta t^n = \frac{\min\limits_{\omega_c(t^n) \in \mathcal{T}_h(t^n),\, S_m^c(t^n) \in \mathcal{T}_{\omega_c}^k(t^n)} |S_m^c(t^n)|}{\sigma_{\mathrm{ALE}} \max\limits_{\omega_c(t^n) \in \mathcal{T}_h(t^n)} \left(\sum\limits_{k=1}^{\#\mathcal{F}_c(t^n)} \ell_{cv(k)} \right)}, \tag{4.3}$$

where σ_{ALE} is accounting also for the grid velocity, and yet to be defined in (4.16).

4.2 Arbitrary Lagrangian–Eulerian description

In this section, we introduce an ALE description of the problem. A key feature of this framework is the construction of a smooth and regular change of coordinates that maps the initial (stationary) domain $\Omega(0)$ to the current (moving) domain $\Omega(t)$:

$$\Omega(0) \times [0, T_{\max}] \ni (\boldsymbol{\xi}, t) \mapsto \boldsymbol{x}(\boldsymbol{\xi}, t) \in \Omega(t),$$

where $\boldsymbol{\xi} = (\xi_x, \xi_y)$ denotes the *reference* coordinate, and $\boldsymbol{x} := \boldsymbol{x}(\boldsymbol{\xi}, t)$ the corresponding *physical* coordinate.

4.2.1 Grid velocity and geometric conservation law

Assuming sufficient regularity of the mapping (continuous in time and piecewise continuously differentiable in $\boldsymbol{\xi}$), the grid velocity at the physical point $\boldsymbol{x}$ is given by:

$$\boldsymbol{w}^g(\boldsymbol{x}, t) = \partial_t \boldsymbol{x}(\boldsymbol{\xi}, t).$$

Let $\rho(\cdot, t) \in \mathscr{C}^\infty(\Omega(t))$ be any smooth scalar field. We define its pullback to the reference frame as $\widetilde{\rho} := \rho \circ \boldsymbol{x}$, so that:

$$\rho\big(\boldsymbol{x}(\boldsymbol{\xi}, t), t\big) =: \widetilde{\rho}(\boldsymbol{\xi}, t).$$

Then, the fundamental ALE identity for total time derivatives reads:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \rho\big(\boldsymbol{x}(\boldsymbol{\xi}, t), t\big) &:= \partial_t \rho\big(\boldsymbol{x}(\boldsymbol{\xi}, t), t\big) + \boldsymbol{w}^g\big(\boldsymbol{x}(\boldsymbol{\xi}, t), t\big) \cdot \nabla_{\boldsymbol{x}} \rho\big(\boldsymbol{x}(\boldsymbol{\xi}, t), t\big) \\ &= \partial_t \widetilde{\rho}(\boldsymbol{\xi}, t). \end{aligned} \tag{4.4}$$

The geometric conservation law (GCL), together with its discrete counterpart, is a basic consistency requirement for moving-mesh methods. Its role is to ensure that the numerical method correctly accounts for the motion of the control volumes. In particular, a constant state should remain constant when the mesh moves: in the opposite case, the scheme may generate artificial waves only because the grid is deforming.

At the continuous level, the GCL is first understood as a geometric identity for moving domains. Applying Reynolds' transport formula to the constant function 1 gives

$$\frac{\mathrm{d}}{\mathrm{d}t}|\Omega(t)| = \frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega(t)} \mathrm{d}\boldsymbol{x} = \int_{\partial\Omega(t)} \boldsymbol{w}^g \cdot \mathbf{n}\, \mathrm{d}s = \int_{\Omega(t)} \nabla_{\boldsymbol{x}} \cdot \boldsymbol{w}^g \, \mathrm{d}\boldsymbol{x}. \tag{4.5}$$

Thus, the variation of the volume of the moving domain is entirely determined by the grid velocity $\boldsymbol{w}^g$. The same property can be written locally by using the ALE mapping. Let $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{\xi}, t)$ be the ALE mapping, and let $\mathbb{J}(\boldsymbol{\xi}, t) := \nabla_{\boldsymbol{\xi}} \boldsymbol{x}(\boldsymbol{\xi}, t)$ be its Jacobian matrix. The determinant of this matrix, denoted by $\mathsf{J}(\boldsymbol{\xi}, t) := \det \mathbb{J}(\boldsymbol{\xi}, t)$, represents the local change of volume between the reference and physical domains. Indeed, for any smooth scalar field ρ, we have the following change of variables formula:

$$\int_{\Omega(t)} \rho(\boldsymbol{x}, t)\, \mathrm{d}\boldsymbol{x} = \int_{\Omega(0)} \rho(\boldsymbol{x}(\boldsymbol{\xi}, t), t)\mathsf{J}(\boldsymbol{\xi}, t)\, \mathrm{d}\boldsymbol{\xi}.$$

Therefore, the local version of the GCL is an evolution equation for the volume factor J. Since the grid velocity is defined by

$$\boldsymbol{w}^g(\boldsymbol{x}(\boldsymbol{\xi}, t), t) = \partial_t \boldsymbol{x}(\boldsymbol{\xi}, t),$$

we obtain, by differentiating the mapping with respect to $\boldsymbol{\xi}$,

$$\partial_t \mathbb{J}(\boldsymbol{\xi}, t) = \partial_t \nabla_{\boldsymbol{\xi}} \boldsymbol{x}(\boldsymbol{\xi}, t) = \nabla_{\boldsymbol{\xi}} \partial_t \boldsymbol{x}(\boldsymbol{\xi}, t) = \nabla_{\boldsymbol{\xi}} \boldsymbol{w}^g(\boldsymbol{x}(\boldsymbol{\xi}, t), t). \tag{4.6}$$

Then, using Jacobi's formula for the derivative of a determinant, we obtain

$$\partial_t \mathsf{J} = \mathsf{J} \operatorname{tr}\left(\mathbb{J}^{-1} \partial_t \mathbb{J}\right).$$

Moreover, from (4.6) and by the chain rule, one has

$$\partial_t \mathbb{J}(\boldsymbol{\xi}, t) = \nabla_{\boldsymbol{x}} \boldsymbol{w}^g(\boldsymbol{x}(\boldsymbol{\xi}, t), t) \nabla_{\boldsymbol{\xi}} \boldsymbol{x}(\boldsymbol{\xi}, t) = \nabla_{\boldsymbol{x}} \boldsymbol{w}^g(\boldsymbol{x}(\boldsymbol{\xi}, t), t)\, \mathbb{J}(\boldsymbol{\xi}, t).$$

Substituting this identity into Jacobi's formula gives

$$\partial_t \mathsf{J} = \mathsf{J} \operatorname{tr}\left(\mathbb{J}^{-1} \nabla_{\boldsymbol{x}} \boldsymbol{w}^g \mathbb{J}\right),$$

then by invoking the cyclic invariance property of the trace, we obtain

$$\operatorname{tr}\left(\mathbb{J}^{-1} \nabla_{\boldsymbol{x}} \boldsymbol{w}^g \mathbb{J}\right) = \operatorname{tr}\left(\nabla_{\boldsymbol{x}} \boldsymbol{w}^g \mathbb{J} \mathbb{J}^{-1}\right) = \operatorname{tr}\left(\nabla_{\boldsymbol{x}} \boldsymbol{w}^g\right).$$

Since the trace of the velocity gradient is the divergence, this finally yields

$$\partial_t \mathsf{J}(\boldsymbol{\xi}, t) = \mathsf{J}(\boldsymbol{\xi}, t)\, \nabla_{\boldsymbol{x}} \cdot \boldsymbol{w}^g(\boldsymbol{x}(\boldsymbol{\xi}, t), t). \tag{4.7}$$

This relation is the local form of the continuous GCL. It states that the local variation of the physical volume element is exactly balanced by the divergence of the grid velocity. It means that this identity must be preserved at the discrete level in order to avoid artificial changes of volume induced only by the mesh motion.

At the discrete level, the D-GCL is the analogous compatibility condition between the numerical evolution of the cell areas and the discrete motion of the mesh faces. More precisely, the variation of each moving cell area must be consistent with the area swept by its boundary during the time step.

4.2.2 Adapted basis functions

In ALE methods, the numerical scheme is typically formulated in a reference domain where the mesh remains fixed over time. Within this framework, basis functions are defined with respect to the reference coordinates and are assumed to be time-independent. In particular, their total derivative with respect to time is taken to be zero. This choice indeed significantly simplifies the derivation of the semi-discrete equations: when physical variables are expressed as interpolations (or reconstructions) in terms of these basis functions, the time evolution is carried solely by the degrees of freedom (the coefficients of the expansion), while the basis functions themselves remain constant in time.

Sub-resolution basis functions

Before establishing the equivalence between the ALE subcell finite volume formulation and the ALE-DG scheme, it is essential to ensure that the indicator basis functions used in the sub-resolution reconstruction follow the motion of the mesh. This property guarantees that the subcell decomposition is consistent over time and that integrals computed over moving subcells remain coherent with the reference configuration.

Proposition 4.2.1 (*Invariance of indicator basis functions*). *Assume that the ALE mapping* $\boldsymbol{x}_c(\cdot,t):\omega_c(0)\to\omega_c(t)$ *is affine on each cell* $\omega_c(0)$ *and that each subcell is transported by this mapping. The sub-resolution spanning set of functions*

$$\boldsymbol{\Phi}_{\omega_c(t)} := \{\phi_m^c\}_{m\in[\![1,N_s]\!]}, \quad \textit{with} \quad \phi_m^c(\cdot,t) := \pi_c^k \mathbb{1}_m^c(\cdot,t), \qquad \boldsymbol{x}\in\omega_c(t), \tag{4.8}$$

follows the trajectory of the subcell $S_m^c(t)$*, i.e.*

$$\phi_m^c(\boldsymbol{x}(\boldsymbol{\xi},t),t) = \widetilde{\phi}_m^c(\boldsymbol{\xi}),$$

for all $\boldsymbol{x}(\boldsymbol{\xi},t)\in S_m^c(t)$ *and* $t\in[0,T_{\max}]$.

Proof. Since the ALE mapping $\boldsymbol{x}_c(\cdot,t):\omega_c(0)\to\omega_c(t)$ is affine on the cell $\omega_c(0)$, its Jacobian determinant is constant on $\omega_c(0)$ and satisfies

$$\mathsf{J}_c(t) = \det\nabla_{\boldsymbol{\xi}}\boldsymbol{x}_c(\boldsymbol{\xi},t) = \frac{|\omega_c(t)|}{|\omega_c(0)|}. \tag{4.9}$$

Moreover, since each subcell is transported by the ALE mapping, we have $S_m^c(t)=\boldsymbol{x}_c(S_m^c(0),t)$. By definition, $\phi_m^c(\cdot,t)$ is the L^2-orthogonal projection of the indicator function $\mathbb{1}_m^c(\cdot,t)$ onto $\mathbb{P}^k(\omega_c(t))$. Hence, for any $\varphi\in\mathbb{P}^k(\omega_c(t))$, we have

$$\int_{\omega_c(t)}\phi_m^c(\boldsymbol{x},t)\,\varphi(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x} = \int_{S_m^c(t)}\varphi(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x}. \tag{4.10}$$

Let $\widetilde{\varphi}\in\mathbb{P}^k(\omega_c(0))$ and define its transported counterpart on $\omega_c(t)$ by $\varphi(\boldsymbol{x}_c(\boldsymbol{\xi},t),t)=\widetilde{\varphi}(\boldsymbol{\xi})$. This implies that φ is constant along the ALE trajectories, namely

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi(\boldsymbol{x}_c(\boldsymbol{\xi},t),t) = \frac{\mathrm{d}}{\mathrm{d}t}\widetilde{\varphi}(\boldsymbol{\xi}) = 0.$$

Moreover, since $\boldsymbol{x}_c(\cdot,t)$ is affine on $\omega_c(0)$, its inverse is affine as well. Hence $\varphi(\boldsymbol{x},t)=\widetilde{\varphi}(\boldsymbol{x}_c^{-1}(\boldsymbol{x},t))\in\mathbb{P}^k(\omega_c(t))$. Applying the change of variables $\boldsymbol{x}=\boldsymbol{x}_c(\boldsymbol{\xi},t)$ to the left-hand side of (4.10), we obtain

$$\int_{\omega_c(t)}\phi_m^c(\boldsymbol{x},t)\,\varphi(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x} = \mathsf{J}_c(t)\int_{\omega_c(0)}\phi_m^c(\boldsymbol{x}_c(\boldsymbol{\xi},t),t)\,\widetilde{\varphi}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}.$$

Similarly, using $S_m^c(t)$ which is obtained through the transformation of $S_m^c(0)$ via the mapping $\boldsymbol{x}_c$, the right-hand side becomes

$$\int_{S_m^c(t)}\varphi(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x} = \mathsf{J}_c(t)\int_{S_m^c(0)}\widetilde{\varphi}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi},$$

therefore, since $\mathsf{J}_c(t)>0$, identity (4.10) gives

$$\int_{\omega_c(0)}\phi_m^c(\boldsymbol{x}_c(\boldsymbol{\xi},t),t)\,\widetilde{\varphi}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} = \int_{S_m^c(0)}\widetilde{\varphi}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}. \tag{4.11}$$

On the other hand, by definition of the reference sub-resolution basis function $\widetilde{\phi}_m^c := \pi_c^k\mathbb{1}_m^c(\cdot,0)$, we have, for all $\widetilde{\varphi}\in\mathbb{P}^k(\omega_c(0))$,

$$\int_{\omega_c(0)}\widetilde{\phi}_m^c(\boldsymbol{\xi})\,\widetilde{\varphi}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} = \int_{S_m^c(0)}\widetilde{\varphi}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}. \tag{4.12}$$

Subtracting (4.12) from (4.11) yields

$$\int_{\omega_c(0)}\Big(\phi_m^c(\boldsymbol{x}_c(\boldsymbol{\xi},t),t) - \widetilde{\phi}_m^c(\boldsymbol{\xi})\Big)\,\widetilde{\varphi}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} = 0$$

for all $\widetilde{\varphi} \in \mathbb{P}^k(\omega_c(0))$. Since the ALE mapping is affine, the pullback $\phi_m^c(\boldsymbol{x}_c(\boldsymbol{\xi},t),t)$ belongs to $\mathbb{P}^k(\omega_c(0))$. Hence

$$\tau(\boldsymbol{\xi}) := \phi_m^c(\boldsymbol{x}_c(\boldsymbol{\xi},t),t) - \widetilde{\phi}_m^c(\boldsymbol{\xi}) \in \mathbb{P}^k(\omega_c(0)).$$

Taking $\widetilde{\varphi} = \tau$ gives $\|\tau\|^2_{L^2(\omega_c(0))} = 0$, therefore $\tau \equiv 0$ on $\omega_c(0)$, which proves $\phi_m^c(\boldsymbol{x}_c(\boldsymbol{\xi},t),t) = \widetilde{\phi}_m^c(\boldsymbol{\xi})$, for all $\boldsymbol{x}(\boldsymbol{\xi},t) \in S_m^c(t)$ and $t \in [0, T_{\max}]$. □

Remark 4.2.2 (*Invariance on the whole reference cell*). *The proposition is stated in terms of the trajectory of the subcell $S_m^c(t)$. However, the proof gives the stronger result $\phi_m^c(\boldsymbol{x}_c(\boldsymbol{\xi},t),t) = \widetilde{\phi}_m^c(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \omega_c(0)$. Thus, the indicator basis functions are invariant along the ALE trajectories on the whole cell $\omega_c(t)$, and not only on the subcell $S_m^c(t)$.*

Time-dependent physical basis functions

In our current implementation, the local basis functions $\Psi_c^{(i,j)}(\boldsymbol{x})$ are defined in physical (Eulerian) coordinates as tensor products of rescaled monomials:

$$\Psi_c^{(i,j)}(\boldsymbol{x}) := \left(\frac{x - x_{\mathrm{b}}^c}{h_x^c}\right)^i \left(\frac{y - y_{\mathrm{b}}^c}{h_y^c}\right)^j.$$

These basis functions are fixed in space and do not evolve in time, which makes them suitable for classical Eulerian schemes. However, in an ALE framework, the mesh may deform over time. It is then natural to introduce basis functions which follow the motion of the mesh. Let

$$\boldsymbol{x}_c(\cdot,t) : \omega_c(0) \ni \boldsymbol{\xi} \to \boldsymbol{x}_c(\boldsymbol{\xi},t) \in \omega_c(t),$$

be again the ALE mapping from the reference cell $\omega_c(0)$ to the physical cell $\omega_c(t)$. We denote by

$$\boldsymbol{\xi}_c(\cdot,t) : \omega_c(t) \ni \boldsymbol{x} \to \boldsymbol{\xi}_c(\boldsymbol{x},t) \in \omega_c(0),$$

the inverse mapping, so that $\boldsymbol{\xi}_c(\boldsymbol{x}_c(\boldsymbol{\xi},t),t) = \boldsymbol{\xi}$. Let now $\widetilde{\Psi}_c^{(i,j)}$ be a polynomial basis function defined on the reference cell $\omega_c(0)$. The associated time-dependent physical basis function is defined by

$$\Psi_c^{(i,j)}(\boldsymbol{x},t) := \widetilde{\Psi}_c^{(i,j)}(\boldsymbol{\xi}_c(\boldsymbol{x},t)), \qquad \boldsymbol{x} \in \omega_c(t). \tag{4.13}$$

This construction means that the value of the physical basis function at a moving physical point $\boldsymbol{x} = \boldsymbol{x}_c(\boldsymbol{\xi},t)$ is prescribed by the fixed reference basis function evaluated at the reference coordinate $\boldsymbol{\xi}$.

Proposition 4.2.3 (*Invariance of physical basis functions*). *Let $\widetilde{\Psi}_c^{(i,j)} \in \mathbb{P}^k(\omega_c(0))$ be a reference basis function on the fixed cell $\omega_c(0)$, and define the time-dependent physical basis function by*

$$\Psi_c^{(i,j)}(\boldsymbol{x},t) := \widetilde{\Psi}_c^{(i,j)}(\boldsymbol{\xi}_c(\boldsymbol{x},t)), \qquad \boldsymbol{x} \in \omega_c(t). \tag{4.14}$$

Then, under the ALE mapping $\boldsymbol{x} = \boldsymbol{x}_c(\boldsymbol{\xi},t)$, we have

$$\Psi_c^{(i,j)}(\boldsymbol{x}_c(\boldsymbol{\xi},t),t) = \widetilde{\Psi}_c^{(i,j)}(\boldsymbol{\xi}), \qquad \forall \boldsymbol{\xi} \in \omega_c(0), \quad t \in [0, T_{\max}].$$

In other words, the basis functions $\Psi_c^{(i,j)}$ follow the trajectory of the cell $\omega_c(t)$.

Proof. By definition of the time-dependent physical basis function, we have

$$\Psi_c^{(i,j)}(\boldsymbol{x},t) = \widetilde{\Psi}_c^{(i,j)}(\boldsymbol{\xi}_c(\boldsymbol{x},t)).$$

Evaluating this identity along the ALE trajectory $\boldsymbol{x} = \boldsymbol{x}_c(\boldsymbol{\xi},t)$ gives

$$\Psi_c^{(i,j)}(\boldsymbol{x}_c(\boldsymbol{\xi},t),t) = \widetilde{\Psi}_c^{(i,j)}\left(\boldsymbol{\xi}_c(\boldsymbol{x}_c(\boldsymbol{\xi},t),t)\right).$$

Since $\boldsymbol{\xi}_c(\cdot,t)$ is the inverse mapping of $\boldsymbol{x}_c(\cdot,t)$, we have $\boldsymbol{\xi}_c(\boldsymbol{x}_c(\boldsymbol{\xi},t),t) = \boldsymbol{\xi}$, therefore we get

$$\Psi_c^{(i,j)}(\boldsymbol{x}_c(\boldsymbol{\xi},t),t) = \widetilde{\Psi}_c^{(i,j)}(\boldsymbol{\xi}),$$

for all $\boldsymbol{\xi} \in \omega_c(0)$ and $t \in [0, T_{\max}]$. This proves that the physical basis functions are constant in reference coordinates along the ALE trajectories, and therefore follow the mesh motion. □

4.2.3 ALE integral formulation

Multiplying equation (3.1) by any test function $\varphi(\cdot,t) \in \mathbb{P}^k(\mathcal{T}_h(t))$ satisfying $\frac{\mathrm{d}}{\mathrm{d}t}\varphi(\boldsymbol{x}(\boldsymbol{\xi},t),t) = 0$, and integrating over the moving cell $\omega_c(t)$ yields:

$$\int_{\omega_c(t)} \varphi\, \partial_t \boldsymbol{v}\, \mathrm{d}\boldsymbol{x} + \int_{\omega_c(t)} \varphi\, \nabla_{\boldsymbol{x}} \cdot \mathbb{F}(\boldsymbol{v}, b)\, \mathrm{d}\boldsymbol{x} = \int_{\omega_c(t)} \varphi\, \boldsymbol{B}[b](\boldsymbol{v})\, \mathrm{d}\boldsymbol{x}.$$

Using the ALE identity (4.4) and the geometric conservation relation (4.7), we can write:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\int_{\omega_c(t)} \boldsymbol{v}\,\varphi\,\mathrm{d}\boldsymbol{x} &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{\omega_c(0)} \boldsymbol{v}\,\varphi \mathrm{J}\,\mathrm{d}\boldsymbol{\xi} = \int_{\omega_c(0)} \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}\varphi \mathrm{J}\,\mathrm{d}\boldsymbol{\xi} + \int_{\omega_c(0)} \boldsymbol{v}\,\varphi \partial_t \mathrm{J}\,\mathrm{d}\boldsymbol{\xi} \\
&= \int_{\omega_c(0)} \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}\,\varphi\,\mathrm{J}\,\mathrm{d}\boldsymbol{\xi} + \int_{\omega_c(0)} \boldsymbol{v}\,\varphi \mathrm{J}\,\nabla_{\boldsymbol{x}}\cdot\boldsymbol{w}^g\,\mathrm{d}\boldsymbol{\xi} = \int_{\omega_c(t)} \varphi\,\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}\,\mathrm{d}\boldsymbol{x} + \int_{\omega_c(t)} \varphi\,\boldsymbol{v}\,\nabla_{\boldsymbol{x}}\cdot\boldsymbol{w}^g\,\mathrm{d}\boldsymbol{x} \\
&= \int_{\omega_c(t)} \varphi\partial_t\boldsymbol{v}\,\mathrm{d}\boldsymbol{x} + \int_{\omega_c(t)} \varphi\,\boldsymbol{w}^g\cdot\nabla_{\boldsymbol{x}}\boldsymbol{v}\,\mathrm{d}\boldsymbol{x} + \int_{\omega_c(t)} \varphi\,\boldsymbol{v}\,\nabla_{\boldsymbol{x}}\cdot\boldsymbol{w}^g\,\mathrm{d}\boldsymbol{x} \\
&= \int_{\omega_c(t)} \varphi\,\partial_t\boldsymbol{v}\,\mathrm{d}\boldsymbol{x} + \int_{\omega_c(t)} \varphi\,\nabla_{\boldsymbol{x}}\cdot(\boldsymbol{v}\otimes\boldsymbol{w}^g)\,\mathrm{d}\boldsymbol{x}.
\end{aligned}$$

Therefore, we get:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\omega_c(t)} \boldsymbol{v}\,\varphi\,\mathrm{d}\boldsymbol{x} = \int_{\omega_c(t)} \varphi\,\partial_t\boldsymbol{v}\,\mathrm{d}\boldsymbol{x} + \int_{\omega_c(t)} \varphi\,\nabla_{\boldsymbol{x}}\cdot(\boldsymbol{v}\otimes\boldsymbol{w}^g)\,\mathrm{d}\boldsymbol{x},$$

and the original equation becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\omega_c(t)} \boldsymbol{v}\varphi\,\mathrm{d}\boldsymbol{x} + \int_{\omega_c(t)} \varphi\,\nabla_{\boldsymbol{x}}\cdot\mathbb{G}(\boldsymbol{v},b,\boldsymbol{w}^g)\,\mathrm{d}\boldsymbol{x} = \int_{\omega_c(t)} \varphi\,\boldsymbol{B}[b](\boldsymbol{v})\,\mathrm{d}\boldsymbol{x}, \tag{4.15}$$

where we set the modified ALE flux tensor $\mathbb{G}(\boldsymbol{v},b,\boldsymbol{w}^g) := \mathbb{F}(\boldsymbol{v},b) - \boldsymbol{v}\otimes\boldsymbol{w}^g$. s.t.

$$\mathbb{G}(\boldsymbol{v},b,\boldsymbol{w}^g) = \left(\mathbb{F}^x(\boldsymbol{v},b) - w_x^g\boldsymbol{v},\ \mathbb{F}^y(\boldsymbol{v},b) - w_y^g\boldsymbol{v}\right) = \begin{pmatrix} q_x - w_x^g\eta & q_y - w_y^g\eta \\ u_xq_x + \frac{\mathrm{g}\eta}{2}(\eta-2b) - w_x^g q_x & u_xq_y - w_y^g q_x \\ u_yq_x - w_x^g q_y & u_yq_y + \frac{\mathrm{g}\eta}{2}(\eta-2b) - w_y^g q_y \end{pmatrix}.$$

Remark 4.2.4 (*Eigenstructure of the ALE flux*). *The Jacobian matrix associated with the ALE flux has the same eigenvectors as the Eulerian shallow-water flux, while the eigenvalues are shifted by the normal grid velocity. More precisely, the eigenvalues are*

$$\lambda_0 = (\boldsymbol{u} - \boldsymbol{w}^g)\cdot\mathbf{n}, \qquad \lambda_\pm = (\boldsymbol{u} - \boldsymbol{w}^g)\cdot\mathbf{n} \pm \sqrt{\mathrm{g}H}.$$

Therefore, a natural ALE spectral radius is given by

$$\sigma_{\mathrm{ALE}} = |(\boldsymbol{u} - \boldsymbol{w}^g)\cdot\mathbf{n}| + \sqrt{\mathrm{g}H}. \tag{4.16}$$

4.3 Discrete formulations and ALE extension

For the sake of clarity, let us now introduce the following shortcuts, for a face Γ_{cv} being shared with elements $\omega_c(t)$ and $\omega_v(t)$:

$$\mathbb{F}^*_{cv} := \mathbb{F}(\boldsymbol{v}_h, b_h)_{|\Gamma_{cv}}\cdot\mathbf{n}_{cv} \quad \text{and} \quad \mathbb{G}^*_{cv} := \mathbb{G}(\boldsymbol{v}_h, b_h, \boldsymbol{w}^g)_{|\Gamma_{cv}}\cdot\mathbf{n}_{cv}$$

We consider now that the face neighbors of cell $\omega_c(t)$ are the $\omega_{v(k)}$, $k \in [\![1, \#\mathcal{F}_c(t)]\!]$. To approximate the ALE normal flux $\mathbb{G}^*_{cv(k)}$ on the k-th interface $\Gamma_{cv(k)}$, we use a consistent numerical flux of the form $\mathbb{G}^* = \mathbb{F}^* - (\boldsymbol{w}^g\cdot\mathbf{n})\mathbb{V}^*$, where $\mathbb{F}^*$ is the numerical flux associated with the Eulerian flux $\mathbb{F}$, and $\mathbb{V}^*$ is also a numerical flux consistent with $\boldsymbol{v}$:

$$\mathbb{F}^*\left(\boldsymbol{v}^-,\boldsymbol{v}^+,b^-,b^+,\mathbf{n}\right) := \frac{\mathbb{F}(\boldsymbol{v}^-,b^-)+\mathbb{F}(\boldsymbol{v}^+,b^+)}{2}\cdot\mathbf{n} - \frac{\sigma}{2}\left(\boldsymbol{v}^+ - \boldsymbol{v}^-\right), \tag{4.17}$$

$$\mathbb{V}^*\left(\boldsymbol{v}^-,\boldsymbol{v}^+,b^-,b^+,\mathbf{n}\right) := \frac{\boldsymbol{v}^+ + \boldsymbol{v}^-}{2} - \frac{\mathbb{F}(\boldsymbol{v}^+,b^+)-\mathbb{F}(\boldsymbol{v}^-,b^-)}{2\sigma}\cdot\mathbf{n}. \tag{4.18}$$

4.3.1 DG formulation and DG-ALE formulation

A semi-discrete global DG formulation of the NSW equations with topography source term may reads as: find $\boldsymbol{v}_h := (\eta_h, \boldsymbol{q}_h) \in \mathbf{P}^k(\mathcal{T}_h(t))$ such that, for all $\varphi_h \in \mathbb{P}^k(\mathcal{T}_h(t))$,

$$\big(\partial_t \boldsymbol{v}_h, \varphi_h\big)_{\mathcal{T}_h(t)} - \big(\mathbb{F}(\boldsymbol{v}_h, b_h), \nabla_{\boldsymbol{x}}\varphi_h\big)_{\mathcal{T}_h(t)} + \big\langle \mathbb{F}_\partial^{*,\mathrm{DG}}, \varphi_h\big\rangle_{\mathcal{T}_h(t)} = \big(\boldsymbol{B}[b_h](\boldsymbol{v}_h), \varphi_h\big)_{\mathcal{T}_h(t)}, \tag{4.19}$$

which locally traduces as: for all $\omega_c(t) \in \mathcal{T}_h(t)$, find $\boldsymbol{v}_h^c := (\eta_h^c, \boldsymbol{q}_h^c) \in \mathbf{P}^k(\omega_c(t))$ such that, for all $\varphi_h^c \in \mathbb{P}^k(\omega_c(t))$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\omega_c(t)} \boldsymbol{v}_h^c\varphi_h^c \,\mathrm{d}\boldsymbol{x} - \int_{\omega_c(t)} \mathbb{F}(\boldsymbol{v}_h^c, b_h^c)\cdot\nabla_{\boldsymbol{x}}\varphi_h^c \,\mathrm{d}\boldsymbol{x} + \sum_{\Gamma\in\mathcal{F}_c(t)}\int_\Gamma \mathbb{F}_\Gamma^{*,\mathrm{DG}}\varphi_h^c \,\mathrm{d}s = \int_{\omega_c(t)} \boldsymbol{B}[b_h^c](\boldsymbol{v}_h^c)\varphi_h^c \,\mathrm{d}\boldsymbol{x}. \tag{4.20}$$

Considering the ALE framework introduced earlier, a semi-discrete global DG-ALE formulation of the NSW equations can be obtained by substituting the classical flux by its ALE version. It reads: find $\boldsymbol{v}_h := (\eta_h, \boldsymbol{q}_h) \in \mathbf{P}^k(\mathcal{T}_h(t))$ such that, for all transported test functions $\varphi_h \in \mathbb{P}^k(\mathcal{T}_h(t))$ satisfying $\frac{\mathrm{d}}{\mathrm{d}t}\varphi_h(\boldsymbol{x}(\boldsymbol{\xi}, t), t) = 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(\boldsymbol{v}_h, \varphi_h\big)_{\mathcal{T}_h(t)} - \big(\mathbb{G}(\boldsymbol{v}_h, b_h, \boldsymbol{w}_h^g), \nabla_{\boldsymbol{x}}\varphi_h\big)_{\mathcal{T}_h(t)} + \big\langle \mathbb{G}_\partial^{*,\mathrm{DG}}, \varphi_h\big\rangle_{\mathcal{T}_h(t)} = \big(\boldsymbol{B}[b_h](\boldsymbol{v}_h), \varphi_h\big)_{\mathcal{T}_h(t)}, \tag{4.21}$$

that can be again locally translated as: for all $\omega_c(t) \in \mathcal{T}_h(t)$, find $\boldsymbol{v}_h^c := (\eta_h^c, \boldsymbol{q}_h^c) \in \mathbf{P}^k(\omega_c(t))$ such that, for all $\varphi_h^c \in \mathbb{P}^k(\omega_c(t))$ satisfying $\frac{\mathrm{d}}{\mathrm{d}t}\varphi_h^c(\boldsymbol{x}(\boldsymbol{\xi}, t), t) = 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\omega_c(t)} \boldsymbol{v}_h^c\varphi_h^c \,\mathrm{d}\boldsymbol{x} - \int_{\omega_c(t)} \mathbb{G}(\boldsymbol{v}_h^c, b_h^c, \boldsymbol{w}_h^{g,c})\cdot\nabla_{\boldsymbol{x}}\varphi_h^c \,\mathrm{d}\boldsymbol{x} + \sum_{\Gamma\in\mathcal{F}_c(t)}\int_\Gamma \mathbb{G}_\Gamma^{*,\mathrm{DG}}\varphi_h^c \,\mathrm{d}s = \int_{\omega_c(t)} \boldsymbol{B}[b_h^c](\boldsymbol{v}_h^c)\varphi_h^c \,\mathrm{d}\boldsymbol{x}. \tag{4.22}$$

4.3.2 Monolithic DG/FV-ALE subcell formulation

Let us now follow a similar procedure as the classical DG/FV subcell reconstruction, but in the context of the ALE framework. The proof is similar to the non-ALE case, the reader can skip it if he is familiar with the subcell reconstruction procedure. The main idea is to recast the DG scheme (4.22) into a set of N_s subcell schemes, one for each subcell $S_m^c(t)$, $m \in [\![1, N_s]\!]$, where the solution is approximated by its mean value $\overline{\boldsymbol{v}}_m^c$ on each subcell.

Theorem 4.3.1 (*Equivalence between DG-ALE and subcell FV-ALE formulations*). *For the Saint-Venant equations, considering any $\varphi \in \mathbb{P}^k(\omega_c(t))$ s.t. $\frac{\mathrm{d}}{\mathrm{d}t}\varphi(\boldsymbol{x}(\boldsymbol{\xi}, t), t) = 0$, the classical unlimited DG-ALE scheme in cell $\omega_c(t)$*

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\omega_c(t)} \boldsymbol{v}_h^c\varphi \,\mathrm{d}\boldsymbol{x} - \int_{\omega_c(t)} \mathbb{G}(\boldsymbol{v}_h^c, b_h^c, \boldsymbol{w}_h^{g,c})\cdot\nabla_{\boldsymbol{x}}\varphi \,\mathrm{d}\boldsymbol{x} + \sum_{k=1}^{\#\mathcal{F}_c}\int_{\Gamma_{cv(k)}} \mathbb{G}_{cv(k)}^{*,\mathrm{DG}}\varphi \,\mathrm{d}s = \int_{\omega_c(t)} \boldsymbol{B}(\boldsymbol{v}_h^c, \nabla_{\boldsymbol{x}} b_h^c)\varphi \,\mathrm{d}\boldsymbol{x}, \tag{4.23}$$

is equivalent to the following formulation for all $m \in [\![1, N_s]\!]$

$$\frac{\mathrm{d}}{\mathrm{d}t}(|S_m^c(t)|\,\overline{\boldsymbol{v}}_m^c) = -\left(\sum_{S_p^v(t)\in\mathcal{V}_m^c(t)}\int_{\Gamma_{mp}^c} \mathbb{G}_h^c\cdot\mathbf{n}_{mp} \,\mathrm{d}s - \sum_{k=1}^{\#\mathcal{F}_c}\int_{\Gamma_{cv(k)}} \left(\mathbb{G}_h^c\cdot\mathbf{n}_{cv(k)} - \mathbb{G}_{cv(k)}^{*,\mathrm{DG}}\right)\widetilde{\phi}_m^c \,\mathrm{d}s\right) + |S_m^c(t)|\,\overline{\boldsymbol{B}}_m^c,$$

that can be recast into N_s FV-like subcell schemes as

$$\frac{\mathrm{d}}{\mathrm{d}t}(|S_m^c(t)|\,\overline{\boldsymbol{v}}_m^c) = -\sum_{S_p^v(t)\in\mathcal{V}_m^c} \ell_{mp}\widehat{\mathbb{G}}_{mp} + |S_m^c(t)|\,\overline{\mathbf{B}}_m^c, \qquad \forall m \in [\![1, N_s]\!], \tag{4.24}$$

allowed by the DG-ALE reconstructed flux that writes as

$$\widehat{\mathbb{G}}_c = \mathbb{G}_c - \mathbb{A}_c^\top \mathscr{L}_c^{-1}\partial\mathbb{G}_c, \tag{4.25}$$

where:

- $\mathbb{R}^{N_f^c} \ni \mathbb{G}_c$ *is the vector of interior fluxes s.t. for all $S_m^c(t) \in \mathcal{T}_{\omega_c}^k((t)$,*

$$(\ell_{mp}\mathbb{G}_{mp}^c)_{p\in[\![1,N_f^c]\!]} := \int_{\Gamma_{mp}^c} \mathbb{G}_h^c\cdot\mathbf{n}_{mp} \,\mathrm{d}s,$$

– $\mathscr{M}_{N_k \times N_f^c}(\mathbb{R}) \ni \mathbb{A}_c$ *stands for the adjacency matrix of cell* $\omega_c(t)$*, where*

$$\left(\mathbb{A}^c_{mp}\right)_{(m,p)\in[\![1,N_k]\!]\times[\![1,N_f^c]\!]} = \varepsilon^c_{mp},$$

– $\mathbb{R}^{N_s} \ni \partial\mathbb{G}_c$ *contains the following boundary contribution*

$$(\partial\mathbb{G}^c_m)_{m\in[\![1,N_s]\!]} = \int_{\partial\omega_c(t)} \left(\mathbb{G}^c_h \cdot \mathbf{n}_{\partial\omega_c(t)} - \mathbb{G}^{*,\mathrm{DG}}_{\partial\omega_c(t)}\right) \widehat{\phi}^c_m(s)\,\mathrm{d}s,$$

where $\widehat{\phi}^c_m$ *is defined as follows:*

$$\widehat{\phi}^c_m(s) := \begin{cases} \phi^c_m(s) & \text{if } s \in \partial\omega_c(t) \setminus \partial S^c_m(t), \\ \phi^c_m(s) - 1 & \text{if } s \in \partial\omega_c(t) \cap \partial S^c_m(t). \end{cases}$$

– $\mathscr{M}_{N_f^c \times N_s}(\mathbb{R}) \ni \mathscr{L}_c^{-1}$ *is the inverse of the Laplacian matrix of the interior subgrid graph* $\mathbb{L}_c := \mathbb{A}_c^t\mathbb{A}_c$ *on the orthogonal of its kernel. For any* $\lambda \neq 0$*, this generalized inverse writes:*

$$\mathscr{L}_c^{-1} = (\mathbb{L}_c + \lambda\Pi)^{-1} - \frac{1}{\lambda}\Pi \tag{4.26}$$

with $\Pi = \frac{1}{N_s}(\mathbf{1}\otimes\mathbf{1}) \in \mathscr{M}_{N_k}(\mathbb{R})$.

Proof. Let us consider $\mathbb{G}_h := \pi_h^k\mathbb{G}(\boldsymbol{v}_h, b_h, \boldsymbol{w}_h^{g,c})$ and $\boldsymbol{B}_h := \pi_h^k\boldsymbol{B}[b_h^c](\boldsymbol{v}_h^c)$, the L^2-projections of the flux function and source term onto $\mathbb{P}^k(\mathcal{T}_h)$. By replacing now the flux function and the source term with their L^2-projections in (4.20), we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\omega_c(t)} \boldsymbol{v}_h^c\varphi_h^c\,\mathrm{d}\boldsymbol{x} - \int_{\omega_c(t)} \mathbb{G}_h^c\cdot\nabla_{\boldsymbol{x}}\varphi_h^c\,\mathrm{d}\boldsymbol{x} + \sum_{k=1}^{\#\mathcal{F}_c}\int_{\Gamma_{cv(k)}} \mathbb{G}^{*,\mathrm{DG}}_{cv(k)}\varphi_h^c\,\mathrm{d}s = \int_{\omega_c(t)} \boldsymbol{B}_h^c\varphi_h^c\,\mathrm{d}\boldsymbol{x}, \quad \forall\varphi_h^c \in \mathbb{P}^k(\omega_c(t)).$$

Equivalently, using an integration by part, we get, for all $\varphi_h^c \in \mathbb{P}^k(\omega_c(t))$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\omega_c(t)} \boldsymbol{v}_h^c\varphi_h^c\,\mathrm{d}\boldsymbol{x} - \sum_{k=1}^{\#\mathcal{F}_c}\int_{\Gamma_{cv(k)}} \left(\mathbb{G}_h^c\cdot\mathbf{n}_{cv(k)} - \mathbb{G}^{*,\mathrm{DG}}_{cv(k)}\right)\varphi_h^c\,\mathrm{d}s + \int_{\omega_c(t)} \varphi_h^c\nabla_{\boldsymbol{x}}\cdot\mathbb{G}_h^c\,\mathrm{d}\boldsymbol{x} = \int_{\omega_c(t)} \boldsymbol{B}_h^c\varphi_h^c\,\mathrm{d}\boldsymbol{x}. \tag{4.27}$$

Now, since $\boldsymbol{v}_h^c$, $\nabla_{\boldsymbol{x}}\cdot\mathbb{G}_h^c$ and $\boldsymbol{B}_h^c$ belong to $\mathbf{P}^k(\omega_c(t))$, we can test (4.27) with the sub-resolution spanning functions ϕ_m^c. Let us remark that here, unlike in Chapter 3, no tilde is used for these functions, since the tilde notation is reserved for quantities written on the reference configuration in the ALE framework. This gets us, for all $m \in [\![1, N_s]\!]$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\omega_c(t)} \boldsymbol{v}_h^c\phi_m^c\,\mathrm{d}\boldsymbol{x} - \sum_{k=1}^{\#\mathcal{F}_c}\int_{\Gamma_{cv(k)}} \left(\mathbb{G}_h^c\cdot\mathbf{n}_{cv(k)} - \mathbb{G}^{*,\mathrm{DG}}_{cv(k)}\right)\phi_m^c\,\mathrm{d}s + \int_{\omega_c(t)} \phi_m^c\nabla_{\boldsymbol{x}}\cdot\mathbb{G}_h^c\,\mathrm{d}\boldsymbol{x} = \int_{\omega_c(t)} \boldsymbol{B}_h^c\phi_m^c\,\mathrm{d}\boldsymbol{x},$$

thus we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\omega_c(t)} \boldsymbol{v}_h^c\,\mathrm{d}\boldsymbol{x} = \sum_{k=1}^{\#\mathcal{F}_c}\int_{\Gamma_{cv(k)}} \left(\mathbb{G}_h^c\cdot\mathbf{n}_{cv(k)} - \mathbb{G}^{*,\mathrm{DG}}_{cv(k)}\right)\phi_m^c\,\mathrm{d}s - \int_{\omega_c(t)} \nabla_{\boldsymbol{x}}\cdot\mathbb{G}_h^c\,\mathrm{d}\boldsymbol{x} + \int_{\omega_c(t)} \boldsymbol{B}_h^c\,\mathrm{d}\boldsymbol{x}.$$

We can then rewrite this with submean values to get

$$\frac{\mathrm{d}}{\mathrm{d}t}(|S_m^c(t)|\,\overline{\boldsymbol{v}}_m^c) = \left(\sum_{S_p^v(t)\in\mathcal{V}_m^c(t)}\int_{\Gamma^c_{mp}} \mathbb{G}_h^c\cdot\mathbf{n}_{mp}\,\mathrm{d}s - \sum_{k=1}^{\#\mathcal{F}_c}\int_{\Gamma_{cv(k)}} \left(\mathbb{G}_h^c\cdot\mathbf{n}_{cv(k)} - \mathbb{G}^{*,\mathrm{DG}}_{cv(k)}\right)\phi_m^c\,\mathrm{d}s\right) + |S_m^c(t)|\,\overline{\boldsymbol{B}}_m^c \tag{4.28}$$

for all $m \in [\![1, N_s]\!]$, where $\overline{\boldsymbol{B}}_m^c$ and $\overline{\boldsymbol{v}}_m^c$ are respectively the submean values of $\boldsymbol{B}_h$ and $\boldsymbol{v}_h$ on subcell $S_m^c(t)$, ie. $\overline{\boldsymbol{B}}_m^c := \frac{1}{|S_m^c(t)|}\int_{S_m^c(t)} \boldsymbol{B}_h^c\,\mathrm{d}\boldsymbol{x}$ and $\overline{\boldsymbol{v}}_m^c := \frac{1}{|S_m^c(t)|}\int_{S_m^c(t)} \boldsymbol{v}_h^c\,\mathrm{d}\boldsymbol{x}$. As in the classical non-ALE case, we need to

impose the equality between reconstructed flux $\widehat{\mathbb{G}}_n$ and DG numerical flux $\mathbb{G}_n^{*,\mathrm{DG}}$ on the boundary $\partial\omega_c(t)$, i.e. $\widehat{\mathbb{G}}_{n|\partial\omega} = \mathbb{G}_n^{*,\mathrm{DG}}$. Noticing that numerical flux term

$$\sum_{S_p^v(t)\in\mathscr{V}_m^c(t)} \int_{\Gamma_{mp}^c} \widehat{\mathbb{G}}_n \,\mathrm{d}s = \sum_{S_p^v(t)\in\mathscr{V}_m^c(t)} \int_{\Gamma_{mp}^c} \mathbb{G}_h^c \cdot \mathbf{n}_{mp} \,\mathrm{d}s - \sum_{k=1}^{\#\mathscr{T}_c} \int_{\Gamma_{cv(k)}^c} \left(\mathbb{G}_h^c \cdot \mathbf{n}_{cv(k)} - \mathbb{G}_{cv(k)}^*\right) \phi_m^c \,\mathrm{d}s,$$

can be equivalently written as

$$\int_{\partial S_m^c(t)} \widehat{\mathbb{G}}_n \,\mathrm{d}s = \int_{\partial S_m^c(t)} \mathbb{G}_h^c \cdot \mathbf{n} \,\mathrm{d}s - \int_{\partial\omega_c(t)} \left(\mathbb{G}_h^c \cdot \mathbf{n}_{\partial\omega_c(t)} - \mathbb{G}_{\partial\omega_c(t)}^{*,\mathrm{DG}}\right) \phi_m^c \,\mathrm{d}s, \tag{4.29}$$

we directly get

$$\int_{\partial S_m^c(t)\setminus\partial\omega_c(t)} \widehat{\mathbb{G}}_n \,\mathrm{d}s = \int_{\partial S_m^c(t)\setminus\partial\omega_c(t)} \mathbb{G}_h^c \cdot \mathbf{n} \,\mathrm{d}s - \int_{\partial\omega_c(t)} \left(\mathbb{G}_h^c \cdot \mathbf{n}_{\partial\omega_c(t)} - \mathbb{G}_{\partial\omega_c(t)}^{*,\mathrm{DG}}\right) \widehat{\phi}_m^c \,\mathrm{d}s, \tag{4.30}$$

where $\widehat{\phi}_m^c$ is defined above. As in the non-ALE case, we have

$$\int_{\Gamma_{mp}^c} \widehat{\mathbb{G}}_n \,\mathrm{d}s = \varepsilon_{mp}^c \widehat{\mathbb{G}}_{mp} \quad \text{and} \quad \int_{\Gamma_{mp}^c} \mathbb{G}_h^c \cdot \mathbf{n} \,\mathrm{d}s = \varepsilon_{mp}^c \mathbb{G}_{mp}.$$

Now, considering the vectors $\mathbb{G}_c \in \mathbb{R}^{N_f^c}$ and $\widehat{\mathbb{G}}_c \in \mathbb{R}^{N_f^c}$ defined earlier gets us to the following:

$$\mathbb{A}_c \widehat{\mathbb{G}}_c = \mathbb{A}_c \mathbb{G}_c - \partial \mathbb{G}_c.$$

$\mathbb{A}_c$ being not necessarily a square matrix and since $\ker \mathbb{A}_c^\top \neq \mathbf{0}$, we again make use of the graph Laplacian technique employed in [2, 267], allowing us to get the following reconstructed flux

$$\widehat{\mathbb{G}}_c = \mathbb{G}_c - \mathbb{A}_c^\top \mathscr{L}_c^{-1} \partial \mathbb{G}_c,$$

what we wanted. □

4.4 Computation of the ALE grid velocity

In a general ALE framework, the grid velocity is not directly prescribed by the underlying physics and must therefore be chosen so as to maintain mesh regularity and preserve element quality throughout the deformation. At the same time, it must satisfy the continuous GCL (4.7) in order to ensure consistency between the mesh motion and the discretization of the governing PDE.

4.4.1 From boundary motion to mesh motion

In this context, the mesh motion is determined once the velocity is prescribed on the moving part of the boundary. More precisely, let $\partial\Omega_{\mathrm{mov}}(t) \subset \partial\Omega(t)$ denote the portion of the boundary whose motion is imposed by the geometry, and let $\partial\Omega_{\mathrm{fix}} := \partial\Omega(t) \setminus \partial\Omega_{\mathrm{mov}}(t)$ denote the fixed part. The grid velocity is then required to satisfy

$$\boldsymbol{w}^g(\boldsymbol{x},t) = \mathbf{0} \quad \text{on } \partial\Omega_{\mathrm{fix}}, \qquad \boldsymbol{w}^g(\boldsymbol{x},t) = \partial_t \boldsymbol{x}_{\mathrm{mov}}(t) \quad \text{on } \partial\Omega_{\mathrm{mov}}(t),$$

where $\boldsymbol{x}_{\mathrm{mov}}(\cdot,t)$ denotes a parametrization of the prescribed moving boundary.

The remaining task is therefore to construct an interior extension of these boundary data such that the ALE mapping

$$\Omega(0) \ni \boldsymbol{\xi} \mapsto \boldsymbol{x}(\boldsymbol{\xi},t) \in \Omega(t)$$

remains smooth and non-degenerate. In one space dimension, such a construction may sometimes be written explicitly (see Subsection 4.4.4). On two-dimensional unstructured meshes, however, the grid velocity is more naturally obtained as the solution of an auxiliary problem.

4.4.2 On the choice of the mesh velocity

The definition of the grid velocity $\boldsymbol{w}^g$ is a central ingredient of the ALE formulation. Several choices are possible, leading to different behaviors in terms of mesh deformation and numerical robustness:

- If one considers the *Eulerian* framework, the mesh is kept fixed, *i.e* $\boldsymbol{w}^g = \mathbf{0}$. This completely avoids mesh deformation and is therefore very robust from a geometric point of view. However, all transport is then resolved through the numerical fluxes, and the mesh cannot follow moving boundaries, free-surfaces, or fluid-structure interfaces. This corresponds to the classical fixed-grid formulation discussed in Chapter 3, and does not provide a suitable framework when the computational domain itself evolves in time.
- A natural alternative is then the purely *Lagrangian* choice $\boldsymbol{w}^g = \boldsymbol{u}$, where $\boldsymbol{u}$ denotes the fluid velocity. In this case, the mesh follows the motion of the fluid particles, so that the transport is described in a frame moving with the flow. The relative advective velocity $\boldsymbol{u} - \boldsymbol{w}^g$ then vanishes, which reduces the numerical diffusion associated with advection and allows material interfaces to be tracked accurately. However, in practical situations involving shear, strong gradients or recirculations, this approach can rapidly lead to severe mesh distortion, what we would like to avoid. Also, let us remark that in order to do so, one needs to have node solvers for the fluid velocity, which increases the complexity of the numerical scheme.

The ALE framework provides an intermediate strategy in which the mesh velocity is chosen independently of the fluid velocity. The goal is not to make the mesh follow every fluid particle, but rather to move it only where it is useful, typically near moving boundaries or interfaces, while keeping the deformation sufficiently smooth in the interior of the domain. This idea goes back to the early ALE formulations introduced in finite difference and finite element contexts [91, 146, 149], and has since become a useful tool for moving-domain and fluid-structure interaction problems [92, 141, 142, 277]. In the present work, this is achieved by defining $\boldsymbol{w}^g$ as a smooth extension of the prescribed boundary velocity, obtained as the solution of an auxiliary elliptic problem. Such a choice distributes the deformation over the mesh, reduces the risk of cell distortion compared with a purely Lagrangian motion, and remains compatible with moving-boundary configurations. It is particularly relevant for high-order discretizations, for which the accuracy of quadrature formulas, polynomial reconstructions, and numerical fluxes depends strongly on the quality of the deformed elements, and where the CFL condition may become very restrictive if the mesh is excessively distorted.

Let $\partial\Omega_{\text{ext}}$ denote the fixed external boundary of the domain, and let $\partial\Omega_{\text{mov}}(t)$ denote the portion of the boundary where the motion is prescribed. We consider two standard elliptic models to compute the ALE grid velocity field $\boldsymbol{w}^g := (w_x^g, w_y^g)$. Such elliptic mesh-motion strategies are classical in ALE computations, where the motion imposed on the boundary has to be extended smoothly inside the computational domain while preserving the quality of the moving mesh [92, 277].

The harmonic extension model

The simplest approach is to define the grid velocity as the harmonic extension of the boundary displacement, which amounts to solving the following *Laplace equation*:

$$\begin{cases} -\nabla_{\boldsymbol{x}} \cdot \left(\kappa(\boldsymbol{x}) \nabla_{\boldsymbol{x}} \boldsymbol{w}^g\right) = 0 & \text{in } \Omega(t), \\ \boldsymbol{w}^g = \mathbf{0} & \text{on } \partial\Omega_{\text{ext}}, \\ \boldsymbol{w}^g = \partial_t \boldsymbol{x}_{\text{mov}}(t) & \text{on } \partial\Omega_{\text{mov}}(t), \end{cases} \tag{4.31}$$

where $\boldsymbol{x}_{\text{mov}}(t)$ denotes the prescribed physical position of the moving boundary. This type of Laplacian extension is commonly used in ALE formulations to propagate a boundary velocity into the interior of the domain in a smooth way [98, 277]. The diffusion coefficient $\kappa(\boldsymbol{x}) > 0$ controls the local mesh stiffness. Setting $\kappa = 1$ yields the pure Laplace equation, which provides a smooth interior velocity field but may struggle to preserve element quality in regions with strong mesh grading or large boundary motion. To protect highly refined elements from excessive distortion, a common strategy consists in using a variable diffusion coefficient, in the spirit of Winslow-type mesh smoothing and related elliptic mesh deformation techniques [124, 145, 279]. In practice, this amounts to choosing $\kappa(\boldsymbol{x})$ as a local stiffness parameter, often depending on the mesh size, so that smaller elements are made more resistant to deformation.

The pseudo-elasticity model

Because the harmonic extension (4.31) minimizes the gradient of each velocity component independently, it offers no intrinsic resistance to shear or local compression, which may still cause small elements near the moving

boundary to invert under severe deformations. To overcome this limitation, one can treat the fluid mesh as an artificial elastic solid governed by the equations of *linear elasticity*. This pseudo-elastic point of view has also been used in ALE and fluid-structure interaction computations, in particular to propagate boundary displacements inside the domain while preserving the quality of highly deformed meshes [164, 250, 277].

$$\begin{cases} -\nabla_{\boldsymbol{x}} \cdot \boldsymbol{\sigma}(\boldsymbol{w}^g) = \mathbf{0} & \text{in } \Omega(t), \\ \boldsymbol{\sigma}(\boldsymbol{w}^g) := 2\mu(\boldsymbol{x})\boldsymbol{\varepsilon}(\boldsymbol{w}^g) + \lambda(\boldsymbol{x})(\nabla_{\boldsymbol{x}} \cdot \boldsymbol{w}^g)\mathbb{I}_2 & \text{in } \Omega(t), \\ \boldsymbol{\varepsilon}(\boldsymbol{w}^g) := \frac{1}{2}\left\{\nabla_{\boldsymbol{x}}\boldsymbol{w}^g + (\nabla_{\boldsymbol{x}}\boldsymbol{w}^g)^\top\right\} & \text{in } \Omega(t), \\ \boldsymbol{w}^g = \mathbf{0}, & \text{on } \partial\Omega_{\text{ext}}, \\ \boldsymbol{w}^g = \partial_t \boldsymbol{x}_{\text{mov}}(t) & \text{on } \partial\Omega_{\text{mov}}(t). \end{cases} \tag{4.32}$$

Here, $\boldsymbol{\sigma}(\boldsymbol{w}^g)$ denotes the Cauchy stress tensor and $\boldsymbol{\varepsilon}(\boldsymbol{w}^g)$ is the symmetric strain tensor. The Lamé parameters μ and λ characterize the fictitious material properties. Choosing a Poisson's ratio ν close to 0.5 mimics an incompressible behaviour, thus strongly penalizing local volume variations and helping to prevent element inversion. Similarly to the harmonic extension, μ and λ can be scaled locally so that the smallest elements are made stiffer than the larger ones, a strategy commonly used in elasticity-based mesh-motion methods for large-displacement configurations [246, 250, 256].

Remark 4.4.1 (*Boundary conditions*). *In systems* (4.31) *and* (4.32)*, homogeneous Dirichlet boundary conditions* ($\boldsymbol{w}^g = \mathbf{0}$) *are prescribed on the entire external boundary* $\partial\Omega_{\text{ext}}$*, which strictly fixes the nodes. In practice, however, to accommodate large internal deformations and prevent severe element shearing near the boundaries, it is often more robust to enforce a slip condition on straight walls. This is achieved by constraining only the normal component of the grid velocity* ($\boldsymbol{w}^g \cdot \mathbf{n} = 0$) *while leaving the tangential component free, thus allowing the mesh nodes to slide along the boundary.*

In Figure 4.1, we compare two ALE mesh-motion strategies on a simple benchmark: a circular internal boundary is prescribed an oscillatory motion inside the computational mesh, while the outer boundary is held fixed. The interior node displacement is then computed either by a Laplacian extension or by a pseudo-elasticity model. While both approaches follow the imposed motion correctly, the pseudo-elasticity model distributes the deformation more uniformly and preserves mesh quality better near the moving boundary.

4.4.3 Finite element formulations for the grid velocity

Since the mesh update is performed at the vertices, it is natural to approximate the grid velocity field $\boldsymbol{w}^g(\boldsymbol{x}, t) \in \mathbb{R}^2$ with nodal (vertex-based) continuous *Lagrange elements*. We therefore discretize the elliptic models with $\mathbb{P}^1$ finite elements on the current triangulation $\mathcal{T}_h(t)$.

For this purpose we introduce the standard continuous piecewise affine space

$$\mathsf{W}_h(t) := \left\{ w \in \mathscr{C}^0(\overline{\Omega}) \mid \forall \omega_c(t) \in \mathcal{T}_h(t),\ w_{|\omega_c(t)} \in \mathbb{P}^1(\omega_c(t)) \right\},$$

and the corresponding vector-valued space $\mathbf{W}_h(t) := (\mathsf{W}_h(t))^2$. Let $\{\varphi_i\}_{i \in [\![1, N_v]\!]}$ be the nodal $\mathbb{P}^1$ basis associated with the vertices $\{\boldsymbol{x}_i(t)\}_{i \in [\![1, N_v]\!]}$, satisfying $\varphi_i(\boldsymbol{x}_j(t), t) = \delta_{ij}$. We approximate the grid velocity as

$$\boldsymbol{w}_h^g(\boldsymbol{x}, t) = \sum_{i=1}^{N_v} \underline{\boldsymbol{w}}_i^g(t)\, \varphi_i(\boldsymbol{x}),$$

so that the degrees of freedom $\{\underline{\boldsymbol{w}}_i^g\}_{i \in [\![1, N_v]\!]}$ are exactly the nodal grid velocities in the x and y-directions.

We define the discrete Dirichlet space and the associated test space with homogeneous boundary conditions as follows:

$$\mathbf{W}_{h,\mathrm{D}}(t) := \left\{ \boldsymbol{w}_h \in \mathbf{W}_h(t) \mid \boldsymbol{w}_h = \mathbf{0} \text{ on } \partial\Omega_{\text{fix}},\ \boldsymbol{w}_h = \partial_t \boldsymbol{x}_{\text{mov}}(t) \text{ on } \partial\Omega_{\text{mov}}(t) \right\},$$
$$\mathbf{W}_{h,0}(t) := \left\{ \boldsymbol{w}_h \in \mathbf{W}_h(t) \mid \boldsymbol{w}_h = \mathbf{0} \text{ on } \partial\Omega_{\text{fix}} \cup \partial\Omega_{\text{mov}}(t) \right\}.$$

The $\mathbb{P}^1$ FE weak formulation for the harmonic extension (4.31) then reads: find $\boldsymbol{w}_h^g(\cdot, t) \in \mathbf{W}_{h,\mathrm{D}}(t)$ such that for all $\boldsymbol{z}_h \in \mathbf{W}_{h,0}(t)$,

$$\left(\kappa(\boldsymbol{x})\nabla_{\boldsymbol{x}}\boldsymbol{w}_h^g(\boldsymbol{x}, t), \nabla_{\boldsymbol{x}}\boldsymbol{z}_h(\boldsymbol{x}, t)\right)_{\Omega(t)} = 0.$$

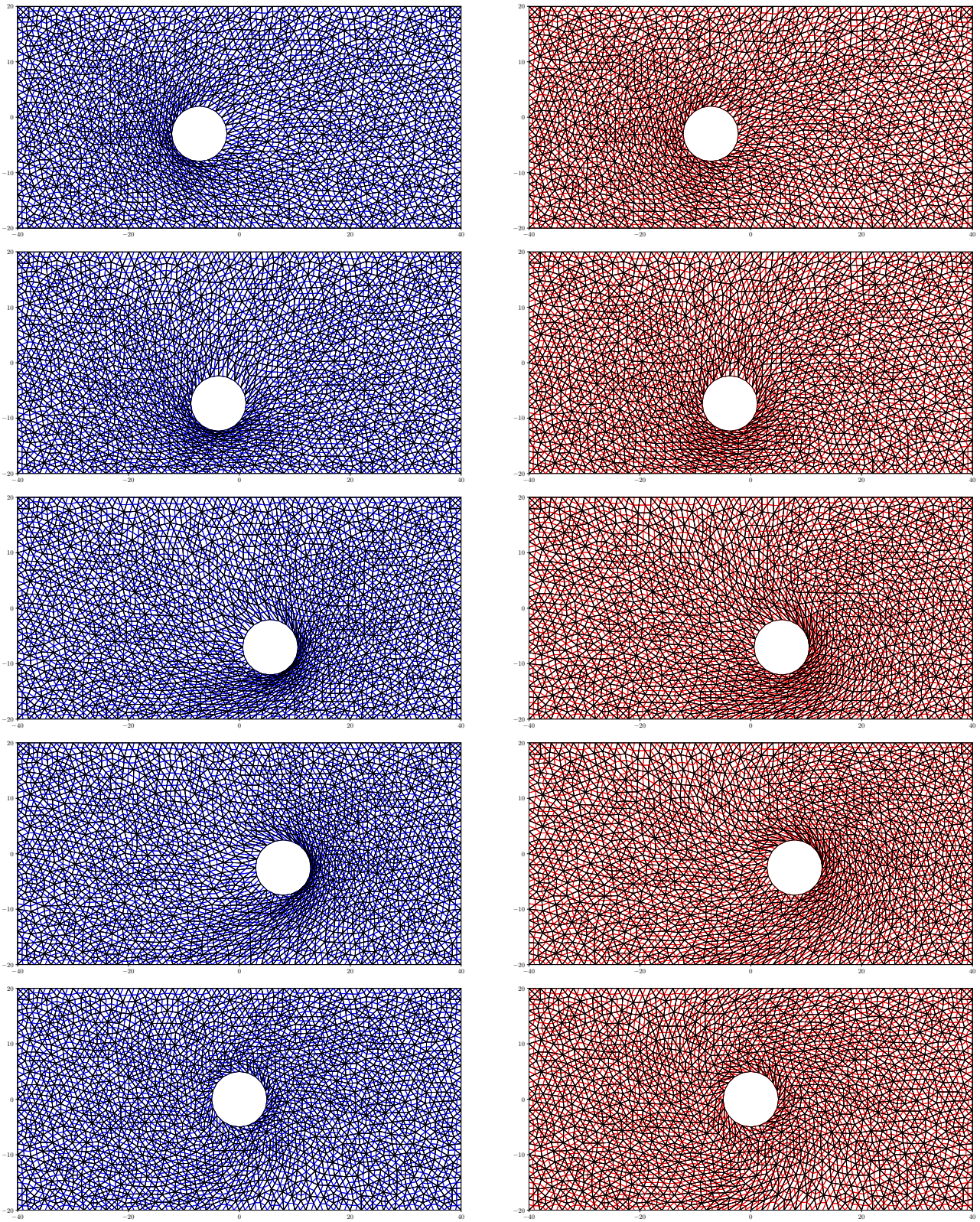


Figure 4.1: Evolution of the ALE mesh/submesh around an oscillating cylinder at five consecutive time steps. Left: standard Laplacian smoothing (animation). Right: pseudo-elasticity approach (animation).

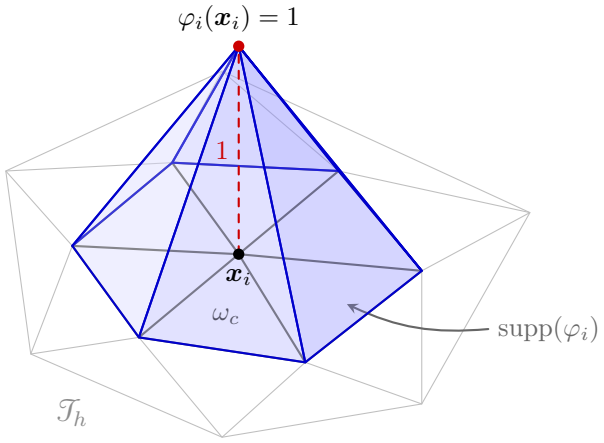


Figure 4.2: Typical continuous piecewise-linear ($\mathbb{P}^1$) Lagrange basis function $\varphi_i(\boldsymbol{x})$ associated with the degree of freedom $\boldsymbol{x}_i$ on a two-dimensional triangular mesh $\mathcal{T}_h$. The function takes the value 1 at its associated node and vanishes at all other vertices, ensuring $\mathsf{W}_h \subset \mathsf{H}^1(\Omega)$.

Let us now derive the weak formulation associated with the pseudo-elasticity model (4.32) at the discrete level. Multiplying the equation by a vector-valued test function $\boldsymbol{z}_h \in \mathbf{W}_{h,0}(t)$ and integrating over the current domain $\Omega(t)$ yields

$$-\int_{\Omega(t)} \big(\nabla_{\boldsymbol{x}} \cdot \boldsymbol{\sigma}(\boldsymbol{w}_h^g)\big) \cdot \boldsymbol{z}_h(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x} = 0,$$

and applying an integration by parts gives

$$\int_{\Omega(t)} \boldsymbol{\sigma}(\boldsymbol{w}_h^g) : \nabla_{\boldsymbol{x}} \boldsymbol{z}_h(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x} - \int_{\partial\Omega(t)} \big(\boldsymbol{\sigma}(\boldsymbol{w}_h^g)\mathbf{n}\big) \cdot \boldsymbol{z}_h(\boldsymbol{x},t)\,\mathrm{d}s = 0.$$

Since the test functions satisfy $\boldsymbol{z}_h = \mathbf{0}$ on the domain boundaries $\partial\Omega_{\mathrm{ext}} \cup \partial\Omega_{\mathrm{mov}}(t)$, the boundary integral vanishes. Next, using the expression of the Cauchy stress tensor and decomposing the gradient of the test function into its symmetric part $\boldsymbol{\varepsilon}(\boldsymbol{z}_h)$ and skew-symmetric part $\boldsymbol{\vartheta}(\boldsymbol{z}_h) = \frac{1}{2}\{\nabla_{\boldsymbol{x}}\boldsymbol{z}_h - (\nabla_{\boldsymbol{x}}\boldsymbol{z}_h)^\top\}$, we can expand the tensor inner product as follows:

$$\boldsymbol{\sigma}(\boldsymbol{w}_h^g) : \nabla_{\boldsymbol{x}}\boldsymbol{z}_h = \boldsymbol{\sigma}(\boldsymbol{w}_h^g) : \big\{\boldsymbol{\varepsilon}(\boldsymbol{z}_h) + \boldsymbol{\vartheta}(\boldsymbol{z}_h)\big\}.$$

Since $\boldsymbol{\sigma}(\boldsymbol{w}_h^g)$ is a purely symmetric tensor and $\boldsymbol{\vartheta}(\boldsymbol{z}_h)$ is skew-symmetric, their Frobenius inner product inherently vanishes (i.e., $\boldsymbol{\sigma}(\boldsymbol{w}_h^g) : \boldsymbol{\vartheta}(\boldsymbol{z}_h) = 0$). This simplifies the contraction to $\boldsymbol{\sigma}(\boldsymbol{w}_h^g) : \nabla_{\boldsymbol{x}}\boldsymbol{z}_h = \boldsymbol{\sigma}(\boldsymbol{w}_h^g) : \boldsymbol{\varepsilon}(\boldsymbol{z}_h)$. Substituting the constitutive relation of $\boldsymbol{\sigma}(\boldsymbol{w}_h^g)$ into the variational equation, we obtain

$$\int_{\Omega(t)} \Big(2\mu(\boldsymbol{x})\boldsymbol{\varepsilon}(\boldsymbol{w}_h^g) + \lambda(\boldsymbol{x})(\nabla_{\boldsymbol{x}} \cdot \boldsymbol{w}_h^g)\mathbb{I}_2\Big) : \boldsymbol{\varepsilon}(\boldsymbol{z}_h)\,\mathrm{d}\boldsymbol{x} = 0.$$

Noting that $\mathbb{I}_2 : \boldsymbol{\varepsilon}(\boldsymbol{z}_h) = \mathrm{tr}\big(\boldsymbol{\varepsilon}(\boldsymbol{z}_h)\big) = \nabla_{\boldsymbol{x}} \cdot \boldsymbol{z}_h$, the equation can be rewritten entirely in terms of the strain tensor and the divergence operator.
Finally, the $\mathbb{P}^1$ finite element formulation reads: find the mesh velocity $\boldsymbol{w}_h^g(\cdot,t) \in \mathbf{W}_{h,\mathrm{D}}(t)$ such that, for all $\boldsymbol{z}_h \in \mathbf{W}_{h,0}(t)$,

$$\big(2\mu(\boldsymbol{x})\boldsymbol{\varepsilon}(\boldsymbol{w}_h^g), \boldsymbol{\varepsilon}(\boldsymbol{z}_h)\big)_{\Omega(t)} + \big(\lambda(\boldsymbol{x})\nabla_{\boldsymbol{x}} \cdot \boldsymbol{w}_h^g, \nabla_{\boldsymbol{x}} \cdot \boldsymbol{z}_h(\boldsymbol{x},t)\big)_{\Omega(t)} = 0.$$

Remark 4.4.2 (*Discrete linear systems*). *The discretization of the Laplace problem leads to a standard system where the global stiffness matrix* $\mathbb{K}_\Delta(t) \in \mathscr{M}_{N_v}(\mathbb{R})$ *is defined by*

$$(\mathbb{K}_\Delta)_{(i,j)\in[\![1,N_v]\!]^2}(t) = \int_{\Omega(t)} \kappa(\boldsymbol{x})\nabla_{\boldsymbol{x}}\varphi_i(\boldsymbol{x},t) \cdot \nabla_{\boldsymbol{x}}\varphi_j(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x}.$$

Since the harmonic problem is vector-valued but uncoupled by components, the discrete system is solved component-wise for x and y. In contrast, the pseudo-elasticity model inherently couples the x- and y-components of the grid velocity. The corresponding discrete problem yields a single, larger linear system governed by a fully coupled stiffness matrix $\mathbb{K}_{elas}(t) \in \mathscr{M}_{2N_v}(\mathbb{R})$ *that interleaves the nodal components. In both cases, the variable stiffness scaling (e.g.* $\kappa_{|\omega_c} = |\omega_c|^{-1}$*) is introduced directly at the local element assembly level.*

Once the linear system is solved, the grid velocity at vertex $\boldsymbol{x}_i$ is reconstructed as

$$\boldsymbol{w}_h^g(\boldsymbol{x}_i, t) = \underline{\boldsymbol{w}}_i^g(t) = (\underline{\boldsymbol{w}}_x^g)_i(t)\, \boldsymbol{e}_x + (\underline{\boldsymbol{w}}_y^g)_i(t)\, \boldsymbol{e}_y, \qquad \boldsymbol{x}_i \in \mathcal{V}_h(t),$$

which provides the grid velocity at each mesh vertex and is directly used for the ALE geometric update.

4.4.4 Reduction to explicit affine mapping for pseudo-1D configurations

To gain further insight into the ALE mesh motion driven by the two-dimensional Laplace equation (4.31), it is instructive to study its behaviour in a pseudo-one-dimensional configuration. We consider one of the rectangular fluid regions located on one side of a horizontally moving surface obstacle, such as a floating pontoon, further discussed in Chapter 5 and Chapter 6. In this idealized setting, the physical domain is invariant along the y-direction, and the obstacle imposes a uniform boundary motion along the x-axis.

Denoting the moving interface by

$$\partial\Omega_{\text{mov}}(t) = \{\boldsymbol{x} = (x_{\text{mov}}(t), y) \mid y \in [0, L_y]\},$$

and assuming that the opposite outer boundary $\partial\Omega_{\text{ext}}$ is fixed at x_{ext}, the fluid domain under consideration is the time-dependent rectangle

$$\Omega(t) = (x_{\text{mov}}(t), x_{\text{ext}}) \times (0, L_y).$$

Under these geometric constraints, the prescribed boundary motion is purely horizontal and independent of y. Therefore, the grid velocity has the form

$$\boldsymbol{w}^g(\boldsymbol{x}, t) = (w_x^g(x, t), 0)^\top, \qquad \partial_y w_x^g = 0.$$

For this reduction, we take $\kappa = 1$ and impose slip conditions on the horizontal walls: $w_y^g = 0$ and $\partial_y w_x^g = 0$. Fixing both velocity components on these walls would generally exclude the nonzero, y-independent profile considered here. Consequently, the two-dimensional Laplace problem (4.31) reduces to the following one-dimensional scalar boundary value problem:

$$\begin{cases} -\partial_{xx} w_x^g = 0, \quad x \in (x_{\text{mov}}(t), x_{\text{ext}}), \\ w_x^g(x_{\text{ext}}, t) = 0, \\ w_x^g(x_{\text{mov}}(t), t) = \partial_t x_{\text{mov}}(t). \end{cases} \tag{4.33}$$

Since the second spatial derivative vanishes, the solution of (4.33) is affine in space. By imposing the two boundary conditions, one obtains the explicit expression

$$w_x^g(\boldsymbol{x}, t) = \partial_t x_{\text{mov}}(t) \left(\frac{x_{\text{ext}} - x}{x_{\text{ext}} - x_{\text{mov}}(t)} \right). \tag{4.34}$$

This linear velocity profile describes a uniform contraction or expansion of the fluid domain along the x-axis. The moving boundary velocity is recovered at $x = x_{\text{mov}}(t)$, while the grid velocity vanishes at the fixed boundary $x = x_{\text{ext}}$.

Restricting this affine function to an arbitrary moving cell $\omega_c(t)$ and rewriting it in terms of the velocities at the two cell boundaries gives

$$w_x^g(\boldsymbol{x}, t)_{|\omega_c(t)} = w_{x,c-1/2}^g(t) \left(\frac{x_{c+1/2}(t) - x}{h_x^c(t)} \right) + w_{x,c+1/2}^g(t) \left(\frac{x - x_{c-1/2}(t)}{h_x^c(t)} \right), \tag{4.35}$$

where

$$w_{x,c\pm1/2}^g(t) = \partial_t x_{c\pm1/2}(t), \qquad h_x^c(t) = x_{c+1/2}(t) - x_{c-1/2}(t),$$

and $h_x^c(t)$ denotes the cell width in the x-direction.

Interestingly, this local affine velocity profile matches exactly the explicit mapping strategy developed in [141]. In purely one-dimensional approaches, the mesh displacement is entirely governed by explicit interpolations between node velocities. Equation (4.35) shows that such explicit one-dimensional interpolations can be interpreted as the analytical restriction of the two-dimensional harmonic extension problem in a pseudo-one-dimensional geometry.

To avoid stretching the whole fluid domain up to x_{ext}, a common practical modification used in [141] consists in restricting the deformation to a layer of size $\varepsilon > 0$ adjacent to the moving obstacle. This is achieved by introducing a smooth *cut-off function* $\mu \in \mathscr{C}_0^\infty(\mathbb{R})$ satisfying $\mu(0) = 1$ and having compact support, typically $\operatorname{supp}(\mu) \subset [-1, 1]$; see [141] for an explicit definition. In this localized formulation, the horizontal component of the mesh mapping from the reference configuration $\boldsymbol{\xi} = (\xi_x, \xi_y)^\top \in \Omega(0)$ is defined by

$$x(\boldsymbol{\xi}, t) := \xi_x + \mu\left(\frac{\xi_x - x_{\text{mov}}(0)}{\varepsilon}\right)\left(x_{\text{mov}}(t) - x_{\text{mov}}(0)\right), \tag{4.36}$$

where $x_{\text{mov}}(0)$ is the initial position of the moving interface.

This localized mapping should be understood as a *practical modification* of the global affine construction: the displacement is prescribed only in a neighbourhood of the moving interface, while the rest of the mesh remains fixed. Elements located farther than ε from the obstacle are therefore strictly Eulerian, in the sense that $x(\boldsymbol{\xi}, t) = \xi_x$. At the formal level, replacing the compactly supported cut-off by the corresponding linear profile over the whole interval recovers the global affine mapping (4.34).

Finally, let us consider the associated time-dependent mapping $\boldsymbol{x}(\boldsymbol{\xi}, t)$ from a reference cell $\omega_c(0)$ onto its deformed counterpart $\omega_c(t)$. Under the global affine pseudo-one-dimensional assumption, the horizontal component of the mapping is affine in χ_x, while the vertical coordinate is left unchanged.

In the particular case of a purely one-dimensional strip cell, the Jacobian determinant formula (4.9) reduces to $\mathsf{J}_c(t) = h_x^c(t)/h_x^c(0)$. Differentiating this last expression with respect to time gives

$$\partial_t \mathsf{J}_c(t) = \frac{\partial_t h_x^c(t)}{h_x^c(0)} = \frac{\boldsymbol{w}^g_{x,c+1/2}(t) - \boldsymbol{w}^g_{x,c-1/2}(t)}{h_x^c(0)}.$$

On the other hand, since $\boldsymbol{w}_x^g$ is affine on $\omega_c(t)$, one has

$$\partial_x \boldsymbol{w}_x^g(\boldsymbol{x}, t)_{|\omega_c(t)} = \frac{\boldsymbol{w}^g_{x,c+1/2}(t) - \boldsymbol{w}^g_{x,c-1/2}(t)}{h_x^c(t)}.$$

Using $\mathsf{J}_c(t) = h_x^c(t)/h_x^c(0)$, we obtain

$$\partial_t \mathsf{J}_c(t) = \mathsf{J}_c(t)\, \partial_x \boldsymbol{w}_x^g(\boldsymbol{x}, t)_{|\omega_c(t)}. \tag{4.37}$$

This is exactly the geometric conservation law (4.7) restricted to the pseudo-one-dimensional setting.

The same conclusion remains valid on an unstructured simplicial mesh, provided that the deformation is induced by the same pseudo-one-dimensional affine mapping. In that case, the quantity $h_x^c(t)/h_x^c(0)$ has to be replaced by the element area ratio $|\omega_c(t)|/|\omega_c(0)|$.

Remark 4.4.3 (*Relevance to pontoon obstacles*). *This pseudo-one-dimensional interpretation is particularly relevant for the pontoon configurations introduced in Chapter 5 and Chapter 6. In such a setting, the horizontal motion of the floating body may be reproduced in an ALE framework by viewing the two fluid subdomains located on each side of the pontoon as two spring-like mesh regions. This provides a natural way to construct a genuinely two-dimensional ALE configuration for a floating object undergoing horizontal motion, while keeping the mesh deformation mechanism consistent with the harmonic extension point of view adopted in (4.31).*

To verify the explicit mapping strategy, a simple test case is performed on a rectangular domain: the right boundary is forced to oscillate horizontally, moving back and forth over a complete cycle. This motion compresses and then stretches the adjacent elements. Figure 4.3 displays eight snapshots of this oscillation. It shows that the explicit mapping correctly absorbs the boundary displacement without creating tangled cells or negative volumes.

Remark 4.4.4 (*Localized explicit mapping*). *From a numerical perspective, one could also consider a simpler explicit linear mapping spanning the whole domain to absorb the boundary displacement uniformly. Such a global linear profile evenly distributes the compression or expansion across all elements from the moving obstacle to the fixed boundary. However, this uniform deformation is often undesirable in practical fluid-structure interaction configurations. Indeed, the mesh is usually refined in the immediate vicinity of the moving obstacle in order to capture accurately the local flow dynamics, while it is significantly coarser farther away. A global linear mapping would force all cells, including the smallest ones close to the solid boundary, to absorb part of the displacement, which may lead to severe mesh distortion or restrictive CFL conditions. The localized mapping described above offers a useful alternative: the refined elements immediately adjacent to the moving boundary may be transported almost rigidly with the obstacle, so that they undergo little internal deformation. The required grid compression is then transferred to larger and more robust elements farther away from the structure, which helps preserving the local mesh quality throughout the simulation.*

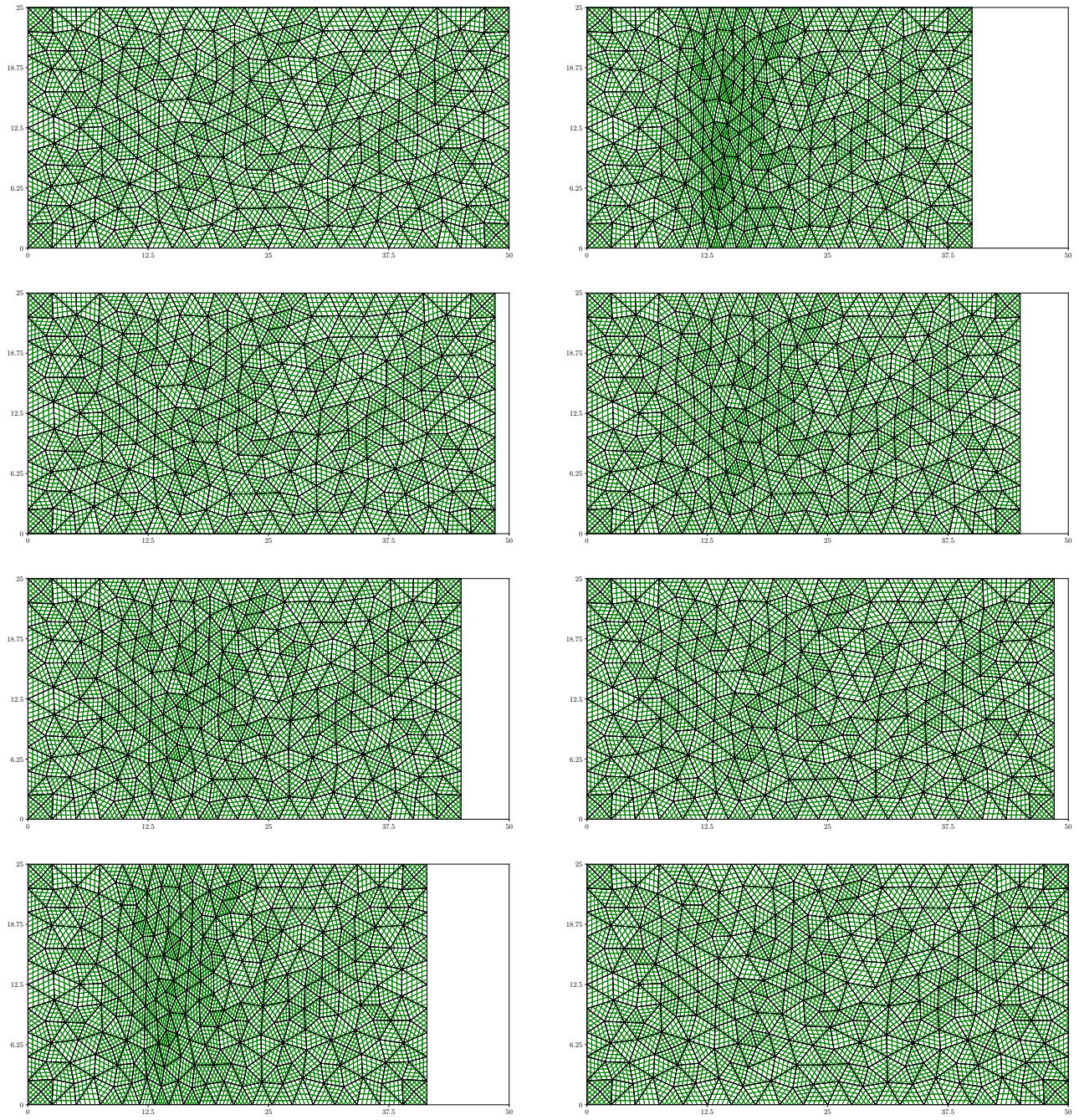


Figure 4.3: Snapshots of the ALE mesh and submesh under the pseudo-one-dimensional explicit mapping over a complete deformation cycle.

4.5 Local subcell monolithic DG/FV–ALE scheme

The previous sections were devoted to the ALE form of the governing equations, the definition of the moving control volumes, the geometric conservation requirements, and the construction of the grid velocity. These ingredients allow us to extend the local subcell monolithic DG/FV method to moving meshes in a *natural way*: once the physical flux is replaced by the ALE flux and the control volumes are updated according to the mesh motion, the Eulerian convex subcell limiting strategy can be applied almost unchanged. This is a central point of the present chapter: since the Eulerian formulation of Chapter 3 was designed so that the high-order DG update could be rewritten locally as a conservative FV-like balance over subcells, the ALE formulation preserves the same structure on moving control volumes, up to the time variation of the subcell volumes and faces. To the best of our knowledge, this places the resulting method among one of the first *subcell-based high-order schemes* formulated in an ALE setting, while retaining the main features of the original construction.

Indeed, following the methodology introduced for the Eulerian framework in Chapter 3, the development of our high-order ALE scheme relies on the reformulation of the DG-ALE equations into a *subcell FV-like approach*. This is achieved by introducing the ALE reconstructed fluxes $\widehat{\mathbb{G}}$ alongside a well-balanced first-order FV-ALE corrected flux, evaluated directly on the moving mesh.

To preserve both high-order accuracy in smooth regions and robustness near discontinuities, we combine again the high-order DG-ALE reconstructed flux $\widehat{\mathbb{G}}_{mp}$ and the first-order robust FV-ALE numerical flux $\mathbb{G}^{*,\mathrm{FV}}_{mp}$ in a convex manner. For every moving subface $\Gamma^c_{mp}(t) \in \mathscr{F}^c_m$ of each subcell $S^c_m(t)$, the blended ALE numerical flux is defined as:

$$\widetilde{\mathbb{G}}_{mp} = \mathbb{G}^{*,\mathrm{FV}}_{mp} + \Theta_{mp}\left(\widehat{\mathbb{G}}_{mp} - \mathbb{G}^{*,\mathrm{FV}}_{mp}\right). \tag{4.38}$$

We still employ a scalar coefficient $\Theta_{mp} \in [0,1]$ exactly as detailed in Section 3.3. Using the shortcut $\Delta\mathbb{G}_{mp} := \widehat{\mathbb{G}}_{mp} - \mathbb{G}^{*,\mathrm{FV}}_{mp}$ to denote the difference between the high-order and first-order ALE fluxes, setting $\Theta_{mp} = 0$ yields a robust first-order FV-ALE scheme, while $\Theta_{mp} = 1$ fully recovers the high-order DG-ALE scheme.

Using a forward Euler time integration from t^n to t^{n+1}, the discrete scheme reads:

Definition 4.5.1 (*Subcell monolithic DG/FV–ALE scheme*). *The subcell monolithic DG/FV convex property preserving scheme on moving meshes, supplemented with a first-order time integration, is given by*

$$|S^{c,n+1}_m|\overline{\boldsymbol{v}}^{c,n+1}_m = |S^{c,n}_m|\overline{\boldsymbol{v}}^{c,n}_m - \Delta t^n \sum_{S^{v,n}_p \in \mathscr{V}^{c,n}_m} \ell^n_{mp}\widetilde{\mathbb{G}}_{mp} + \Delta t^n |S^{c,n}_m|\overline{\boldsymbol{B}}^{c,n}_m, \tag{4.39}$$

where $S^{c,n}_m := |S^c_m(t^n)|$ is the time-dependent subcell volume, $\ell^n_{mp} := \ell_{mp}(t^n)$ is the moving interface length, and $\widetilde{\mathbb{G}}_{mp}$ is the blended ALE flux defined in (4.38). Also, the topography source term $\overline{\boldsymbol{B}}^{c,n}_m$ is discretized following the exact same projection and averaging procedures outlined in Section 3.3. However, all spatial integrations are now performed over the moving control volume $S^c_m(t^n)$ at the current time level.

Proposition 4.5.2 (*Godunov-like reformulation of the first-order FV–ALE scheme*). *Let us consider the first-order part of the subcell monolithic DG/FV–ALE scheme (4.39), obtained by setting $\Theta_{mp} = 0$ on every moving subface. For each $S^{v,n}_p \in \mathscr{V}^{c,n}_m$, we introduce the normal grid velocity*

$$w^{g,n}_{mp} := \boldsymbol{w}^{g,n}_{mp} \cdot \mathbf{n}^n_{mp},$$

and a local ALE wave-speed estimate $\sigma^{\mathrm{ALE},n}_{mp} > 0$ bounding the spectral radius of the ALE flux across $\Gamma^{c,n}_{mp}$. We then define the interior first-order FV–ALE Riemann intermediate state by

$$\boldsymbol{v}^{*,\mathrm{ALE},n}_{mp} := \overline{\boldsymbol{v}}^{c,n}_m - \frac{\mathbb{G}^{*,\mathrm{FV},n}_{mp} - \mathbb{G}\left(\overline{\boldsymbol{v}}^{c,n}_m, \overline{b}^{c,n}_m, \boldsymbol{w}^{g,n}_{mp}\right)\cdot \mathbf{n}^n_{mp}}{\sigma^{\mathrm{ALE},n}_{mp}}. \tag{4.40}$$

Assume that the mesh motion satisfies the fully-discrete geometric conservation law (4.49), with the same face grid velocity as the one entering the ALE numerical flux. Then, the first-order FV–ALE update can be rewritten as

$$\begin{aligned}\overline{\boldsymbol{v}}^{c,n+1}_m = &\left(1 - \frac{\Delta t^n}{|S^{c,n+1}_m|}\sum_{S^{v,n}_p \in \mathscr{V}^{c,n}_m} \ell^n_{mp}\sigma^{\mathrm{ALE},n}_{mp}\right)\overline{\boldsymbol{v}}^{c,n}_m \\ &+ \frac{\Delta t^n}{|S^{c,n+1}_m|}\sum_{S^{v,n}_p \in \mathscr{V}^{c,n}_m} \ell^n_{mp}\sigma^{\mathrm{ALE},n}_{mp}\boldsymbol{v}^{*,\mathrm{ALE},n}_{mp} + \Delta t^n \frac{|S^{c,n}_m|}{|S^{c,n+1}_m|}\overline{\boldsymbol{B}}^{c,n}_m.\end{aligned} \tag{4.41}$$

Consequently, the homogeneous part of (4.41) is a convex combination of states defined at time t^n, provided that

$$\Delta t^n \sum_{S_p^{v,n} \in \mathscr{V}_m^{c,n}} \ell_{mp}^n \sigma_{mp}^{\mathrm{ALE},n} \leq |S_m^{c,n+1}|. \tag{4.42}$$

Equivalently, using the fully-discrete geometric conservation law, this local CFL condition reads

$$\Delta t^n \sum_{S_p^{v,n} \in \mathscr{V}_m^{c,n}} \ell_{mp}^n \left(\sigma_{mp}^{\mathrm{ALE},n} - w_{mp}^{g,n}\right) \leq |S_m^{c,n}|. \tag{4.43}$$

Proof. Starting from (4.40), the first-order FV–ALE numerical flux can be written as

$$\mathbb{G}_{mp}^{*,\mathrm{FV},n} = \mathbb{G}\left(\overline{\boldsymbol{v}}_m^{c,n}, \overline{b}_m^{c,n}, \boldsymbol{w}_{mp}^{g,n}\right) \cdot \mathbf{n}_{mp}^n + \sigma_{mp}^{\mathrm{ALE},n}\left(\overline{\boldsymbol{v}}_m^{c,n} - \boldsymbol{v}_{mp}^{*,\mathrm{ALE},n}\right).$$

The first-order scheme, obtained by setting $\widetilde{\mathbb{G}}_{mp} = \mathbb{G}_{mp}^{*,\mathrm{FV}}$ in (4.39), therefore gives

$$\begin{aligned} |S_m^{c,n+1}|\overline{\boldsymbol{v}}_m^{c,n+1} = |S_m^{c,n}|\overline{\boldsymbol{v}}_m^{c,n} &- \Delta t^n \sum_{S_p^{v,n} \in \mathscr{V}_m^{c,n}} \ell_{mp}^n \mathbb{G}\left(\overline{\boldsymbol{v}}_m^{c,n}, \overline{b}_m^{c,n}, \boldsymbol{w}_{mp}^{g,n}\right) \cdot \mathbf{n}_{mp}^n \\ &- \Delta t^n \sum_{S_p^{v,n} \in \mathscr{V}_m^{c,n}} \ell_{mp}^n \sigma_{mp}^{\mathrm{ALE},n}\left(\overline{\boldsymbol{v}}_m^{c,n} - \boldsymbol{v}_{mp}^{*,\mathrm{ALE},n}\right) + \Delta t^n |S_m^{c,n}| \overline{\boldsymbol{B}}_m^{c,n}. \end{aligned}$$

We now split the ALE flux into its Eulerian and grid-transport contributions:

$$\mathbb{G}\left(\overline{\boldsymbol{v}}_m^{c,n}, \overline{b}_m^{c,n}, \boldsymbol{w}_{mp}^{g,n}\right) \cdot \mathbf{n}_{mp}^n = \mathbb{F}\left(\overline{\boldsymbol{v}}_m^{c,n}, \overline{b}_m^{c,n}\right) \cdot \mathbf{n}_{mp}^n - w_{mp}^{g,n} \overline{\boldsymbol{v}}_m^{c,n}.$$

Since the physical state and the topography mean value are constant with respect to the subfaces of $S_m^{c,n}$, and since the outward normals satisfy the geometrical identity

$$\sum_{S_p^{v,n} \in \mathscr{V}_m^{c,n}} \ell_{mp}^n \mathbf{n}_{mp}^n = \mathbf{0},$$

the Eulerian contribution vanishes when summed over the closed boundary of the subcell. Hence,

$$\sum_{S_p^{v,n} \in \mathscr{V}_m^{c,n}} \ell_{mp}^n \mathbb{G}\left(\overline{\boldsymbol{v}}_m^{c,n}, \overline{b}_m^{c,n}, \boldsymbol{w}_{mp}^{g,n}\right) \cdot \mathbf{n}_{mp}^n = -\overline{\boldsymbol{v}}_m^{c,n} \sum_{S_p^{v,n} \in \mathscr{V}_m^{c,n}} \ell_{mp}^n w_{mp}^{g,n}.$$

It follows that

$$\begin{aligned} |S_m^{c,n+1}|\overline{\boldsymbol{v}}_m^{c,n+1} = &\left(|S_m^{c,n}| + \Delta t^n \sum_{S_p^{v,n} \in \mathscr{V}_m^{c,n}} \ell_{mp}^n w_{mp}^{g,n} - \Delta t^n \sum_{S_p^{v,n} \in \mathscr{V}_m^{c,n}} \ell_{mp}^n \sigma_{mp}^{\mathrm{ALE},n}\right) \overline{\boldsymbol{v}}_m^{c,n} \\ &+ \Delta t^n \sum_{S_p^{v,n} \in \mathscr{V}_m^{c,n}} \ell_{mp}^n \sigma_{mp}^{\mathrm{ALE},n} \boldsymbol{v}_{mp}^{*,\mathrm{ALE},n} + \Delta t^n |S_m^{c,n}| \overline{\boldsymbol{B}}_m^{c,n}. \end{aligned}$$

By means of the fully-discrete geometric conservation law (4.49), we have

$$|S_m^{c,n}| + \Delta t^n \sum_{S_p^{v,n} \in \mathscr{V}_m^{c,n}} \ell_{mp}^n w_{mp}^{g,n} = |S_m^{c,n+1}|.$$

Substituting this identity into the previous expression and dividing by $|S_m^{c,n+1}|$ yields (4.41). The weights associated with $\overline{\boldsymbol{v}}_m^{c,n}$ and the intermediate states $\boldsymbol{v}_{mp}^{*,\mathrm{ALE},n}$ sum to one. They are non-negative under (4.42), which proves the convex-combination property of the homogeneous part of the scheme. Finally, replacing $|S_m^{c,n+1}|$ in (4.42) by means of (4.49) directly gives the equivalent condition (4.43). □

Remark 4.5.3 (*On the role of the fully-discrete geometric conservation law*). *The fully-discrete GCL is not only a sufficient technical assumption in Proposition 4.5.2; it is necessary to obtain an exact fully-discrete convex-combination form for the scheme (4.39) as it is presently written. Indeed, let us introduce the local geometric defect*

$$\epsilon_m^{c,n} := |S_m^{c,n+1}| - |S_m^{c,n}| - \Delta t^n \sum_{S_p^{v,n} \in \mathscr{V}_m^{c,n}} \ell_{mp}^n w_{mp}^{g,n}. \tag{4.44}$$

Without assuming (4.49), the same computations as in the previous proof lead to

$$\overline{\boldsymbol{v}}_m^{c,n+1} = \left(1 - \frac{\epsilon_m^{c,n}}{|S_m^{c,n+1}|} - \frac{\Delta t^n}{|S_m^{c,n+1}|} \sum_{S_p^{v,n} \in \mathcal{V}_m^{c,n}} \ell_{mp}^n \sigma_{mp}^{\mathrm{ALE},n}\right) \overline{\boldsymbol{v}}_m^{c,n} + \frac{\Delta t^n}{|S_m^{c,n+1}|} \sum_{S_p^{v,n} \in \mathcal{V}_m^{c,n}} \ell_{mp}^n \sigma_{mp}^{\mathrm{ALE},n} \boldsymbol{v}_{mp}^{*,\mathrm{ALE},n} + \Delta t^n \frac{|S_m^{c,n}|}{|S_m^{c,n+1}|} \overline{\boldsymbol{B}}_m^{c,n}. \tag{4.45}$$

The weights of the homogeneous part now sum to $1 - \frac{\epsilon_m^{c,n}}{|S_m^{c,n+1}|}$, rather than one. This obstruction can also be seen by considering a constant state $\overline{\boldsymbol{v}}_m^{c,n} = \boldsymbol{v}^0 \neq \boldsymbol{0}$ together with a constant topography. By consistency of the first-order ALE flux, all the intermediate states then satisfy $\boldsymbol{v}_{mp}^{*,\mathrm{ALE},n} = \boldsymbol{v}^0$, while (4.45) gives

$$\overline{\boldsymbol{v}}_m^{c,n+1} = \left(1 - \frac{\epsilon_m^{c,n}}{|S_m^{c,n+1}|}\right) \boldsymbol{v}^0.$$

Consequently, the constant state is exactly preserved if and only if $\epsilon_m^{c,n} = 0$. Since any convex combination of identical states necessarily returns the same state, an exact fully-discrete Godunov-like reformulation cannot hold in general when the fully-discrete GCL is violated. At the semi-discrete level, no such difficulty occurs since the semi-discrete GCL (4.48) is exactly satisfied. At the fully-discrete level, Proposition 4.6.2 below identifies mesh motions for which this defect vanishes automatically. Outside these cases, (4.45) retains a geometric consistency error unless a compatible geometric time quadrature or a reconstruction of the face grid velocities is used.

Remark 4.5.4 (*Blending coefficients*). *Let us now return to the complete blended flux (4.38). Using $\Delta\mathbb{G}_{mp} := \widehat{\mathbb{G}}_{mp} - \mathbb{G}_{mp}^{*,\mathrm{FV}}$, we define the interior blended ALE Riemann intermediate state as*

$$\widetilde{\boldsymbol{v}}_{mp}^{*,\mathrm{ALE},n} := \overline{\boldsymbol{v}}_m^{c,n} - \frac{\widetilde{\mathbb{G}}_{mp}^n - \mathbb{G}\left(\overline{\boldsymbol{v}}_m^{c,n}, \overline{b}_m^{c,n}, \boldsymbol{w}_{mp}^{g,n}\right) \cdot \mathbf{n}_{mp}^n}{\sigma_{mp}^{\mathrm{ALE},n}} = \boldsymbol{v}_{mp}^{*,\mathrm{ALE},n} - \Theta_{mp}^n \frac{\Delta\mathbb{G}_{mp}^n}{\sigma_{mp}^{\mathrm{ALE},n}}. \tag{4.46}$$

Under the same fully-discrete GCL assumption as in Proposition 4.5.2, the complete monolithic DG/FV–ALE scheme consequently admits the Godunov-like reformulation

$$\overline{\boldsymbol{v}}_m^{c,n+1} = \left(1 - \frac{\Delta t^n}{|S_m^{c,n+1}|} \sum_{S_p^{v,n} \in \mathcal{V}_m^{c,n}} \ell_{mp}^n \sigma_{mp}^{\mathrm{ALE},n}\right) \overline{\boldsymbol{v}}_m^{c,n} + \frac{\Delta t^n}{|S_m^{c,n+1}|} \sum_{S_p^{v,n} \in \mathcal{V}_m^{c,n}} \ell_{mp}^n \sigma_{mp}^{\mathrm{ALE},n} \widetilde{\boldsymbol{v}}_{mp}^{*,\mathrm{ALE},n} + \Delta t^n \frac{|S_m^{c,n}|}{|S_m^{c,n+1}|} \overline{\boldsymbol{B}}_m^{c,n}. \tag{4.47}$$

The convex weights and the associated CFL condition are therefore exactly the same as in the first-order reformulation; only the Riemann intermediate states are modified by the antidiffusive correction $-\Theta_{mp}^n \Delta\mathbb{G}_{mp}^n / \sigma_{mp}^{\mathrm{ALE},n}$. Consequently, the ALE blending coefficients may be determined as in the Eulerian framework by enforcing the required admissibility constraints on $\widetilde{\boldsymbol{v}}_{mp}^{*,\mathrm{ALE},n}$. This amounts to replacing the Eulerian flux difference $\Delta\mathbb{F}_{mp}$ by its ALE counterpart $\Delta\mathbb{G}_{mp}$, alongside the use of the ALE wave-speed estimate. Also, to prevent severe accuracy losses caused by abrupt transitions from high to low-order, the same mean and minimum values blending smoothers, as well as the linear smooth extrema detector (3.38), are applied to Θ_{mp} at the end of each time step. In particular, on a constant topography and with vanishing sources and compatible boundary traces, constant states are preserved for arbitrary Θ_{mp} if both fluxes reduce on each subface to the same consistent ALE flux. Indeed, $\Delta\mathbb{G}_{mp} = 0$ face by face and the remaining grid-transport balance is exactly (4.49). Equality of the sums of the high-order and first-order fluxes alone would not suffice when Θ_{mp} varies between subfaces.

4.6 Conservation properties of the ALE formulation

In this section, we analyze some of the fundamental conservation properties of the semi-discrete DG-ALE formulation (4.22).

4.6.1 Semi and fully-discrete geometric conservation laws

As already mentioned, the geometric conservation law plays a central role in ALE formulations. For any moving subcell $S_m^c(t) \in \mathcal{T}_{\omega_c}^k(t)$, the continuous GCL gives

$$\frac{\mathrm{d}}{\mathrm{d}t}|S_m^c(t)| = \int_{S_m^c(t)} \nabla_{\boldsymbol{x}} \cdot \boldsymbol{w}^g \,\mathrm{d}\boldsymbol{x} = \sum_{S_p^v(t)\in\mathcal{V}_m^c(t)} \int_{\Gamma_{mp}^c(t)} \boldsymbol{w}^g \cdot \mathbf{n}_{mp} \,\mathrm{d}s.$$

This identity is purely geometrical: it states that the variation of the moving subcell area is exactly equal to the flux of the grid velocity through its boundary.

Proposition 4.6.1 (*Semi-discrete geometric conservation law*). *Consider the semi-discrete FV-like ALE balance (4.28) obtained from the DG-ALE formulation on a moving subcell $S_m^c(t)$:*

$$\frac{\mathrm{d}}{\mathrm{d}t}(|S_m^c(t)|\,\overline{\boldsymbol{v}}_m^c) = -\sum_{S_p^v(t)\in\mathcal{V}_m^c(t)} \int_{\Gamma_{mp}^c} \mathbb{G}_h^c \cdot \mathbf{n}_{mp} \,\mathrm{d}s + \sum_{k=1}^{\#\mathcal{F}_c} \int_{\Gamma_{cv(k)}} \left(\mathbb{G}_h^c \cdot \mathbf{n}_{cv(k)} - \mathbb{G}_{cv(k)}^{*,\mathrm{DG}}\right) \phi_m^c \,\mathrm{d}s + |S_m^c(t)|\,\overline{\boldsymbol{B}}_m^c.$$

Assume that $\overline{\boldsymbol{v}}_m^c(t) \equiv \boldsymbol{v}^0 = (\eta^0, q_x^0, q_y^0)^\top \neq \mathbf{0}$ is a constant state and $b = b^0$ is constant. Then the preservation of this constant state by the semi-discrete ALE formulation is equivalent to the semi-discrete geometric conservation law

$$\frac{\mathrm{d}}{\mathrm{d}t}|S_m^c(t)| = \sum_{S_p^v(t)\in\mathcal{V}_m^c(t)} \int_{\Gamma_{mp}^c(t)} \boldsymbol{w}^g \cdot \mathbf{n}_{mp} \,\mathrm{d}s. \tag{4.48}$$

Proof. Since $b = b^0$ is constant, one has $\nabla_{\boldsymbol{x}} b^0 = \mathbf{0}$. Hence, for the bathymetry source term, $\boldsymbol{B}[b^0](\boldsymbol{v}^0) = (0, -\mathrm{g}\eta^0 \nabla_{\boldsymbol{x}} b^0)^\top = \mathbf{0}$, and therefore $\overline{\boldsymbol{B}}_m^c = \mathbf{0}$. We recall that, in the projected DG-ALE formulation, the polynomial flux used in the cell is defined by

$$\mathbb{G}_h^c := \pi_c^k \mathbb{G}(\boldsymbol{v}_h^c, b_h^c, \boldsymbol{w}_h^{g,c}).$$

For the constant state $\boldsymbol{v}_h^c = \boldsymbol{v}^0$ and the constant bathymetry $b_h^c = b^0$, this gives

$$\mathbb{G}(\boldsymbol{v}^0, b^0, \boldsymbol{w}_h^{g,c}) = \mathbb{F}(\boldsymbol{v}^0, b^0) - \boldsymbol{v}^0 \otimes \boldsymbol{w}_h^{g,c}.$$

Since $\mathbb{F}(\boldsymbol{v}^0, b^0)$ is constant in space and since the grid velocity $\boldsymbol{w}_h^{g,c}$ is represented by a polynomial field belonging to the approximation space used for the projection, the ALE flux already belongs to $\mathbf{P}^k(\omega_c(t))^2$. Hence the L^2-projection is exact and one has $\mathbb{G}_h^c = \mathbb{G}(\boldsymbol{v}^0, b^0, \boldsymbol{w}_h^{g,c})$ for all $\omega_c(t) \in \mathcal{T}_h(t)$. In particular, on any subface $\Gamma_{mp}(t) \in \mathcal{F}_m^c(t)$ with normal $\mathbf{n}_{mp}$,

$$\mathbb{G}_h^c \cdot \mathbf{n}_{mp} = \mathbb{F}(\boldsymbol{v}^0, b^0) \cdot \mathbf{n}_{mp} - \boldsymbol{v}^0 \left(\boldsymbol{w}_h^{g,c} \cdot \mathbf{n}_{mp}\right).$$

For the constant state $\boldsymbol{v}^0$ and the constant bathymetry b^0, the ALE flux satisfies

$$\mathbb{G}(\boldsymbol{v}^0, b^0, \boldsymbol{w}^g) \cdot \mathbf{n}_{mp} = \mathbb{F}(\boldsymbol{v}^0, b^0) \cdot \mathbf{n}_{mp} - \boldsymbol{v}^0 \left(\boldsymbol{w}^g \cdot \mathbf{n}_{mp}\right).$$

Moreover, by consistency of the DG numerical flux, one has on each external face $\Gamma_{cv(k)}(t) \in \mathcal{F}_c(t)$ and normal $\mathbf{n}_{cv(k)}$ with $k \in [\![1, \#\mathcal{F}_c]\!]$:

$$\mathbb{G}_{cv(k)}^{*,\mathrm{DG}} = \mathbb{G}(\boldsymbol{v}^0, b^0, \boldsymbol{w}^g) \cdot \mathbf{n}_{cv(k)} = \mathbb{G}_h^c \cdot \mathbf{n}_{cv(k)}.$$

Therefore, $\mathbb{G}_h^c \cdot \mathbf{n}_{cv(k)} - \mathbb{G}_{cv(k)}^{*,\mathrm{DG}} = \mathbf{0}$ for all $k \in [\![1, \#\mathcal{F}_c]\!]$ and all $\omega_c(t) \in \mathcal{T}_h(t)$, and the boundary correction term vanishes. The balance now reduces to

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(|S_m^c(t)|\boldsymbol{v}^0\right) = -\sum_{S_p^v(t)\in\mathcal{V}_m^c(t)} \int_{\Gamma_{mp}^c(t)} \mathbb{G}(\boldsymbol{v}^0, b^0, \boldsymbol{w}^g) \cdot \mathbf{n}_{mp} \,\mathrm{d}s.$$

Using the decomposition of the ALE flux, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(|S_m^c(t)|\boldsymbol{v}^0\right) = -\sum_{S_p^v(t)\in\mathcal{V}_m^c(t)} \int_{\Gamma_{mp}^c(t)} \mathbb{F}(\boldsymbol{v}^0, b^0) \cdot \mathbf{n}_{mp} \,\mathrm{d}s + \boldsymbol{v}^0 \sum_{S_p^v(t)\in\mathcal{V}_m^c(t)} \int_{\Gamma_{mp}^c(t)} \boldsymbol{w}^g \cdot \mathbf{n}_{mp} \,\mathrm{d}s.$$

Since $\mathbb{F}(\boldsymbol{v}^0, b^0)$ is constant in space, its contribution over the closed boundary of the subcell vanishes:

$$\sum_{S_p^v(t)\in\mathcal{V}_m^c(t)} \int_{\Gamma_{mp}^c(t)} \mathbb{F}(\boldsymbol{v}^0, b^0)\cdot \mathbf{n}_{mp}\,\mathrm{d}s = \mathbb{F}(\boldsymbol{v}^0, b^0)\left(\sum_{S_p^v(t)\in\mathcal{V}_m^c(t)} \int_{\Gamma_{mp}^c(t)} \mathbf{n}_{mp}\,\mathrm{d}s\right) = \mathbf{0}.$$

Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(|S_m^c(t)|\boldsymbol{v}^0\right) = \boldsymbol{v}^0 \sum_{S_p^v(t)\in\mathcal{V}_m^c(t)} \int_{\Gamma_{mp}^c(t)} \boldsymbol{w}^g\cdot\mathbf{n}_{mp}\,\mathrm{d}s,$$

and since $\boldsymbol{v}^0 \neq \mathbf{0}$ is constant in time, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}|S_m^c(t)| = \sum_{S_p^v(t)\in\mathcal{V}_m^c(t)} \int_{\Gamma_{mp}^c(t)} \boldsymbol{w}^g\cdot\mathbf{n}_{mp}\,\mathrm{d}s,$$

which is exactly the semi-discrete geometric conservation law. □

Integrating (4.48) over one time step gives the exact geometric balance

$$|S_m^c(t^{n+1})| - |S_m^c(t^n)| = \int_{t^n}^{t^{n+1}} \sum_{S_p^v(t)\in\mathcal{V}_m^c(t)} \int_{\Gamma_{mp}^c(t)} \boldsymbol{w}^g\cdot\mathbf{n}_{mp}\,\mathrm{d}s\,\mathrm{d}t.$$

This identity is still exact, but it is not yet a fully discrete GCL. To obtain a fully discrete geometric conservation law, one must also prescribe a time quadrature for the right-hand side which is exactly compatible with the way the new subcell volumes are computed.

For instance, a forward Euler discretization of the geometric fluxes would lead to the formal relation

$$|S_m^{c,n+1}| - |S_m^{c,n}| = \Delta t^n \sum_{S_p^{v,n}\in\mathcal{V}_m^{c,n}} \ell_{mp}^n \boldsymbol{w}_{mp}^{g,n}\cdot\mathbf{n}_{mp}^n, \tag{4.49}$$

where ℓ_{mp}^n and $\mathbf{n}_{mp}^n$ denote respectively the length and the outward unit normal of the moving subface $\Gamma_{mp}^{c,n}$, and where $\boldsymbol{w}_{mp}^{g,n}$ denotes the numerical grid velocity used in the ALE flux through this subface. More precisely, it has to be understood as a face-averaged grid velocity satisfying

$$\boldsymbol{w}_{mp}^{g,n}\cdot\mathbf{n}_{mp}^n := \frac{1}{\ell_{mp}^n}\int_{\Gamma_{mp}^{c,n}} \boldsymbol{w}_h^{g,n}\cdot\mathbf{n}_{mp}^n\,\mathrm{d}s.$$

In the case where $\boldsymbol{w}_h^{g,n}$ is represented by continuous $\mathbb{P}^1$ finite elements, and if $\boldsymbol{x}_{mp,A}^n$ and $\boldsymbol{x}_{mp,B}^n$ denote the two vertexes belonging to $\Gamma_{mp}^{c,n}$, this face average reduces to

$$\boldsymbol{w}_{mp}^{g,n}\cdot\mathbf{n}_{mp}^n = \frac{1}{2}\left(\boldsymbol{w}^{g,n}(\boldsymbol{x}_{mp,A}^n) + \boldsymbol{w}^{g,n}(\boldsymbol{x}_{mp,B}^n)\right)\cdot\mathbf{n}_{mp}^n.$$

Identity (4.49) should be understood as a discrete compatibility condition between the mesh update and the ALE flux discretization. It is the geometric analogue of the conservation requirement imposed on the numerical physical fluxes. In particular, the same face grid velocity must be used both in the ALE numerical flux and in the discrete update of the control-volume measure.

However, in the present setting, (4.49) is not enforced algebraically. The mesh nodes are updated according to

$$\boldsymbol{x}_i^{n+1} = \boldsymbol{x}_i^n + \Delta t^n \boldsymbol{w}_i^{g,n},$$

and if the new subcell volumes are recomputed directly from the deformed geometry, then the exact volume variation generally contains higher-order terms in Δt^n. However, these terms vanish for particular mesh motions. For the piecewise-affine geometry used here, the following result gives their exact expression.

Proposition 4.6.2 (*Exact geometric defect for an affine mesh update*). *Assume that the continuous $\mathbb{P}^1$ grid velocity $\boldsymbol{w}_h^{g,n}$ is used both in the above Euler nodal update and in the exact face averages entering (4.49). Assume that the subcell vertices have fixed barycentric coordinates in their parent triangle, and that the mesh remains non-degenerate and orientation-preserving during the step. Then the geometric defect (4.44) satisfies*

$$\epsilon_m^{c,n} = (\Delta t^n)^2 |S_m^{c,n}|\,\det\left(\nabla_{\boldsymbol{x}}\boldsymbol{w}_h^{g,n}\big|_{\omega_c(t^n)}\right). \tag{4.50}$$

Consequently, for $\Delta t^n > 0$, the D-GCL is exactly satisfied on every subcell of $\omega_c(t^n)$ if and only if the determinant vanishes.

Proof. The affine map $\boldsymbol{x} \mapsto \boldsymbol{x} + \Delta t^n \boldsymbol{w}_h^{g,n}(\boldsymbol{x})$ transports the parent triangle and its subcells. Since the velocity gradient is constant on $\omega_c(t^n)$, the two-dimensional determinant identity gives, with all derivatives evaluated on this triangle,

$$|S_m^{c,n+1}| = |S_m^{c,n}| \det\left(\mathbb{I} + \Delta t^n \nabla_{\boldsymbol{x}} \boldsymbol{w}_h^{g,n}\right) = |S_m^{c,n}| \left(1 + \Delta t^n \, \nabla_{\boldsymbol{x}} \cdot \boldsymbol{w}_h^{g,n} + (\Delta t^n)^2 \det\left(\nabla_{\boldsymbol{x}} \boldsymbol{w}_h^{g,n}\right)\right).$$

On the other hand, the divergence theorem and the exact face averages yield

$$\sum_{S_p^{v,n} \in \mathcal{V}_m^{c,n}} \ell_{mp}^n \boldsymbol{w}_{mp}^{g,n} \cdot \mathbf{n}_{mp}^n = |S_m^{c,n}| \, \nabla_{\boldsymbol{x}} \cdot \boldsymbol{w}_h^{g,n}.$$

Substituting these two identities into the definition (4.44) gives

$$\begin{aligned}
\epsilon_m^{c,n} &= |S_m^{c,n+1}| - |S_m^{c,n}| - \Delta t^n \sum_{S_p^{v,n} \in \mathcal{V}_m^{c,n}} \ell_{mp}^n \boldsymbol{w}_{mp}^{g,n} \cdot \mathbf{n}_{mp}^n \\
&= |S_m^{c,n}| \left(\Delta t^n \, \nabla_{\boldsymbol{x}} \cdot \boldsymbol{w}_h^{g,n} + (\Delta t^n)^2 \det\left(\nabla_{\boldsymbol{x}} \boldsymbol{w}_h^{g,n}\right)\right) - \Delta t^n |S_m^{c,n}| \, \nabla_{\boldsymbol{x}} \cdot \boldsymbol{w}_h^{g,n} \\
&= (\Delta t^n)^2 |S_m^{c,n}| \, \det\left(\nabla_{\boldsymbol{x}} \boldsymbol{w}_h^{g,n}\right),
\end{aligned}$$

which proves (4.50). □

Proposition 4.6.3 (*D-GCL preservation by Laplacian mesh motion*). *Let us assume that the geometric hypotheses of Proposition 4.6.2 hold. Let $\boldsymbol{w}_h^{g,n}$ solve the coercive $\mathbb{P}^1$ discretization of (4.31), with the same positive coefficient κ for both components, zero velocity on $\partial\Omega_{\text{ext}}$, and a spatially uniform velocity on $\partial\Omega_{\text{mov}}(t^n)$, namely*

$$\boldsymbol{w}_h^{g,n}(\boldsymbol{x}) = \partial_t \boldsymbol{x}_{\text{mov}}(t^n), \quad \forall \boldsymbol{x} \in \partial\Omega_{\text{mov}}(t^n).$$

Then the D-GCL (4.49) is satisfied exactly on every subcell.

Proof. We fix the time level t^n and write $\boldsymbol{w}_h^{g,n} = (w_{x,h}^{g,n}, w_{y,h}^{g,n})^\top$. Since the same coefficient κ is used for both components, testing the discrete Laplace formulation with $\boldsymbol{z}_h = z_h \boldsymbol{e}_j$, for $j \in \{x, y\}$, gives

$$\int_{\Omega(t^n)} \kappa \, \nabla_{\boldsymbol{x}} w_{j,h}^{g,n} \cdot \nabla_{\boldsymbol{x}} z_h \, \mathrm{d}\boldsymbol{x} = 0, \quad \forall \, \boldsymbol{z}_h = z_h \boldsymbol{e}_j \in \mathbf{W}_{h,0}(t^n).$$

Since both components have the same homogeneous test boundary conditions, the same scalar function z_h can be used in the two equations. The prescribed boundary velocity is spatially uniform. Consequently, $\partial_t x_{\text{mov}}(t^n)$ and $\partial_t y_{\text{mov}}(t^n)$ are constants with respect to $\boldsymbol{x}$. Multiplying the equation for $w_{x,h}^{g,n}$ by $\partial_t y_{\text{mov}}(t^n)$ and subtracting $\partial_t x_{\text{mov}}(t^n)$ times the equation for $w_{y,h}^{g,n}$ gives

$$\int_{\Omega(t^n)} \kappa \, \nabla_{\boldsymbol{x}} \left(\partial_t y_{\text{mov}}(t^n) w_{x,h}^{g,n} - \partial_t x_{\text{mov}}(t^n) w_{y,h}^{g,n}\right) \cdot \nabla_{\boldsymbol{x}} z_h \, \mathrm{d}\boldsymbol{x} = 0.$$

The scalar combination inside the parentheses belongs to $\mathsf{W}_h(t^n)$ and vanishes on both boundaries. Indeed, both velocity components vanish on $\partial\Omega_{\text{ext}}$, while on $\partial\Omega_{\text{mov}}(t^n)$ the prescribed data give

$$\partial_t y_{\text{mov}}(t^n) w_{x,h}^{g,n} - \partial_t x_{\text{mov}}(t^n) w_{y,h}^{g,n} = \partial_t y_{\text{mov}}(t^n) \partial_t x_{\text{mov}}(t^n) - \partial_t x_{\text{mov}}(t^n) \partial_t y_{\text{mov}}(t^n) = 0.$$

Multiplying this combination by either $\boldsymbol{e}_x$ or $\boldsymbol{e}_y$ therefore gives an element of $\mathbf{W}_{h,0}(t^n)$. It can thus be used as a scalar test function in both equations. Choosing

$$z_h = \partial_t y_{\text{mov}}(t^n) w_{x,h}^{g,n} - \partial_t x_{\text{mov}}(t^n) w_{y,h}^{g,n}$$

in the preceding identity yields

$$\int_{\Omega(t^n)} \kappa \left| \nabla_{\boldsymbol{x}} \left(\partial_t y_{\text{mov}}(t^n) w_{x,h}^{g,n} - \partial_t x_{\text{mov}}(t^n) w_{y,h}^{g,n}\right) \right|^2 \mathrm{d}\boldsymbol{x} = 0.$$

By coercivity on the homogeneous test space, we obtain

$$\partial_t y_{\text{mov}}(t^n) w_{x,h}^{g,n} - \partial_t x_{\text{mov}}(t^n) w_{y,h}^{g,n} = 0 \quad \text{in } \Omega(t^n). \tag{4.51}$$

If the prescribed boundary velocity is nonzero, this identity shows that the interior grid velocity is everywhere parallel to it. To deduce the D-GCL from Proposition 4.6.2, it remains to show that the grid-velocity gradient has zero determinant on every triangle. We distinguish two cases. Assume first that $\partial_t \boldsymbol{x}_{\mathrm{mov}}(t^n) \neq \mathbf{0}$. Differentiating (4.51) on each triangle $\omega_c(t^n)$ gives

$$\partial_t y_{\mathrm{mov}}(t^n) \nabla_{\boldsymbol{x}} w_{x,h}^{g,n} - \partial_t x_{\mathrm{mov}}(t^n) \nabla_{\boldsymbol{x}} w_{y,h}^{g,n} = \mathbf{0}.$$

Since the two coefficients are not both zero, this is a non-trivial linear relation between the rows of

$$\nabla_{\boldsymbol{x}} \boldsymbol{w}_h^{g,n} = \begin{pmatrix} \partial_x w_{x,h}^{g,n} & \partial_y w_{x,h}^{g,n} \\ \partial_x w_{y,h}^{g,n} & \partial_y w_{y,h}^{g,n} \end{pmatrix}.$$

This matrix therefore has rank at most one, and its determinant vanishes on every triangle. If instead $\partial_t \boldsymbol{x}_{\mathrm{mov}}(t^n) = \mathbf{0}$, both components solve the homogeneous discrete Laplace problem with zero boundary data. Uniqueness then gives $\boldsymbol{w}_h^{g,n} = \mathbf{0}$, so the determinant also vanishes in this case. Finally, Proposition 4.6.2 yields

$$\epsilon_m^{c,n} = (\Delta t^n)^2 |S_m^{c,n}| \det\left(\nabla_{\boldsymbol{x}} \boldsymbol{w}_h^{g,n} \big|_{\omega_c(t^n)} \right) = 0.$$

By the definition (4.44), this is precisely the D-GCL (4.49). □

This includes a translating obstacle following a curved path, for instance an elliptical trajectory, with an unstructured mesh and variable κ. The result concerns a spatially uniform boundary translation, not arbitrary boundary deformations or rotation of the obstacle about its own centre. With slip conditions, the conclusion requires the scalar combination in (4.51) to retain homogeneous boundary conditions and coercivity; this is not automatic on oblique walls.

Remark 4.6.4 (*Pseudo-elastic extension*). *For the pseudo-elastic extension (4.32), D-GCL preservation is less direct than for the Laplacian model. Indeed, the elastic stress couples the two velocity components. A spatially uniform horizontal velocity on the moving boundary, with zero velocity on the fixed boundary, does not necessarily produce a horizontal velocity inside the domain. The scalar linear-combination argument of Proposition 4.6.3 therefore does not apply, and the vanishing of the grid-velocity gradient determinant is not guaranteed by these boundary data alone. To see what must be checked, consider these horizontal Dirichlet data and construct a candidate velocity by setting $w_{y,h}^{g,n} = 0$ and solving the horizontal elastic equation for $w_{x,h}^{g,n}$. This only enforces one of the two equations. The candidate solves the full elastic problem if and only if it also satisfies the vertical equation, which reduces to*

$$\int_{\Omega(t^n)} \left(\mu \partial_y w_{x,h}^{g,n} \, \partial_x z_{y,h} + \lambda \partial_x w_{x,h}^{g,n} \, \partial_y z_{y,h} \right) \mathrm{d}\boldsymbol{x} = 0, \quad \forall \boldsymbol{z}_h = z_{y,h} \boldsymbol{e}_y \in \mathbf{W}_{h,0}(t^n). \tag{4.52}$$

Thus, even when the vertical velocity is set to zero, gradients of the horizontal velocity remain in the vertical equation. Condition (4.52) must therefore be checked separately; it does not follow from the horizontal solve. If it holds, uniqueness identifies the candidate with the full elastic solution. Its gradient then has a zero second row, so its determinant vanishes and, under the geometric hypotheses of Proposition 4.6.2, the D-GCL follows. If this condition fails, the solution is not purely horizontal; its determinant must be examined directly, since a nonzero vertical component does not by itself imply a nonzero determinant. A particular case where the horizontal structure is guaranteed is the rectangular strip of Section 4.4.4, with constant Lamé coefficients $\mu > 0$, $\lambda \geq 0$, the prescribed velocities at its vertical ends, and slip on its horizontal walls: $w_y^g = 0$ and zero tangential elastic traction. The horizontal affine profile (4.34), with zero vertical component, has constant stress and satisfies both elastic equations and these boundary conditions. Since it also belongs to the $\mathbb{P}^1$ space on any triangulation, uniqueness makes it the discrete elastic solution, and the D-GCL is preserved under the same geometric hypotheses.

Remark 4.6.5 (*Time integration and prescribed positions*). *Under the geometric hypotheses of Proposition 4.6.2 at every stage, explicit RK methods, including SSP-RK, preserve the stage and final D-GCL if the coordinates and $|S_m^c| \overline{\boldsymbol{v}}_m^c$ use the same RK coefficients and either:*

- *all nodal velocities remain parallel to one fixed direction during the step, or*
- *the scalar Laplacian extension of 1 on the moving boundary and 0 on the fixed boundary is computed once at t^n. Its nodal values are kept fixed and multiplied by the uniform boundary velocity at each stage, whose direction may change.*

Recomputing the Laplacian at every stage is not sufficient in general.

4.6.2 Well-balancing property

Preserving motionless steady states is a fundamental requirement for any numerical scheme solving the shallow-water equations, as it ensures that the scheme can accurately represent equilibrium solutions without introducing spurious numerical artifacts. In the ALE framework, the moving mesh adds an additional layer of complexity to maintaining this well-balancing property, as the numerical fluxes and source terms must be carefully balanced to account for the mesh motion.

Proposition 4.6.6 (*ALE well-balancing property*). *Assume that the monolithic DG/FV-ALE subcell scheme* (4.39) *is advanced in time with a first-order Euler time-marching algorithm. Assume also that the discrete geometric conservation law* (4.49) *is satisfied on each moving subcell* $S_m^c(t)$ *with the same mesh velocity as the one used in the ALE numerical fluxes, namely*

$$|S_m^{c,n+1}| - |S_m^{c,n}| = \Delta t^n \sum_{S_p^{v,n} \in \mathcal{V}_m^{c,n}} \ell_{mp}^n \, \boldsymbol{w}_{mp}^{g,n} \cdot \mathbf{n}_{mp}^n.$$

If the source term integrals are exactly computed at motionless steady states, then the monolithic DG/FV-ALE subcell scheme preserves the lake-at-rest equilibria. More precisely,

$$\forall n \in \mathbb{N}, \quad \forall \eta^e \in \mathbb{R}, \qquad (\eta_h^n = \eta^e \quad \text{and} \quad \boldsymbol{q}_h^n = \mathbf{0}) \Longrightarrow \left(\eta_h^{n+1} = \eta^e \quad \text{and} \quad \boldsymbol{q}_h^{n+1} = \mathbf{0}\right).$$

Proof. Let us consider a moving subcell $S_m^c(t) \in \mathcal{T}_{\omega_c}^k(t)$. The monolithic DG/FV-ALE subcell scheme writes

$$|S_m^{c,n+1}| \overline{\boldsymbol{v}}_m^{c,n+1} = |S_m^{c,n}| \overline{\boldsymbol{v}}_m^{c,n} - \Delta t^n \sum_{S_p^{v,n} \in \mathcal{V}_m^{c,n}} \ell_{mp}^n \widetilde{\mathbb{G}}_{mp}^{\,n} + \Delta t^n |S_m^{c,n}| \overline{\boldsymbol{B}}_m^{c,n}.$$

We assume that, at time t^n, the numerical solution is a motionless steady state, namely $\eta_h^n = \eta^e$ and $\boldsymbol{q}_h^n = \mathbf{0}$. We first consider the free-surface equation. Since the mass component of the physical flux is $\boldsymbol{q}_h \cdot \mathbf{n}$, the physical mass flux vanishes at lake at rest. Therefore, on each subcell interface $\Gamma_{mp}^{c,n}$, the only remaining contribution in the ALE mass flux is the geometric transport term:

$$\widetilde{\mathbb{G}}_{mp}^{\,\eta,n} = -\eta^e \, \boldsymbol{w}_{mp}^{g,n} \cdot \mathbf{n}_{mp}^n.$$

The first component of the update then gives

$$|S_m^{c,n+1}| \overline{\eta}_m^{c,n+1} = |S_m^{c,n}| \eta^e - \Delta t^n \sum_{S_p^{v,n} \in \mathcal{V}_m^{c,n}} \ell_{mp}^n \left(-\eta^e \, \boldsymbol{w}_{mp}^{g,n} \cdot \mathbf{n}_{mp}^n\right) = \eta^e \left(|S_m^{c,n}| + \Delta t^n \sum_{S_p^{v,n} \in \mathcal{V}_m^{c,n}} \ell_{mp}^n \boldsymbol{w}_{mp}^{g,n} \cdot \mathbf{n}_{mp}^n \right).$$

Using the discrete geometric conservation law (4.49), we get

$$|S_m^{c,n+1}| \overline{\eta}_m^{c,n+1} = \eta^e |S_m^{c,n+1}|.$$

Since $|S_m^{c,n+1}| > 0$, it follows that $\overline{\eta}_m^{c,n+1} = \eta^e$. We now consider the momentum equations. The blended ALE flux can be decomposed as

$$\widetilde{\mathbb{G}}_{mp} = \widetilde{\mathbb{F}}_{mp} - \widetilde{\mathbb{V}}_{mp} \left(\boldsymbol{w}_{mp}^{g} \cdot \mathbf{n}_{mp}\right),$$

where $\widetilde{\mathbb{F}}_{mp}$ denotes the blended Eulerian numerical flux, and $\widetilde{\mathbb{V}}_{mp}$ is a numerical interface value consistent with $\boldsymbol{v}$. At lake at rest, both the high-order reconstructed states and the low-order hydrostatic FV reconstructed states satisfy

$$\breve{\eta}^- = \breve{\eta}^+ = \eta^e, \qquad \breve{\boldsymbol{q}}^- = \breve{\boldsymbol{q}}^+ = \mathbf{0}.$$

Therefore, by consistency of both interface values entering the blended flux, the momentum components of the ALE transport state vanish:

$$\left(\widetilde{\mathbb{V}}_{mp}\right)^{q_x} = \left(\widetilde{\mathbb{V}}_{mp}\right)^{q_y} = 0.$$

Consequently, the ALE geometric correction does not contribute to the momentum equations:

$$\left(\widetilde{\mathbb{G}}_{mp}\right)^{\boldsymbol{q}} = \left(\widetilde{\mathbb{F}}_{mp}\right)^{\boldsymbol{q}}.$$

Thus, at a motionless steady state, the ALE momentum flux reduces exactly to the Eulerian hydrostatic momentum flux. Invoking the Eulerian well-balancing result of Proposition 3.4.8, the blended Eulerian fluxes are then exactly balanced by the discrete bathymetry source term. Hence the momentum update gives

$$|S_m^{c,n+1}| \overline{\boldsymbol{q}}_m^{c,n+1} = \mathbf{0}.$$

Since $|S_m^{c,n+1}| > 0$, we obtain $\overline{\boldsymbol{q}}_m^{c,n+1} = \mathbf{0}$. We have therefore shown that, for every moving subcell $S_m^c(t)$,

$$\overline{\eta}_m^{c,n+1} = \eta^e, \qquad \overline{\boldsymbol{q}}_m^{c,n+1} = \mathbf{0}.$$

Equivalently, the updated subcell representation is the projection of the polynomial state

$$\eta_h^{c,n+1} = \eta^e, \qquad \boldsymbol{q}_h^{c,n+1} = \mathbf{0}.$$

Indeed, the constant function η^e and the zero momentum field belong to $\mathbb{P}^k(\omega_c)$, and their subcell mean values are exactly $\overline{\pi}_c^k \eta^e = \eta^e$ and $\overline{\pi}_c^k \mathbf{0} = \mathbf{0}$. Therefore, reconstructing the polynomial degrees of freedom from these subcell mean values gives back the same polynomial state. Hence,

$$\eta_h^{c,n+1} = \eta^e, \qquad \boldsymbol{q}_h^{c,n+1} = \mathbf{0}, \qquad \forall \omega_c(t) \in \mathcal{T}_h(t).$$

This proves the preservation of lake-at-rest steady states by the monolithic DG/FV-ALE subcell scheme. □

4.7 Numerical illustrations

In this section, we present some preliminary numerical results obtained with the monolithic DG/FV-ALE subcell scheme (4.39) for the two-dimensional shallow-water equations. We first verify the preservation of a motionless steady state over a non-flat bottom on a moving unstructured mesh, using the Laplacian grid velocity (4.31). The following test cases are *homemade* and designed to validate the operational behavior of the ALE implementation in controlled settings. Their main purpose is to verify that the moving-grid formulation is effectively coupled to the shallow-water solver, and that the ALE fluxes are able to convert a prescribed boundary motion into a physically meaningful wave response. For these wave-generation experiments, we use (4.34) in the pseudo-1D configurations, and the pseudo-elastic grid velocity (4.32) in the genuinely 2D configurations.

4.7.1 Well-balancing on a moving unstructured mesh

PARAMETERS

Computational domain: $\Omega = [0,2] \times [0,1]$
Number of cells: $N_{\text{el}} = 506$ cells
Boundary conditions: fixed solid walls

Polynomial degree: $k = 3$
Time-marching order: $p = 1$
Final time: $T_{\max} = 20$ sec

We consider a fully immersed lake-at-rest configuration similar to that presented in Chapter 3, on the same coarse triangulation. The initial state is $\eta(\boldsymbol{x}, 0) = 1$ and $\boldsymbol{q}(\boldsymbol{x}, 0) = \mathbf{0}$, over the fixed bottom

$$b(\boldsymbol{x}) = \begin{cases} 0.5e^{-100((x-1.2)^2+150(y-0.7)^2)} & \text{if } x > 0.68, \\ -0.5e^{-100((x-0.45)^2+150(y-0.4)^2)} & \text{otherwise.} \end{cases}$$

Unlike the subsequent wavemaker experiments, the physical boundary is kept fixed: only interior mesh vertices are displaced. No physical motion is therefore imposed on the fluid, whose exact solution remains stationary. Preserving this equilibrium requires the hydrostatic flux–source balance and the discrete geometric conservation law to hold simultaneously. Their compatibility within the high-order monolithic DG/FV subcell formulation is an original aspect of the present ALE extension; this experiment provides a numerical verification of Proposition 4.6.6, beyond the preservation of constant states on a flat bottom.

The nine interior vertices initially satisfying $|x - 1| < 0.13$ and $|y - 0.5| < 0.13$ undergo the same prescribed translation i.e.

$$\boldsymbol{x}_{\text{mov}}(t) - \boldsymbol{x}_{\text{mov}}(0) = \big(0.10\sin(2\pi t),\ 0.06(1 - \cos(2\pi t))\big)^\top.$$

At each time step, we solve the $\mathbb{P}^1$ Laplacian extension (4.31) with $\kappa = 1$, zero grid velocity on the outer boundary, and the common secant velocity $(\boldsymbol{x}_{\text{mov}}(t^{n+1}) - \boldsymbol{x}_{\text{mov}}(t^n))/\Delta t^n$ at these interior vertices. The two components have the same scalar harmonic extension, so the linearity argument of Proposition 4.6.3 still applies: the grid-velocity gradient has rank at most one on each triangle. The Euler geometric update therefore satisfies (4.49), although the direction of the imposed translation varies during each period.

The errors are measured on the evolving subcell mean values, with

$$E_\infty^\eta(t^n) := \max_{c,m} \big|\overline{\eta}_m^{c,n} - 1\big|, \qquad E_\infty^{q_i}(t^n) := \max_{c,m} \big|\overline{q}_{i,m}^{c,n}\big|, \quad i \in \{x, y\}.$$

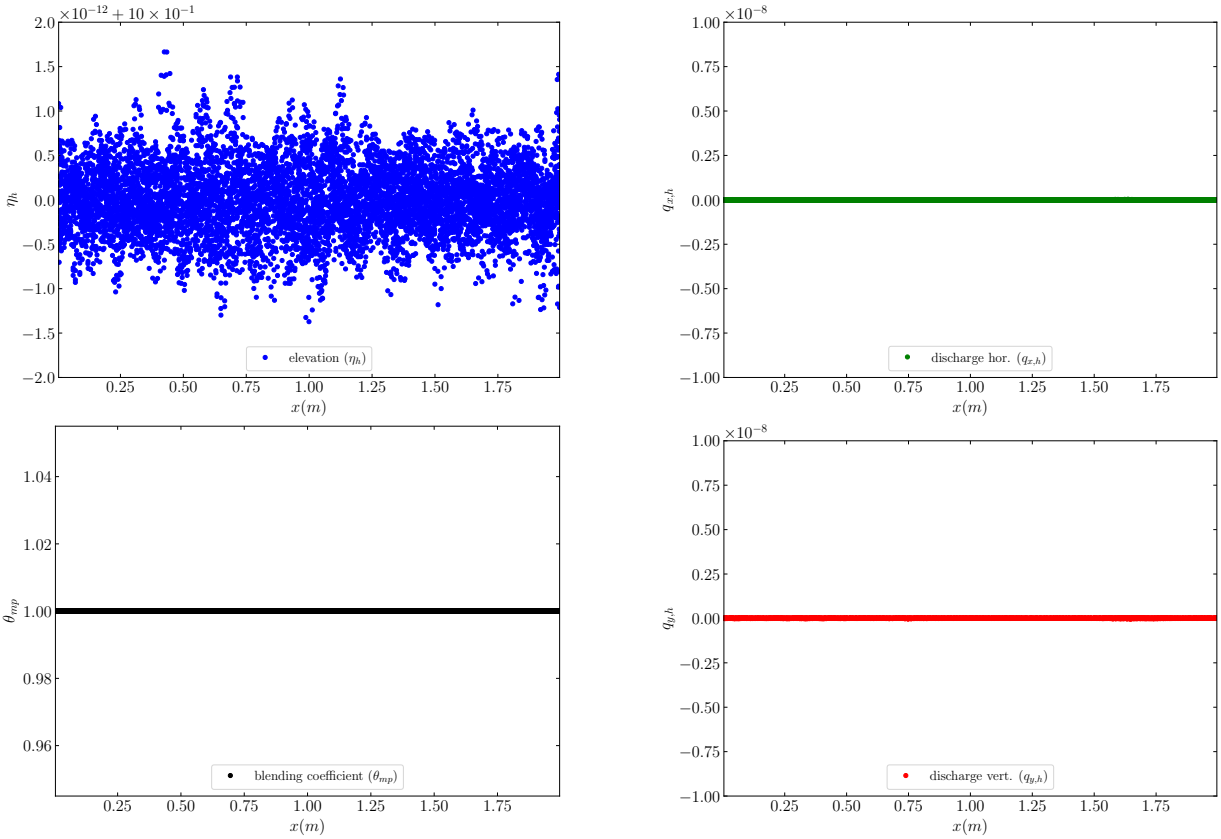


Figure 4.4: ALE well-balancing – $\mathbb{P}^3$ solution at $t = 20$ sec, from left to right and top to bottom: free-surface elevation, horizontal discharge, blending coefficient and vertical discharge density profiles on the deformed mesh (animation).

	$t = 20$ sec	$\max_n$
E_∞^η	1.6653×10^{-12}	1.6653×10^{-12}
$E_\infty^{q_x}$	3.4551×10^{-11}	3.5944×10^{-11}
$E_\infty^{q_y}$	4.1932×10^{-11}	4.2398×10^{-11}
D-GCL	5.9365×10^{-14}	1.2098×10^{-13}

Table 4.1: ALE well-balancing – $\mathbb{P}^3$ errors on 506 moving cells over 85720 Euler time steps.

We also monitor the relative geometric defect $\max_{c,m} |\epsilon_m^{c,n}|/|S_m^{c,n}|$, where $\epsilon_m^{c,n}$ is defined in (4.44). The table below reports the final errors and their maxima over all time steps.
Table 4.1 shows that the fully immersed equilibrium is preserved up to accumulated floating-point errors, with no appreciable parasitic discharge, similarly to the Eulerian case. The deformation is not negligible: the largest vertex displacement is 0.125, and the smallest ratio $|\omega_c(t^n)|/|\omega_c(0)|$ is 0.4137. We also measure $\max_{n,c} |\det(\nabla_{\boldsymbol{x}} \boldsymbol{w}_h^{g,n}|_{\omega_c(t^n)})| = 2.6645 \times 10^{-15}$, consistently with the rank-one argument above. Only the prescribed vertices follow a periodic trajectory; the freely moving interior vertices need not return exactly to their initial positions after each period, since the harmonic extension is recomputed on the current mesh.

Figure 4.4 displays the final density profiles, with the same presentation as the Eulerian well-balancing experiment: all subcell mean values are plotted against their current horizontal coordinate, without selecting a particular transverse slice.

4.7.2 Periodic wave generation by an oscillating piston

PARAMETERS

Computational domain: $\Omega = [0, 50] \times [0, 25]$ — *Polynomial degree*: $k = 2$
Number of cells: $N_{\text{el}} = 512$ cells — *Time-marching order*: $p = 3$
Boundary conditions: Dirichlet (left), solid walls (top/bottom), piston (right) — *Final time*: $T_{\max} = 20$ sec

We now consider a simple ALE benchmark designed to isolate the effect of a prescribed boundary motion on an initially motionless shallow-water state. The computational domain is a rectangular channel, with a fixed Dirichlet boundary on the left, solid walls on the top and bottom boundaries, and a horizontally moving boundary on the right. This moving boundary is used as a spring-like piston, generating periodic waves by compressing and expanding the fluid domain. The initial condition is chosen as a flat lake-at-rest state without topography,

$$b(\boldsymbol{x}) = 0, \qquad \eta(\boldsymbol{x}, 0) = 10, \qquad \boldsymbol{q}(\boldsymbol{x}, 0) = \boldsymbol{0}.$$

Therefore, in the absence of mesh motion, the solution would remain exactly stationary. Any wave observed in the numerical solution is then only generated by the ALE displacement of the right boundary. This makes the test case useful to verify that the moving-grid formulation is effectively coupled to the shallow-water solver, and that the ALE fluxes are able to convert a prescribed boundary motion into a physically meaningful wave response.

The horizontal displacement imposed on the right boundary is prescribed by

$$d(t) = r(t) A_{\text{p}} \sin(\omega_{\text{p}} t), \qquad \omega_{\text{p}} = \frac{2\pi}{T_{\text{p}}},$$

with amplitude $A_{\text{p}} = -3$ and period $T_{\text{p}} = 4$ sec. The negative sign means that the first displacement of the moving boundary is directed towards the interior of the domain. To avoid an impulsive start at $t = 0$, we use the smooth ramp function

$$r(t) = \begin{cases} \dfrac{1}{2}\left(1 - \cos\left(\dfrac{\pi t}{T_{\text{r}}}\right)\right), & \text{if } t < T_{\text{r}}, \\ 1, & \text{otherwise}, \end{cases} \qquad T_{\text{r}} = \frac{T_{\text{p}}}{2}.$$

The velocity imposed on the moving boundary is then obtained by differentiating the previous displacement law,

$$\partial_t d(t) = r'(t) A_{\text{p}} \sin(\omega_{\text{p}} t) + r(t) A_{\text{p}} \omega_{\text{p}} \cos(\omega_{\text{p}} t), \tag{4.53}$$

with

$$r'(t) = \begin{cases} \dfrac{\pi}{2T_{\text{r}}} \sin\left(\dfrac{\pi t}{T_{\text{r}}}\right), & \text{if } t < T_{\text{r}}, \\ 0, & \text{otherwise}. \end{cases}$$

Thus, the prescribed grid velocity on the right boundary is $\boldsymbol{w}^g = (\partial_t d(t), 0)^\top$.
This test case is deliberately elementary, as it is designed mainly to validate the operational behavior of the ALE implementation in a controlled setting. Starting from a perfectly flat free-surface, the oscillating boundary acts

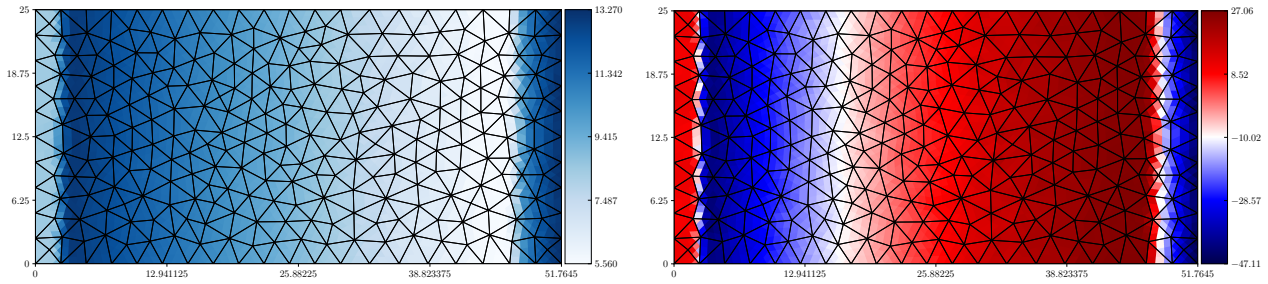


Figure 4.5: Oscillating piston test case − At $t = 7.6$ sec, from left to right: free-surface elevation and horizontal discharge q_x (animation).

as a periodic wavemaker. This provides a direct way to observe whether the mesh motion, the ALE numerical fluxes, and the DG/FV subcell stabilization are consistently combined. Figure 4.5 shows a two-dimensional snapshot of the free-surface elevation and horizontal discharge at $t = 7.6$ sec. We observe that the solution remains symmetric on this coarse grid with respect to the channel centerline, as expected from the geometry and from the purely horizontal prescribed motion. The resulting wave fronts remain sharply captured, without visible spurious oscillations, even after several cycles of mesh compression and expansion. Figure 4.6 reports corresponding one-dimensional density profiles during the second part of the first period. The solution remains free of spurious oscillations, and the wave amplitude is consistent with the expected physical response to the imposed boundary motion. An animation of the solution is available here.

4.7.3 Wave run-up over a sloping beach generated by an impulsively stopped piston

PARAMETERS

Computational domain: $\Omega = [0, 50] \times [0, 25]$
Number of cells: $N_{\text{el}} = 2060$ cells
Boundary conditions: Dirichlet (left), solid walls (top/bottom), piston (right)
Polynomial degree: $k = 1$
Time-marching order: $p = 3$
Final time: $T_{\max} = 20$ sec

We now consider a second ALE benchmark, built from the same piston mechanism as in the previous test case, but with a different objective. The right boundary first moves as a spring-like piston, generating a single incoming wave, and is then stopped abruptly. After this stopping time, the computational domain remains in its compressed configuration and the wave propagates over a non-flat bathymetry. This test is therefore more demanding than the previous one, since it combines the ALE moving mesh formulation with a non-trivial topography and a wet/dry transition generated by a sloping beach.

The bathymetry is defined as a simple plane beach on the left part of the domain:

$$b(\boldsymbol{x}) = \begin{cases} b_\ell \left(1 - \dfrac{x}{x_{\text{b}}}\right), & \text{if } x \leq x_{\text{b}}, \\ 0, & \text{otherwise,} \end{cases} \qquad x_{\text{b}} = 15, \qquad b_\ell = 20.$$

Thus the topography decreases linearly from $b = 20$ at $x = 0$ to $b = 0$ at $x = 15$, and remains flat for $x > 15$. The initial condition is chosen as a lake-at-rest state with a dry part induced by the beach,

$$\eta(\boldsymbol{x}, 0) = \max\big(b(\boldsymbol{x}), 10\big), \qquad \boldsymbol{q}(\boldsymbol{x}, 0) = \boldsymbol{0}.$$

Hence the right part of the domain is initially wet with a flat free-surface $\eta = 10$, while the left part contains a dry area where the topography is above the water level.

The imposed motion of the right boundary follows the same ramped sinusoidal law (4.53) as before, but is stopped at time $t_{\text{s}} = 1.25$ sec. This creates a single-wave through the initial compression of the mesh, after which the mesh remains fixed in its deformed state.
This benchmark is intended to check several features at once. First, during the early stage $t \in [0, 1.25]$ sec, the wave is generated only by the mesh compression induced by the piston. Thus, as in the previous test case, the fluid is initially at rest and the perturbation comes exclusively from the ALE boundary motion. Secondly, once the piston has stopped, the mesh is no longer moving and the generated wave evolves over a non-flat

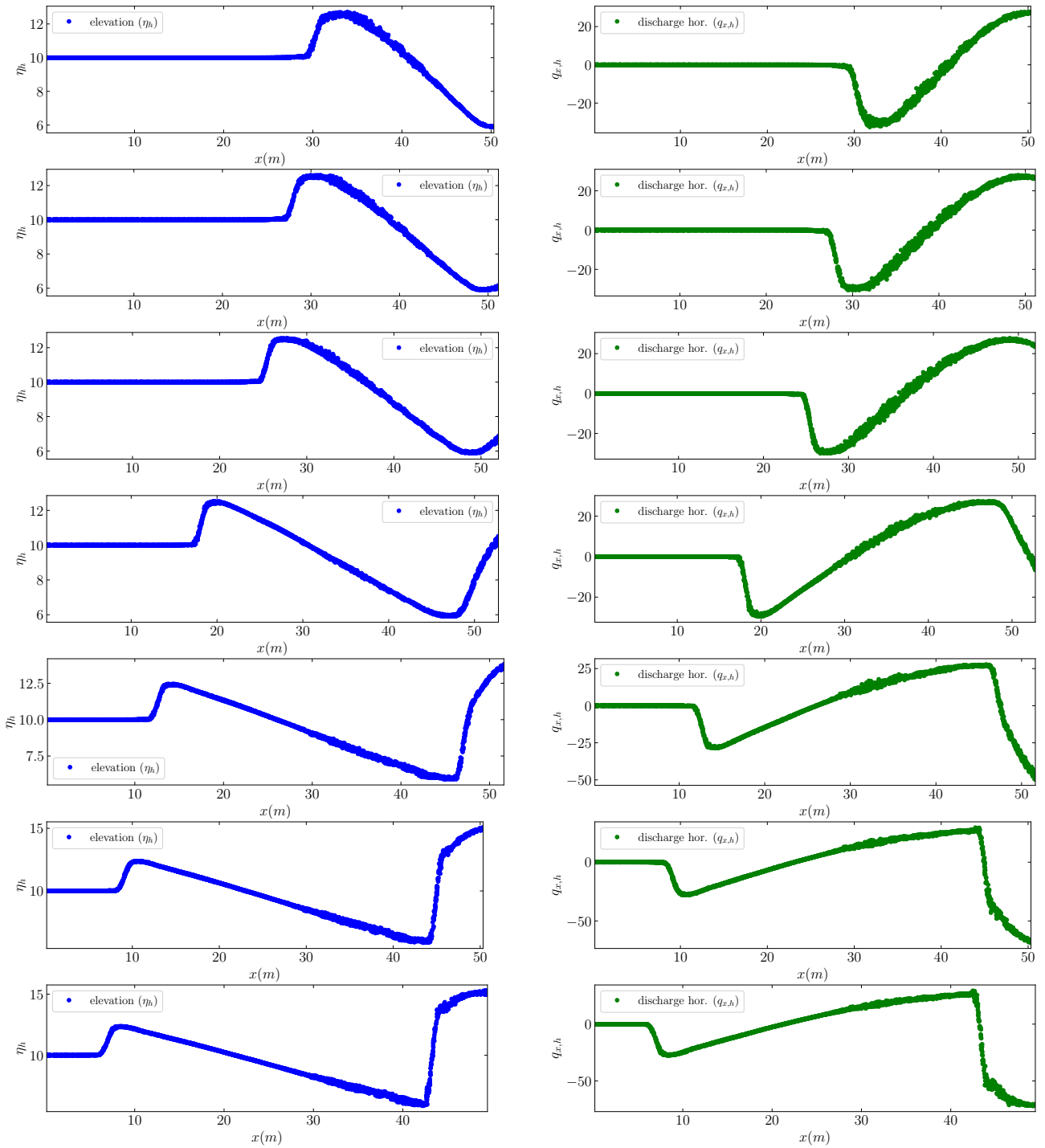


Figure 4.6: Oscillating piston test case − One-dimensional snapshots of the free-surface elevation (left column) and horizontal discharge (right column) during [2.5, 4] sec (animation).

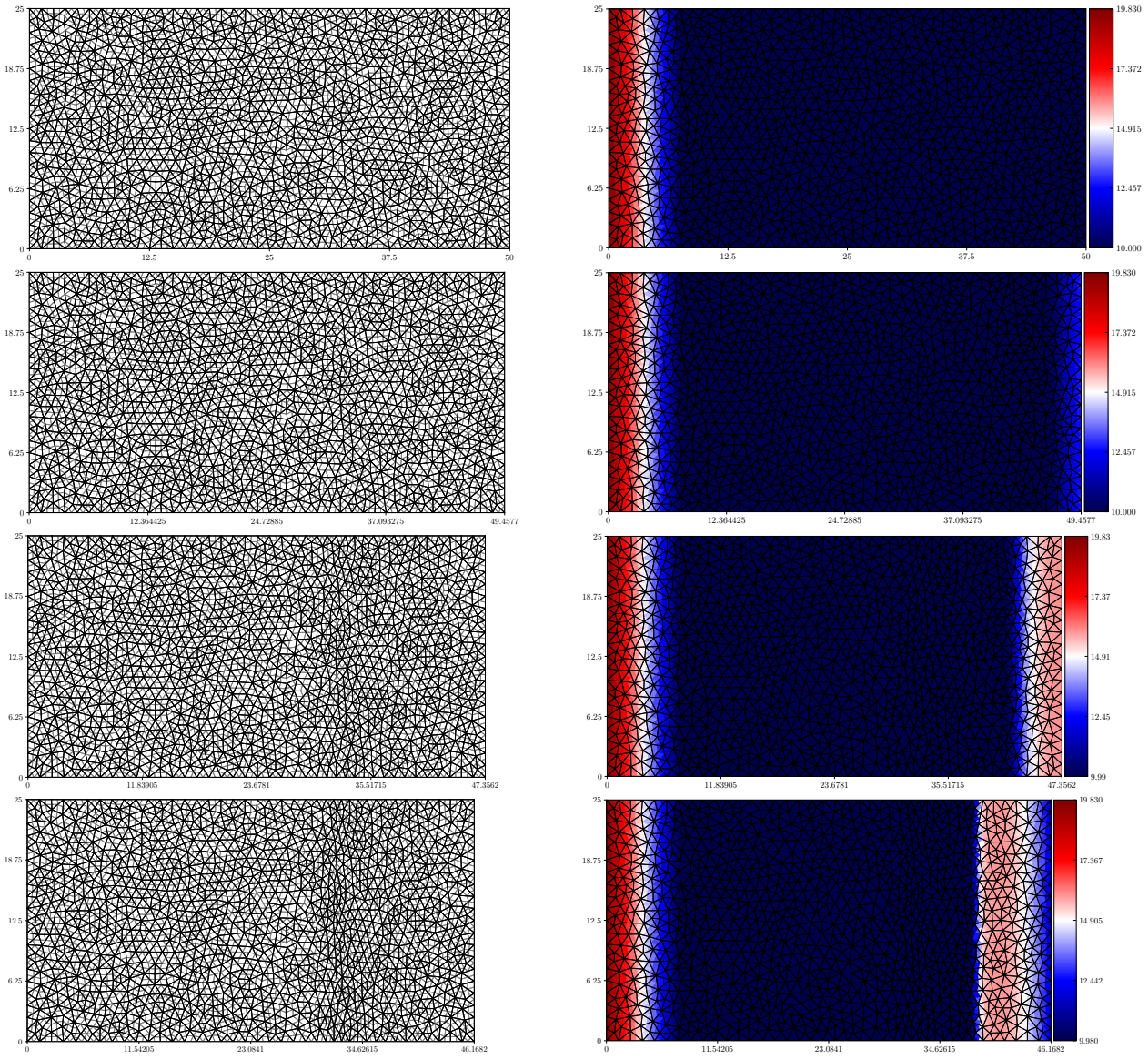


Figure 4.7: Stopped piston with sloping beach − Snapshots of the ALE mesh deformation (left column) and free-surface elevation (right column) during the piston motion, for $t \in [0, 1.25]$ sec (animation).

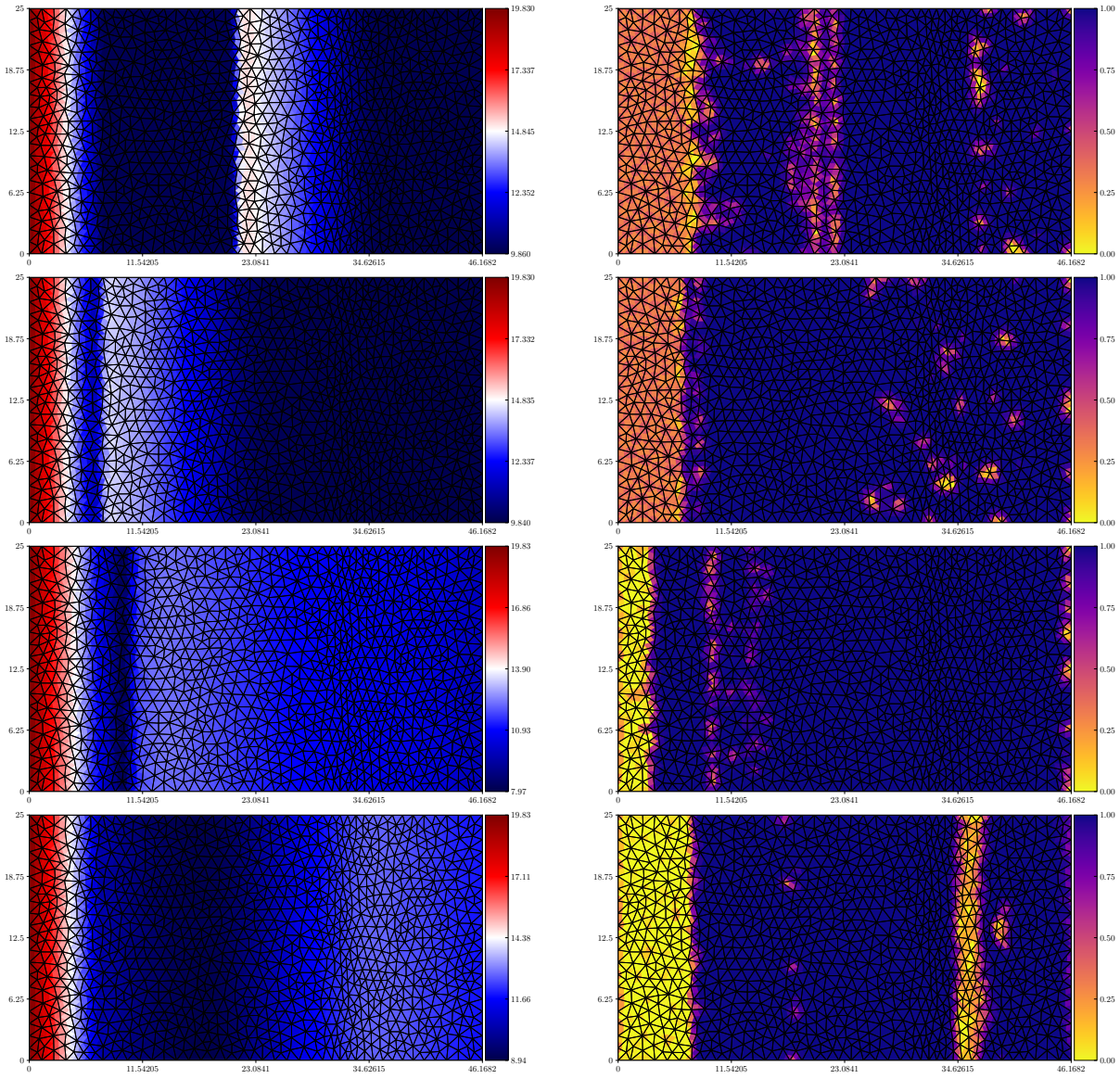


Figure 4.8: Stopped piston with sloping beach – Two-dimensional snapshots of the wave propagation and run-up over the sloping beach, with map of blending coefficients, for $t \in [2.5, 7.5]$ sec (animation).

bathymetry. This allows us to observe whether the transition between the moving-grid phase and the fixed deformed configuration is handled without introducing non-physical oscillations.

Figure 4.7 shows the mesh deformation during the first part of the simulation, before the piston stops. The deformation remains regular and localized near the moving side of the domain, while the quality of the mesh is preserved. Figure 4.8 then displays the subsequent evolution of the wave as it travels towards the beach. The solution has the expected qualitative behavior: the wave generated by the piston propagates leftward, interacts with the sloping bathymetry, and finally runs up over the initially dry part of the domain. This test therefore indicates that the ALE formulation remains compatible with the well-balanced and positivity-preserving mechanisms used for topography and wet/dry fronts. The map of blending coefficients also behaves as in the Eulerian setting: the limiting remains localized only in dry regions and/or where the solution develops sharp gradients, while the smooth parts of the flow keep a high-order treatment.
The corresponding one-dimensional profiles are reported in Figure 4.9. They make the run-up dynamics easier to observe, especially near the wet/dry transition. The front remains sharply captured and does not generate visible spurious oscillations when it reaches the sloping part of the domain. This test (with the simulation available here) therefore gives a useful validation of the ALE treatment of topography, dry zones and mesh-induced wave generation in the same numerical experiment.

4.7.4 Wave generation by an oscillating cylindrical wavemaker

We now consider a genuinely two-dimensional ALE test case in which the moving boundary is no longer located on the exterior of the domain, but is instead an internal cylindrical boundary. This configuration should not be interpreted as a floating body model (as there is no water in this area); it may be viewed instead as a simplified cylindrical wavemaker, or as a vertical pile oscillating horizontally in a shallow basin. The computational mesh contains a circular hole, and the nodes located on this internal boundary have a prescribed velocity, while the external boundary is kept fixed. This provides a simple but demanding test for the ALE implementation, since the mesh deformation is now propagated from an internal moving boundary toward the surrounding fixed outer boundary.

The initial condition is again chosen as a flat lake-at-rest state without topography,

$$b(\boldsymbol{x}) = 0, \qquad \eta(\boldsymbol{x}, 0) = 10, \qquad \boldsymbol{q}(\boldsymbol{x}, 0) = \boldsymbol{0}.$$

Thus, without motion of the internal boundary, the numerical solution would remain stationary. The waves observed in this test case are therefore entirely generated by the ALE displacement of the cylindrical boundary. This makes the benchmark useful to assess whether the grid velocity extension, the moving-boundary ALE fluxes, and the subcell stabilization remain robust when the mesh deformation is no longer essentially one-dimensional. The reader can find a simulation on similar benchmarks with different settings on this link.

Unidirectional oscillation

PARAMETERS

Computational domain: $([-40, 40] \times [-20, 20]) \setminus \{\boldsymbol{x} \in \mathbb{R}^2 \mid |\boldsymbol{x}| < 5\}$
Number of cells: $N_{\text{el}} = 2340$ cells
Boundary conditions: solid walls (outside), moving boundary (inside)
Polynomial degree: $k = 2$
Time-marching order: $p = 3$
Final time: $T_{\max} = 3$ sec

The velocity imposed on the cylindrical boundary is prescribed as

$$\boldsymbol{w}^g_{|\partial\Omega_{\text{mov}}(t)} = \left(A_x \cos\left(\frac{2\pi}{T} t\right), 0 \right)^\top,$$

with $A_x = -8$, $A_y = 0$, and $T = 5$ sec. In the present experiment, the cylinder therefore oscillates only in the horizontal direction, and the negative sign in A_x only fixes the first direction of motion. The velocity is imposed strongly on the internal boundary nodes, while the remaining boundary nodes are kept fixed, so that all the deformation is generated by the motion of the cylindrical hole. The interior grid velocity is then obtained by solving the ALE mesh-motion problem using the pseudo-elasticity model introduced previously, which allows the mesh to act as an elastic medium.
Figure 4.11 shows the evolution of the free-surface elevation during the simulation. The horizontal oscillation of the internal cylindrical boundary generates cylindrical shock waves that propagate away from the moving obstacle and interact with the surrounding channel. Compared with the previous piston configuration, this test is less directional and produces a more genuinely two-dimensional wave field, since the perturbation is created

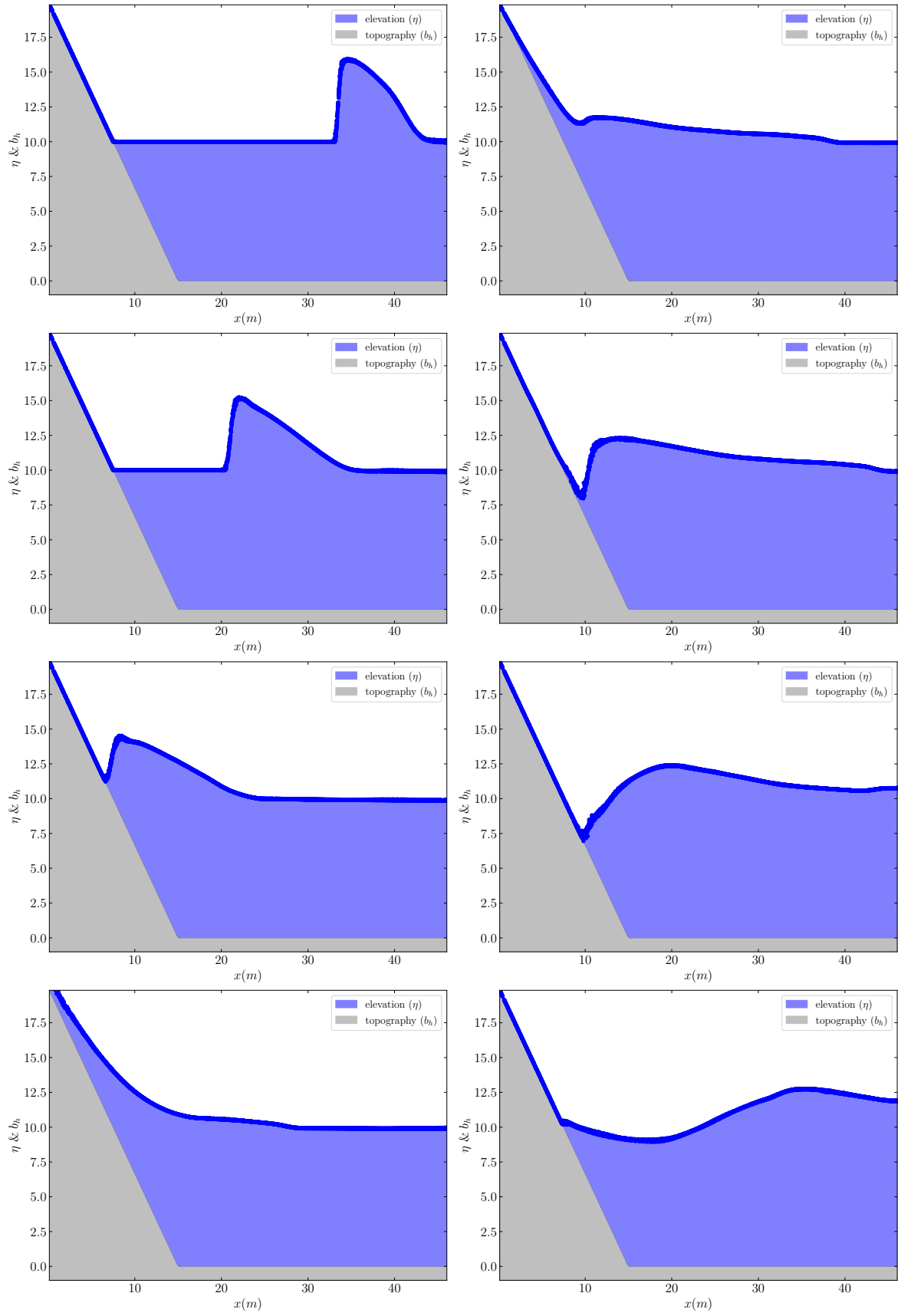


Figure 4.9: Stopped piston with sloping beach – Columnwise snapshots of the free-surface elevation density profile for $t \in [1.6, 7.6]$ sec (animation).

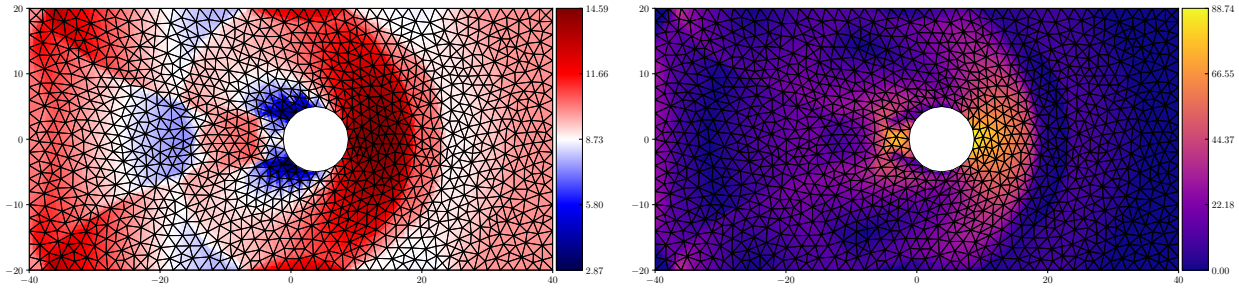


Figure 4.10: Oscillating cylindrical wavemaker (horizontal motion) – At final time: free-surface elevation (left) and discharge norm (right).

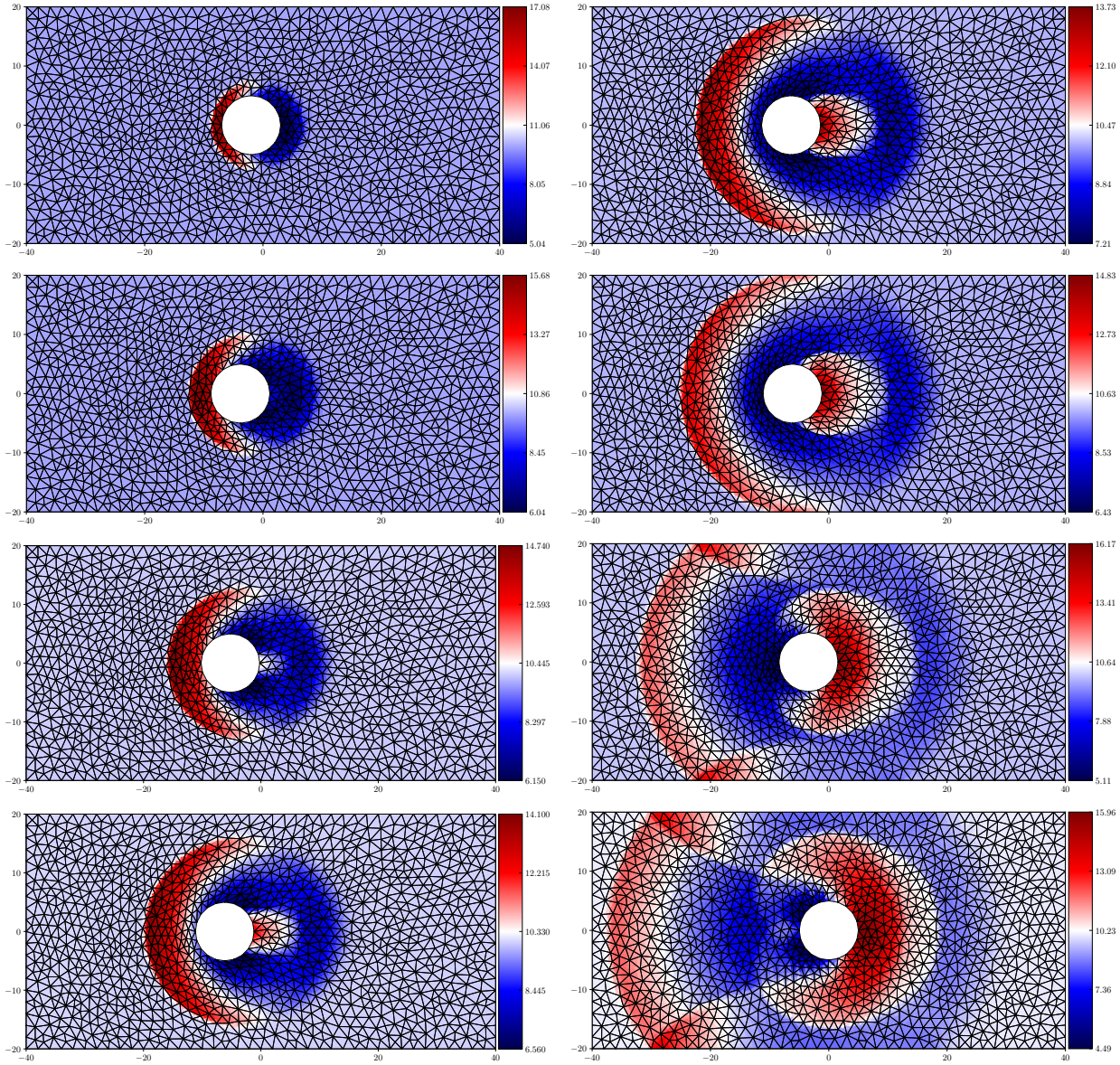


Figure 4.11: Oscillating cylindrical wavemaker (horizontal motion) – Two-dimensional snapshots of the free-surface elevation during the period $[0.5, 2.5]$ sec, ordered columnwise from top to bottom.

from an internal moving boundary rather than from one side of the domain. It is therefore a useful intermediate benchmark between the simple external-piston test and a future fluid-structure configuration involving a moving immersed object.

The final-time solution reported in Figure 4.10 confirms that the ALE formulation remains stable under this internal mesh deformation. Indeed, the method sharply preserves the several cylindrical wave fronts generated by the obstacle motion, without visible parasitic oscillations near the moving boundary or in the wake of the propagating shocks. The discharge remains coherent with the wave pattern and follows the expected radial structure induced by the oscillating cylinder. This test therefore confirms that the monolithic DG/FV-ALE subcell scheme can handle internal moving boundaries, preserve clean cylindrical shock structures, and generate stable two-dimensional wave responses.

Circular oscillation

PARAMETERS

Computational domain: $([-40, 40] \times [-20, 20]) \setminus \{\boldsymbol{x} \in \mathbb{R}^2 \mid |\boldsymbol{x}| < 5\}$
Number of cells: $N_{\text{el}} = 2340$ cells
Boundary conditions: solid walls (outside), moving boundary (inside)
Polynomial degree: $k = 1$
Time-marching order: $p = 3$
Final time: $T_{\max} = 8$ sec

We now consider a second configuration in which the internal cylindrical boundary is no longer forced only in the horizontal direction, but follows a circular motion. The imposed grid velocity on the moving boundary is now given by

$$\boldsymbol{w}^g_{|\partial\Omega_{\text{mov}}(t)} = \left(A_x \cos\left(\frac{2\pi}{T} t\right), A_y \sin\left(\frac{2\pi}{T} t\right) \right)^\top,$$

with $A_x = -8$, $A_y = -4$, and $T = 2.5$ sec.

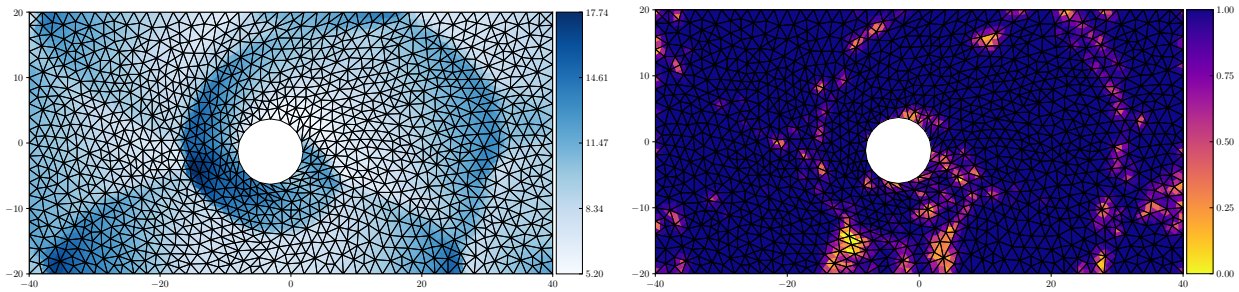


Figure 4.12: Oscillating cylindrical wavemaker (circular motion) − At final time, from left to right: free-surface elevation and blending map (animation).

Figure 4.13 shows the time evolution of the free-surface elevation during the circular motion of the internal boundary, while Figure 4.12 reports the final-time elevation together with the corresponding blending map. In contrast with the previous unidirectional oscillation, the cylindrical wavemaker now follows a genuinely two-dimensional trajectory over several periods, which produces a more complex pattern of vortexes-like waves. The snapshots show that the wave pattern remains well organized during the computation, even though the direction of the imposed deformation changes continuously. The monolithic DG/FV-ALE scheme sharply captures the resulting shocks without visible parasitic oscillations near the moving cylinder or in the interaction regions between successive fronts. The blending map at final time also shows the expected behavior of the stabilization: despite the presence of several steep gradients, the low-order correction remains localized while most of the domain keeps a high-order DG treatment. This confirms that the ALE fluxes, the mesh-motion solver and the subcell stabilization remain consistently coupled for internal moving boundaries undergoing genuinely two-dimensional motions, without introducing excessive numerical diffusion. Again, an animation of the solution is available here.

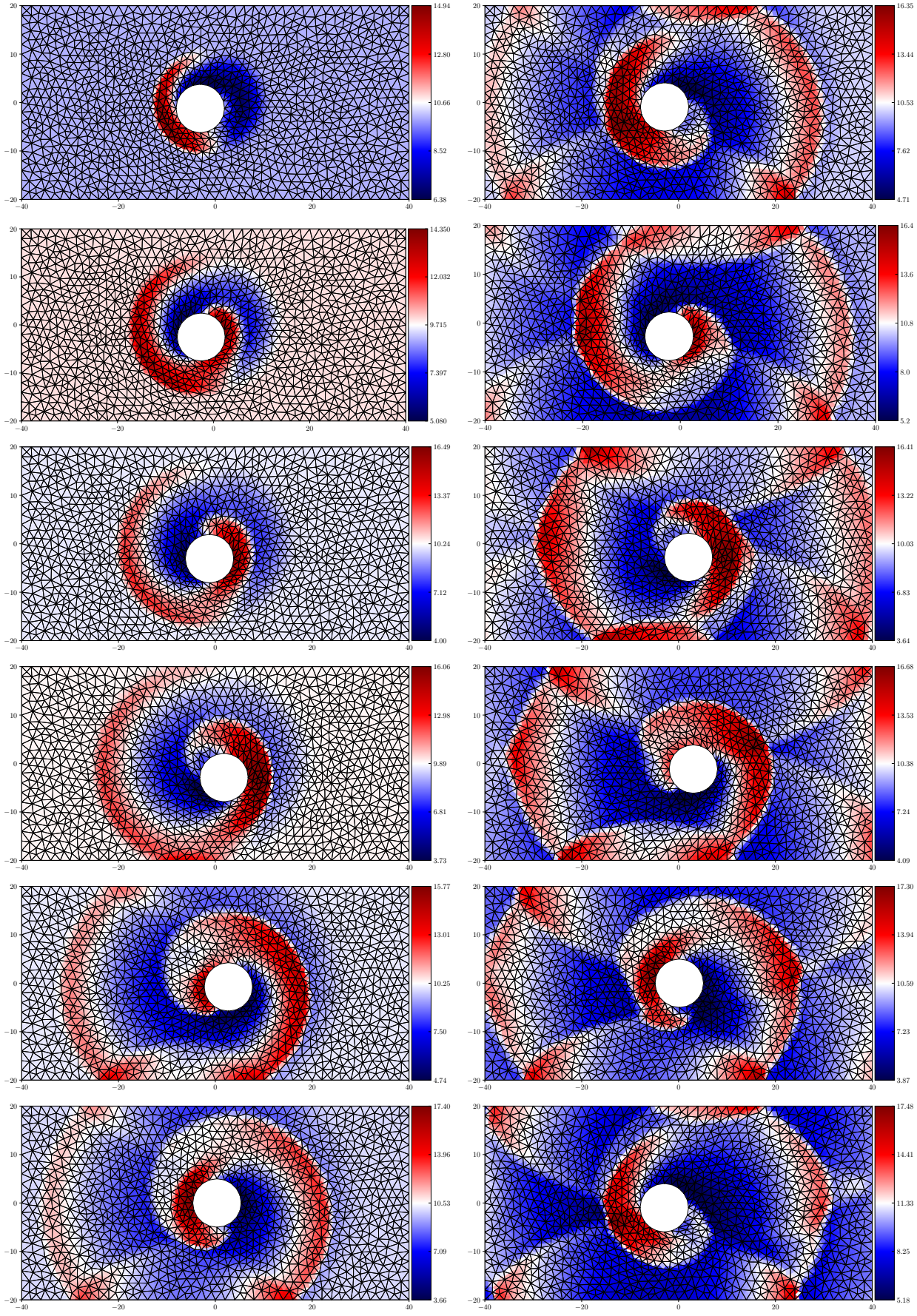


Figure 4.13: Oscillating cylindrical wavemaker (circular motion) – Columnwise from top to bottom: two-dimensional snapshots of the free-surface elevation during the period $[0.5, 7.5]$ sec (animation).

Part II

Waves interacting with floating structures in shallow-water regimes

Hark now, hear the sailors cry, smell the sea and feel the sky
Let your soul and spirit fly, into the mystic
FROM: ***"Into the Mystic"***, Van Morrison – *Moondance* (1970)

CHAPTER 5

Numerical modeling of a (pseudo)-one-dimensional wave-structure interactions problem with floating object

Some elements presented here, together with the main developments of Chapter 6, are expected to form the basis of two research articles devoted to wave-structure interactions.

In this chapter, we consider a one-dimensional nonlinear shallow-water flow interacting with a rigid structure of fixed horizontal extension. The derivation follows the ideas developed for floating structures in Bt in [19], but is adapted to the NSW setting used in this thesis, as a major step towards 2D simulations. Two configurations are treated in the same framework: a fixed surface obstacle, for which the vertical position is prescribed, and a freely floating object whose motion is restricted to the vertical direction and determined by a Newton equation. In both cases, the domain is split into two exterior shallow-water regions and one interior wetted region, where the constraint imposed by the body leads to an elliptic Bernoulli pressure problem.

The main novelty lies in the numerical strategy which targets high-order polynomial approximations through coupling the local monolithic DG/FV subcell method in the exterior regions with a SWIP-DG discretization of the interior elliptic pressure equation. The transmission conditions at the contact points enforce the exchange between the exterior flow, the pressure under the body, and, in the freely floating case, the vertical motion of the object and the mean interior discharge. The chapter therefore provides a high-order monolithic DG/FV–SWIP-DG discretization of a coupled NSW-based wave-structure interactions model, together with a pseudo-one-dimensional extension in two space dimensions with a HHO method.

Contents

5.1 Shallow-water free-surface flows 156
5.2 Floating structure in a shallow-water flow 156
5.3 Discrete formulation 161
5.3.1 SWIP-DG method for the elliptic problem on Bernoulli pressure 163
5.3.2 General SSP-RK discretization of the ODEs for the coupled system 164
5.3.3 Local subcell monolithic DG/FV method for the exterior region 165
5.3.4 Semi-discrete DG/SWIP-DG formulation 165
5.3.5 Fully-discrete monolithic DG/FV–SWIP-DG–SSP-RK formulation 166
5.4 Numerical illustrations 166
5.4.1 Geometry of the floating object 167
5.4.2 Single-wave interaction with an obstacle over an emerged topography 168
5.4.3 Wave generation by a prescribed vertical motion 169
5.4.4 Wave interaction with a freely heaving floating structure 175
5.4.5 Return to equilibrium test with analytical reference 180

I was sitting in the Hollywood Hawaiian Hotel
I was listening to the air conditioner hum; it went mm, mm, mm, mm...
FROM: ***"Desperados Under the Eaves"***, Warren Zevon – *Warren Zevon* (1976)

5.1 Shallow-water free-surface flows

We recall the one-dimensional NSW equations detailed in General shallow-water model in one space dimension, which reads as

$$\partial_t \boldsymbol{v} + \partial_x \boldsymbol{F}(\boldsymbol{v}, b) = \boldsymbol{B}[b](\boldsymbol{v}) + \boldsymbol{P}[\underline{\mathrm{p}}](\boldsymbol{v}), \tag{5.1}$$

supplemented with some initial-data $\boldsymbol{v}(\cdot, 0) \coloneqq \boldsymbol{v}^0$. One can observe the occurrence of the pressure term $\boldsymbol{P}[\underline{\mathrm{p}}](\boldsymbol{v})$ since we consider a free-surface flow interacting with a surface obstacle.

5.2 Floating structure in a shallow-water flow

We consider a floating non-deformable object, denoted by $\mathcal{O}_{bj}$, of mass m_O, inertia coefficient i_O and center of mass $\mathcal{M}_O^G$, which is partly immersed in an inviscid, incompressible and irrotational fluid. We assume that there are only two time-independent *contact points* between the fluid, the air, and the object, denoted by χ_l and χ_r, with, in our case, a symmetric object such that $\chi_\mathrm{l} \coloneqq -l < 0 < \chi_\mathrm{r} \coloneqq l$, $l > 0$, i.e. $\chi_\mathrm{r} - \chi_\mathrm{l} = 2l$. We also denote by δ_O the vertical deviation of the object from its equilibrium position.

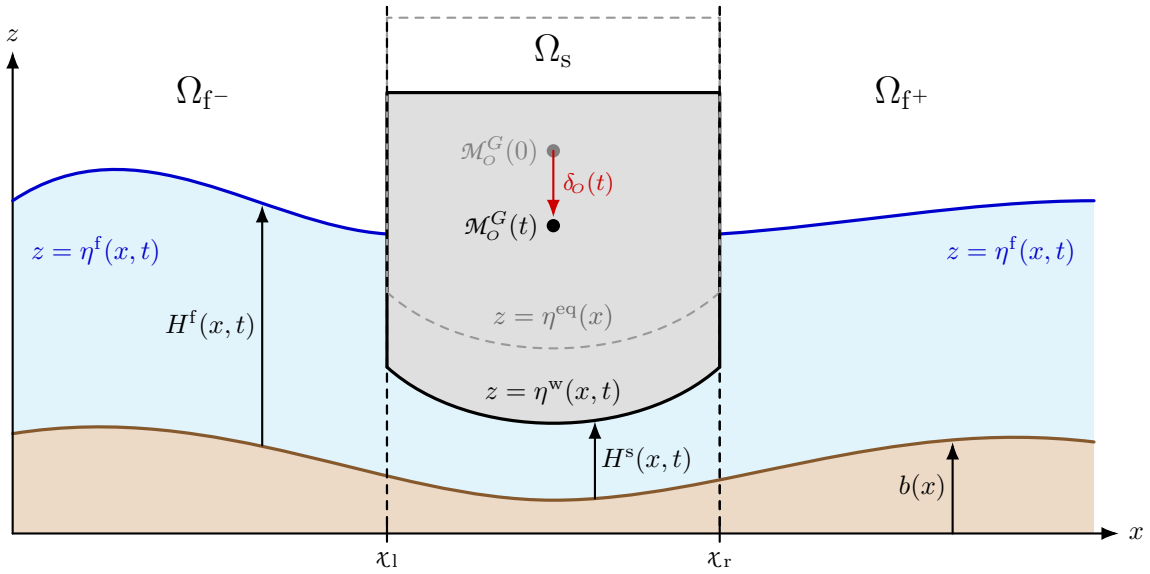


Figure 5.1: Schematic representation of the one-dimensional wave-structure interactions problem.

We define the solid domain and the fluid domain as two subdomains:

$$\Omega_\mathrm{s} \coloneqq (\chi_\mathrm{l}, \chi_\mathrm{r}), \quad \Omega_\mathrm{f} \coloneqq \Omega_{\mathrm{f}^-} \cup \Omega_{\mathrm{f}^+}, \quad \text{with} \quad \Omega_{\mathrm{f}^-} \coloneqq (-\infty, \chi_\mathrm{l}), \quad \Omega_{\mathrm{f}^+} \coloneqq (\chi_\mathrm{r}, +\infty).$$

Here, Ω_s represents the *interior* domain (under the floating object) and Ω_f the *exterior* domain (exposed to the air). The boundary of the interior region is $\partial\Omega_\mathrm{fs} = \{\chi_\mathrm{l}, \chi_\mathrm{r}\}$. In each region, we denote by $H^\bullet$ the water height, by $u^\bullet$ the horizontal velocity, and by $q^\bullet$ the corresponding discharge, with

$$\eta^\bullet \coloneqq H^\bullet + b, \qquad q^\bullet \coloneqq H^\bullet u^\bullet, \qquad \bullet \in \{\mathrm{f}^-, \mathrm{f}^+, \mathrm{s}\}.$$

Here f^- and f^+ refer respectively to the left and right connected components of the exterior fluid domain. The conservative variables are gathered into the vector

$$\boldsymbol{v}^\bullet \coloneqq \begin{pmatrix} \eta^\bullet \\ q^\bullet \end{pmatrix}, \qquad \bullet \in \{\mathrm{f}^-, \mathrm{f}^+, \mathrm{s}\}.$$

Let $\mu \in \mathsf{H}^1(\Omega_\mathrm{f}) \times \mathsf{H}^1(\Omega_\mathrm{s})$ be a sufficiently regular function that admits well-defined one-sided traces at the contact points $x = \chi_{\mathrm{l/r}}$. For each $\bullet \in \{\mathrm{f}^-, \mathrm{s}, \mathrm{f}^+\}$, we denote by $\mu^\bullet \coloneqq \mu_{|\Omega_\bullet}$ the restriction of μ to $\Omega_\bullet$. For $\bullet \in \{\mathrm{f}, \mathrm{s}\}$, we define the left and right traces associated with the exterior and interior regions by

$$\mu_-^\bullet \coloneqq \begin{cases} \lim\limits_{x \to \chi_\mathrm{l}^-} \mu(x) & \text{if } \bullet = \mathrm{f}, \\ \lim\limits_{x \to \chi_\mathrm{l}^+} \mu(x) & \text{if } \bullet = \mathrm{s}, \end{cases} \qquad \mu_+^\bullet \coloneqq \begin{cases} \lim\limits_{x \to \chi_\mathrm{r}^+} \mu(x) & \text{if } \bullet = \mathrm{f}, \\ \lim\limits_{x \to \chi_\mathrm{r}^-} \mu(x) & \text{if } \bullet = \mathrm{s}. \end{cases}$$

The corresponding jumps and averages are then defined, for $\bullet \in \{\mathrm{f}, \mathrm{s}\}$, by

$$[\![\mu^\bullet]\!] := \mu^\bullet_+ - \mu^\bullet_-, \qquad \langle \mu^\bullet \rangle := \frac{1}{2}\left(\mu^\bullet_+ + \mu^\bullet_-\right).$$

We assume hydrostatic pressure in each region:

$$\underline{\mathrm{p}}(x,z,t) = \begin{cases} \underline{\mathrm{p}}^{\mathrm{f}}(x,t) - \rho \mathrm{g}(z - \eta^{\mathrm{f}}(x,t)) & \text{in } \Omega_{\mathrm{f}}, \\ \underline{\mathrm{p}}^{\mathrm{s}}(x,t) - \rho \mathrm{g}(z - \eta^{\mathrm{s}}(x,t)) & \text{in } \Omega_{\mathrm{s}}, \end{cases} \tag{5.2}$$

where $\underline{\mathrm{p}}^{\mathrm{f}}(x,t)$ is the atmospheric pressure and $\underline{\mathrm{p}}^{\mathrm{s}}(x,t)$ is the pressure exerted by the solid object on the fluid from below.

The governing equations for the fluid–structure interaction problem are derived from the shallow-water equations in each domain, relying on the derivation from [19] for Boussinesq equations with appropriate boundary conditions and matching conditions at the contact points $\varkappa_{\mathrm{l}}$ and $\varkappa_{\mathrm{r}}$.

Constraint in the solid region Ω_{s}

In this region the fluid surface of the fluid coincides with the bottom of the object in the interior region,

$$\eta^{\mathrm{s}} = \eta^{\mathrm{w}} \quad \text{ on } \partial\Omega_{\mathrm{s}} \times (0,T), \tag{5.3}$$

where η^{w} parametrizes the wet extremity of the object, and is defined as

$$\eta^{\mathrm{w}}(x,t) = \eta^{\mathrm{eq}}(x) + \delta_{\mathrm{O}}(t) \quad \text{ on } \Omega_{\mathrm{s}} \times (0,T), \tag{5.4}$$

where $\eta^{\mathrm{eq}}(x) := H^{\mathrm{eq}} + b(x)$, H^{eq} being the height of the object at equilibrium, and assuming that the bottom of the object is symmetric, i.e. η^{w} is an even function. Let us remark that the second conservation law in (5.1)

$$\partial_t q^{\mathrm{s}} + \partial_x \left(u^{\mathrm{s}} q^{\mathrm{s}} + \frac{1}{2}\mathrm{g}\eta^{\mathrm{s}}(\eta^{\mathrm{s}} - 2b) \right) = -\mathrm{g}\eta^{\mathrm{s}}\partial_x b - \frac{1}{\rho} H^{\mathrm{s}} \partial_x \underline{\mathrm{p}}^{\mathrm{s}}, \tag{5.5}$$

can be rewritten as

$$\partial_t u^{\mathrm{s}} + \frac{1}{\rho}\partial_x \Pi^{\mathrm{s}} = 0,$$

where $\Pi^{\mathrm{s}} := \underline{\mathrm{p}}^{\mathrm{s}} - \mathrm{p}_{\mathrm{atm}} + \rho \mathrm{g}\zeta^{\mathrm{s}} + \frac{1}{2}\rho (u^{\mathrm{s}})^2$ is the Bernoulli pressure in the interior region. Using (5.4), along with the relation $u^{\mathrm{s}} = q^{\mathrm{s}}/H^{\mathrm{w}}$, the system

$$\begin{cases} \partial_t \eta^{\mathrm{w}} + \partial_x q^{\mathrm{s}} = 0, \\ \partial_t u^{\mathrm{s}} + \frac{1}{\rho}\partial_x \Pi^{\mathrm{s}} = 0, \end{cases} \tag{5.6}$$

can be equivalently rewritten as

$$\begin{cases} \partial_x q^{\mathrm{s}} = -\dot{\delta}_{\mathrm{O}}, \\ \partial_t \left(\frac{q^{\mathrm{s}}}{H^{\mathrm{w}}} \right) = -\frac{1}{\rho}\partial_x \Pi^{\mathrm{s}}. \end{cases} \tag{5.7}$$

that allows to express the interior discharge q^{s} in terms of the vertical displacement δ_{O} of the object, as

$$q^{\mathrm{s}}(x,t) = -x\dot{\delta}_{\mathrm{O}}(t) + \langle q^{\mathrm{s}} \rangle (t), \tag{5.8}$$

where $\langle q^{\mathrm{s}} \rangle$ is only time-dependent.

Remark 5.2.1 (*Equilibrium configuration*). *If the object is at equilibrium, then $\partial_t H^{\mathrm{w}} = 0$ and*

$$\partial_t q^{\mathrm{s}} = -\frac{H^{\mathrm{w}}}{\rho}\partial_x \Pi^{\mathrm{s}}.$$

These equations gives us an elliptic problem on Π^{s}, as we have $-\frac{H^{\mathrm{w}}}{\rho}\partial_x \Pi^{\mathrm{s}} = \dot{\delta}_{\mathrm{O}} = 0$ such that

$$-\partial_x \left(\frac{H^{\mathrm{w}}}{\rho} \partial_x \Pi^{\mathrm{s}} \right) = 0,$$

but also an evolution equation on $\langle q^{\mathrm{s}}\rangle$. Indeed, remarking that if we integrate over the interior region, we have

$$\int_{\chi_1}^{\chi_{\mathrm{r}}} \partial_t \left(\frac{q^{\mathrm{s}}}{H^{\mathrm{w}}}\right) \,\mathrm{d}x = -\frac{[\![\Pi^{\mathrm{s}}]\!]}{\rho},$$

and we get

$$\alpha(\chi_{\mathrm{l/r}})\frac{\mathrm{d}\langle q^{\mathrm{s}}\rangle}{\mathrm{d}t} = -\frac{[\![\Pi^{\mathrm{s}}]\!]}{\rho}, \qquad \alpha(\chi_{\mathrm{l/r}}) := \int_{\chi_1}^{\chi_{\mathrm{r}}} \frac{\mathrm{d}x}{H^{\mathrm{w}}}.$$

Remark 5.2.2 (*Flat-bottomed object*). *When the object moves, we have $H^{\mathrm{w}} = H^{\mathrm{eq}} + \delta_{\mathrm{o}}$, where H^{eq} is the height of the object at equilibrium, and therefore if we consider a flat-bottomed object, we have $\partial_x H^{\mathrm{eq}} = 0$.*

Having the following relations

$$H^{\mathrm{w}}\partial_t\left(\frac{q^{\mathrm{s}}}{H^{\mathrm{w}}}\right) = -\frac{H^{\mathrm{w}}}{\rho}\partial_x\Pi^{\mathrm{s}}, \qquad \frac{q^{\mathrm{s}}}{H^{\mathrm{w}}}\partial_t H^{\mathrm{w}} = -\frac{q^{\mathrm{s}}}{H^{\mathrm{w}}}\partial_x q^{\mathrm{s}},$$

gives us

$$\partial_t q^{\mathrm{s}} + \frac{1}{2H^{\mathrm{w}}}\partial_x (q^{\mathrm{s}})^2 = -\frac{H^{\mathrm{w}}}{\rho}\partial_x\Pi^{\mathrm{s}},$$

and

$$\begin{aligned}
\partial_x\left(\frac{H^{\mathrm{w}}}{\rho}\partial_x\Pi^{\mathrm{s}}\right) &= \partial_x\partial_t q^{\mathrm{s}} + \partial_x\left(\frac{1}{2H^{\mathrm{w}}}\partial_x(q^{\mathrm{s}})^2\right) = -\ddot{\delta_{\mathrm{o}}} + \frac{1}{2}\partial_x\left(\frac{\partial_x(q^{\mathrm{s}})^2}{H^{\mathrm{w}}}\right)\\
&= -\ddot{\delta_{\mathrm{o}}} + \frac{\partial_x^2(q^{\mathrm{s}})^2}{2H^{\mathrm{w}}} + \frac{1}{2}\partial_x(q^{\mathrm{s}})^2\partial_x\left(\frac{1}{H^{\mathrm{w}}}\right) = -\ddot{\delta_{\mathrm{o}}} + \frac{\dot{\delta_{\mathrm{o}}}^2}{H^{\mathrm{w}}} - \frac{1}{2}\partial_x(q^{\mathrm{s}})^2\left(\frac{\partial_x H^{\mathrm{w}}}{(H^{\mathrm{w}})^2}\right)\\
&= -\ddot{\delta_{\mathrm{o}}} + \frac{\dot{\delta_{\mathrm{o}}}^2}{H^{\mathrm{w}}} - \frac{1}{2}\left(2x\dot{\delta_{\mathrm{o}}}\frac{\partial_x H^{\mathrm{eq}}}{(H^{\mathrm{eq}}+\delta_{\mathrm{o}})^2}\right).
\end{aligned}$$

For the evolution equation on $\langle q^{\mathrm{s}}\rangle$, we first observe that $[\![(q^{\mathrm{s}})^2]\!]_{\Omega_{\mathrm{s}}} = -4l\dot{\delta_{\mathrm{o}}}\langle q^{\mathrm{s}}\rangle$. If we take $\alpha(l) = \frac{1}{2l}\int_{-l}^{l}(H^{\mathrm{w}})^{-1}\,\mathrm{d}x$, one can see that

$$\frac{1}{2l}\int_{-l}^{l}\frac{q^{\mathrm{s}}}{H^{\mathrm{w}}}\,\mathrm{d}x = \frac{1}{2l}\int_{-l}^{l}\frac{-x\dot{\delta_{\mathrm{o}}}+\langle q^{\mathrm{s}}\rangle}{H^{\mathrm{w}}}\,\mathrm{d}x = -\frac{\dot{\delta_{\mathrm{o}}}}{2l}\int_{-l}^{l}\frac{x}{H^{\mathrm{w}}}\,\mathrm{d}x + \frac{\langle q^{\mathrm{s}}\rangle}{2l}\int_{-l}^{l}\frac{\mathrm{d}x}{H^{\mathrm{w}}} = \frac{\langle q^{\mathrm{s}}\rangle}{2l}\int_{-l}^{l}\frac{\mathrm{d}x}{H^{\mathrm{w}}} = \alpha(\delta_{\mathrm{o}})\langle q^{\mathrm{s}}\rangle.$$

Finally, since

$$\int_{-l}^{l}\partial_t\left(\frac{q^{\mathrm{s}}}{H^{\mathrm{w}}}\right)\,\mathrm{d}x = -\frac{[\![\Pi^{\mathrm{s}}]\!]}{\rho} \qquad \text{and} \qquad \partial_t\left(\alpha(\delta_{\mathrm{o}})\langle q^{\mathrm{s}}\rangle\right) = -\frac{[\![\Pi^{\mathrm{s}}]\!]}{\rho},$$

we get the ordinary differential equation

$$2l\left(\alpha(\delta_{\mathrm{o}})\frac{\mathrm{d}\langle q^{\mathrm{s}}\rangle}{\mathrm{d}t} + \alpha'(\delta_{\mathrm{o}})\dot{\delta_{\mathrm{o}}}\langle q^{\mathrm{s}}\rangle\right) = -\frac{[\![\Pi^{\mathrm{s}}]\!]}{\rho}. \tag{5.9}$$

Constraint in the fluid region Ω_{f}

The pressure constraint is assumed to coincide with the constant atmospheric pressure $\mathrm{p}_{\mathrm{atm}}$,

$$\underline{\mathrm{p}}(x,t) = \mathrm{p}_{\mathrm{atm}} \quad \text{for } t \geq 0, \qquad \text{in } \Omega_{\mathrm{f}} \times (0,T). \tag{5.10}$$

Therefore, the right-hand side in the second equation of (5.1) vanishes and the equations reduces to the standard prebalanced NSW equations,

$$\begin{cases} \partial_t\eta^{\mathrm{f}} + \partial_x q^{\mathrm{f}} = 0 \\ \partial_t q^{\mathrm{f}} + \partial_x\left(u^{\mathrm{f}}q^{\mathrm{f}} + \frac{1}{2}\mathrm{g}\eta^{\mathrm{f}}(\eta^{\mathrm{f}} - 2b)\right) = -\mathrm{g}\eta^{\mathrm{f}}\partial_x b \end{cases} \qquad \text{in } \Omega_{\mathrm{f}} \times (0,T).$$

Matching conditions on the contact points $\chi_{\mathrm{l/r}}$

Some continuity conditions at the contact points $\chi_{\mathrm{l/r}} = \pm l$ must also be enforced:

- the horizontal discharge is continuous across the contact points, i.e. $q(\pm l, t) = \mp l\dot{\delta_o} + \langle q^{\mathrm{s}}\rangle(t)$,
- the Bernoulli pressure is continuous across the contact points, i.e. $\Pi^{\mathrm{f}}(\pm l) = \Pi^{\mathrm{s}}(\pm l)$.

Remark 5.2.3 (*Transmission problem and conservative form*). *As shown in [19] for the Boussinesq equations, the problem can be reformulated as a transmission problem, where the equations in the interior and exterior regions are coupled through the matching conditions at the contact points. The discharge continuity condition can be expressed as*

$$\begin{cases} \langle q\rangle = \langle q^{\mathrm{s}}\rangle, \\ [\![q]\!] = -2l\dot{\delta_o}, \end{cases}$$

where $\langle q^{\mathrm{s}}\rangle$ is the solution of the following differential equation:

$$\alpha(\delta_o)\frac{\mathrm{d}\langle q^{\mathrm{s}}\rangle}{\mathrm{d}t} + \alpha'(\delta_o)\dot{\delta_o}\langle q^{\mathrm{s}}\rangle = -\frac{[\![\Pi^{\mathrm{s}}]\!]}{2l\rho},$$

with $\alpha(\delta_o) = \frac{1}{2l}\int_{-l}^{l}\frac{\mathrm{d}x}{H^{\mathrm{w}}}$ and $\alpha'(\delta_o) = -\frac{1}{2l}\int_{-l}^{l}\frac{\mathrm{d}x}{(H^{\mathrm{w}})^2}$. By noticing that

$$\alpha(\delta)\frac{\mathrm{d}\langle q^{\mathrm{s}}\rangle}{\mathrm{d}t} + \alpha'(\delta)\dot{\delta}\langle q^{\mathrm{s}}\rangle = \frac{\mathrm{d}}{\mathrm{d}t}\left[\alpha(\delta)\langle q^{\mathrm{s}}\rangle\right],$$

the equation can also be rewritten in the conservative form

$$\frac{\mathrm{d}\mathfrak{q}^{\mathrm{s}}}{\mathrm{d}t} = -\frac{[\![\Pi^{\mathrm{s}}]\!]}{2l\rho},$$

using the quantity $\mathfrak{q}^{\mathrm{s}} := \alpha(\delta)\langle q^{\mathrm{s}}\rangle$ as the conservative variable. This form is advantageous for numerical purposes, as it removes variable coefficients from the derivative term and allows $\langle q^{\mathrm{s}}\rangle$ to be recovered by a simple division by $\alpha(\delta)$.

Vertical motion of the floating structure

We now describe the evolution of the vertical displacement δ_o of the floating structure: the motion is assumed to be purely vertical, so that the dynamics reduces to a single scalar variable. The vertical dynamics here is governed by Newton's second law, which balances inertia, gravity, and hydrodynamic pressure forces exerted by the fluid on the wetted surface of the object. Denoting by $\Pi^{\mathrm{s}}(t,x)$ the Bernoulli pressure in the interior region $\Omega_{\mathrm{s}} = (-l, l)$, we write

$$m_o\ddot{\delta_o}(t) + m_o\mathrm{g} = \int_{-l}^{l}\underline{\mathrm{p}}_i(t,x)\,\mathrm{d}x, \tag{5.11}$$

where $\underline{\mathrm{p}}_i(t,x)$ denotes the pressure exerted by the fluid at the interface with the structure. At equilibrium, the pressure is hydrostatic and satisfies

$$\underline{\mathrm{p}}_i^{\mathrm{eq}}(x) = -\rho\mathrm{g}\big(H^{\mathrm{eq}}(x) - H_0\big),$$

so that the mass of the object is given by Archimedes' principle:

$$m_o = \rho\int_{-l}^{l}\big(H_0 - H^{\mathrm{eq}}(x)\big)\,\mathrm{d}x.$$

Subtracting the equilibrium state from (5.11), we obtain the dynamical equation

$$m_o\ddot{\delta_o}(t) = \int_{-l}^{l}\Big(\underline{\mathrm{p}}_i(t,x) + \rho\mathrm{g}\big(H^{\mathrm{eq}}(x) - H_0\big)\Big)\,\mathrm{d}x. \tag{5.12}$$

Reminding that the Bernoulli pressure writes as $\Pi^{\mathrm{s}}(t,x) := \underline{\mathrm{p}}_i(t,x) + \rho\mathrm{g}\eta^{\mathrm{w}}(t,x)$, and using that $\eta^{\mathrm{w}}(t,x) = \delta_o(t) + \eta^{\mathrm{eq}}(x)$, we can rewrite (5.12) as

$$m_o\ddot{\delta_o}(t) = \int_{-l}^{l}\Pi^{\mathrm{s}}(t,x)\,\mathrm{d}x - \rho\mathrm{g}\int_{-l}^{l}\eta^{\mathrm{w}}(t,x)\,\mathrm{d}x + \rho\mathrm{g}\int_{-l}^{l}\big(H^{\mathrm{eq}}(x) - H_0\big)\,\mathrm{d}x. \tag{5.13}$$

Using again the equilibrium relation, the last two terms combine and yield a restoring force proportional to $\delta_O(t)$. After simplification, one obtains the following second-order ordinary differential equation:

$$\tau_{\text{buoy}}^2 \ddot{\delta_O}(t) + \delta_O(t) = \frac{1}{2\rho g l} \int_{-l}^{l} \Pi^{\text{s}}(t,x)\, \mathrm{d}x, \tag{5.14}$$

where the buoyancy time scale τ_{buoy} is defined by

$$\tau_{\text{buoy}}^2 := \frac{m_O}{2\rho g l}. \tag{5.15}$$

Remark 5.2.4 (*Buoyancy time scale*). *Buoyancy refers here to the upward force exerted by the fluid on the immersed structure, resulting from the hydrostatic pressure distribution. According to Archimedes' principle, this force is equal to the weight of the displaced fluid and depends on the submerged volume of the object. In the present setting, a vertical displacement of the structure modifies this submerged volume, which induces a restoring force tending to bring the system back to equilibrium.*

Equation (5.14) shows that the vertical motion of the structure behaves as a forced harmonic oscillator, where the forcing term is given by the spatial average of the Bernoulli pressure in the interior domain. This equation provides the coupling between the elliptic problem (through Π^{s}) and the rigid-body motion.

Remark 5.2.5 (*First-order formulation*). *For numerical purposes, it is also convenient to rewrite* (5.14) *as a first-order system:*

$$\begin{cases} \dot{\delta_O}(t) = w_O(t), \\ \dot{w_O}(t) = \tau_{\text{buoy}}^{-2} \left(\dfrac{1}{2\rho g l} \displaystyle\int_{-l}^{l} \Pi^{\text{s}}(x,t)\, \mathrm{d}x - \delta_O(t) \right), \end{cases}$$

where $w_O(t)$ denotes the vertical velocity of the structure.

Full one-dimensional system

We formulate below the coupled system of equations governing the fluid-structure interaction problem:

$$\begin{cases} \partial_t \eta^{\text{f}} + \partial_x q^{\text{f}} = 0 & \text{in } (0,T)\times\Omega_{\text{f}}, \\ \partial_t q^{\text{f}} + \partial_x \left(u^{\text{f}} q^{\text{f}} + \frac{g\eta^{\text{f}}}{2}(\eta^{\text{f}} - 2b) \right) = -g\eta^{\text{f}} \partial_x b & \text{in } (0,T)\times\Omega_{\text{f}}, \\ -\partial_x \left(\frac{H^{\text{w}}}{\rho} \partial_x \Pi^{\text{s}} \right) = \ddot{\delta_O} + \frac{\dot{\delta_O}^2}{H^{\text{w}}} - x\dot{\delta_O} \frac{\partial_x H^{\text{eq}}}{(H^{\text{eq}}+\delta_O)^2} & \text{in } (0,T)\times\Omega_{\text{s}}, \\ \Pi^{\text{s}} = \Pi^{\text{f}} & \text{on } (0,T)\times\partial\Omega_{\text{fs}}, \\ \alpha(\delta_O) \frac{\mathrm{d}\langle q^{\text{s}}\rangle}{\mathrm{d}t} + \alpha'(\delta_O)\dot{\delta_O}\langle q^{\text{s}}\rangle = -\frac{[\![\Pi^{\text{s}}]\!]}{2l\rho} & \text{in } (0,T), \\ \tau_{\text{buoy}}^2 \ddot{\delta_O}(t) + \delta_O(t) = \frac{1}{2\rho g l} \int_{-l}^{l} \Pi^{\text{s}}(x,t)\, \mathrm{d}x & \text{in } (0,T), \\ q^{\text{f}} = q^{\text{s}} = -x\dot{\delta_O} + \langle q^{\text{s}}\rangle & \text{on } (0,T)\times\partial\Omega_{\text{fs}}. \end{cases} \tag{5.16}$$

A pseudo-one-dimensional extension in a two-dimensional setting

Going further than [19], it is possible to consider a pseudo-one-dimensional extension (extruded) in a two-dimensional geometrical setting, to assess the robustness of the numerical strategy on unstructured two-dimensional meshes, while keeping a configuration that remains invariant in the transverse direction.

More precisely, the interior region Ω_{s} is an infinite strip, while the exterior region Ω_{f} consists of two half-planes and $\partial\Omega_{\text{fs}}$ is the union of two straight lines, i.e. $\Omega_{\text{s}} = (-l,l)\times\mathbb{R}$ and $\Omega_{\text{f}} = \Omega_{\text{f}^-} \cup \Omega_{\text{f}^+}$ with $\Omega_{\text{f}^\pm} = \{(x,y)\in\mathbb{R}^2 \mid \pm x > l\}$, and $\partial\Omega_{\text{fs}} = \partial\Omega_{\text{fs}^-} \cup \partial\Omega_{\text{fs}^+}$ with $\partial\Omega_{\text{fs}^\pm} = \{\pm l\}\times\mathbb{R}$.

Denoting by $\boldsymbol{q}^{\text{f}} = (q_x^{\text{f}}, q_y^{\text{f}})$ the horizontal discharge vector, the coupled system can be written in the following

two-dimensional form, for $\bullet \in \{\mathrm{f}^-, \mathrm{f}^+\}$:

$$\begin{cases} \partial_t \eta^{\mathrm{f}} + \nabla_{\boldsymbol{x}} \cdot \boldsymbol{q}^{\mathrm{f}} = 0 & \text{in } (0,T)\times\Omega_{\mathrm{f}}, \\ \partial_t \boldsymbol{q}^{\mathrm{f}} + \nabla_{\boldsymbol{x}} \cdot (\boldsymbol{u}^{\mathrm{f}} \otimes \boldsymbol{q}^{\mathrm{f}} + \frac{\mathrm{g}\eta^{\mathrm{f}}}{2}(\eta^{\mathrm{f}} - 2b)\mathbb{I}_2) = -\mathrm{g}\eta^{\mathrm{f}}\nabla_{\boldsymbol{x}} b & \text{in } (0,T)\times\Omega_{\mathrm{f}}, \\ -\nabla_{\boldsymbol{x}} \cdot (\frac{H^{\mathrm{w}}}{\rho}\nabla_{\boldsymbol{x}}\Pi^{\mathrm{s}}) = \ddot{\delta_{\mathrm{o}}} + \frac{\dot{\delta_{\mathrm{o}}}^2}{H^{\mathrm{w}}} - x\dot{\delta_{\mathrm{o}}}\frac{\partial_x H^{\mathrm{eq}}}{(H^{\mathrm{eq}}+\delta_{\mathrm{o}})^2} & \text{in } (0,T)\times\Omega_{\mathrm{s}}, \\ \Pi^{\mathrm{s}} = \Pi^{\mathrm{f}} & \text{on } (0,T)\times\partial\Omega_{\mathrm{fs}}, \\ \alpha(\delta_{\mathrm{o}})\frac{\mathrm{d}\langle q_x^{\mathrm{s}}\rangle}{\mathrm{d}t} + \alpha'(\delta_{\mathrm{o}})\dot{\delta_{\mathrm{o}}}\langle q_x^{\mathrm{s}}\rangle = -\frac{[\![\Pi^{\mathrm{s}}]\!]}{2l\rho} & \text{in } (0,T), \\ \tau_{\mathrm{buoy}}^2\ddot{\delta_{\mathrm{o}}}(t) + \delta_{\mathrm{o}}(t) = \frac{1}{2\rho\mathrm{g}l}\int_{-l}^{l}\Pi^{\mathrm{s}}(\boldsymbol{x},t)\,\mathrm{d}x & \text{in } (0,T), \\ \eta^{\mathrm{f}} = \eta^{\mathrm{w}}, \qquad \boldsymbol{q}^{\mathrm{f}}\cdot\mathbf{n}_{\mathrm{fs}\bullet} = \left(-x\dot{\delta_{\mathrm{o}}} + \langle q_x^{\mathrm{s}}\rangle\right)\boldsymbol{e}_x\cdot\mathbf{n}_{\mathrm{fs}\bullet} & \text{on } (0,T)\times\partial\Omega_{\mathrm{fs}\bullet}. \end{cases} \tag{5.17}$$

5.3 Discrete formulation

In what follows, we introduce a generic arbitrary polynomial order method in both the one-dimensional and two-dimensional settings. The numerical strategy is based on a domain decomposition approach, where the coupled problems are solved in each subdomain using dedicated numerical methods, and the transmission conditions at the contact points are enforced through a suitable coupling procedure.

The computational domain is composed of three distinct regions: the left exterior domain Ω_{f^-}, the interior domain Ω_{s} occupied by the floating structure, and the right exterior domain Ω_{f^+}. Due to the heterogeneous nature of the coupled problem, different numerical strategies are employed in each subdomain in the 1D case:

- In the exterior domains $\Omega_{\mathrm{f}^\pm}$, where the governing equations are hyperbolic, we rely on a monolithic DG/FV subcell approach introduced in Chapter 1,
- In the interior domain Ω_{s}, where the motion of the structure induces an elliptic-type constraint on the pressure field, we resort to a *Symmetric Weighted Interior Penalty Discontinuous Galerkin* (SWIP-DG) method.

We highlight that the mesh has to be compatible with the domain discretization, and each sub-domain requires a discrete setting. We choose to keep the same approximation spaces for the sake of simplicity, but it is obviously possible and straightforward in our setting to use different polynomial spaces in each subdomain.

For the 2D pseudo-one-dimensional extension (5.17), even though the solution remains invariant in the transverse direction, the numerical approximation is carried out on genuinely two-dimensional unstructured meshes. The exterior regions are discretized using the two-dimensional version of the local monolithic DG/FV subcell method introduced in Chapter 3, while the elliptic problem under the structure is approximated by a HHO method: all the details of this 2D approach are temporarily postponed to Chapter 6, where the discrete well-posedness of the associated boundary value problem is studied. The coupling with the vertical motion of the structure is handled through the same Newton law as in the one-dimensional case, using the discrete approximation of the interior pressure.

Discrete settings for the hyperbolic part: DG framework

Let the exterior domain $\Omega_{\mathrm{f}} = \Omega_{\mathrm{f}^-} \cup \Omega_{\mathrm{f}^+}$ be partitioned into two distinct meshes $\mathcal{T}_h^\bullet = \{\omega_j^\bullet\}_{j=1}^{N_{\mathrm{el}}^\bullet}$, where $\bullet \in \{\mathrm{f}^-, \mathrm{f}^+\}$. These partitions are characterized by the mesh size $h^\bullet := \max_{\omega_j^\bullet \in \mathcal{T}_h^\bullet} h_j^\bullet$ where $h_j^\bullet := |\omega_j^\bullet|$. For any $\omega_j^\bullet \in \mathcal{T}_h^\bullet$, we define $\omega_j^\bullet := [x_{j-\frac{1}{2}}^\bullet, x_{j+\frac{1}{2}}^\bullet]$, and write $x_j^\bullet$ for its barycenter.

Given an integer $k \geq 0$, we define the local broken-polynomial spaces

$$\mathbb{P}^k(\mathcal{T}_h^\bullet) := \left\{ v \in L^2(\Omega_\bullet) \;\middle|\; \forall \omega_i^\bullet \in \mathcal{T}_h^\bullet,\ v_{|\omega_i^\bullet} \in \mathbb{P}^k(\omega_i^\bullet) \right\},$$

where $\mathbb{P}^k(\omega_i^\bullet)$ denotes the space of univariate polynomials of degree at most k on $\omega_i^\bullet$. For any $v_h^\bullet \in \mathbb{P}^k(\mathcal{T}_h^\bullet)$, and for all $\omega_i^\bullet \in \mathcal{T}_h^\bullet$, we use the notation $v_h^{j,\bullet} := v_{h|\omega_j^\bullet}^\bullet$. The corresponding local polynomial basis is denoted by

$$\boldsymbol{\Psi}_{\omega_j^\bullet} := \{\psi_m^{j,\bullet}\}_{m\in[\![1,k+1]\!]}.$$

Each $v_h^{j,\bullet}$ admits the expansion

$$v_h^{j,\bullet} = \sum_{m=1}^{k+1} \underline{v}_m^{j,\bullet} \psi_m^{j,\bullet},$$

where the collection $\underline{v}^{j,\bullet} := \left\{\underline{v}_m^{j,\bullet}\right\}_{m=1}^{k+1}$ defines the local degrees of freedom. The global degrees of freedom are denoted by

$$\underline{v}^{\bullet} := \left\{\underline{v}^{j,\bullet}\right\}_{j=1}^{N_{\mathrm{el}}^{\bullet}}.$$

The L^2-orthogonal projector onto $\mathbb{P}^k(\mathcal{T}_h^{\bullet})$ is denoted $\pi_h^{k,\bullet}$, and acts locally on each element $\omega_j^{\bullet}$ via the local counterpart $\pi_j^{k,\bullet}$. The broken gradient operator

$$\nabla_h^k : \mathbb{P}^k(\mathcal{T}_h^{\bullet}) \to \mathbb{P}^{k-1}(\mathcal{T}_h^{\bullet})$$

is defined piecewise on each cell, such that for all $v_h^{\bullet} \in \mathbb{P}^k(\mathcal{T}_h^{\bullet})$,

$$(\nabla_h^k v_h^{\bullet})|\omega_j^{\bullet} := \partial_x(v_h^{j,\bullet}) \qquad \forall \omega_j^{\bullet} \in \mathcal{T}_h^{\bullet}.$$

Approximate solutions in the exterior domains are sought as pairs

$$\boldsymbol{v}_h^{\bullet} := (\eta_h^{\bullet}, q_h^{\bullet}) \in (\mathbb{P}^k(\mathcal{T}_h^{\bullet}))^2,$$

and we may use the simpler notation $\mathbf{P}^k(\mathcal{T}_h^{\bullet}) := (\mathbb{P}^k(\mathcal{T}_h^{\bullet}))^2$.

As before, for interface values at cell boundaries, we denote left and right traces as

$$\boldsymbol{v}_{j+\frac{1}{2}}^{-,\bullet} := \lim_{\omega_j^{\bullet} \ni x \to x_{j+\frac{1}{2}}^{\bullet}} \boldsymbol{v}_h^{j,\bullet}(x), \quad \boldsymbol{v}_{j+\frac{1}{2}}^{+,\bullet} := \lim_{\omega_j^{\bullet} \ni x \to x_{j+\frac{1}{2}}^{\bullet}} \boldsymbol{v}_h^{j+1,\bullet}(x).$$

Let us also define the following shorthands for the *exterior scalar-products* of functions $v, w \in L^2(\mathcal{T}_h^{\bullet})$ and $\mu, \nu \in L^2(\partial\mathcal{T}_h^{\bullet})$:

$$(v,w)_{\mathcal{T}_h^{\bullet}} := \sum_{\omega \in \mathcal{T}_h^{\bullet}} \int_{\omega_i^{\mathrm{s}}} v(x,t) w(x,t)\, \mathrm{d}x, \qquad \langle \mu, \nu \rangle_{\mathcal{T}_h^{\bullet}} := \sum_{\omega \in \mathcal{T}_h^{\bullet}} [\![\mu\nu]\!]_{\partial\omega},$$

the extension to vector or tensor-valued functions being straightforward.

Discrete settings for the hyperbolic part: DG/FV subcells framework

Let each cell $\omega_j^{\bullet} \in \mathcal{T}_h^{\bullet}$ be subdivided into $k+1$ disjoint open subcells, forming the local sub-mesh

$$\mathcal{T}_{\omega_j}^{k,\bullet} := \left\{S_m^{j,\bullet}\right\}_{m \in [\![1,k+1]\!]},$$

such that

$$\overline{\omega}_j^{\bullet} = \bigcup_{m=1}^{k+1} \overline{S_m^{j,\bullet}}.$$

Each subcell $S_m^{j,\bullet}$ is defined as the interval

$$S_m^{j,\bullet} := \left[\widetilde{x}_{m-\frac{1}{2}}^{j,\bullet}, \widetilde{x}_{m+\frac{1}{2}}^{j,\bullet}\right],$$

with $\widetilde{x}_{\frac{1}{2}}^{j,\bullet} = x_{j-\frac{1}{2}}^{\bullet}$ and $\widetilde{x}_{k+\frac{3}{2}}^{j,\bullet} = x_{j+\frac{1}{2}}^{\bullet}$. When considering a sequence of neighboring elements $\omega_{j-1}^{\bullet}$, $\omega_j^{\bullet}$, and $\omega_{j+1}^{\bullet}$, we adopt the convenient convention $S_0^{j,\bullet} := S_{k+1}^{j-1,\bullet}$ and $S_{k+2}^{j,\bullet} := S_1^{j+1,\bullet}$.

We define the space of piecewise-constant functions over each sub-mesh as

$$\mathbb{P}^0(\mathcal{T}_{\omega_j}^{k,\bullet}) := \left\{ v \in L^{\infty}(\omega_j^{\bullet}) \mid \forall S_m^{j,\bullet} \in \mathcal{T}_{\omega_j}^{k,\bullet},\ v_{|S_m^{j,\bullet}} \in \mathbb{P}^0(S_m^{j,\bullet}) \right\}.$$

Let $\{\mathbb{1}_m^{j,\bullet} \mid m \in [\![1,k+1]\!]\}$ denote the indicator functions of the subcells in $\mathcal{T}_{\omega_j}^{k,\bullet}$, with $\mathbb{1}_m^{j,\bullet}(x) = 1$ if $x \in S_m^{j,\bullet}$ and 0 otherwise. The mean-value projector onto $\mathbb{P}^0(\mathcal{T}_{\omega_j}^{k,\bullet})$ is defined by

$$\mathbb{P}^k(\omega_j^{\bullet}) \ni v_h^{j,\bullet} \mapsto \overline{\pi}_j^{k,\bullet} \circ v_h^{j,\bullet} := \sum_{m=1}^{k+1} \overline{v}_m^{j,\bullet} \mathbb{1}_m^{j,\bullet} =: \overline{v}_h^{j,\bullet} \in \mathbb{P}^0(\mathcal{T}_{\omega_j}^{k,\bullet}),$$

where

$$\overline{v}^{j,\bullet} := \left\{\overline{v}_m^{j,\bullet},\ m \in [\![1, k+1]\!]\right\}$$

denotes the collection of subcell mean-values of $v_h^{j,\bullet}$ over $\mathcal{T}_{\omega_j}^{k,\bullet}$.

The global sub-mesh over $\Omega_{\mathrm{f}\bullet}$ is then given by

$$\widetilde{\mathcal{T}}_h^{k,\bullet} := \bigcup_{\omega_j^\bullet \in \mathcal{T}_h^\bullet} \mathcal{T}_{\omega_j}^{k,\bullet}, \qquad \mathbb{P}^0(\widetilde{\mathcal{T}}_h^{k,\bullet}) := \bigtimes_{\omega_j^\bullet \in \mathcal{T}_h^\bullet} \mathbb{P}^0(\mathcal{T}_{\omega_j}^{k,\bullet}).$$

The global projection operator onto $\mathbb{P}^0(\mathcal{T}_h^{k,\bullet})$ is denoted $\overline{\pi}_h^{k,\bullet}$, and the projected function is written $\overline{v}_h^{\pm} \in \mathbb{P}^0(\mathcal{T}_h^{k,\bullet})$ with components $\overline{v}_m^{j,\bullet} := (\overline{v}_h^{\pm})_{|S_m^{j,\bullet}}$. Piecewise-constant functions belonging to $\mathbb{P}^0(\widetilde{\mathcal{T}}_h^{k,\bullet})$ are denoted with a $^-$and a subscript h, and for any $\overline{v}_h \in \mathbb{P}^0(\widetilde{\mathcal{T}}_h^{k,\bullet})$, and for all $S_m^{j,\bullet} \in \widetilde{\mathcal{T}}_h^{k,\bullet}$, we have the identity $\overline{v}_m^{j,\bullet} = \overline{v}_{h|S_m^{i,\bullet}}^{j,\bullet}$.

Discrete settings for the elliptic part

Let the interior domain Ω_{s} be partitioned into a mesh $\mathcal{T}_h^{\mathrm{s}} = \{\omega_j^{\mathrm{s}}\}_{j=1}^{N_{\mathrm{el}}^{\mathrm{s}}}$, with $\omega_j^{\mathrm{s}} = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$, and mesh size $h_j := |\omega_j^{\mathrm{s}}|$. The set of interior and boundary faces is denoted by $\mathcal{F}_h^{\mathrm{s}} = \mathcal{F}_h^{\mathrm{s},0} \cup \mathcal{F}_h^{\mathrm{s},\partial}$, where

$$\mathcal{F}_h^{\mathrm{s},0} := \left\{x \in \Omega_{\mathrm{s}} \mid x = x_{j+\frac{1}{2}},\ 1 \le j < N_{\mathrm{el}}^{\mathrm{s}}\right\}$$

is the set of interior faces, and $\mathcal{F}_h^{\mathrm{s},\partial} := \{-l, l\}$ is the set of boundary points of the interior domain. For every face $\Gamma \in \mathcal{F}_h^{\mathrm{s}}$, we denote by h_Γ the local mesh size associated with Γ, defined as the minimum of the diameters of the mesh elements sharing Γ. We also introduce a positive penalty parameter $\xi > 0$, which will be used to ensure the coercivity of the discrete bilinear form.

We denote by $\mathbb{P}^k(\mathcal{T}_h^{\mathrm{s}})$ the broken polynomial space

$$\mathbb{P}^k(\mathcal{T}_h^{\mathrm{s}}) := \left\{v \in L^2(\Omega_{\mathrm{s}}) \mid \forall \omega \in \mathcal{T}_h^{\mathrm{s}},\, v_{|\omega} \in \mathbb{P}^k(\omega)\right\}.$$

We set $\kappa := H^{\mathrm{w}}/\rho > 0$. For an interior face $\Gamma \in \mathcal{F}_h^{\mathrm{s},0}$ shared by two neighboring cells, we denote by κ_Γ^- and κ_Γ^+ the traces of κ from the left and right cells. The jump and weighted average operators are defined by

$$[\![v]\!]_\Gamma := v_\Gamma^+ - v_\Gamma^-, \qquad \{\!\{v\}\!\}_{\lambda,\Gamma} := \lambda_\Gamma^- v_\Gamma^- + \lambda_\Gamma^+ v_\Gamma^+, \tag{5.18}$$

with the SWIP-DG weights

$$\lambda_\Gamma^- := \frac{\kappa_\Gamma^+}{\kappa_\Gamma^- + \kappa_\Gamma^+}, \qquad \lambda_\Gamma^+ := \frac{\kappa_\Gamma^-}{\kappa_\Gamma^- + \kappa_\Gamma^+}.$$

In what follows, and whenever no ambiguity is possible, we may omit the subscript Γ in both $[\![\cdot]\!]_\Gamma$ and $\{\!\{\cdot\}\!\}_{\lambda,\Gamma}$. At boundary vertices, the definition of jumps and averages depends on the boundary conditions that are prescribed. For the weak imposition of Dirichlet boundary conditions on a boundary face $\Gamma \in \mathcal{F}_h^{\mathrm{s},\partial}$, we use the convention

$$[\![v]\!]_\Gamma := v_{|\Gamma}, \qquad \{\!\{\kappa\partial_x v\}\!\}_{\lambda,\Gamma} := \mathrm{n}_\Gamma \left(\kappa\partial_x v\right)_{|\Gamma},$$

where n_Γ is the outward unit normal at the boundary points, namely $\mathrm{n}_{-l} = -1$ and $\mathrm{n}_l = 1$.

The κ-dependent penalty parameter $\gamma_{\kappa,\Gamma}$ is defined by

$$\gamma_{\kappa,\Gamma} := \begin{cases} \dfrac{2\kappa_\Gamma^-\kappa_\Gamma^+}{\kappa_\Gamma^- + \kappa_\Gamma^+}, & \text{if } \Gamma \in \mathcal{F}_h^{\mathrm{s},0}, \\ \kappa_{|\Gamma}, & \text{if } \Gamma \in \mathcal{F}_h^{\mathrm{s},\partial}. \end{cases} \tag{5.19}$$

5.3.1 SWIP-DG method for the elliptic problem on Bernoulli pressure

We aim to discretize the following second-order elliptic problem with Dirichlet boundary conditions using the SWIP-DG method:

$$\begin{cases} -\partial_x\left(\dfrac{H^{\mathrm{w}}}{\rho}\partial_x \Pi^{\mathrm{s}}\right) = \ddot{\delta_{\mathrm{o}}} + \dfrac{\dot{\delta_{\mathrm{o}}}^2}{H^{\mathrm{w}}} - \dfrac{1}{2}\left(2x\dot{\delta_{\mathrm{o}}}\dfrac{\partial_x H^{\mathrm{eq}}}{(H^{\mathrm{eq}} + \delta_{\mathrm{o}})^2}\right) & \text{in } \Omega_{\mathrm{s}}, \\ \Pi^{\mathrm{s}}(\pm l) = \Pi^{\mathrm{f}}(\pm l) & \text{on } \partial\Omega_{\mathrm{fs}}. \end{cases} \tag{5.20}$$

The SWIP-DG method enhances the classical SIP-DG scheme by incorporating local weightings in the numerical fluxes and penalty terms, which improves stability and robustness when dealing with heterogeneous diffusion coefficients. In our context, where $\kappa := H^{\mathrm{w}}/\rho$ varies in space, this weighted treatment yields a symmetric and coercive discrete operator, provided that the penalty parameter is chosen large enough.

SWIP bilinear form and weak formulation

The SWIP-DG formulation seeks $\Pi_h^{\mathrm{s}} \in \mathbb{P}^k(\mathcal{T}_h^{\mathrm{s}})$ such that, for all $w_h \in \mathbb{P}^k(\mathcal{T}_h^{\mathrm{s}})$,

$$\mathscr{A}_h^{\mathrm{s}}(\Pi_h^{\mathrm{s}}, w_h) = \ell_h(w_h),$$

where the bilinear form $\mathscr{A}_h^{\mathrm{s}}(\cdot,\cdot)$ on $\mathbb{P}^k(\mathcal{T}_h^{\mathrm{s}}) \times \mathbb{P}^k(\mathcal{T}_h^{\mathrm{s}})$ and the linear form $\ell_h(\cdot)$ on $\mathbb{P}^k(\mathcal{T}_h^{\mathrm{s}})$ read:

$$\begin{aligned}
\mathscr{A}_h^{\mathrm{s}}(\Pi_h^{\mathrm{s}}, w_h) :=& \sum_{\omega_i^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}} \int_{\omega_i^{\mathrm{s}}} \kappa\, \partial_x \Pi_h^{\mathrm{s}}\, \partial_x w_h \,\mathrm{d}x + \sum_{\Gamma \in \mathcal{F}_h^{\mathrm{s}}} \xi \int_\Gamma \gamma_{\kappa,\Gamma} h_\Gamma^{-1} [\![\Pi_h^{\mathrm{s}}]\!][\![w_h]\!]\,\mathrm{d}s \\
&- \sum_{\Gamma \in \mathcal{F}_h^{\mathrm{s}}} \int_\Gamma \{\!\{\kappa\, \partial_x \Pi_h^{\mathrm{s}}\}\!\}_\lambda\, [\![w_h]\!]\,\mathrm{d}s - \sum_{\Gamma \in \mathcal{F}_h^{\mathrm{s}}} \int_\Gamma [\![\Pi_h^{\mathrm{s}}]\!]\, \{\!\{\kappa\, \partial_x w_h\}\!\}_\lambda \,\mathrm{d}s, \qquad (5.21)\\
\ell_h(w_h) :=& \sum_{\omega_i^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}} \int_{\omega_i^{\mathrm{s}}} \left\{ \ddot{\delta}_o + \frac{\dot{\delta}_o^{\,2}}{H^{\mathrm{w}}} - \frac{1}{2}\left(2x\dot{\delta}_o \frac{\partial_x H^{\mathrm{eq}}}{(H^{\mathrm{eq}} + \delta_o)^2} \right) \right\} w_h \,\mathrm{d}x \\
&- \sum_{\Gamma \in \mathcal{F}_h^{\mathrm{s},\partial}} \int_\Gamma \{\!\{\kappa\, \partial_x w_h\}\!\}_\lambda\, \Pi_{|\Gamma}^{\mathrm{f}}\,\mathrm{d}s + \sum_{\Gamma \in \mathcal{F}_h^{\mathrm{s},\partial}} \xi \int_\Gamma \gamma_{\kappa,\Gamma} h_\Gamma^{-1} \Pi_{|\Gamma}^{\mathrm{f}} w_{h|\Gamma}\,\mathrm{d}s. \qquad (5.22)
\end{aligned}$$

Remark 5.3.1 (*SIP or SWIP*). *The SIP bilinear form was introduced by Arnold [9], and contains a penalty term of the form*

$$\frac{\sigma}{h_\Gamma} [\![\Pi_h^{\mathrm{s}}]\!]_\Gamma [\![w_h]\!]_\Gamma,$$

which ensures stability and coercivity on general meshes, especially in higher-order settings. The concept of weakly enforcing both boundary conditions and inter-element continuity through penalty terms dates back to the 1970s. Foundational contributions include the work of Nitsche [222], Babuška [14], Babuška and Zlámal [15], Douglas and Dupont [94], Baker [16], and Wheeler [275]. The SWIP bilinear form (5.21) generalizes the classical SIP-DG approach by introducing weighted averages and adapted penalty terms depending on the local coefficient $\kappa := H^{\mathrm{w}}/\rho$. This ensures better stability and coercivity properties, especially when the diffusion coefficient exhibits strong heterogeneity or large gradients. Such modifications are crucial in our application where the water-height H^{w} may vary significantly due to interaction with the moving object.

5.3.2 General SSP-RK discretization of the ODEs for the coupled system

We now describe a general strong stability-preserving Runge–Kutta (SSP-RK) method for the time discretization of the ordinary differential equations arising in the coupled fluid-structure system (see Time discretization and time-marching algorithm for details). These correspond to the evolution of the conservative interior discharge $\mathfrak{q}^{\mathrm{s}}$ and the vertical motion of the structure, described by (δ_o, w_o). The ODEs read:

$$\begin{cases} \frac{\mathrm{d}\mathfrak{q}^{\mathrm{s}}}{\mathrm{d}t}(t) = r_q(t), \\ \dot{\delta}_o(t) = w_o(t), \\ \dot{w}_o(t) = r_\delta(t), \end{cases}$$

where

$$\mathfrak{q}^{\mathrm{s}}(t) := \alpha(\delta_o(t))\langle q^{\mathrm{s}}\rangle(t), \qquad r_q(t) := -\frac{[\![\Pi^{\mathrm{s}}]\!](t)}{2l\rho}, \qquad r_\delta(t) := \frac{1}{\tau_{\mathrm{buoy}}^2}\left(\frac{1}{2\rho g l} \int_{-l}^{l} \Pi^{\mathrm{s}}(x,t)\,\mathrm{d}x - \delta_o(t) \right).$$

Introducing the state vector and the right-hand side operator defined by $\boldsymbol{y} := (\mathfrak{q}^{\mathrm{s}}, \delta_o, w_o)^\top$ and $\mathscr{R}(\boldsymbol{y}) := (r_q, w_o, r_\delta)^\top$, the ODE system can be written in compact form as

$$\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t} = \mathscr{R}(\boldsymbol{y}).$$

Let t^n be the current time level and Δt^n the time step. A s-stage SSP–RK scheme reads:

$$\begin{aligned}\boldsymbol{y}^{(0)} &= \boldsymbol{y}^n,\\ \boldsymbol{y}^{(i)} &= \sum_{j=0}^{i-1} \mathsf{a}_{ij}\boldsymbol{y}^{(j)} + \Delta t^n\, \mathsf{b}_i\, \mathscr{R}\big(\boldsymbol{y}^{(i-1)}\big), \qquad \forall i \in [\![1,s]\!],\\ \boldsymbol{y}^{n+1} &= \boldsymbol{y}^{(s)},\end{aligned}$$

the right-hand side $\mathscr{R}(\boldsymbol{y}^{(i-1)})$ being evaluated using the corresponding intermediate values of Π^{s} and $\delta_{\circ}$.

5.3.3 Local subcell monolithic DG/FV method for the exterior region

The construction of the numerical method in the hyperbolic parts of the domain (i.e. $\bullet \in \{\mathrm{f}^-, \mathrm{f}^+\}$) follows the local monolithic DG/FV subcells approach introduced in Chapter 1. We recall that a semi-discrete classical DG formulation of the nonlinear shallow-water equations in both fluid domains can read as: find $\boldsymbol{v}_h^\bullet := (\eta_h^\bullet, q_h^\bullet) \in (\mathbb{P}^k(\mathcal{T}_h^\bullet))^2$ such that, for all $\psi_h^\bullet \in \mathbf{P}^k(\mathcal{T}_h^\bullet)$:

$$\big(\partial_t \boldsymbol{v}_h^\bullet, \psi_h^\bullet\big)_{\mathcal{T}_h^\bullet} - \big(\boldsymbol{F}(\boldsymbol{v}_h^\bullet, b_h^\bullet), \partial_x \psi_h^\bullet\big)_{\mathcal{T}_h^\bullet} + \big\langle \boldsymbol{\mathcal{F}}^*, \psi_h^\bullet\big\rangle_{\partial\mathcal{T}_h^\bullet} = \big(\boldsymbol{B}[b_h^\bullet](\boldsymbol{v}_h^\bullet)\psi_h^\bullet\big)_{\mathcal{T}_h^\bullet} \tag{5.23}$$

with $\boldsymbol{\mathcal{F}}^*$ denoting a suitable DG numerical flux function, and where $b_h^\bullet$ is a discrete approximation of the bottom topography b on $\mathcal{T}_h^\bullet$. As emphasized during the thesis, DG is prone to spurious oscillations in the presence of strong gradients or discontinuities, and need additional design to be domain-preserving. To mitigate this issue, we employ a local subcell FV method, which is applied on the sub-mesh $\widetilde{\mathcal{T}}_h^{k,\bullet}$ associated with each element of the original mesh $\mathcal{T}_h^\bullet$. This approach allows for a more robust treatment of sharp features while maintaining high-order accuracy in smooth regions.

This method is applied independently on the two distinct meshes $\mathcal{T}_h^\bullet$ associated with the left and right exterior domains. Since the elliptic problem inside the structure determines the evolution of the interior pressure Π^{s} and discharge $\langle q^{\mathrm{s}}\rangle$, the corresponding conservative variable $\mathfrak{q}^{\mathrm{s}}$ provides interface data for the hyperbolic solvers. Specifically, on $\mathcal{T}_h^{\mathrm{f}^-}$ (resp. $\mathcal{T}_h^{\mathrm{f}^+}$), a FV-subcells method is used with a right (resp. left) boundary condition prescribed by the solution of the ODE coupled with the elliptic problem.

For the sake of clarity and simplicty, the entire limiting procedure described in Chapter 1 will be simply represented by applying a stabilization operator $\mathcal{S}_h^{n,\star,\bullet}$, defined as:

$$\mathcal{S}_h^{n,\star,\bullet} : \mathbf{P}^k(\mathcal{T}_h^\bullet) \ni \boldsymbol{v}_h^{n,\bullet} \mapsto \mathcal{S}_h^{n,\star,\bullet} \circ \boldsymbol{v}_h^{n,\bullet} = \Big(\overline{\pi}_h^{k,\bullet}\Big)^{-1} \circ \overline{\boldsymbol{v}}_h^{n,\star,\bullet} \in \mathbf{P}^k(\mathcal{T}_h^\bullet),$$

where $\overline{\boldsymbol{v}}_h^{n,\star,\bullet} \in \mathbf{P}^0(\widetilde{\mathcal{T}}_h^\bullet)$ denotes the limited subcell mean-value, the superscript $\star$ indicating that the convex-limiting procedure has been applied, and where $\overline{\pi}_h^{k,\bullet}$ is the mean-value projection operator onto piecewise-constant functions on subcells.

5.3.4 Semi-discrete DG/SWIP-DG formulation

The resulting semi-discrete formulation of the coupled system reads:

Find $(\boldsymbol{v}_h^\bullet, \langle q^{\mathrm{s}}\rangle, \delta_{\circ}, \Pi_h^{\mathrm{s}}) \in \mathscr{C}^1(0,T;\mathbf{P}^k(\mathcal{T}_h^\bullet)) \times \mathscr{C}^1(0,T) \times \mathscr{C}^2(0,T) \times \mathbb{P}^k(\mathcal{T}_h^{\mathrm{s}})$ for all $(\psi_h^\bullet, w_h) \in \mathbf{P}^k(\mathcal{T}_h^\bullet) \times \mathbb{P}^k(\mathcal{T}_h^{\mathrm{s}})$ such that, for $\bullet \in \{\mathrm{f}^-, \mathrm{f}^+\}$ and all $t \in (0,T)$, the following system holds:

$$\begin{cases} \mathscr{A}_h^{\mathrm{s}}(\Pi_h^{\mathrm{s}}, w_h) = \ell_h(w_h) & \forall w_h \in \mathbb{P}^k(\mathcal{T}_h^{\mathrm{s}}),\\ \Pi_{h|\Gamma}^{\mathrm{s}} = \Pi_{h|\Gamma}^{\mathrm{f}} & \forall \Gamma \in \mathcal{F}_h^{\mathrm{s},\partial} \cap \mathcal{F}_h^{\bullet,\partial},\\ \alpha(\delta_{\circ})\frac{\mathrm{d}\langle q^{\mathrm{s}}\rangle}{\mathrm{d}t} + \alpha'(\delta_{\circ})\dot{\delta_{\circ}}\langle q^{\mathrm{s}}\rangle = -\frac{[\![\Pi_h^{\mathrm{s}}]\!]}{2l\rho} & \text{in } (0,T),\\ \tau_{\mathrm{buoy}}^2\ddot{\delta_{\circ}}(t) + \delta_{\circ}(t) = \frac{1}{2\rho g l}\int_{-l}^{l}\Pi_h^{\mathrm{s}}(x,t)\,\mathrm{d}x & \text{in } (0,T),\\ \frac{\mathrm{d}}{\mathrm{d}t}\big(\boldsymbol{v}_h^\bullet, \psi_h^\bullet\big)_{\mathcal{T}_h^\bullet} + \big(\mathcal{U}_h^\bullet \circ \boldsymbol{v}_h^\bullet, \psi_h^\bullet\big)_{\mathcal{T}_h^\bullet} = 0 & \forall \psi_h^\bullet \in \mathbf{P}^k(\mathcal{T}_h^\bullet),\\ \boldsymbol{v}_{h|\Gamma}^\bullet = (\eta_{h|\Gamma}^{\mathrm{w}}, q_{|\Gamma}^{\mathrm{s}})^\top = (\eta_{\mathrm{eq}|\Gamma} + \delta_{\circ}, -x_\Gamma\dot{\delta_{\circ}} + \langle q^{\mathrm{s}}\rangle)^\top & \forall \Gamma \in \mathcal{F}_h^{\mathrm{s},\partial} \cap \mathcal{F}_h^{\bullet,\partial},\\ \boldsymbol{v}_h^{\bullet,0} = (\overline{\pi}_h^{k,\bullet})^{-1} \circ \overline{\boldsymbol{v}}_h^{\bullet,0} & \text{in } \mathcal{T}_h^\bullet \times \{0\}, \end{cases} \tag{5.24}$$

where the discrete nonlinear operator $\mathcal{U}_h^\bullet$ is defined as

$$\big(\mathcal{U}_h^\bullet \circ \boldsymbol{v}_h^\bullet, \psi_h^\bullet\big)_{\mathcal{T}_h^\bullet} := -\big(\boldsymbol{F}\left(\boldsymbol{v}_h^\bullet, b_h^\bullet\right), \psi_h^\bullet\big)_{\mathcal{T}_h^\bullet} + \big\langle \boldsymbol{\mathcal{F}}^*, \psi_h^\bullet\big\rangle_{\partial\mathcal{T}_h^\bullet} - \big(\boldsymbol{B}\left[b_h^\bullet\right](\boldsymbol{v}_h^\bullet), \psi_h^\bullet\big)_{\mathcal{T}_h^\bullet}.$$

5.3.5 Fully-discrete monolithic DG/FV–SWIP-DG–SSP-RK formulation

For clarity, we describe the first-order explicit E uler version; higher-order SSP–Runge–Kutta schemes follow by adjusting the time-stage indices accordingly.

Assume that all required quantities $\big(\boldsymbol{v}_h^{\bullet,n}, \Pi_h^{\mathrm{s},n}, \mathfrak{q}^{\mathrm{s},n}, \langle q^{\mathrm{s}}\rangle^n, \delta_O^n, w_O^n\big)$ are known at time level t^n. The fully-discrete subcell monolithic DG/FV–SWIP-DG formulation then reads as follows:

Find $\big(\boldsymbol{v}_h^{\bullet,n+1}, \Pi_h^{\mathrm{s},n+1}, \mathfrak{q}^{\mathrm{s},n+1}, \langle q^{\mathrm{s}}\rangle^{n+1}, \delta_O^{n+1}, w_O^{n+1}\big) \in \mathbf{P}^k(\mathcal{T}_h^{\bullet}) \times \mathbb{P}^k(\mathcal{T}_h^{\mathrm{s}}) \times \mathbb{R}^4$ such that, for $\bullet \in \{\mathrm{f}^-, \mathrm{f}^+\}$, the following system holds:

$$
\begin{cases}
\mathscr{A}_h^{\mathrm{s}}\big(\Pi_h^{\mathrm{s},n}, w_h\big) = \ell_h^n(w_h) & \forall w_h \in \mathbb{P}^k(\mathcal{T}_h^{\mathrm{s}}),\\
\Pi_{h|\Gamma}^{\mathrm{s},n} = \Pi_{h|\Gamma}^{\mathrm{f},n} & \forall \Gamma \in \mathcal{F}_h^{\mathrm{s},\partial} \cap \mathcal{F}_h^{\bullet,\partial},\\
\mathfrak{q}^{\mathrm{s},n+1} = \mathfrak{q}^{\mathrm{s},n} - \Delta t^n \frac{[\![\Pi_h^{\mathrm{s},n}]\!]}{2l\rho} & \text{in } (t^n, t^{n+1}),\\
\langle q^{\mathrm{s}}\rangle^{n+1} = \frac{\mathfrak{q}^{\mathrm{s},n+1}}{\alpha(\delta_O^{n+1})} & \text{in } (t^n, t^{n+1}),\\
\delta_O^{n+1} = \delta_O^n + \Delta t^n w_O^n & \text{in } (t^n, t^{n+1}),\\
w_O^{n+1} = w_O^n + \frac{\Delta t^n}{\tau_{\mathrm{buoy}}^2}\Big(\frac{1}{2\rho g l}\int_{-l}^{l} \Pi_h^{\mathrm{s},n}(x)\,\mathrm{d}x - \delta_O^n\Big) & \text{in } (t^n, t^{n+1}),\\
\boldsymbol{v}_{h|\Gamma}^{\bullet,n} = \Big(\eta_{h|\Gamma}^{\mathrm{w,n}}, q_{|\Gamma}^{\mathrm{s,n}}\Big)^{\top} = \big(\eta_{\mathrm{eq}|\Gamma} + \delta_O^n, -x_\Gamma w_O^n + \langle q^{\mathrm{s}}\rangle^n\big)^{\top} & \forall \Gamma \in \mathcal{F}_h^{\mathrm{s},\partial} \cap \mathcal{F}_h^{\bullet,\partial},\\
\boldsymbol{v}_h^{\bullet,n+1} = \mathcal{S}_h^{n,\star,\bullet} \circ \big\{\boldsymbol{v}_h^{\bullet,n} - \Delta t^n\big(\mathcal{U}_h^{\bullet} \circ \boldsymbol{v}_h^{\bullet,n}\big)\big\} & \text{in } \mathcal{T}_h^{\bullet},\\
\boldsymbol{v}_h^{\bullet,0} = (\overline{\pi}_h^{k,\bullet})^{-1} \circ \overline{\boldsymbol{v}}_h^{\bullet,0}, & \text{in } \mathcal{T}_h^{\bullet} \times \{0\}.
\end{cases}
\tag{5.25}
$$

Remark 5.3.2 (*Time level of the elliptic solve*). *In the first-order explicit version written in (5.25), the SWIP-DG elliptic problem is solved using the data available at time t^n. The resulting pressure $\Pi_h^{\mathrm{s},n}$ is then used to update the conservative interior discharge, the vertical motion of the object, and the exterior DG/FV unknowns. For higher-order SSP–RK schemes, the same sequence is applied at each Runge–Kutta stage, replacing the superscript n by the corresponding stage index.*

One can now consider the following flowchart for the time-stepping procedure:

(*i*) The SWIP-DG scheme computes the interior Bernoulli pressure Π_h^{s} from the elliptic problem posed in Ω_{s}. The boundary values are imposed by the transmission condition $\Pi_h^{\mathrm{s}} = \Pi_h^{\mathrm{f}}$ at the two contact points,

(*ii*) The conservative interior discharge variable $\mathfrak{q}^{\mathrm{s}} := \alpha(\delta_O)\langle q^{\mathrm{s}}\rangle$ is advanced through the ODE (5.9) driven by the pressure jump $[\![\Pi_h^{\mathrm{s}}]\!]$,

(*iii*) The vertical motion of the structure is then computed through the ODE (5.14) driven by the variables δ_O and w_O, whose acceleration is forced by the averaged Bernoulli pressure under the object,

(*iv*) The exterior hyperbolic problem is finally solved independently on Ω_{f^-} and Ω_{f^+} using the monolithic DG/FV subcell method. At the contact points, the boundary state is prescribed by the solutions of the ODEs for δ_O and $\langle q^{\mathrm{s}}\rangle$.

5.4 Numerical illustrations

We now present several numerical illustrations of the coupled wave-structure interaction model introduced in the previous sections. The goal is to assess the behaviour of the proposed discretization on several configurations. We focus in particular on the consistency between the different numerical components: the monolithic DG/FV subcell method in the exterior regions, the interior elliptic pressure problem, and the time integration of the structure-related ordinary differential equations. Unless stated otherwise, the polynomial degree indicated in each parameter box is used both for the exterior hyperbolic solver and for the interior elliptic solver.

Depending on the test case, we either consider the genuinely one-dimensional formulation described in (5.16), or its pseudo-one-dimensional counterpart introduced in (5.17) on genuinely two-dimensional unstructured meshes. Some configurations are run in both settings. In that case, the comparison is especially relevant since the 1D and 2D computations rely on distinct implementations of both the hyperbolic and elliptic parts, and therefore provide a useful cross-check of the coupled method.

The numerical results and the behaviours observed in the following test cases are consistent with the physical mechanisms reported in the literature for similar wave-structure interaction configurations, see for instance

[19, 20, 141, 142]. Some of the following benchmarks are also designed to reproduce standard configurations considered in these previous works, which allows us to check that the numerical response of the model is physically meaningful, even though no systematic convergence study is carried out here.

The pseudo-one-dimensional computations are especially useful as a first step towards the fully two-dimensional wave-structure interactions problems considered later, in particular in Chapter 6.

5.4.1 Geometry of the floating object

In all the numerical illustrations below, the floating object has vertical side walls and is described through the equilibrium wetted surface η^{eq}. In the one-dimensional configuration, the interior domain is given by $\Omega_{\mathrm{s}} = (-l, l)$ and we define the equilibrium water-height under the object by

$$H^{\mathrm{eq}}(x) := H_{\mathrm{c}} + \alpha_{\mathrm{c}} \left(\frac{x}{l}\right)^2, \qquad x \in (-l, l),$$

where $H_{\mathrm{c}} > 0$ is the minimal water-height under the center of the object and $\alpha_{\mathrm{c}} > 0$ controls the curvature of the object bottom. The corresponding equilibrium wetted surface is then

$$\eta^{\mathrm{eq}}(x) := b(x) + H^{\mathrm{eq}}(x).$$

This choice gives a smooth bell-shaped lower boundary for the object, symmetric with respect to $x = 0$. The parameter α_{c} fixes how much the object bottom rises near the contact points $x = \pm l$. In particular, smaller values of α_{c} correspond to a flatter bottom, while larger values produce a more curved immersed shape.

Recalling that the vertical displacement of the object is denoted by $\delta_{\mathrm{o}}(t)$, the moving wetted surface is prescribed by

$$\eta^{\mathrm{w}}(x,t) = \eta^{\mathrm{eq}}(x) + \delta_{\mathrm{o}}(t), \qquad H^{\mathrm{w}}(x,t) = H^{\mathrm{eq}}(x) + \delta_{\mathrm{o}}(t).$$

The derivative of the equilibrium water-height, which appears in the elliptic right-hand side in (5.16), is explicitly given by

$$\partial_x H^{\mathrm{eq}}(x) = 2\alpha_{\mathrm{c}} \frac{x}{l^2}.$$

Therefore the geometry of the lower boundary is directly included in the interior elliptic problem through the coefficient H^{w} and through the source term depending on $\partial_x H^{\mathrm{eq}}$.

For the pseudo-one-dimensional computations, which allow to pave the way for towards full two-dimensional simulations, the same object geometry is extruded in the transverse direction. The interior region is the strip $\Omega_{\mathrm{s}} = (-l, l) \times \mathbb{R}$, and the equilibrium water-height under the object is defined by

$$H^{\mathrm{eq}}(\boldsymbol{x}) := H_{\mathrm{c}} + \alpha_{\mathrm{c}} \left(\frac{x}{l}\right)^2, \qquad \boldsymbol{x} = (x, y) \in \Omega_{\mathrm{s}}.$$

It is therefore independent of the transverse variable y. The moving wetted surface and water-height are again given by

$$\eta^{\mathrm{w}}(\boldsymbol{x},t) = b(\boldsymbol{x}) + H^{\mathrm{eq}}(\boldsymbol{x}) + \delta_{\mathrm{o}}(t), \qquad H^{\mathrm{w}}(\boldsymbol{x},t) = H^{\mathrm{eq}}(\boldsymbol{x}) + \delta_{\mathrm{o}}(t).$$

This construction ensures that the two-dimensional setup is exactly invariant in the transverse direction at the continuous level. Any deviation from a one-dimensional profile in the pseudo-one-dimensional simulations can therefore be interpreted as a numerical effect of the two-dimensional unstructured discretization. In the numerical tests presented below, the geometrical parameters are $l = 5$ m, $H_{\mathrm{c}} = 5$ m, and $\alpha_{\mathrm{c}} = 1$ m. Therefore, the equilibrium water-height under the object is

$$H^{\mathrm{eq}}(x) = 5 + \left(\frac{x}{5}\right)^2, \qquad x \in (-5, 5),$$

and the moving wetted height is

$$H^{\mathrm{w}}(x,t) = 5 + \left(\frac{x}{5}\right)^2 + \delta_{\mathrm{o}}(t).$$

In particular, the minimum water-height under the object is attained at the center and is equal to $H^{\mathrm{eq}}(0) = 5$ m, while at the contact points one has $H^{\mathrm{eq}}(\pm 5) = 6$ m. The derivative entering the elliptic right-hand side is then

$$\partial_x H^{\mathrm{eq}}(x) = \frac{2x}{25}.$$

Remark 5.4.1 (*Mass of the object*). *With the present choice of parameters, Archimedes' principle gives*

$$m_{\mathcal{O}} = \rho \int_{-5}^{5} \left(10 - 5 - \frac{x^2}{25}\right) \mathrm{d}x = \rho \left[5x - \frac{x^3}{75}\right]_{-5}^{5} = \rho \frac{140}{3},$$

thus the mass of the object is not prescribed as an additional independent parameter, as it is fixed by the equilibrium wetted surface through the displaced fluid volume. In the numerical implementation, the same quantity is evaluated by quadrature on the obstacle mesh. This is equivalent to the explicit computation above for the present polynomial geometry, but it also allows us to treat more general immersed shapes for which no closed-form expression of the integral is available.

5.4.2 Single-wave interaction with an obstacle over an emerged topography

PARAMETERS

Polynomial degree: $k = 6$
Time-marching order: $p = 3$
Final time: $T_{\max} = 100$ sec
Boundary conditions: homogeneous Neumann

Computational domain:
$\Omega_{\mathrm{f}-} = [-40, -5]$ ($N_{\mathrm{el}} = 20$)
$\Omega_{\mathrm{s}} = [-5, 5]$ ($N_{\mathrm{el}} = 10$)
$\Omega_{\mathrm{f}+} = [5, 40]$ ($N_{\mathrm{el}} = 20$)

Before considering a moving object, we first study a configuration in which the floating structure is kept fixed. In this test, the vertical displacement is prescribed as

$$\delta_{\mathcal{O}}(t) = 0, \qquad \dot{\delta}_{\mathcal{O}}(t) = 0, \qquad \ddot{\delta}_{\mathcal{O}}(t) = 0,$$

so that the wetted surface of the object remains equal to its equilibrium position,

$$\eta^{\mathrm{w}}(x, t) = \eta^{\mathrm{eq}}(x).$$

Consequently, the Newton equation (5.14) vanishes. The purpose of this first test is to isolate the interaction between an incoming shallow-water wave, the fixed interior obstacle region, and a non-flat bathymetry. It also provides a useful check of the long-time relaxation of the coupled solver towards a motionless state.

The bathymetry is chosen as an emerged localized bump located in the left part of the computational domain:

$$b(x) = b_{\mathrm{m}} \exp\left(-\left(\frac{x - x_{\mathrm{b}}}{\sigma_{\mathrm{b}}}\right)^2\right),$$

where $\sigma_{\mathrm{b}} = 4$ and $b_{\mathrm{m}} = 15$ is chosen large enough so that part of the topography emerges above the still-water level. The bump is placed away from the floating structure, in the left exterior domain, so that the bottom is essentially flat in a neighbourhood of the contact points $x = \pm l$. This choice makes the test more demanding than a flat-bottom configuration, while keeping the interaction with the object clearly separated from the bathymetric effect.

The initial condition consists of a localized incoming wave initialized in the right exterior domain and propagating towards the obstacle, defined as

$$\eta(x, 0) = \eta_0 + A_\eta \exp\left(-\left(\frac{x - x_c}{\sigma_\eta}\right)^2\right), \qquad \eta_0 = 10, \quad A_\eta = 1, \quad x_c = 20, \quad \sigma_\eta = 3,$$

and the corresponding initial discharge is chosen as:

$$q(x, 0) = -2H(x, 0)\left(\sqrt{\mathrm{g}H(x, 0)} - \sqrt{\mathrm{g}H_0(x)}\right).$$

When the wave reaches the right contact point of the obstacle, we observe the backward propagation of a reflected component, while part of the information is transmitted through the region under the object. This transmitted component then reaches the left contact point before interacting with the emerged bathymetric bump. This configuration is therefore useful to check several features of the numerical coupling: transmission conditions, the resolution of the elliptic Bernoulli pressure problem in the interior region, and the ability of the exterior DG/FV subcell solver to propagate incident, transmitted, and reflected waves without producing spurious oscillations.
Let us also emphasize that this test is deliberately performed on a very coarse mesh: the two exterior domains are discretized with only 20 cells, while the interior domain under the object contains 10 cells. We use a

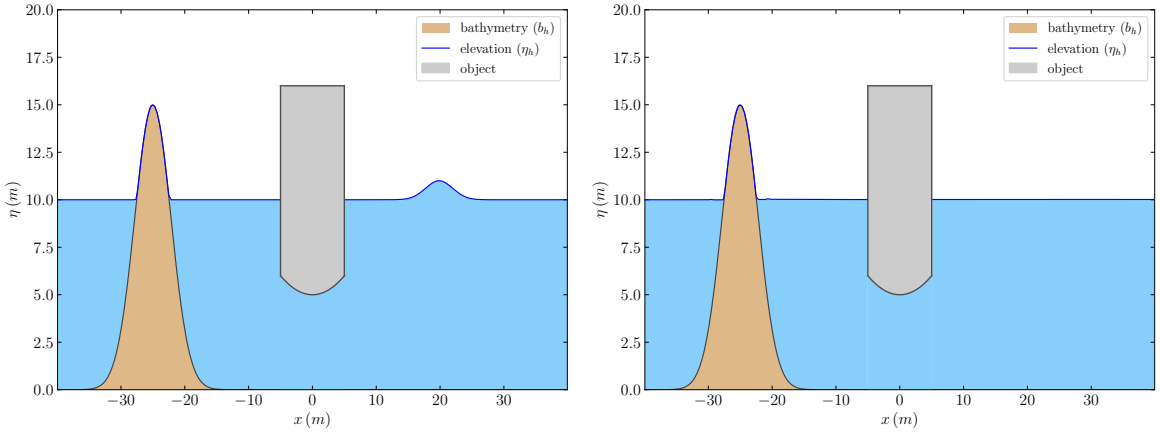


Figure 5.2: Wave interacting with a fixed object and topography − Initial state (left) and final state at $t = 100$ sec (animation).

polynomial degree $k = 6$: this choice makes the test quite demanding as the method has to represent sharp waves, contact interactions, and the relaxation towards equilibrium. The observed overall robustness, together with the convergence towards a motionless steady state, are therefore meaningful indicators of the operability of the coupled DG/FV–SWIP-DG discretization.
The first stages of the computation are shown in Figure 5.3. The strongest interactions occur during this initial time interval, which explains why the snapshots are focused on the time range $0 \leq t \leq 5$ s.

After these interactions, the simulation is continued until $t = 100$ sec over more than 1270000 time steps. This long time range is not used to display additional transient dynamics, but to check the long-time stability of the coupled system. In particular, the region between the floating object and the emerged topography takes more time to stabilize, which is expected because several reflected and transmitted components remain trapped for a while between the two obstacles. The final state is reported in Figure 5.2. It shows that the discharge eventually becomes negligible and that the solution returns towards a lake-at-rest configuration, despite the coarse discretization used in the exterior and interior regions. This complements the well-balanced testing discussed in Chapter 1, with the addition of a surface obstacle. A simulation of this benchmark during the time range $0 \leq t \leq 10$ sec is available here.

5.4.3 Wave generation by a prescribed vertical motion

PARAMETERS	COMMON SETUP	1D COMPUTATIONAL DOMAIN	2D COMPUTATIONAL DOMAIN
	Polynomial degree: $k = 3$	$\Omega_{\mathrm{f}-} = [-40, -5]$ ($N_{\mathrm{el}} = 70$)	$\Omega_{\mathrm{f}-} = [-40, -5] \times [-20, 20]$ ($N_{\mathrm{el}} = 894$)
	Time-marching order: $p = 3$	$\Omega_{\mathrm{s}} = [-5, 5]$ ($N_{\mathrm{el}} = 25$)	$\Omega_{\mathrm{s}} = [-5, 5] \times [-20, 20]$ ($N_{\mathrm{el}} = 368$)
	Final time: $T_{\max} = 5$ sec	$\Omega_{\mathrm{f}+} = [5, 40]$ ($N_{\mathrm{el}} = 70$)	$\Omega_{\mathrm{f}+} = [5, 40] \times [-20, 20]$ ($N_{\mathrm{el}} = 892$)

We now consider a configuration in which the vertical motion of the floating structure is prescribed. More precisely, we impose

$$\delta_{\mathrm{O}}(t) = A_\delta \sin(2\pi t), \qquad \dot{\delta}_{\mathrm{O}}(t) = 2\pi A_\delta \cos(2\pi t), \qquad \ddot{\delta}_{\mathrm{O}}(t) = -4\pi^2 A_\delta \sin(2\pi t),$$

with $A_\delta = 2$. Again, in this test, the Newton equation governing the vertical motion of the center of mass, namely (5.14), is not solved, as the displacement, velocity and acceleration of the object are directly imposed through the interior equations.

This configuration provides an intermediate test between the obstacle considered precedently and a fully coupled fluid-structure system. The elliptic problem for the Bernoulli pressure is solved in the interior domain through (5.21)–(5.22), and the kinematic constraint

$$q^{\mathrm{s}}(x, t) = -x\dot{\delta}_{\mathrm{O}}(t) + \langle q^{\mathrm{s}} \rangle(t)$$

is imposed under the object. However, the feedback of the pressure on the vertical motion is removed.

The prescribed motion can be interpreted as a wave-making mechanism for the nonlinear shallow-water equations. Indeed, the discharge matching condition at the contact points gives

$$q^{\mathrm{f}}(\pm l, t) = \mp l\dot{\delta}_{\mathrm{O}}(t) + \langle q^{\mathrm{s}} \rangle(t),$$

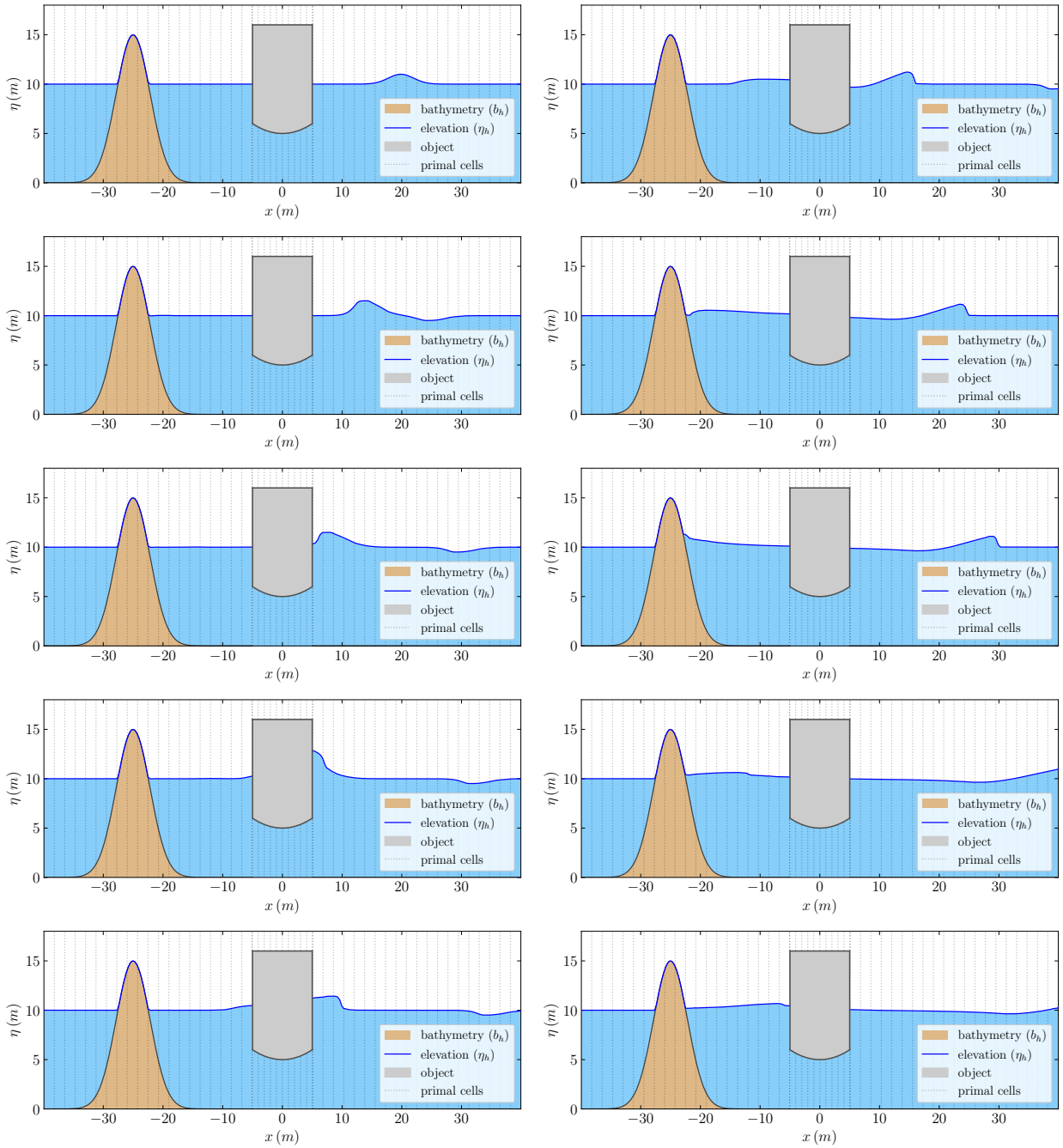


Figure 5.3: Wave interacting with a fixed object and topography − Time evolution of the free-surface elevation η during the period [0,10] sec, ordered columnwise from top to bottom (animation).

so that the moving object generates outgoing waves in the exterior domains. In practice, the resulting time series of the horizontal discharge can also be used as boundary data for a standard shallow-water computation on a larger domain.

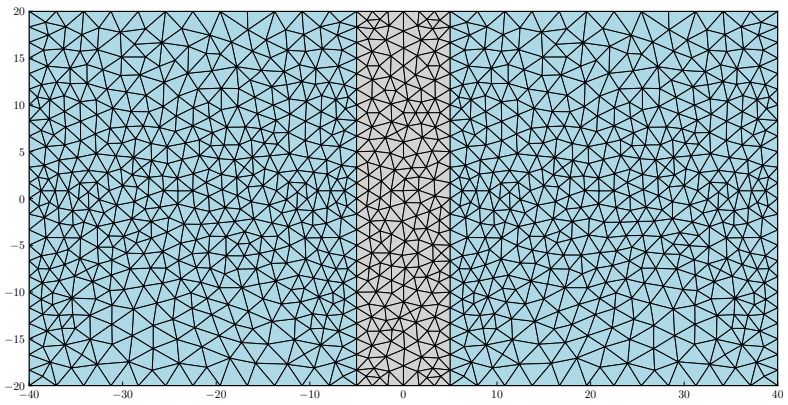


Figure 5.4: Unstructured meshes for $\Omega_{\mathrm{f}-}$ (left), Ω_{s} (center) and $\Omega_{\mathrm{f}+}$ (right) used during the simulation.

The initial state is chosen as a motionless configuration, with the exterior fluid at rest and the object initially at its equilibrium. Once the prescribed motion is activated, the wetted surface relation

$$\eta^{\mathrm{w}}(x,t) = \eta^{\mathrm{eq}}(x) + \delta_{\mathcal{O}}(t)$$

induces a non-zero interior discharge through the term $\partial_x q^{\mathrm{s}} = -\dot{\delta}_{\mathcal{O}}$, which is then transmitted to the exterior fluid through the coupling conditions.

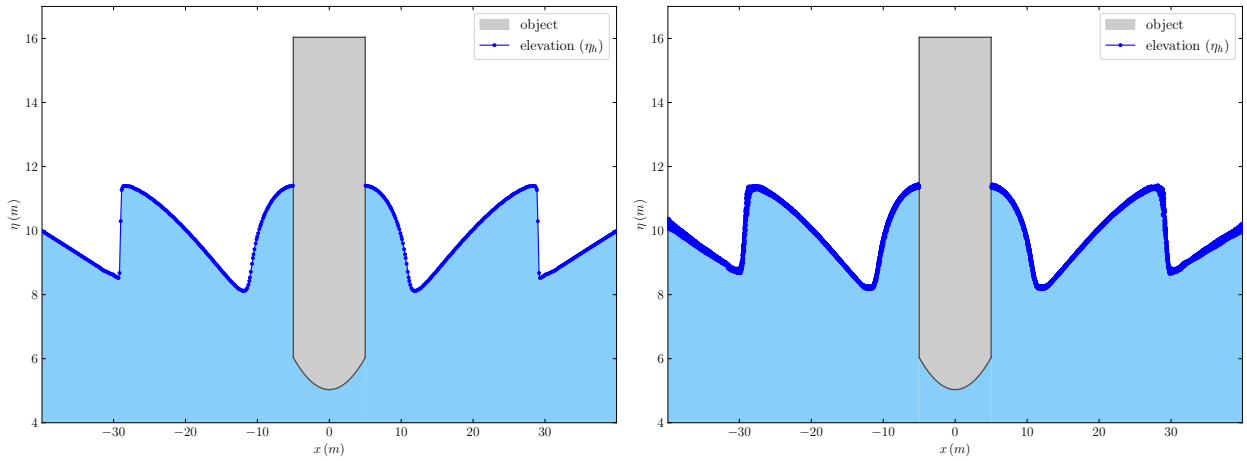


Figure 5.5: Wave generator (1D & 2D) – At time $t = 5$ s: comparison of the free-surface elevation η between the one-dimensional solver (left) and the two-dimensional solver (right). The 2D result is displayed as a density profile, gathering all the numerical values along the longitudinal direction (animation 1D & animation 2D).

We first perform this experiment in the one-dimensional setting. The same configuration is then reproduced in the pseudo-one-dimensional setting on an unstructured two-dimensional mesh displayed in Figure 5.4.
We observe on Figure 5.5 that the 1D and 2D profiles are in very good agreement: the waves generated by the imposed motion of the structure propagate away from the obstacle with the same profiles and at the same velocity. We observe that the unstructured mesh does not introduce transverse artefacts in this pseudo-one-dimensional regime. Both simulations are available here in 1D and here in 2D.
Figure 5.6 shows the time evolution of the one-dimensional solution over one full oscillation period after the first waves have already been generated. The oscillation of the object produces successive outgoing waves on both sides of the structure. The discharge response follows the imposed interface condition and changes sign according to the direction of the vertical motion. The steep fronts generated near the contact points are transported in a stable way, without spurious oscillations.
The corresponding pseudo-one-dimensional computations are reported in Figures 5.7 and 5.8. In Figure 5.7, the free-surface elevation is shown over the second oscillation period, namely between $t = 2$ sec and $t = 4$

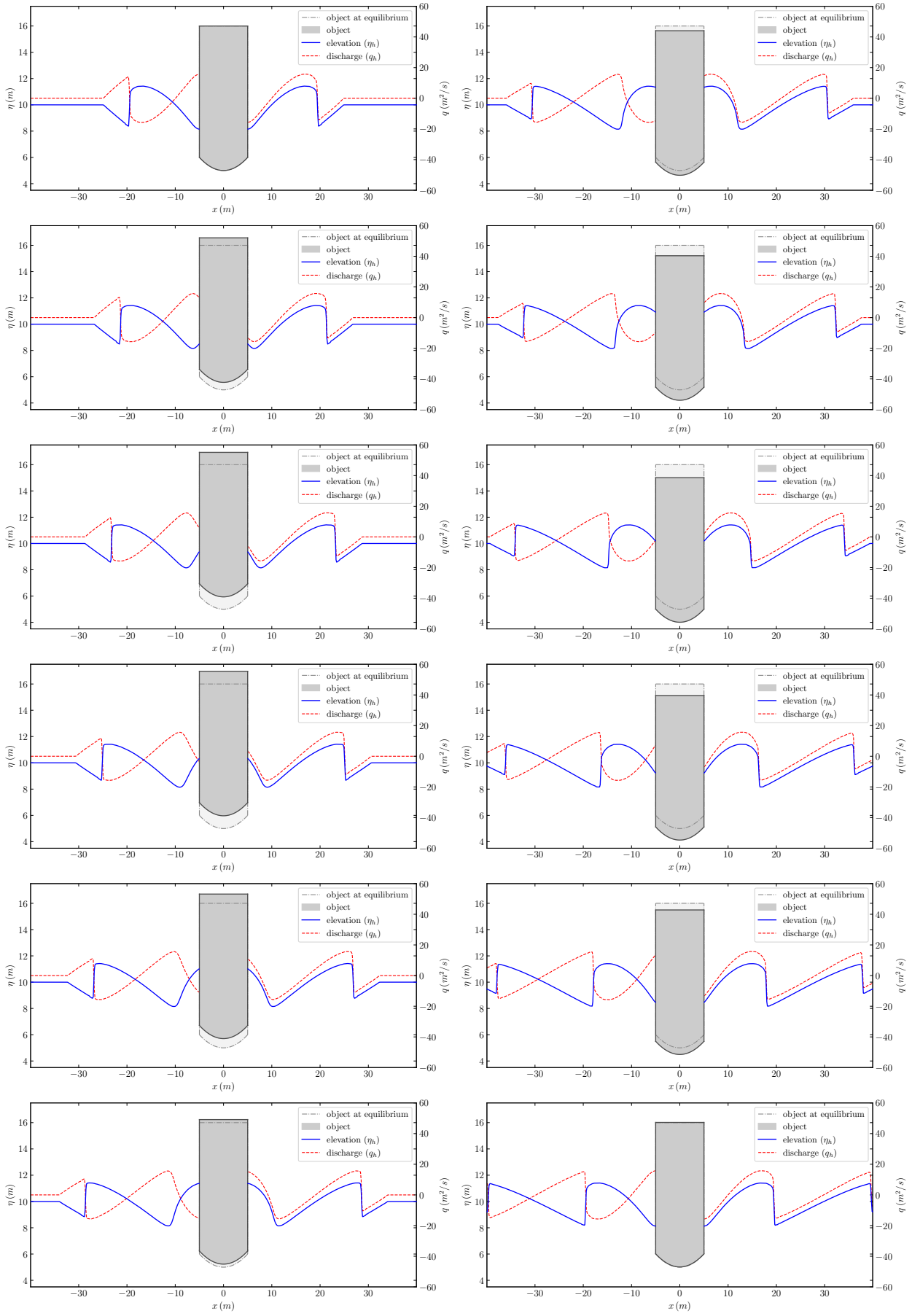


Figure 5.6: Wave generator (1D) – Time evolution of the free-surface elevation η and the horizontal discharge q during the period $[2, 4]$ sec, ordered columnwise from top to bottom (animation).

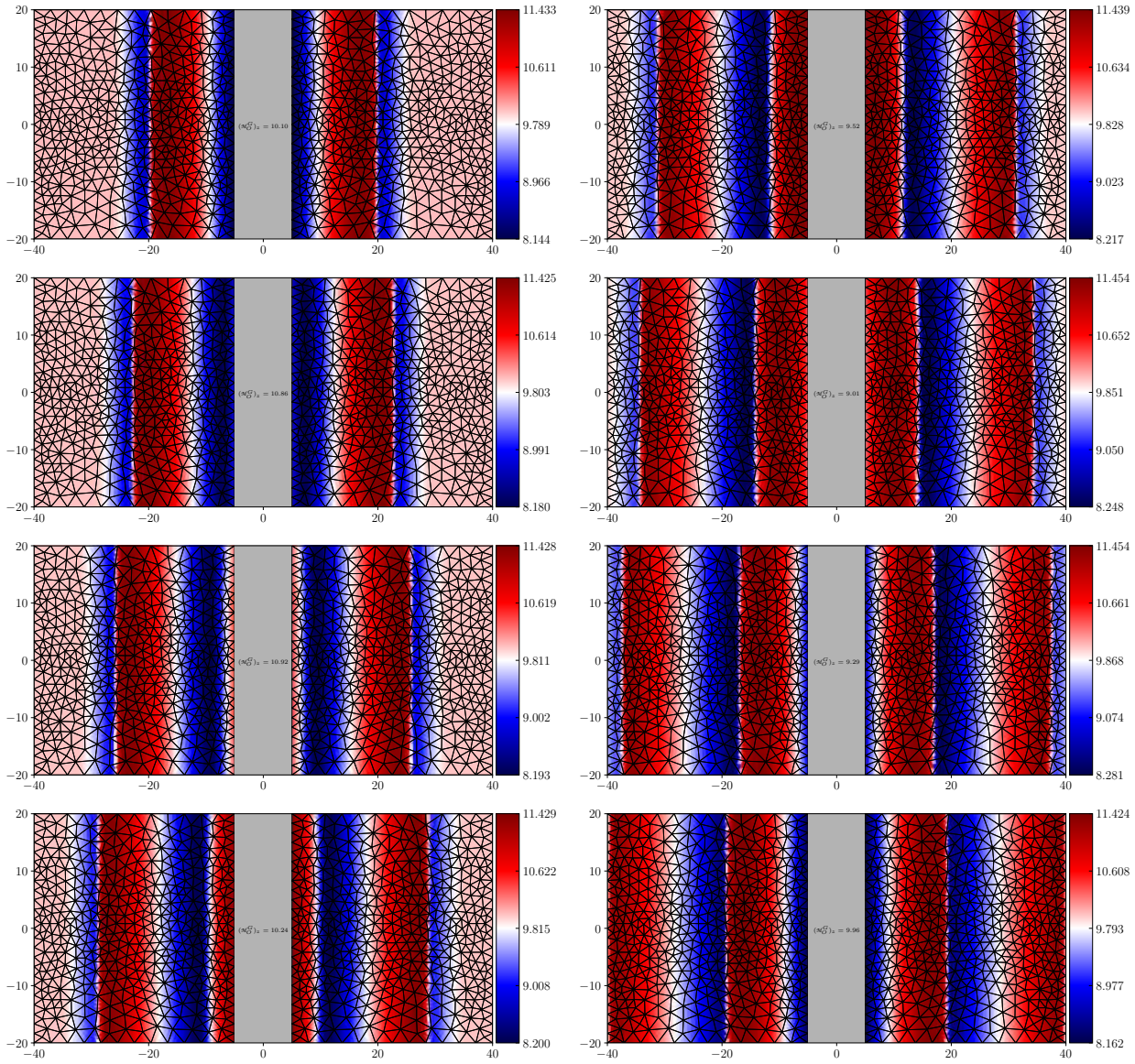


Figure 5.7: Wave generator (2D) − Time evolution of the free-surface elevation η during the period $(2, 4)$ sec, ordered columnwise from top to bottom. The label displayed at the center of each panel indicates the vertical coordinate $(\mathcal{M}_O^G)_z$ of the center of mass of the floating object (animation 2D).

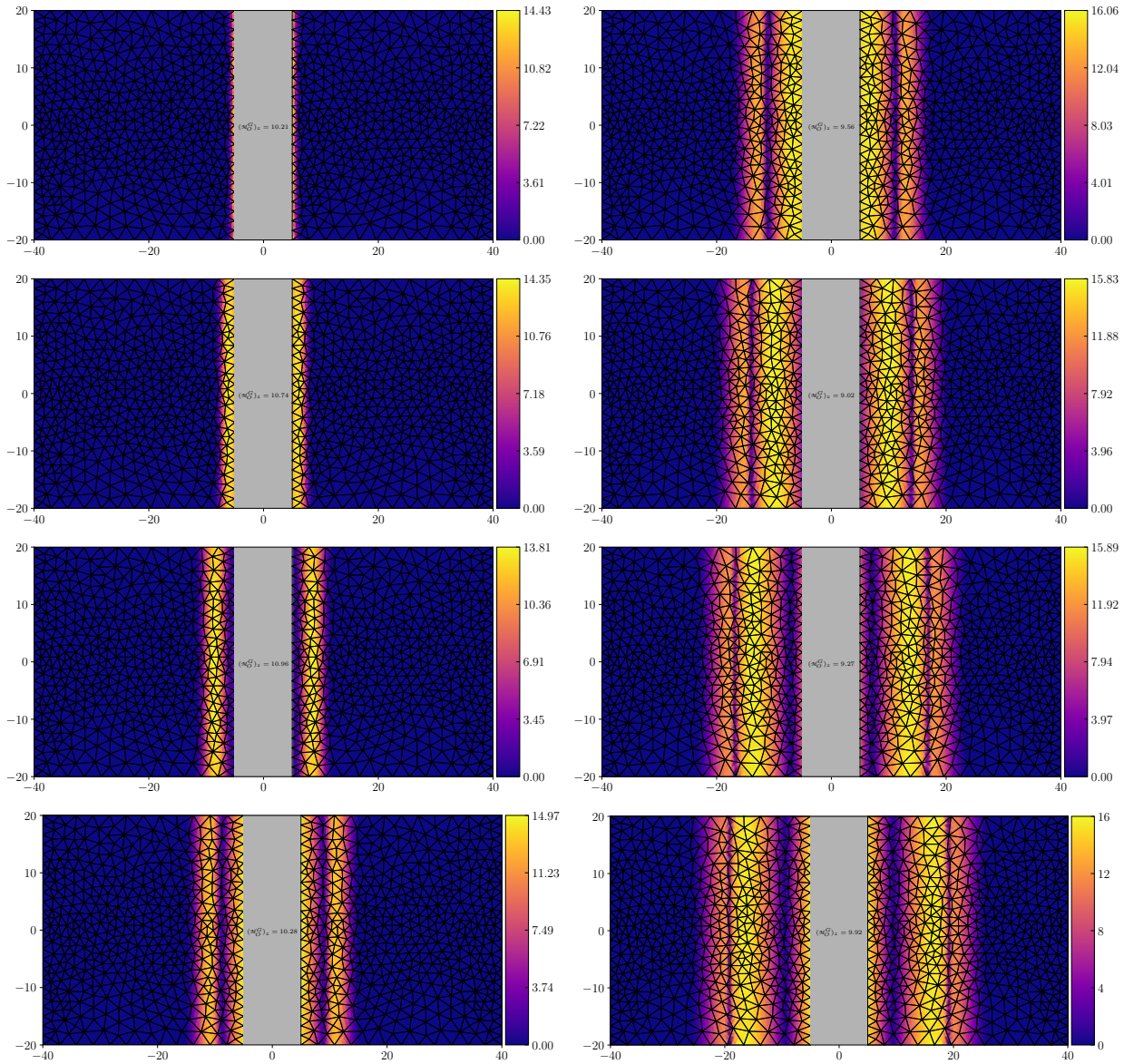


Figure 5.8: Wave generator (2D) – Columnwise from top to bottom: time evolution of the discharge norm during the period $(0, 2)$ sec. The label displayed at the center of each panel indicates the vertical coordinate $(\mathcal{M}_O^G)_z$ of the center of mass of the floating object (animation 2D).

sec, so that the wave field has already been generated before the beginning of the displayed sequence. The free-surface remains essentially invariant in the transverse direction, and the wave fronts are aligned with the obstacle interfaces. In Figure 5.8, the discharge norm is shown over the first period (between $t = 0$ sec and $t = 2$ sec) to highlight the initial transient regime generation of the flow by the prescribed motion of the object. No parasitic oscillation near the contact points or inside the exterior domains are observed. This confirms that the two-dimensional implementation preserves the one-dimensional dynamics even on the unstructured mesh.

It is also encouraging that the same qualitative behaviour is recovered with two different elliptic discretizations: the SWIP-DG method in the one-dimensional computation and the HHO method in the two-dimensional computation. The agreement between these two independent implementations gives a useful consistency check before considering the fully coupled case.

5.4.4 Wave interaction with a freely heaving floating structure

We now gather two fully coupled wave-structure interactions tests in which the floating object is free to move vertically under the action of the hydrodynamic pressure. In contrast with the previous tests, the displacement δ_O is then now computed from the Newton equation (5.14). We compare two different incoming perturbations: a surge wave and a localized single-wave, which produces a transient interaction before coming back towards its initial equilibrium.

Shock-wave interaction

PARAMETERS

Polynomial degree: $k = 2$
Time-marching order: $p = 3$
Final time: $T_{\max} = 100$ sec
Boundary conditions: homogeneous Neumann

Computational domain:
$\Omega_{f-} = [-40, -5] \times [-20, 20]$ ($N_{el} = 894$)
$\Omega_{s} = [-5, 5] \times [-20, 20]$ ($N_{el} = 368$)
$\Omega_{f+} = [5, 40] \times [-20, 20]$ ($N_{el} = 892$)

The initial condition is a Riemann problem which produces a shock wave propagating towards the floating structure:

$$\eta(\boldsymbol{x}, 0) = \begin{cases} 15 & x < -30, \\ 10 & x \geq -30, \end{cases} \qquad \boldsymbol{q}(\boldsymbol{x}, 0) = \boldsymbol{0},$$

producing a rightward propagating shock wave that reaches the floating structure. The object is initially at equilibrium, so that

$$\delta_O(0) = \dot{\delta}_O(0) = \ddot{\delta}_O(0) = 0.$$

The subsequent motion of the object is obtained by solving the coupled system consisting of the exterior nonlinear shallow-water equations, the interior elliptic Bernoulli pressure problem, and the vertical Newton equation for δ_O: the incoming shock modifies the Bernoulli pressure under the structure, this pressure drives the vertical displacement through the Newton equation, and the resulting motion feeds back into the exterior flow through the transmission conditions at the contact points.

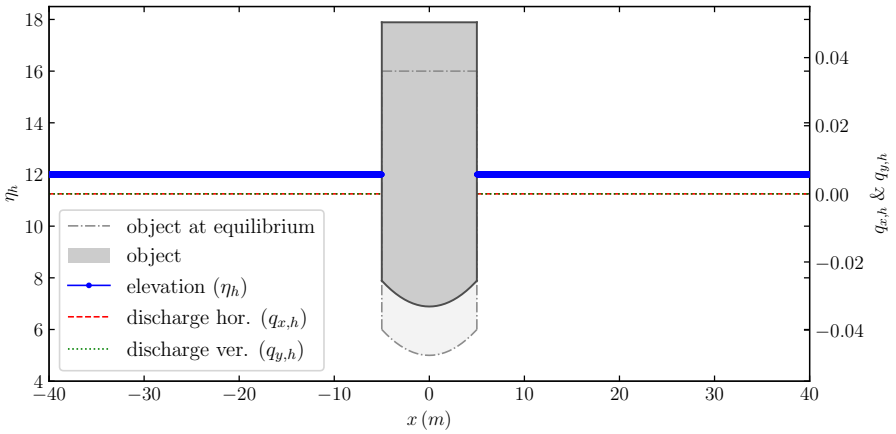


Figure 5.9: Shock-wave interactions with a vertically freely heaving floating object − Elevation and discharge density profiles at the final time $t = 100$ sec.

The one-dimensional free-surface profiles reported in Figure 5.11 focus on the time interval where most of the interactions take place. The incoming shock first reaches the left contact point and induces a rapid increase of the Bernoulli pressure under the object, inducing an upward motion. this motion modifies the wetted height H^{w} and the discharge imposed at both contact points. This produces a depression wave on both sides of the structure, generating both reflected and transmitted components, with a wave pattern also reported in [20].

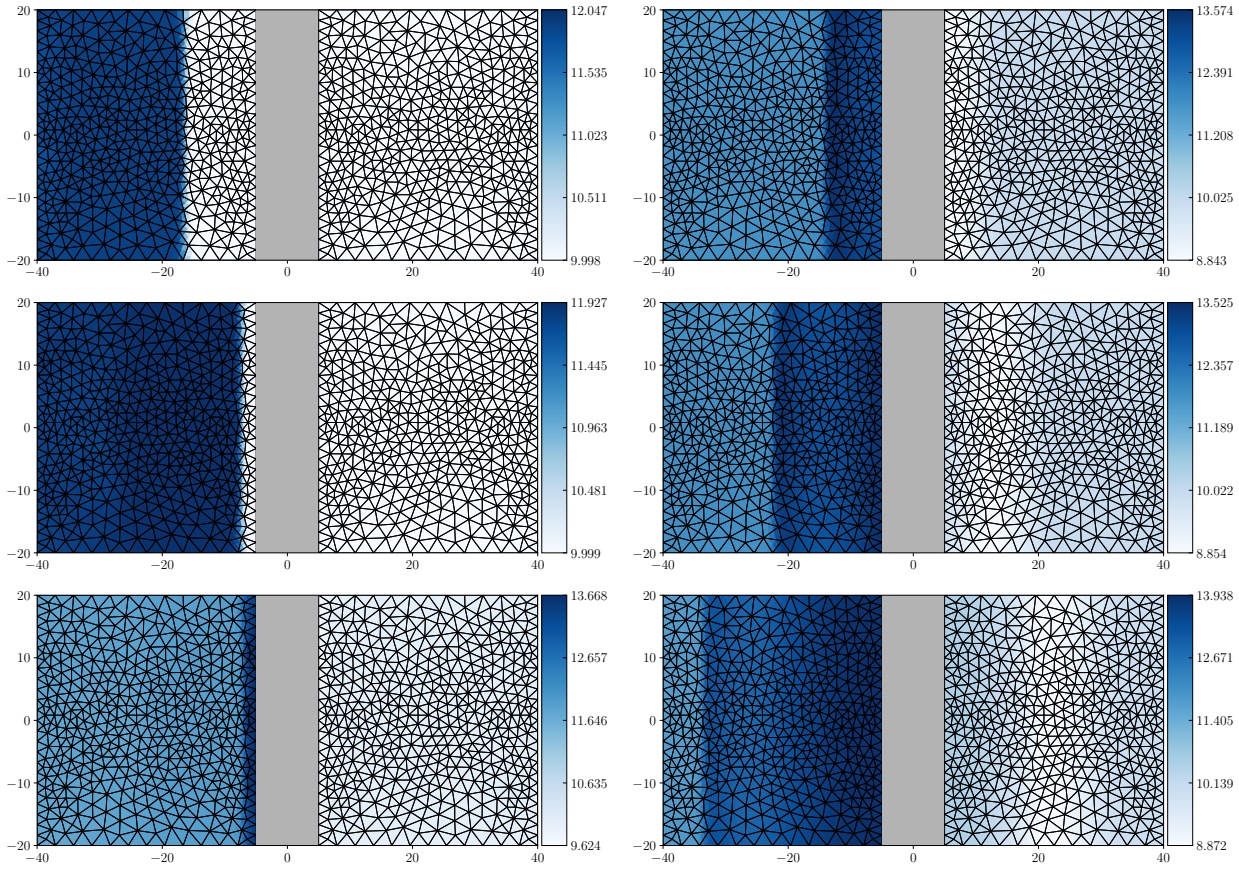


Figure 5.10: Shock-wave interactions with a vertically freely heaving floating object − Time evolution of the free-surface elevation η during the period $[1.2, 5]$ sec, ordered columnwise from top to bottom (animation).

The corresponding two-dimensional snapshots are displayed in Figure 5.10 with a similar behavior. The time interval $[1.2, 5]$ sec is chosen to highlight the part of the computation where the shock reaches the floating structure and where the strongest pressure and displacement variations occur. Although the computation is genuinely two-dimensional, the solution remains essentially invariant in the transverse direction. In particular, the wave fronts remain aligned with the obstacle interfaces, and no visible transverse distortion appears near the contact lines.

After approximately 10 sec, the transients components with large amplitudes have left the neighbourhood of the object and the two exterior states become much smoother. The left and right elevations are not identical, since the initial dam-break creates a larger amount of water on the left side. The remaining imbalance then drives a slow transfer of water through the region below the floating object. At the final time $t = 100$ sec, reported in Figure 5.9, the elevation has relaxed towards an almost uniform level close to 12 m, corresponding to intermediate level enforced by the dam-break dynamics. The new equilibrium position of the object remains slightly lifted with respect to its equilibrium position, consistently with the new averaged hydrostatic loading induced by this final water level. An animation of the simulation during the time interval [0,10] sec is available here.

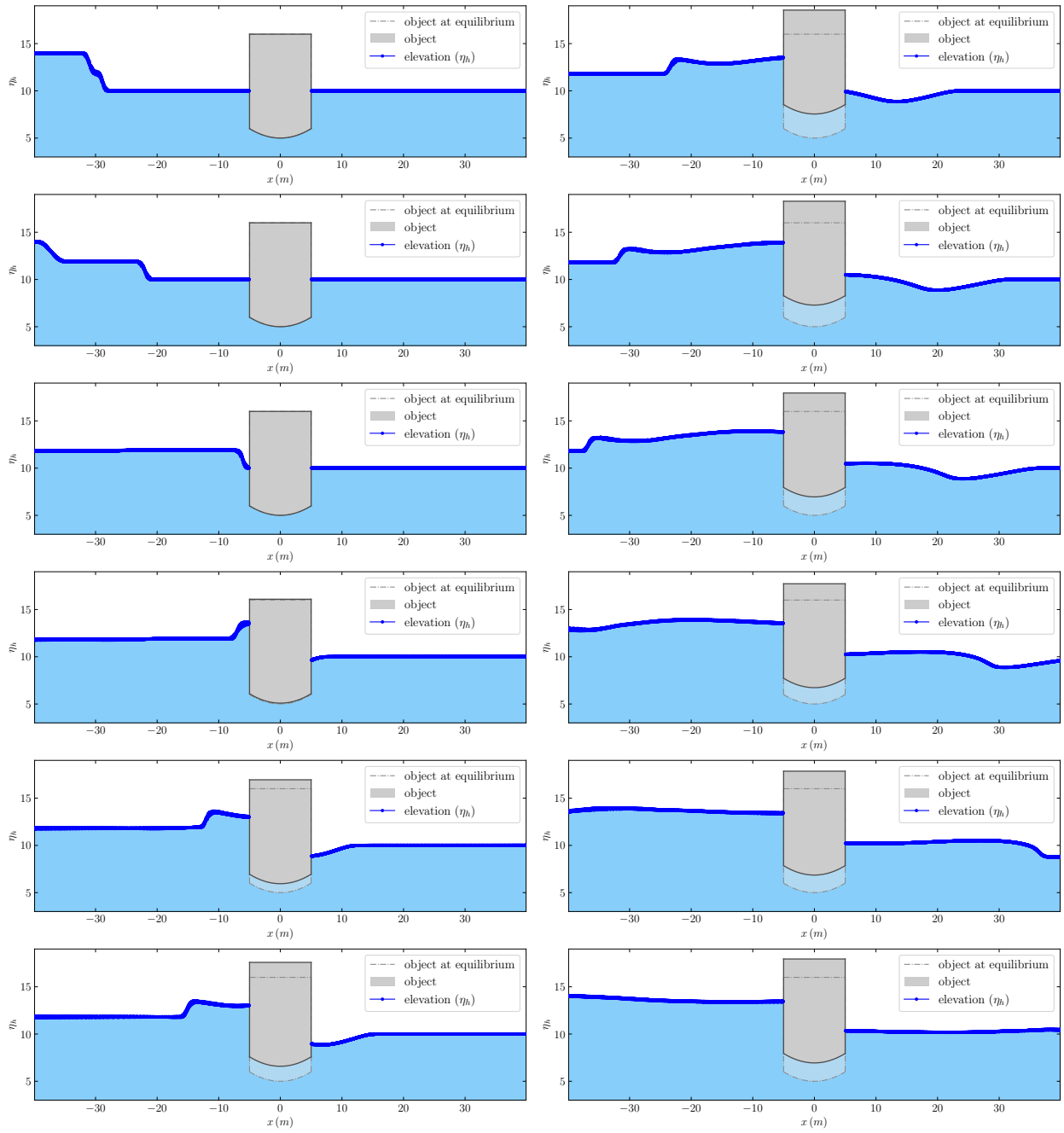


Figure 5.11: Shock-wave interactions with a vertically freely heaving floating object − Columnwise snapshots of the free-surface elevation density profile during the period $[0.1, 9]$ sec, ordered from top to bottom (animation).

Single-wave interaction

PARAMETERS

Polynomial degree: $k = 5$
Time-marching order: $p = 3$
Final time: $T_{\max} = 50$ sec
Boundary conditions: homogeneous Neumann

Computational domain:
$\Omega_{\mathrm{f}-} = [-40, -5] \times [-20, 20]$ ($N_{\mathrm{el}} = 78$)
$\Omega_{\mathrm{s}} = [-5, 5] \times [-20, 20]$ ($N_{\mathrm{el}} = 26$)
$\Omega_{\mathrm{f}+} = [5, 40] \times [-20, 20]$ ($N_{\mathrm{el}} = 78$)

The numerical setting is similar to the previous test, but the initial discontinuity is replaced by a localized single-wave propagating towards the object.

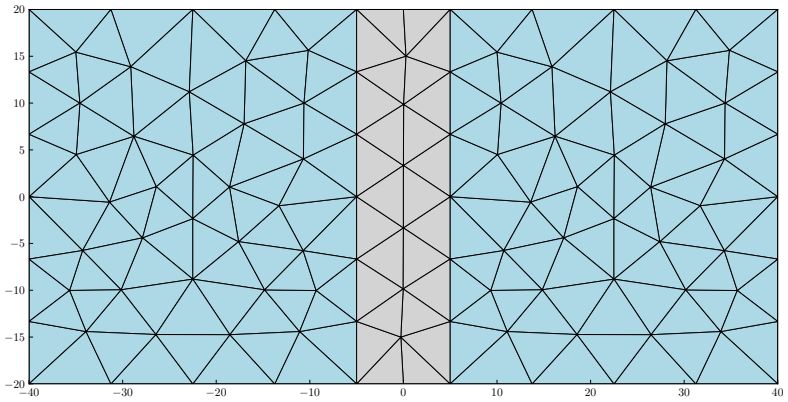


Figure 5.12: Single-wave interactions with a freely heaving floating object − Unstructured meshes for $\Omega_{\mathrm{f}-}$ (left), Ω_{s} (center) and $\Omega_{\mathrm{f}+}$ (right) used during the simulation.

The purpose of this test is to check that the proposed numerical method allows to benefit from high-order polynomial approximations for smooth processes, while using a deliberately coarse two-dimensional unstructured mesh. The mesh, displayed in Figure 5.12, contains only 78 elements in the left and right exterior domains, 26 elements under the object, and we use polynomial degree $k = 5$.

The initial condition consists of a localized right-going single-wave placed in the left exterior domain:

$$\eta(\boldsymbol{x}, 0) = H_0 + A_\eta \operatorname{sech}^2 \left(\gamma(x - x_c)\right), \qquad \gamma = \sqrt{\frac{3A_\eta}{4H_0}},$$

with $H_0 = 10$, $A_\eta = 4.35$ and $x_c = -30$. The initial discharge is defined as

$$\boldsymbol{q}(\boldsymbol{x}, 0) = \left(\sqrt{\frac{\mathrm{g}}{H_0}} \left(\eta(\boldsymbol{x}, 0) - H_0\right) \eta(\boldsymbol{x}, 0), 0\right)^\top,$$

while the object is initially at rest.
The one-dimensional density profiles reported in Figure 6.6 show the main stages of the interactions. The profiles remain sharply resolved during the whole interactions, despite the very coarse meshes used in the computation.

Again, the 2D snapshots displayed in Figure 5.15 confirm that the computation preserves the expected pseudo-one-dimensional structure.
The late-time behaviour, shown in Figure 5.14, is different from the dam-break case: since the single-wave does not impose a persistent difference of water level between the two sides of the object, the system relaxes towards its initial equilibrium state after the wave has left the computational region. The object returns close to its initial vertical position, and the free-surface recovers the original still-water level. This final state is a sensitive check of the consistency between the exterior DG/FV subcell solver, the interior HHO pressure solve, and the time integration of the Newton equation.

This test illustrates again the adaptability and potentially computational efficiency of the proposed high-order approximation strategy: even with a small number of elements, the method allows to capture the relevant wave-structure mechanisms with nonlinear profiles and stable long-time behaviour. It also shows that increasing the polynomial degree can provide a rapid gain in accuracy on coarse unstructured meshes.

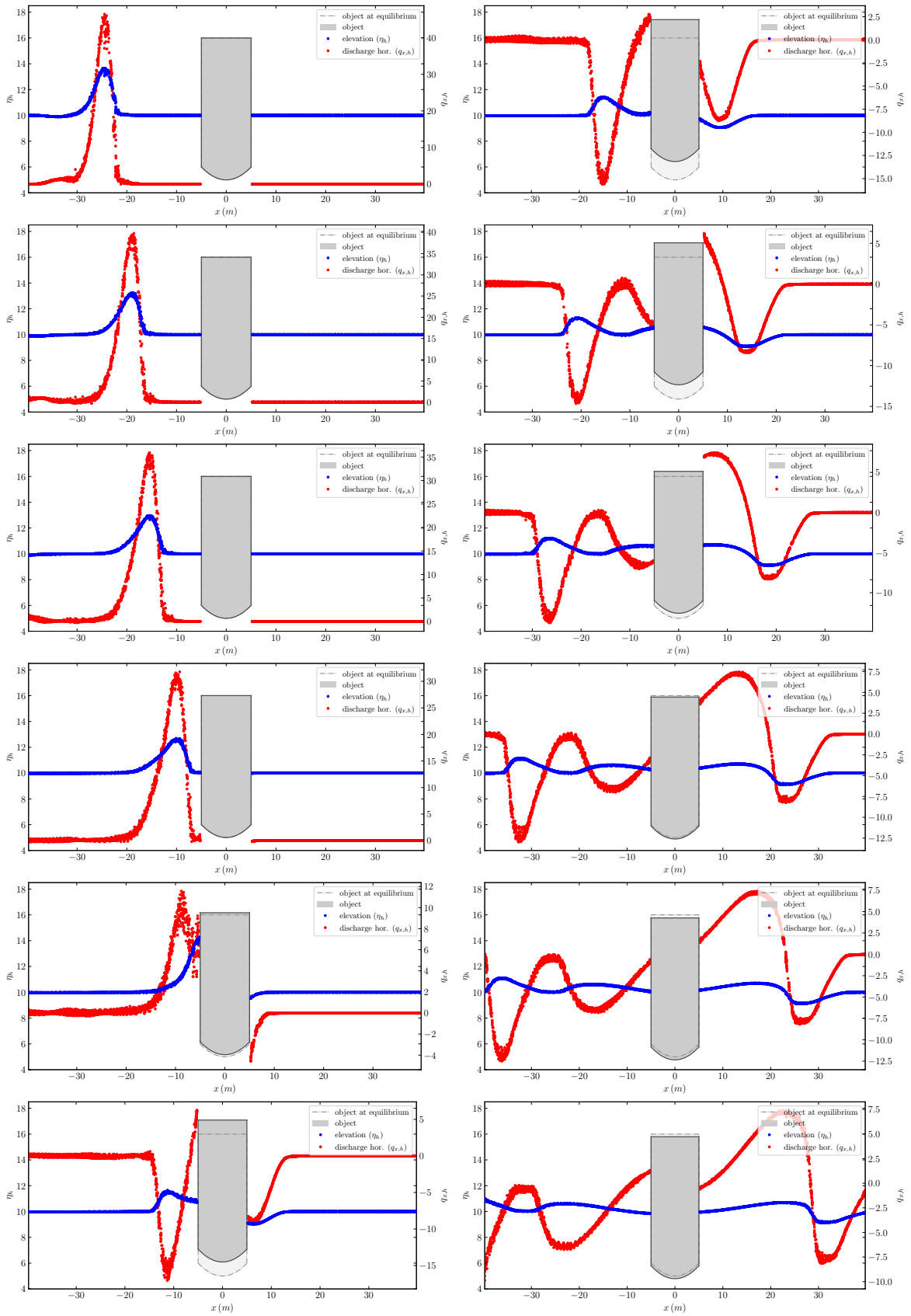


Figure 5.13: Single-wave interactions with a freely heaving floating object − Time evolution of the free-surface elevation η and the horizontal discharge q_x during the period $[0.4, 5.4]$ sec, ordered columnwise from top to bottom (animation).

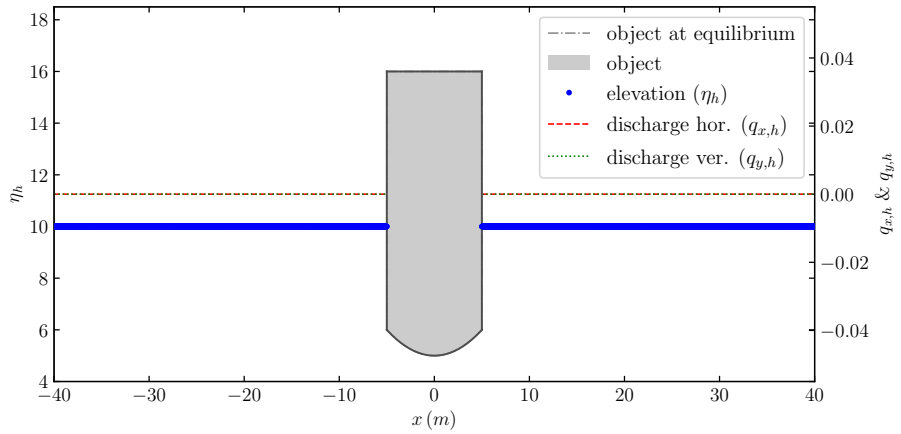


Figure 5.14: Single-wave interactions with a freely heaving floating object − Elevation and discharge profiles at the final time $t = 100$ sec.

5.4.5 Return to equilibrium test with analytical reference

PARAMETERS	COMMON SETUP	1D COMPUTATIONAL DOMAIN	2D COMPUTATIONAL DOMAIN
	Polynomial degree: $k = 3$	$\Omega_{\mathrm{f}-} = [-40, -5]$ ($N_{\mathrm{el}} = 200$)	$\Omega_{\mathrm{f}-} = [-40, -5] \times [-20, 20]$ ($N_{\mathrm{el}} = 894$)
	Time-marching order: $p = 3$	$\Omega_{\mathrm{s}} = [-5, 5]$ ($N_{\mathrm{el}} = 100$)	$\Omega_{\mathrm{s}} = [-5, 5] \times [-20, 20]$ ($N_{\mathrm{el}} = 368$)
	Final time: $T_{\max} = 10$ sec	$\Omega_{\mathrm{f}+} = [5, 40]$ ($N_{\mathrm{el}} = 200$)	$\Omega_{\mathrm{f}+} = [5, 40] \times [-20, 20]$ ($N_{\mathrm{el}} = 892$)

We now consider a return-to-equilibrium configuration for the coupled wave-structure interactions model. This benchmark is based on the analytical reduction obtained in [182] for a symmetric floating structure in the nonlinear shallow-water regime. In contrast with the previous qualitative tests, the present configuration provides a quantitative reference for the vertical motion of the floating object.

The purpose of this benchmark is to use the analytical exterior closure available in this symmetric return-to-equilibrium setting in order to isolate the interior elliptic pressure problem and its coupling with the vertical Newton equation. This allows us to check whether the SWIP-DG discretization of the hydrodynamic pressure and the discrete Newton update reproduce the reduced analytical dynamics.

The object is initially released from a position different from its equilibrium position, while the exterior fluid is initially at rest. For the one-dimensional quantitative comparison with the analytical reduced model, we use a small initial vertical displacement,

$$\delta_o(0) = \delta_o^0 = 0.05, \qquad \dot{\delta_o}(0) = 0.$$

This choice is consistent with the assumptions underlying the analytical return-to-equilibrium reduction, where the exterior closure is obtained on the admissible branch of the contact relation. In particular, the initial vertical deviation is kept small enough to remain in the theoretical flow regime. The subsequent motion is therefore only driven by the hydrostatic restoring force, the added-mass contribution induced by the fluid below the object, and the nonlinear radiation damping associated with the outgoing shallow-water waves.

In the symmetric return-to-equilibrium regime, the average interior discharge vanishes. The interior discharge is then reduced to

$$q^{\mathrm{s}}(x,t) = -x\dot{\delta_o}(t), \qquad x \in (-l, l),$$

and the corresponding discharge transmitted to the exterior domains satisfies

$$q^{\mathrm{f}}(-l,t) = l\dot{\delta_o}(t), \qquad q^{\mathrm{f}}(l,t) = -l\dot{\delta_o}(t).$$

This symmetry is one of the main structural properties of the test case. In particular, the hydrodynamic pressure traces at the two contact points are expected to be identical in the analytical closure.

The analytical reference is obtained by solving the reduced ordinary differential equation

$$(m_o + m_o^{\mathrm{a}}(\delta_o))\,\ddot{\delta_o} = -c_{\mathrm{h}}\delta_o - \nu(\dot{\delta_o}) + \beta(\delta_o)\dot{\delta_o}^{\,2}.$$

Here $c_{\mathrm{h}} = 2\rho g l$ is the *hydrostatic restoring coefficient*,

$$m_o^{\mathrm{a}}(\delta_o) = \rho \int_{-l}^{l} \frac{x^2}{H^{\mathrm{eq}}(x) + \delta_o}\,\mathrm{d}x$$

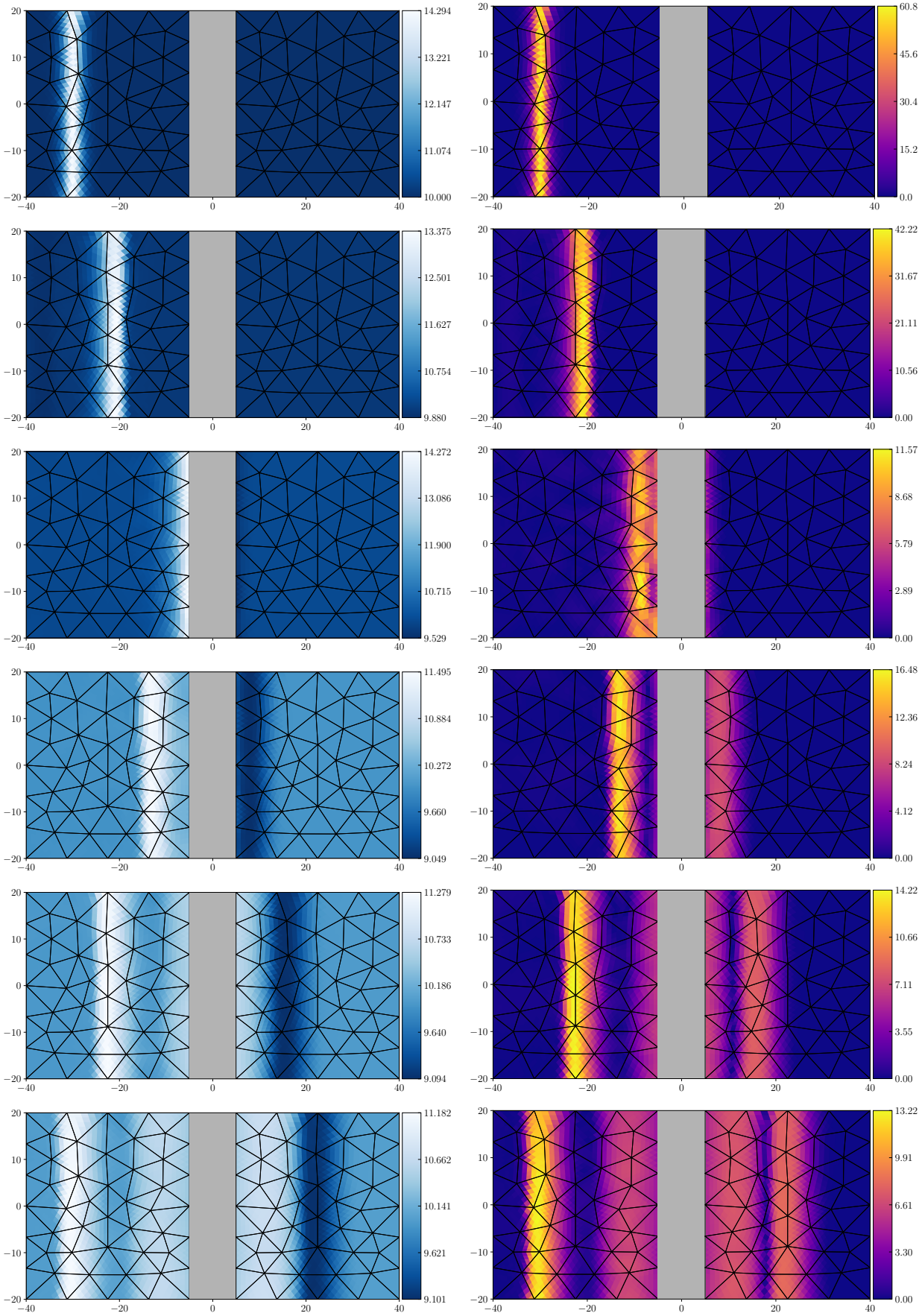


Figure 5.15: Single-wave interactions with a freely heaving floating object – Snapshots of the free-surface elevation (left column) and the discharge norm (right column) during the period $[0, 4.3]$ sec (animation).

is the *added mass*, and

$$\beta(\delta_o) = \rho \int_{-l}^{l} \frac{x}{H^{\mathrm{eq}}(x) + \delta_o} \partial_x \left(\frac{x^2}{H^{\mathrm{eq}}(x) + \delta_o} \right) \mathrm{d}x$$

is the nonlinear coefficient associated with the geometry of the wetted region. These quantities are evaluated by quadrature on the obstacle mesh, using the same equilibrium wetted height H^{eq} as in the interior elliptic problem.

The nonlinear damping term is expressed through the root τ_0 of

$$\tau_0^3 - \sqrt{H_0}\tau_0^2 + \frac{l}{2\sqrt{\mathrm{g}}}\dot{\delta_o} = 0, \qquad \tau_0(0) = \sqrt{H_0}.$$

The corresponding contact elevation is

$$\eta_{\mathrm{c}}^{\mathrm{th}}(t) - H_0 = \tau_0^2 - H_0,$$

and the *damping force* is

$$\nu(\dot{\delta_o}) = 2\rho \mathrm{g} l \left(H_0 - \tau_0^2 \right).$$

This explicit expression of the contact elevation is the key point of the benchmark. It provides an analytical exterior closure for the interior hydrodynamic pressure, depending only on the velocity of the object.

We compare two evolutions; the first one is obtained by integrating the reduced ODE above while the second one is obtained by solving the interior SWIP-DG elliptic pressure problem coupled with the vertical Newton equation. In this second computation, the exterior closure is not reconstructed from the DG/FV exterior domains. Instead, the analytical contact elevation is used to prescribe the Dirichlet data for the hydrodynamic pressure:

$$\Pi^{\mathrm{s}}(-l, t) = \Pi^{\mathrm{s}}(l, t) = \rho \mathrm{g} \left(\eta_{\mathrm{c}}^{\mathrm{th}}(t) - H_0 \right).$$

The displacement is then computed from the hydrodynamic pressure form of the Newton equation,

$$\tau_{\mathrm{buoy}}^2 \ddot{\delta_o} + \delta_o = \frac{1}{2\rho \mathrm{g} l} \int_{-l}^{l} \Pi^{\mathrm{s}}(x, t)\, \mathrm{d}x, \qquad \tau_{\mathrm{buoy}}^2 = \frac{m_o}{2\rho \mathrm{g} l}.$$

This comparison gives a direct validation of the elliptic-Newton component of the method. Indeed, the two computations use different descriptions of the same continuous mechanism: the reduced ODE contains the added-mass and damping terms in closed form, whereas the numerical computation recovers the hydrodynamic pressure by solving the interior elliptic problem and then inserts its integral into the Newton equation.

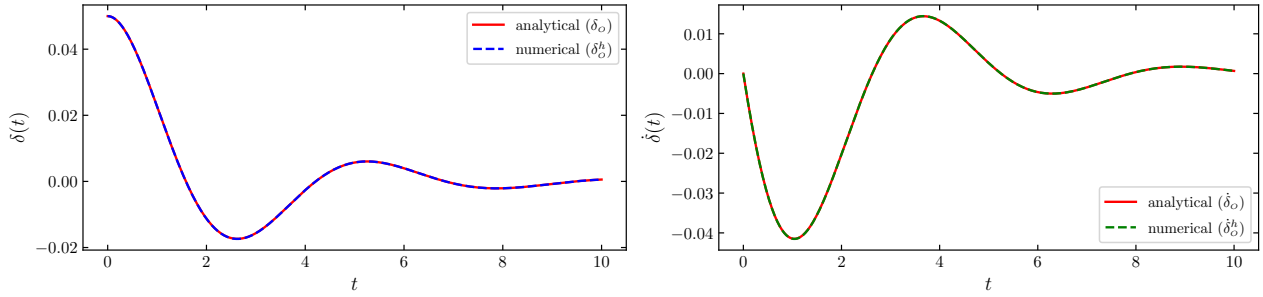


Figure 5.16: Return to equilibrium (1D) – Comparison of the reduced analytical solution and the elliptic-Newton computation for the vertical displacement δ_o and velocity $\dot{\delta_o}$.

The displacement and velocity curves are reported in Figure 5.16. The agreement between the reduced analytical dynamics and the elliptic-Newton computation confirms that the interior pressure solve correctly reproduces the hydrodynamic force entering the vertical Newton equation. This agreement is the main validation result of the present benchmark.
The corresponding errors are shown in Figure 5.17. The smallness of these errors confirms that the SWIP-DG elliptic discretization, the quadrature of the pressure force, and the time update of the vertical displacement are consistent with the analytical return-to-equilibrium model.

We also run the same return-to-equilibrium mechanism in the pseudo-one-dimensional two-dimensional setting on the unstructured meshes displayed in Figure 5.4. To make the motion visible and to stress the coupled solver, we consider larger initial displacements, both upward and downward:

$$\delta_o^0 = 3.5, \qquad \delta_o^0 = -3.5.$$

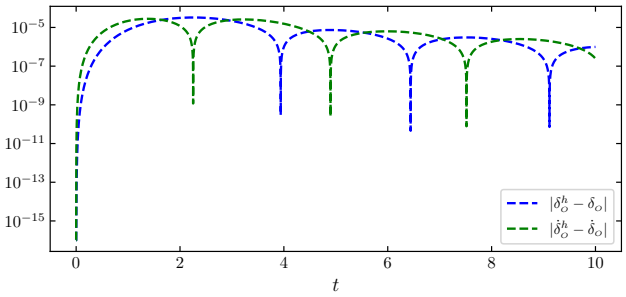


Figure 5.17: Return to equilibrium (1D) – Errors between the reduced analytical solution and the elliptic-Newton computation for δ_O and $\dot{\delta}_O$.

These values are deliberately much larger than the perturbation used in the 1D analytical comparison. They are therefore not used to assess the reduced analytical formula, but rather to illustrate the behaviour of the pseudo-one-dimensional 2D solver when the floating object is released away from its equilibrium position.
The two-dimensional snapshots are displayed in Figure 5.18. In the upward case, the object is initially above its equilibrium position and then moves back under the combined action of gravity, pressure forces and wave radiation. In the downward case, the initial motion is reversed and the fluid is first compressed under the object before the system relaxes. In both cases, the solution remains essentially pseudo-one-dimensional on the unstructured two-dimensional mesh: the free-surface profiles remain aligned in the transverse direction and no transverse instability is generated near the contact lines.

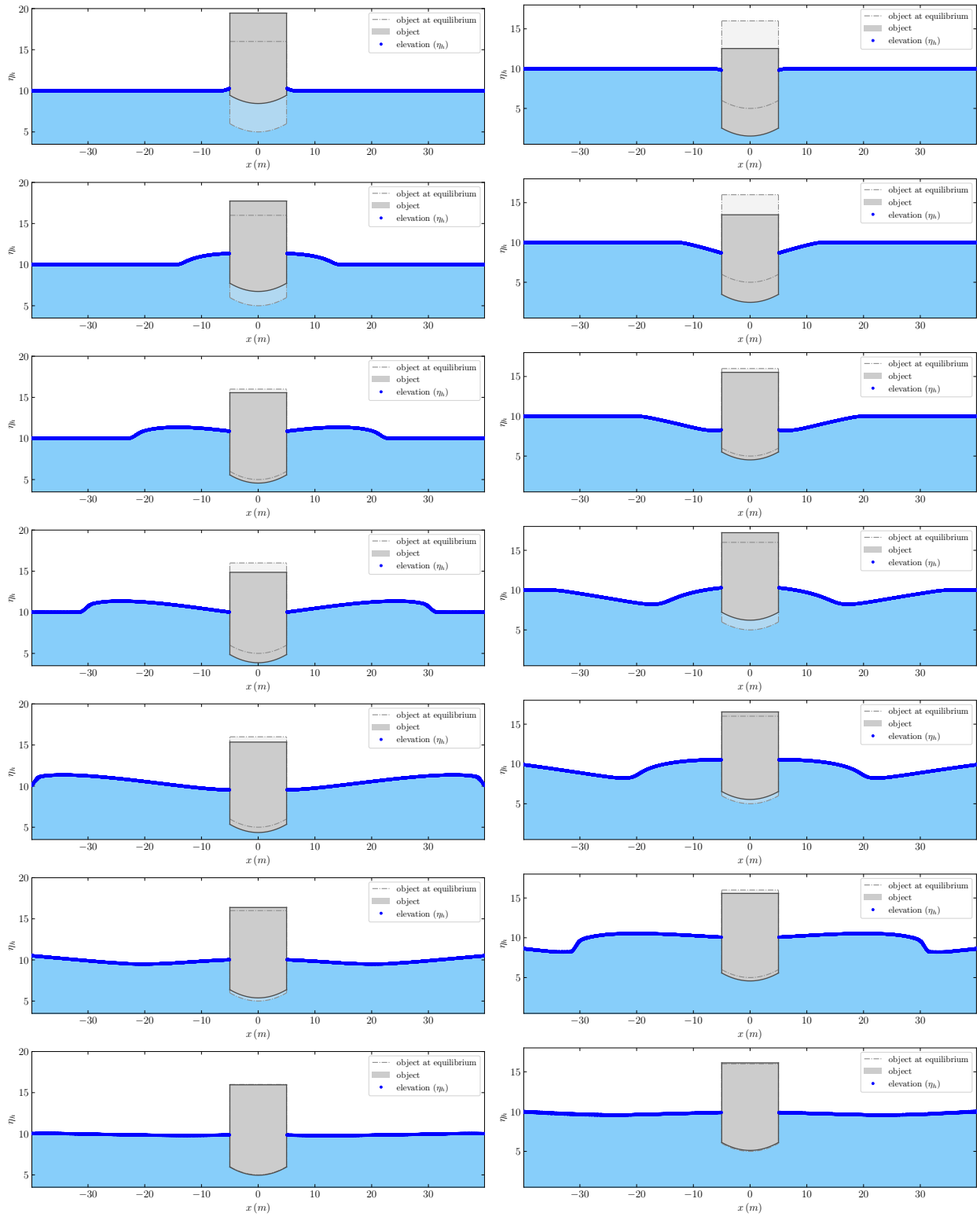


Figure 5.18: Return to equilibrium (2D) – Free-surface elevation profile η during the period $(0, 10]$ sec. Left column: initial upward displacement ${\delta_O}^0 = 3.5$ (upward animation). Right column: initial downward displacement ${\delta_O}^0 = -3.5$ (downward animation).

CHAPTER 6

Coupled subcell monolithic DG/FV–HHO scheme for a new two-dimensional wave-structure interactions model

The main results presented in the following, together with selected developments from Chapter 5, are expected to form the basis of two research articles devoted to wave-structure interactions.

In this chapter we focus on a new two-dimensional model describing the interaction between shallow-water flows and a partially immersed structure. In contrast with the one-dimensional configuration of Chapter 5, the present setting fully resolves the horizontal geometry of the computational domain, leading to a genuinely two-dimensional coupled problem: the fluid motion is governed by the NSW equations in the exterior region, while the structure imposes constraints in the interior region. This leads to an elliptic problem for the interior potential together with an ordinary differential equation for the trace of this potential on the fluid–structure interface. Unlike the one-dimensional case, the coupling is now distributed along a two-dimensional interface, and relies on the introduction of a Dirichlet–to-Neumann operator linking the interior and exterior dynamics.

The NSW equations are discretized using the monolithic DG/FV subcell method developed in Chapter 3, while the interior elliptic problem is approximated with a Hybrid High-Order (HHO) method [86]. The coupling across the interface is enforced through consistent transmission conditions, while the time evolution of the interface trace is integrated using an explicit SSP Runge–Kutta scheme.

Contents

6.1 Wave-structure interaction model 186
6.1.1 Description as a partly congested nonlinear shallow-water system 186
6.1.2 A mixed IBVP in the exterior region for irrotational data 188
6.1.3 An equivalent formulation 189
6.1.4 A special configuration: the infinite pontoon 189
6.2 Discrete formulations 194
6.2.1 Approximation framework 194
6.2.2 SSP-RK discretization of the ODE for the interface potential trace 200
6.2.3 Fast Fourier Transform computation of the DtN operator 201
6.2.4 HHO formulation of the interior elliptic problem 202
6.2.5 Local monolithic DG/FV subcell method for the exterior region 203
6.2.6 Semi-discrete DG–HHO formulation 205
6.2.7 Fully-discrete monolithic DG/FV–HHO–SSP-RK formulation 205
6.3 Numerical experiments 206
6.3.1 Geometry of the structures 206
6.3.2 Wave-pontoon interactions and DtN solver comparison 207
6.3.3 Incident wave interactions with a fixed cylindrical obstacle 210

Toutes les ébènes ont rendez-vous, lambeaux de nuit quand nos ombres s'éteignent
Des routes m'emmènent, je ne sais où, j'avais les yeux perçants avant, je voyais tout
FROM: ***"En passant"***, Jean-Jacques Goldman – *En passant* (1997)

6.1 Wave-structure interaction model

In this section, we detail the mathematical formulation of the wave-structure interactions problem considered in this chapter (see [185]). We describe the physical configuration and the governing equations for the fluid and the structure, then reformulate the problem as a mixed *Initial Boundary Value Problem* (IBVP) in the exterior fluid domain.

6.1.1 Description as a partly congested nonlinear shallow-water system

We consider a fluid layer in $\mathbb{R}^3$ over a flat bottom located at $z = -H_0$, with $H_0 > 0$ and z denoting the vertical coordinate, and in the presence of a partially immersed solid object. Denoting by t the time variable and by $\boldsymbol{x} = (x, y)$ the horizontal variables, we assume that the elevation of the surface of the fluid with respect to the rest state level $z = 0$ is parametrized at time t by a function $\boldsymbol{x} \mapsto \zeta(\boldsymbol{x}, t)$; the domain $\Omega(t)$ occupied by the fluid at time t is therefore

$$\Omega(t) = \{(\boldsymbol{x}, z) \in \mathbb{R}^{d+1} \mid -H_0 < z < \zeta(\boldsymbol{x}, t)\}.$$

The partially immersed object $\mathbf{S}$ considered here is a portion of vertical cylinder (see Figure 6.1),

$$\mathbf{S} = \{(\boldsymbol{x}, z) \in \mathbb{R}^{d+1} \mid \boldsymbol{x} \in \Omega_{\mathrm{s}},\ \zeta^{\mathrm{w}}(\boldsymbol{x}) < z < \zeta_{\mathrm{top}}(\boldsymbol{x})\},$$

where ζ^{w} describes the bottom boundary of the object, and $\zeta_{\mathrm{top}} - \zeta^{\mathrm{w}}$ is the height of the object. The section Ω_{s} of the cylinder therefore coincides with the projection on the horizontal plane of the wet part of the object. We denote by Ω_{f} the part of the horizontal plane associated with the fluid domain, and by $\partial\Omega_{\mathrm{fs}}$ the horizontal projection of the contact line, such that

$$\mathbb{R}^2 = \Omega_{\mathrm{s}} \cup \partial\Omega_{\mathrm{fs}} \cup \Omega_{\mathrm{f}},$$

and we refer to Ω_{s} and Ω_{f} as the interior *structure* and exterior *fluid* regions. We also denote by $\mathbf{n}$ the unit normal vector to $\partial\Omega_{\mathrm{fs}}$ pointing towards Ω_{f}. Throughout this chapter, we denote ζ^{s}, v^{s}, $\underline{\mathrm{p}}^{\mathrm{s}}$, etc. (resp. ζ^{f}, v^{f}, $\underline{\mathrm{p}}^{\mathrm{f}}$, etc.) the quantities ζ, v, $\underline{\mathrm{p}}$, etc. respectively associated with the interior and the exterior domains.

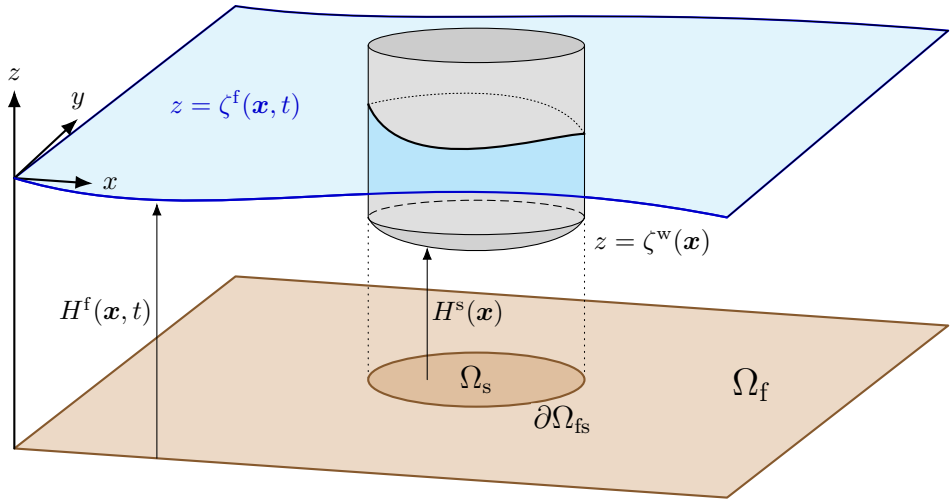


Figure 6.1: Geometry of a partially immersed cylindrical surface obstacle in a shallow-water flow. The free-surface is given by $z = \zeta^{\mathrm{f}}(\boldsymbol{x}, t)$, the wetted surface of the structure by $z = \zeta^{\mathrm{w}}(\boldsymbol{x})$, and the computational domain is split into an exterior region Ω_{f} and an interior region Ω_{s} beneath the structure.

We consider throughout this work that the fluid motion is governed by the nonlinear shallow-water equations. Denoting by $\boldsymbol{u}(\boldsymbol{x}, t)$, with $\boldsymbol{u} = (u_x, u_y)^\top$, the vertically averaged horizontal component of the fluid velocity at time t, and by $\underline{\mathrm{p}}(\boldsymbol{x}, t)$ the pressure at the point $(\boldsymbol{x}, \zeta^{\mathrm{f}}(\boldsymbol{x}, t))$ of the surface, the nonlinear shallow-water equations read:

$$\begin{cases} \partial_t \zeta^{\mathrm{f}} + \nabla_{\boldsymbol{x}} \cdot (H^{\mathrm{f}} \boldsymbol{u}^{\mathrm{f}}) = 0, \\ \partial_t \boldsymbol{u}^{\mathrm{f}} + \boldsymbol{u}^{\mathrm{f}} \cdot \nabla_{\boldsymbol{x}} \boldsymbol{u}^{\mathrm{f}} + \mathrm{g} \nabla_{\boldsymbol{x}} \zeta^{\mathrm{f}} = -\frac{1}{\rho} \nabla_{\boldsymbol{x}} \underline{\mathrm{p}}^{\mathrm{f}}, \end{cases} \qquad (H^{\mathrm{f}} = H_0 + \zeta^{\mathrm{f}}), \tag{6.1}$$

where ρ is the constant density of the fluid. These equations must be completed with constraints in the interior and exterior regions, and matching conditions on $\partial\Omega_{\mathrm{fs}}$:

- *Constraint in the interior region* Ω_{s}. By definition, the surface of the fluid coincides with the bottom of the object in the interior region,

$$\zeta^{\mathrm{s}} = \zeta^{\mathrm{w}} \quad \text{for } t \geq 0, \qquad \text{in } \Omega_{\mathrm{s}} \times (0,T). \tag{6.2}$$

  Since ζ^{w} is independent of time, the constrained shallow-water equations (6.1) have the same structure as the *incompressible* Euler equations and read

$$\begin{cases} \nabla_{\boldsymbol{x}} \cdot \left(H^{\mathrm{s}} \boldsymbol{u}^{\mathrm{s}}\right) = 0, \\ \partial_t \boldsymbol{u}^{\mathrm{s}} + \boldsymbol{u}^{\mathrm{s}} \cdot \nabla_{\boldsymbol{x}} \boldsymbol{u}^{\mathrm{s}} + \mathrm{g} \nabla_{\boldsymbol{x}} \zeta^{\mathrm{w}} = -\frac{1}{\rho} \nabla_{\boldsymbol{x}} \underline{\mathrm{p}}^{\mathrm{s}}, \end{cases} \qquad \text{in } \Omega_{\mathrm{s}} \times (0,T).$$

  Here $H^{\mathrm{s}} = H_0 + \zeta^{\mathrm{w}}$ is independent of time but may depend on the horizontal variable $\boldsymbol{x}$; a flat wetted underside corresponds to the particular case where H^{s} is constant.

- *Constraint in the exterior region* Ω_{f}. In the exterior region, the constrained is on the surface pressure, assumed to coincide with the constant atmospheric pressure $\mathrm{p}_{\mathrm{atm}}$,

$$\underline{\mathrm{p}}^{\mathrm{f}}(\boldsymbol{x},t) = \mathrm{p}_{\mathrm{atm}} \quad \text{for } t \geq 0, \qquad \text{in } \Omega_{\mathrm{f}} \times (0,T), \tag{6.3}$$

  therefore the right-hand side in the second equation of (6.1) vanish and the equations reduce to the standard (i.e. without pressure term) shallow-water equations,

$$\begin{cases} \partial_t \zeta^{\mathrm{f}} + \nabla_{\boldsymbol{x}} \cdot (H^{\mathrm{f}} \boldsymbol{u}^{\mathrm{f}}) = 0, \\ \partial_t \boldsymbol{u}^{\mathrm{f}} + \boldsymbol{u}^{\mathrm{f}} \cdot \nabla_{\boldsymbol{x}} \boldsymbol{u}^{\mathrm{f}} + \mathrm{g} \nabla_{\boldsymbol{x}} \zeta^{\mathrm{f}} = 0, \end{cases} \qquad \text{in } \Omega_{\mathrm{f}} \times (0,T).$$

- *Matching conditions on* $\partial\Omega_{\mathrm{fs}}$. Let us define the Bernoulli pressure as

$$\begin{cases} \Pi^{\mathrm{f}} = \rho \mathrm{g} \zeta^{\mathrm{f}} + \frac{1}{2} \rho |\boldsymbol{u}^{\mathrm{f}}|^2 & \text{in } \Omega_{\mathrm{f}} \times (0,T) \\ \Pi^{\mathrm{s}} = \underline{\mathrm{p}}^{\mathrm{s}} - \mathrm{p}_{\mathrm{atm}} + \rho \mathrm{g} \zeta^{\mathrm{s}} + \frac{1}{2} \rho |\boldsymbol{u}^{\mathrm{s}}|^2 & \text{in } \Omega_{\mathrm{s}} \times (0,T), \end{cases}$$

  as shown very recently in [185], imposing the following two matching conditions ensure the conservation of mass and energy respectively,

$$H^{\mathrm{f}} \boldsymbol{u}^{\mathrm{f}} \cdot \mathbf{n} = H^{\mathrm{s}} \boldsymbol{u}^{\mathrm{s}} \cdot \mathbf{n} \quad \text{on } \partial\Omega_{\mathrm{fs}} \times (0,T), \tag{6.4}$$

$$\Pi^{\mathrm{f}} = \Pi^{\mathrm{s}} \quad \text{on } \partial\Omega_{\mathrm{fs}} \times (0,T), \tag{6.5}$$

  The problem is still underdetermined at this step, hence the following additional matching condition

$$\boldsymbol{u}^{\mathrm{f}} \cdot \mathbf{n}^{\perp} = \boldsymbol{u}^{\mathrm{s}} \cdot \mathbf{n}^{\perp} \quad \text{on } \partial\Omega_{\mathrm{fs}} \times (0,T), \tag{6.6}$$

  ensuring the continuity of the tangential component of the velocity across $\partial\Omega_{\mathrm{fs}}$, therefore preventing the creation of a vortex line there.

Gathering equations (6.1)-(6.6) leads to the system:

$$\begin{cases} \partial_t \zeta^{\mathrm{f}} + \nabla_{\boldsymbol{x}} \cdot (H^{\mathrm{f}} \boldsymbol{u}^{\mathrm{f}}) = 0 & \text{in } \Omega_{\mathrm{f}} \times (0,T), \\ \partial_t \boldsymbol{u}^{\mathrm{f}} + \boldsymbol{u}^{\mathrm{f}} \cdot \nabla_{\boldsymbol{x}} \boldsymbol{u}^{\mathrm{f}} + \mathrm{g} \nabla_{\boldsymbol{x}} \zeta^{\mathrm{f}} = \mathbf{0} & \text{in } \Omega_{\mathrm{f}} \times (0,T), \\ \zeta^{\mathrm{s}} = \zeta^{\mathrm{w}} & \text{in } \Omega_{\mathrm{s}} \times (0,T), \\ \nabla_{\boldsymbol{x}} \cdot \left(H^{\mathrm{s}} \boldsymbol{u}^{\mathrm{s}}\right) = 0 & \text{in } \Omega_{\mathrm{s}} \times (0,T), \\ \partial_t \boldsymbol{u}^{\mathrm{s}} + \boldsymbol{u}^{\mathrm{s}} \cdot \nabla_{\boldsymbol{x}} \boldsymbol{u}^{\mathrm{s}} + \mathrm{g} \nabla_{\boldsymbol{x}} \zeta^{\mathrm{w}} = -\frac{1}{\rho} \nabla_{\boldsymbol{x}} \underline{\mathrm{p}}^{\mathrm{s}} & \text{in } \Omega_{\mathrm{s}} \times (0,T), \\ \Pi^{\mathrm{f}} = \rho \mathrm{g} \zeta^{\mathrm{f}} + \frac{1}{2} \rho |\boldsymbol{u}^{\mathrm{f}}|^2 & \text{in } \Omega_{\mathrm{f}} \times (0,T), \\ \Pi^{\mathrm{s}} = \underline{\mathrm{p}}^{\mathrm{s}} - \mathrm{p}_{\mathrm{atm}} + \rho \mathrm{g} \zeta^{\mathrm{s}} + \frac{1}{2} \rho |\boldsymbol{u}^{\mathrm{s}}|^2 & \text{in } \Omega_{\mathrm{s}} \times (0,T), \\ H^{\mathrm{f}} \boldsymbol{u}^{\mathrm{f}} \cdot \mathbf{n} = H^{\mathrm{s}} \boldsymbol{u}^{\mathrm{s}} \cdot \mathbf{n} & \text{on } \partial\Omega_{\mathrm{fs}} \times (0,T), \\ \Pi^{\mathrm{f}} = \Pi^{\mathrm{s}} & \text{on } \partial\Omega_{\mathrm{fs}} \times (0,T), \\ \boldsymbol{u}^{\mathrm{f}} \cdot \mathbf{n}^{\perp} = \boldsymbol{u}^{\mathrm{s}} \cdot \mathbf{n}^{\perp} & \text{on } \partial\Omega_{\mathrm{fs}} \times (0,T). \end{cases} \tag{6.7}$$

that is the wave-structure interactions problem introduced in [185], where it has been proved that the system (6.7) is locally well-posed in Sobolev spaces for irrotational initial data and admits a unique solution.

6.1.2 A mixed IBVP in the exterior region for irrotational data

A key feature of the equations (6.1)-(6.6) is that irrotationality is propagated from the initial data, see [185]. More precisely, if the initial data is such that $\nabla_{\boldsymbol{x}}^{\perp}\cdot\boldsymbol{u}(0,\cdot)=0$ in Ω_{f} and $\nabla_{\boldsymbol{x}}^{\perp}\cdot\boldsymbol{u}^{\mathrm{s}}(0,\cdot)=0$ in Ω_{s}, then $\nabla_{\boldsymbol{x}}^{\perp}\cdot\boldsymbol{u}(\cdot,t)=0$ in Ω_{f} and $\nabla_{\boldsymbol{x}}^{\perp}\cdot\boldsymbol{u}^{\mathrm{s}}(\cdot,t)=0$ in Ω_{s} for all $t\geq 0$. In particular, there is a scalar potential ϕ^{s} defined in Ω_{s} such that $\boldsymbol{u}^{\mathrm{s}}=\nabla_{\boldsymbol{x}}\phi^{\mathrm{s}}$; since moreover $\nabla_{\boldsymbol{x}}\cdot(H^{\mathrm{s}}\boldsymbol{u}^{\mathrm{s}})=0$ in Ω_{s} (this stems from the first equation in (6.1) and the fact that $\zeta=\zeta^{\mathrm{w}}$ does not depend on time), we deduce that the interior velocity $\boldsymbol{u}^{\mathrm{s}}$ can be deduced from the trace ψ^{s} of ϕ^{s} on $\partial\Omega_{\mathrm{fs}}$ by solving

$$\begin{cases}\nabla_{\boldsymbol{x}}\cdot(H^{\mathrm{s}}\nabla_{\boldsymbol{x}}\phi^{\mathrm{s}})=0 & \text{in } \Omega_{\mathrm{s}},\\ \phi^{\mathrm{s}}=\psi^{\mathrm{s}} & \text{on } \partial\Omega_{\mathrm{fs}},\end{cases}$$

where we recall that $H^{\mathrm{s}}=H_0+\zeta^{\mathrm{s}}=H_0+\zeta^{\mathrm{w}}$. It is therefore convenient to introduce the *Dirichlet-to-Neumann* (DtN) operator Λ classically defined as

$$\Lambda\psi^{\mathrm{s}}:=H^{\mathrm{s}}\nabla_{\boldsymbol{x}}\phi^{\mathrm{s}}\cdot\mathbf{n}\text{ on }\partial\Omega_{\mathrm{fs}},\tag{6.8}$$

and for irrotational data, the wave-structure interactions problem (6.1)-(6.6) can be recast as a mixed problem in the exterior domain Ω_{f}:

$$\begin{cases}\partial_t\zeta^{\mathrm{f}}+\nabla_{\boldsymbol{x}}\cdot(H^{\mathrm{f}}\boldsymbol{u}^{\mathrm{f}})=0,\\ \partial_t\boldsymbol{u}^{\mathrm{f}}+\boldsymbol{u}^{\mathrm{f}}\cdot\nabla_{\boldsymbol{x}}\boldsymbol{u}^{\mathrm{f}}+\mathrm{g}\nabla_{\boldsymbol{x}}\zeta^{\mathrm{f}}=0,\end{cases}\qquad\text{in }\Omega_{\mathrm{f}}\times(0,T),\tag{6.9}$$

with irrotationality condition

$$\nabla_{\boldsymbol{x}}^{\perp}\cdot\boldsymbol{u}^{\mathrm{f}}=0,\qquad\text{in }\Omega_{\mathrm{f}}\times(0,T),\tag{6.10}$$

and boundary conditions

$$\begin{cases}\mathbf{n}\cdot(H\boldsymbol{u})=\Lambda\psi^{\mathrm{s}} & \text{on } \partial\Omega_{\mathrm{fs}}\times(0,T),\\ \mathbf{n}^{\perp}\cdot\boldsymbol{u}^{\mathrm{f}}=\partial_{\tan}\psi^{\mathrm{s}} & \text{on } \partial\Omega_{\mathrm{fs}}\times(0,T),\end{cases}\tag{6.11}$$

where $\partial_{\tan}=\mathbf{n}^{\perp}\cdot\nabla_{\boldsymbol{x}}$ denotes the tangential derivative on $\partial\Omega_{\mathrm{fs}}$, and where ψ^{s} is defined on $(0,T)\times\partial\Omega_{\mathrm{fs}}$ and must solve the forced ODE

$$\partial_t\psi^{\mathrm{s}}=-\mathrm{g}\zeta-\frac{1}{2}|\boldsymbol{u}|^2\qquad\text{on }\partial\Omega_{\mathrm{fs}}\times(0,T).\tag{6.12}$$

Finally, we also impose initial conditions of the form

$$\begin{cases}(\zeta,\boldsymbol{u})_{|_{t=0}}=(\zeta^{\mathrm{in}},\boldsymbol{u}^{\mathrm{in}}) & \text{in } \Omega_{\mathrm{f}},\\ \psi^{\mathrm{s}}{}_{|_{t=0}}=\psi^{\mathrm{s}}_{\mathrm{in}} & \text{on } \partial\Omega_{\mathrm{fs}},\end{cases}\tag{6.13}$$

with of course $\nabla_{\boldsymbol{x}}^{\perp}\cdot\boldsymbol{u}^{\mathrm{in}}=0$ and $\mathbf{n}^{\perp}\cdot\boldsymbol{u}^{\mathrm{in}}=\partial_{\tan}\psi^{\mathrm{s}}_{\mathrm{in}}$.

Considering the previous equations, the wave-structure interactions problem can be reformulated, for irrotational data, as the following mixed initial boundary value problem:

$$\begin{cases}\partial_t\zeta^{\mathrm{f}}+\nabla_{\boldsymbol{x}}\cdot(H^{\mathrm{f}}\boldsymbol{u}^{\mathrm{f}})=0 & \text{in } \Omega_{\mathrm{f}}\times(0,T),\\ \partial_t\boldsymbol{u}^{\mathrm{f}}+\boldsymbol{u}^{\mathrm{f}}\cdot\nabla_{\boldsymbol{x}}\boldsymbol{u}^{\mathrm{f}}+\mathrm{g}\nabla_{\boldsymbol{x}}\zeta^{\mathrm{f}}=\mathbf{0} & \text{in } \Omega_{\mathrm{f}}\times(0,T),\\ \nabla_{\boldsymbol{x}}^{\perp}\cdot\boldsymbol{u}=0 & \text{in } \Omega_{\mathrm{f}}\times(0,T),\\ \mathbf{n}\cdot(H^{\mathrm{f}}\boldsymbol{u}^{\mathrm{f}})=\Lambda\psi^{\mathrm{s}} & \text{on } \partial\Omega_{\mathrm{fs}}\times(0,T),\\ \mathbf{n}^{\perp}\cdot\boldsymbol{u}=\partial_{\tan}\psi^{\mathrm{s}} & \text{on } \partial\Omega_{\mathrm{fs}}\times(0,T),\\ \partial_t\psi^{\mathrm{s}}=-\mathrm{g}\zeta-\frac{1}{2}|\boldsymbol{u}|^2 & \text{on } \partial\Omega_{\mathrm{fs}}\times(0,T),\\ (\zeta^{\mathrm{f}},\boldsymbol{u}^{\mathrm{f}})|_{t=0}=(\zeta^{\mathrm{in}},\boldsymbol{u}^{\mathrm{in}}),\quad \psi^{\mathrm{s}}|_{t=0}=\psi^{\mathrm{s}}_{\mathrm{in}} & \text{in } \Omega_{\mathrm{f}}\times\{0\}\text{ and on }\partial\Omega_{\mathrm{fs}}\times\{0\},\\ \nabla_{\boldsymbol{x}}^{\perp}\cdot\boldsymbol{u}^{\mathrm{in}}=0 & \text{in } \Omega_{\mathrm{f}}\times\{0\},\\ \mathbf{n}^{\perp}\cdot\boldsymbol{u}^{\mathrm{in}}=\partial_{\tan}\psi^{\mathrm{s}}_{\mathrm{in}} & \text{on } \partial\Omega_{\mathrm{fs}}\times\{0\}.\end{cases}\tag{6.14}$$

6.1.3 An equivalent formulation

Since $\boldsymbol{u}\cdot\nabla_{\boldsymbol{x}}\boldsymbol{u} = \nabla_{\boldsymbol{x}}(\frac{1}{2}|\boldsymbol{u}|^2)$ if $\nabla_{\boldsymbol{x}}^{\perp}\cdot\boldsymbol{u} = 0$, it is noted in [185] that the mixed problem 6.14 can be reformulated as

$$\begin{cases} \partial_t\zeta^{\mathrm{f}} + \nabla_{\boldsymbol{x}}\cdot(H^{\mathrm{f}}\boldsymbol{u}^{\mathrm{f}}) = 0, \\ \partial_t\boldsymbol{u}^{\mathrm{f}} + \nabla_{\boldsymbol{x}}\big(\mathrm{g}\zeta^{\mathrm{f}} + \frac{1}{2}|\boldsymbol{u}^{\mathrm{f}}|^2\big) = 0, \end{cases} \qquad \text{in } \Omega_{\mathrm{f}}\times(0,T), \tag{6.15}$$

with boundary conditions

$$\mathbf{n}\cdot(H\boldsymbol{u}) = \Lambda\psi^{\mathrm{s}} \quad \text{ on } \partial\Omega_{\mathrm{fs}}\times(0,T), \tag{6.16}$$

where ψ^{s} satisfies the forced ODE

$$\partial_t\psi^{\mathrm{s}} = -\mathrm{g}\zeta - \frac{1}{2}|\boldsymbol{u}|^2 \quad \text{ on } \partial\Omega_{\mathrm{fs}}\times(0,T), \tag{6.17}$$

and with the initial conditions

$$\begin{cases} (\zeta,\boldsymbol{u})_{|t=0} = (\zeta^{\mathrm{in}},\boldsymbol{u}^{\mathrm{in}}) & \text{in } \Omega_{\mathrm{f}}\times\{0\}, \\ \psi^{\mathrm{s}}{}_{|t=0} = \psi^{\mathrm{s}}_{\mathrm{in}} & \text{on } \partial\Omega_{\mathrm{fs}}\times\{0\}, \end{cases} \tag{6.18}$$

where $\nabla_{\boldsymbol{x}}^{\perp}\cdot\boldsymbol{u}^{\mathrm{in}} = 0$ and $\mathbf{n}^{\perp}\cdot\boldsymbol{u}^{\mathrm{in}} = \partial_{\tan}\psi^{\mathrm{s}}_{\mathrm{in}}$. This formulation is simpler than (6.9)-(6.13) since the irrotationality condition and the second boundary condition in (6.11) have been removed, as they are automatically propagated from the initial conditions.

Let us remark that in Appendix C, we extend the previous formulation to the case of a moving floating structure, in the spirit of what has been done in Chapter 5 and the pioneering work of Bocchi et al. [34, 35].

6.1.4 A special configuration: the infinite pontoon

We consider in this subsection a specific configuration that does not fit in the previous framework. Here, the partially immersed object is an infinite pontoon with rectangular cross section:

$$\mathbf{S} = \{(x,y,z)\in\mathbb{R}^3 \mid -l<x<l,\ \zeta^{\mathrm{w}}<z<\zeta_{\mathrm{top}}\},$$

where $2l>0$ is the length of the object and $\zeta^{\mathrm{w}}<0$ is a constant that denotes the position of the base of the pontoon, and $\zeta_{\mathrm{top}}-\zeta^{\mathrm{w}}$ is its height.

In this case, the interior region Ω_{s} is an infinite strip, while the exterior region Ω_{f} consists of two half-planes and $\partial\Omega_{\mathrm{fs}}$ is the union of two straight lines,

$$\begin{cases} \Omega_{\mathrm{s}} = (-l,l)\times\mathbb{R}, \\ \Omega_{\mathrm{f}} = \Omega_{\mathrm{f}-}\cup\Omega_{\mathrm{f}+} \quad \text{with} \quad \Omega_{\mathrm{f}\pm} = \{(x,y)\in\mathbb{R}^2 \mid \pm x>l\} \\ \partial\Omega_{\mathrm{fs}} = \partial\Omega_{\mathrm{fs}-}\cup\partial\Omega_{\mathrm{fs}+} \quad \text{with} \quad \partial\Omega_{\mathrm{fs}\pm} = \{\pm l\}\times\mathbb{R}. \end{cases}$$

This configuration is interesting for several reasons: (*i*) following the methodology established by Lannes and Bocchi for floating structures in the shallow-water regime, the interaction under the pontoon is formulated as a constrained fluid problem where the hydrodynamic pressure acts as a Lagrange multiplier, leading to the definition of an analytical generalized DtN operator, (*ii*) while Bocchi explicitly expressed the DtN operator in the axisymmetric case using a Fourier-Laplace argument leading to ratios of Hankel functions, we develop here an analogous frequency-domain formulation for an infinite rectangular geometry, where the translation invariance along the y-axis yields an interior DtN operator characterized by classical $\coth(2lD)$ and $\sinh(2lD)^{-1}$ Fourier multipliers, (*iii*) if we consider waves that do not depend on the transverse variable, one recovers a 1D wave-structure interactions problem that can therefore be used for the validation of the numerical code.

We first define a generalized DtN operator associated with this geometry, and provide its explicit expression. Then we reformulate the problem as a transmission problem and show that it degenerates to the expected 1D model when the data do not depend on the transverse variable.

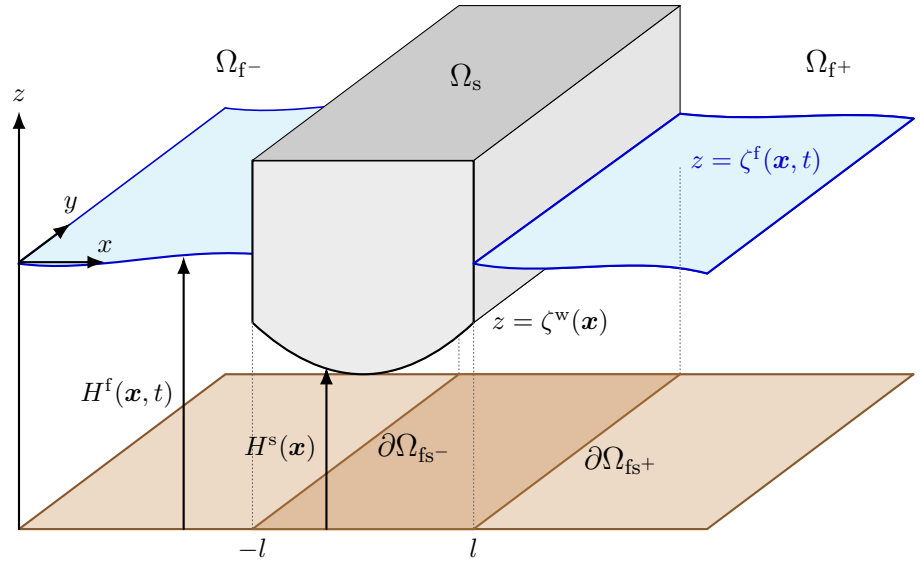


Figure 6.2: Geometry of a partially immersed cylindrical floating structure in a shallow-water flow. The structure is an infinite pontoon with rectangular cross section, and the fluid domain is the union of two half-planes.

A generalized DtN operator

We define here a generalized DtN operator associated with the specific configuration considered here. For sufficiently regular functions ψ_- and ψ_+ defined on $\partial\Omega_{\mathrm{fs}-}$ and $\partial\Omega_{\mathrm{fs}+}$ respectively, we denote by ϕ^{s} the solution of the elliptic problem

$$\begin{cases} \nabla_{\boldsymbol{x}} \cdot (H^{\mathrm{s}} \nabla_{\boldsymbol{x}} \phi^{\mathrm{s}}) = 0 & \text{in } \Omega_{\mathrm{s}}, \\ \phi^{\mathrm{s}} = \psi_{\pm} & \text{on } \partial\Omega_{\mathrm{fs}\pm}, \end{cases} \tag{6.19}$$

with $H^{\mathrm{s}} = H_0 + \zeta^{\mathrm{w}} = H_0 + \zeta^{\mathrm{s}}$. Denoting $\boldsymbol{\psi}^{\mathrm{s}} = (\psi_-, \psi_+)$, we define the generalized DtN operator as

$$\boldsymbol{\Lambda}\boldsymbol{\psi}^{\mathrm{s}} = (\Lambda_-\boldsymbol{\psi}^{\mathrm{s}}, \Lambda_+\boldsymbol{\psi}^{\mathrm{s}}) \quad \text{where} \quad \Lambda_\pm\boldsymbol{\psi}^{\mathrm{s}} = \pm H^{\mathrm{s}}(\partial_x\phi^{\mathrm{s}})_{|_{x=\pm l}}. \tag{6.20}$$

The configuration of the infinite pontoon is interesting because the operators $\Lambda_\pm$ can be computed explicitly.

Let $\mathfrak{F}_y(u) = \mathfrak{F}(y \mapsto u(x,y))$ denote the Fourier transform with respect to the variable y, using the same convention as above. For any function $u \in \mathcal{S}(\mathbb{R})$, we write

$$(\mathfrak{F}_y u)(\xi) = \widehat{u}(\xi) = \int_{\mathbb{R}} e^{-\mathrm{i}y\xi}\, u(y)\, \mathrm{d}y.$$

We define the first-order pseudo-differential operator $D = |D_y|$ as the Fourier multiplier operator associated with the symbol $m(\xi) = |\xi|$. In other words, for any $u \in \mathsf{H}^1(\mathbb{R})$, we have

$$\mathfrak{F}_y(Du)(\xi) = |\xi|\, \widehat{u}(\xi).$$

More generally, if $m : \mathbb{R} \to \mathbb{C}$ is a Fourier symbol, we use the notation

$$(m(D)u)(y) := \mathfrak{F}_y^{-1}\left(m(|\xi|)\, \widehat{u}(\xi)\right)(y),$$

whenever the right-hand side is well defined. In particular, the operators appearing below are understood through the symbols

$$D\coth(2lD) \text{ has symbol } |\xi|\coth(2l|\xi|), \qquad \frac{D}{\sinh(2lD)} \text{ has symbol } \frac{|\xi|}{\sinh(2l|\xi|)}.$$

Both symbols are extended continuously at $\xi = 0$, with value $(2l)^{-1}$. With this convention, D acts as a bounded linear mapping from $\mathsf{H}^s(\mathbb{R})$ to $\mathsf{H}^{s-1}(\mathbb{R})$ for any $s \in \mathbb{R}$.

We can now provide a simple explicit expression for $\Lambda_\pm$ in terms of the Fourier multiplier of the transverse variable y. In the statement below, functions on $\partial\Omega_{\mathrm{fs}\pm}$ are canonically understood as functions on $\mathbb{R}$.

Proposition 6.1.1. *For $\boldsymbol{\psi}^{\mathrm{s}} = (\psi_-, \psi_+)$ with sufficiently regular functions $\psi_\pm$ defined on $\partial\Omega_{\mathrm{fs}^\pm}$, the generalized DtN operator defined in (6.20) satisfies*

$$\Lambda_- \boldsymbol{\psi}^{\mathrm{s}} = H^{\mathrm{s}} D \coth(2lD)\psi_- - \frac{H^{\mathrm{s}} D}{\sinh(2lD)}\psi_+,$$
$$\Lambda_+ \boldsymbol{\psi}^{\mathrm{s}} = -\frac{H^{\mathrm{s}} D}{\sinh(2lD)}\psi_- + H^{\mathrm{s}} D \coth(2lD)\psi_+.$$

Proof. Taking the Fourier transform of the elliptic equation (6.19) with respect to the variable y, one obtains that the Fourier transform $\widehat{\phi}^{\mathrm{s}}(x,\cdot) = \mathfrak{F}_y(\phi^{\mathrm{s}}(x,y))$ solves the ODE

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\widehat{\phi}^{\mathrm{s}}(x,\xi) - |\xi|^2 \widehat{\phi}^{\mathrm{s}}(x,\xi) = 0$$

with boundary conditions $\widehat{\phi}^{\mathrm{s}}(\pm l,\xi) = \widehat{\psi}_\pm(\xi)$. This can be solved explicitly as

$$\widehat{\phi}^{\mathrm{s}}(x,\xi) = -\frac{\sinh((x-l)\xi)}{\sinh(2l\xi)}\widehat{\psi}_-(\xi) + \frac{\sinh((x+l)\xi)}{\sinh(2l\xi)}\widehat{\psi}_+(\xi).$$

Therefore

$$\partial_x\widehat{\phi}^{\mathrm{s}}(x,\xi) = -\xi\frac{\cosh((x-l)\xi)}{\sinh(2l\xi)}\widehat{\psi}_-(\xi) + \xi\frac{\cosh((x+l)\xi)}{\sinh(2l\xi)}\widehat{\psi}_+(\xi),$$

from which we get

$$-\partial_x\widehat{\phi}^{\mathrm{s}}(-l,\xi) = \xi\coth(2l\xi)\widehat{\psi}_-(\xi) - \xi\frac{1}{\sinh(2l\xi)}\widehat{\psi}_+(\xi),$$
$$\partial_x\widehat{\phi}^{\mathrm{s}}(l,\xi) = -\xi\frac{1}{\sinh(2l\xi)}\widehat{\psi}_-(\xi) + \xi\coth(2l\xi)\widehat{\psi}_+(\xi).$$

By definition of the generalized DtN operator, we have

$$\Lambda_-\boldsymbol{\psi}^{\mathrm{s}}(y) = H^{\mathrm{s}}\big(-\partial_x\phi^{\mathrm{s}}(-l,y)\big), \qquad \Lambda_+\boldsymbol{\psi}^{\mathrm{s}}(y) = H^{\mathrm{s}}\big(\partial_x\phi^{\mathrm{s}}(l,y)\big),$$

so taking the inverse Fourier transform of the previous expressions yields, for all $y\in\mathbb{R}$,

$$\Lambda_-\boldsymbol{\psi}^{\mathrm{s}}(y) = H^{\mathrm{s}}\,\mathfrak{F}^{-1}\left(\xi\coth(2l\xi)\widehat{\psi}_-(\xi)\right)(y) - H^{\mathrm{s}}\,\mathfrak{F}^{-1}\left(\xi\frac{1}{\sinh(2l\xi)}\widehat{\psi}_+(\xi)\right)(y),$$
$$\Lambda_+\boldsymbol{\psi}^{\mathrm{s}}(y) = H^{\mathrm{s}}\,\mathfrak{F}^{-1}\left(-\xi\frac{1}{\sinh(2l\xi)}\widehat{\psi}_-(\xi)\right)(y) + H^{\mathrm{s}}\,\mathfrak{F}^{-1}\left(\xi\coth(2l\xi)\widehat{\psi}_+(\xi)\right)(y).$$

Using now the notation $(m(D)f)(y) := \mathfrak{F}_y^{-1}\big(m(|\xi|)(\mathfrak{F}_y f)(\xi)\big)(y)$, we have

$$D\coth(2lD)\psi_\pm = \mathfrak{F}_y^{-1}\left(|\xi|\coth(2l|\xi|)\widehat{\psi}_\pm(\xi)\right),$$
$$\frac{D}{\sinh(2lD)}\psi_\pm = \mathfrak{F}_y^{-1}\left(\frac{|\xi|}{\sinh(2l|\xi|)}\widehat{\psi}_\pm(\xi)\right).$$

We therefore deduce

$$\Lambda_-\boldsymbol{\psi}^{\mathrm{s}} = H^{\mathrm{s}} D\coth(2lD)\psi_- - \frac{H^{\mathrm{s}} D}{\sinh(2lD)}\psi_+,$$
$$\Lambda_+\boldsymbol{\psi}^{\mathrm{s}} = -\frac{H^{\mathrm{s}} D}{\sinh(2lD)}\psi_- + H^{\mathrm{s}} D\coth(2lD)\psi_+,$$

which ends the proof. □

Remark 6.1.2 (*Continuity of the DtN operators*). *The complete symbols $a(\xi) = |\xi|\coth(2l|\xi|)$ and $b(\xi) = |\xi|/\sinh(2l|\xi|)$, with $a(0) = b(0) = (2l)^{-1}$, are smooth and even. The first is of order one and the second is rapidly decreasing at infinity; consequently, $\boldsymbol{\Lambda}$ is continuous from $\mathsf{H}^s(\mathbb{R})^2$ to $\mathsf{H}^{s-1}(\mathbb{R})^2$.*

Reduction to a transmission problem

Since the exterior region has now two connected components, the reductions corresponding to the contents of Sections 6.1.2 and 6.1.3 now lead to a transmission problem rather than an initial boundary value problem:

$$\begin{cases} \partial_t \zeta^{\mathrm{f}} + \nabla_{\boldsymbol{x}} \cdot (H^{\mathrm{f}} \boldsymbol{u}^{\mathrm{f}}) = 0, \\ \partial_t \boldsymbol{u}^{\mathrm{f}} + \nabla_{\boldsymbol{x}} \big(\mathrm{g} \zeta^{\mathrm{f}} + \frac{1}{2} |\boldsymbol{u}^{\mathrm{f}}|^2 \big) = 0, \end{cases} \qquad \text{in } \Omega_{\mathrm{f}^\pm} \times (0,T), \tag{6.21}$$

with transmission conditions

$$\pm \mathbf{e}_x \cdot (H \boldsymbol{u}) = \Lambda_\pm \boldsymbol{\psi}^{\mathrm{s}} \quad \text{ on } \partial\Omega_{\mathrm{fs}} \times (0,T), \tag{6.22}$$

where $\Lambda_\pm$ is as in Proposition 6.1.1 and $\boldsymbol{\psi}^{\mathrm{s}} = (\psi_-, \psi_+)$ must solve

$$\partial_t \psi_\pm = -\mathrm{g}\zeta - \frac{1}{2}|\boldsymbol{u}|^2 \quad \text{ on } \partial\Omega_{\mathrm{fs}^\pm} \times (0,T), \tag{6.23}$$

and with the initial conditions

$$\begin{cases} (\zeta, \boldsymbol{u})_{|t=0} = (\zeta^{\mathrm{in}}, \boldsymbol{u}^{\mathrm{in}}) & \text{in } \Omega_{\mathrm{f}}, \\ \boldsymbol{\psi}^{\mathrm{s}}{}_{|t=0} = \boldsymbol{\psi}^{\mathrm{s}}_{\mathrm{in}} = (\psi^{\mathrm{in}}_-, \psi^{\mathrm{in}}_+) & \text{on } \partial\Omega_{\mathrm{fs}}, \end{cases} \tag{6.24}$$

where $\nabla^\perp_{\boldsymbol{x}} \cdot \boldsymbol{u}^{\mathrm{in}} = 0$ and $\mathbf{e}_y \cdot \boldsymbol{u}^{\mathrm{in}} = \partial_y \psi^{\mathrm{in}}_\pm$ on $\partial\Omega_{\mathrm{fs}^\pm}$.

The problem now takes the form:

$$\begin{cases} \partial_t \zeta^{\mathrm{f}} + \nabla_{\boldsymbol{x}} \cdot (H^{\mathrm{f}} \boldsymbol{u}^{\mathrm{f}}) = 0 & \text{in } \Omega_{\mathrm{f}^\pm} \times (0,T), \\ \partial_t \boldsymbol{u}^{\mathrm{f}} + \nabla_{\boldsymbol{x}} \big(\mathrm{g} \zeta^{\mathrm{f}} + \frac{1}{2} |\boldsymbol{u}^{\mathrm{f}}|^2 \big) = 0, & \text{in } \Omega_{\mathrm{f}^\pm} \times (0,T), \\ \pm \mathbf{e}_x \cdot (H \boldsymbol{u}) = \Lambda_\pm \boldsymbol{\psi}^{\mathrm{s}} & \text{on } \partial\Omega_{\mathrm{fs}} \times (0,T), \\ \partial_t \psi_\pm = -\mathrm{g}\zeta - \frac{1}{2}|\boldsymbol{u}|^2 & \text{on } \partial\Omega_{\mathrm{fs}^\pm} \times (0,T), \\ (\zeta^{\mathrm{f}}, \boldsymbol{u}^{\mathrm{f}})|_{t=0} = (\zeta^{\mathrm{in}}, \boldsymbol{u}^{\mathrm{in}}) & \text{on } \Omega_{\mathrm{f}} \times \{0\}, \\ \boldsymbol{\psi}^{\mathrm{s}}|_{t=0} = (\psi^{\mathrm{in}}_-, \psi^{\mathrm{in}}_+) & \text{on } \partial\Omega_{\mathrm{fs}} \times \{0\}, \\ \nabla^\perp_{\boldsymbol{x}} \cdot \boldsymbol{u}^{\mathrm{in}} = 0, \qquad \mathbf{e}_y \cdot \boldsymbol{u}^{\mathrm{in}} = \partial_y \psi^{\mathrm{in}}_\pm & \text{on } \Omega \times \{0\} \text{ and } \partial\Omega_{\mathrm{fs}^\pm} \times \{0\}. \end{cases} \tag{6.25}$$

Consistency with the one-dimensional case

We check here that the model (6.21)-(6.23) coincides here with the one-dimensional model considered [35, 182] when the waves do not depend on the transverse variable y.

Recall that if $\mathbf{1}$ denotes the constant function equal to 1, and m is continuous, then one has $m(D)\mathbf{1} = m(0)$. Therefore, if $\boldsymbol{\psi}^{\mathrm{s}} = (\psi_-, \psi_+)$ is a constant function, one has by Proposition 6.1.1,

$$\Lambda_- \boldsymbol{\psi}^{\mathrm{s}} = -\frac{H^{\mathrm{s}}}{2l}(\psi_+ - \psi_-) \quad \text{ and } \quad \Lambda_+ \boldsymbol{\psi}^{\mathrm{s}} = \frac{H^{\mathrm{s}}}{2l}(\psi_+ - \psi_-). \tag{6.26}$$

On the other hand, if ζ and $\boldsymbol{u} = (u, v)^\top$ do not depend on y then (6.21)-(6.23) reduce to a one dimensional hyperbolic transmission problem on $(\zeta, \boldsymbol{u})$, or equivalently on (ζ, q), with $q = Hu$, namely,

$$\begin{cases} \partial_t \zeta + \partial_x q = 0, \\ \partial_t q + \partial_x (\mathrm{g}\frac{h^2}{2} + \frac{1}{h} q^2) = 0, \end{cases} \qquad \text{in} \quad (0,T) \times \{\pm x > l\}, \tag{6.27}$$

with transmission conditions

$$q(t, \pm l) = q^{\mathrm{s}}(t) \quad \text{ with } \quad q^{\mathrm{s}}(t) = \frac{H^{\mathrm{s}}}{2l}(\psi_+ - \psi_-) \quad \text{ on } \quad (0,T), \tag{6.28}$$

while, by (6.23), the function of time q^{s} solves

$$\frac{\mathrm{d}q^{\mathrm{s}}}{\mathrm{d}t} = -\frac{H^{\mathrm{s}}}{2l}(\Pi_+ - \Pi_-) \quad \text{ where } \quad \Pi_\pm = \left(\mathrm{g}\zeta + \frac{1}{2}|v|^2 \right)_{|x=\pm l}. \tag{6.29}$$

The system of equations (6.27)-(6.29) coincides with the 1D wave-structure interactions problem considered in [35, 182].

Although the reduction to the one-dimensional case is obtained by assuming invariance with respect to the transverse variable y, the underlying model remains intrinsically two-dimensional. This clearly differs from the pseudo-one-dimensional setting introduced in the Chapter 5, where the equations themselves were purely one-dimensional and only embedded into a two-dimensional setting for numerical purposes.

Remark 6.1.3 (*DtN operator and longitudinal heterogeneities*). *When the object's bottom is uniform in the y-direction, the generalized DtN operator acts as a pure Fourier multiplier with explicit formulations due to translation invariance. Conversely, introducing longitudinal heterogeneities breaks this invariance, transforming the DtN operator into a true pseudo-differential operator with a space-dependent symbol that requires microlocal asymptotic expansions or numerical schemes. Even though the pontoon's motion is restricted to a single degree of freedom (in the case of pure heaving, see Chapter 5) without longitudinal heterogeneities, this test is interesting to validate 2D processes, by considering for instance oblique incident waves. The longitudinal wave number ξ dictates the phase distribution of the hydrodynamic pressure along the y-axis. This induces a spatial cancellation effect on the global excitation force and modifies the fluid's evacuation paths underneath the structure, which should be accurately captured by the ξ-dependence of the generalized DtN operator. This will be further explored in a near future.*

Low-frequency consistency for a transversely varying depth

To exhibit the consistency from the two-dimensional configuration to the purely 1D framework under a transversely varying bathymetry, we may also analyze the asymptotic behavior of the generalized DtN operator in the low-frequency limit, as this reduction corresponds physically to the case where the hydrodynamic fields exhibit no longitudinal variations along the infinite y-axis.

In this framework, we suppose a non-uniform bathymetry that varies across the pontoon width such that $H^{\mathrm{s}} = H^{\mathrm{s}}(x)$ for $x \in (-l, l)$, while remaining uniform along the y-direction. The interior potential $\psi^{\mathrm{s}}(x, y)$ satisfies the elliptic variable-coefficient equation:

$$\partial_x\big(H^{\mathrm{s}}(x)\partial_x\psi^{\mathrm{s}}\big) + H^{\mathrm{s}}(x)\partial_y^2\psi^{\mathrm{s}} = 0 \quad \text{in } (-l, l)\times\mathbb{R}$$

Applying the partial Fourier transform with respect to y, the equation is recast in the frequency domain as a Sturm-Liouville problem parameterized by the longitudinal wavenumber ξ:

$$\partial_x\big(H^{\mathrm{s}}(x)\partial_x\widehat{\psi}^{\mathrm{s}}\big) - \xi^2 H^{\mathrm{s}}(x)\widehat{\psi}^{\mathrm{s}} = 0,$$

subject to the non-homogeneous Dirichlet boundary conditions $\widehat{\psi}^{\mathrm{s}}(-l,\xi) = \widehat{\varrho}_-(\xi)$ and $\widehat{\psi}^{\mathrm{s}}(l,\xi) = \widehat{\varrho}_+(\xi)$. The purely one-dimensional regime represents the spatial limit where the solution is invariant along the longitudinal axis, implying $\partial_y\psi^{\mathrm{s}} = 0$, which restricts the analysis to the zero-wavenumber mode $\xi \to 0$. In this asymptotic limit, the governing differential equation simplifies directly to:

$$\partial_x\big(H^{\mathrm{s}}(x)\partial_x\widehat{\psi}^{\mathrm{s}}\big) = \mathcal{O}(\xi^2).$$

Integrating this relationship once yields a spatially constant horizontal flux profile across the domain

$$H^{\mathrm{s}}(x)\partial_x\widehat{\psi}^{\mathrm{s}}(x) = \mathsf{c}_0 + \mathcal{O}(\xi^2),$$

isolating the gradient as $\partial_x\widehat{\psi}^{\mathrm{s}}(x) = \mathsf{c}_0/H^{\mathrm{s}}(x)$ and integrating over the width interval from $-l$ to l, we enforce the boundary values to determine the explicit value of the constant:

$$\widehat{\psi}^{\mathrm{s}}(l) - \widehat{\psi}^{\mathrm{s}}(-l) = \widehat{\varrho}_+ - \widehat{\varrho}_- = \mathsf{c}_0\int_{-l}^{l}\frac{\mathrm{d}t}{H^{\mathrm{s}}(t)}.$$

Solving for c_0 establishes the direct link between the boundary data differential and the internal flux:

$$\mathsf{c}_0 = \left(\int_{-l}^{l}\frac{\mathrm{d}t}{H^{\mathrm{s}}(t)}\right)^{-1}(\widehat{\varrho}_+ - \widehat{\varrho}_-).$$

By evaluating the definition of the outgoing normal fluxes at the respective boundaries in the limit $\xi \to 0$, we find that $\widehat{\mathcal{N}}^- = -\mathsf{c}_0$ and $\widehat{\mathcal{N}}^+ = \mathsf{c}_0$. Gathering these relations yields the low-frequency matrix representation in Fourier space:

$$\begin{pmatrix} \widehat{\mathcal{N}}^- \\ \widehat{\mathcal{N}}^+ \end{pmatrix} = \left(\int_{-l}^{l} \frac{\mathrm{d}t}{H^{\mathrm{s}}(t)} \right)^{-1} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \widehat{\varrho}_- \\ \widehat{\varrho}_+ \end{pmatrix} + \mathcal{O}(\xi^2).$$

For the zero-frequency mode, equivalently for fields that are invariant with respect to y, this relation reduces in physical variables to

$$\begin{pmatrix} \mathcal{N}^- \\ \mathcal{N}^+ \end{pmatrix} = \left(\int_{-l}^{l} \frac{\mathrm{d}t}{H^{\mathrm{s}}(t)} \right)^{-1} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \varrho_- \\ \varrho_+ \end{pmatrix}.$$

This demonstrates that the low-frequency limit of the 2D DtN formulation automatically condenses the spatial heterogeneity into a single global value matching the integrated compliance of the fluid layer, as in the 1D case.

In other terms, the zero-frequency mode does not see the detailed two-dimensional structure of the flow along the y-direction. It only measures the effective resistance encountered by the fluid when it crosses the region below the pontoon from $x = -l$ to $x = l$. This resistance is encoded by the quantity $\int_{-l}^{l} \frac{\mathrm{d}t}{H^{\mathrm{s}}(t)}$, hence, thinner regions, where H^{s} is small, contribute more strongly to the global resistance, while deeper regions contribute less. In the constant-depth case $H^{\mathrm{s}}(x) = H^{\mathrm{s}}$, this formula reduces to

$$\left(\int_{-l}^{l} \frac{\mathrm{d}t}{H^{\mathrm{s}}} \right)^{-1} = \frac{H^{\mathrm{s}}}{2l},$$

and we recover the coefficient appearing in (6.26). Therefore, the low-frequency limit of the generalized two-dimensional DtN operator is consistent with the one-dimensional transmission law.

Remark 6.1.4 (*Transverse filtering induced by the generalized DtN operator*). *The previous low-frequency consistency is also consistent with the explicit Fourier representation of Proposition 6.1.1. In the constant-depth case, the off-diagonal terms of the generalized DtN operator, which describe the influence of one side of the pontoon on the other one, are associated with the Fourier multiplier* $-H^{\mathrm{s}} \frac{|\xi|}{\sinh(2l|\xi|)}$. *This multiplier has a nonzero low-frequency limit, since*

$$\lim_{|\xi| \to 0} \frac{|\xi|}{\sinh(2l|\xi|)} = \frac{1}{2l},$$

whereas its high-frequency behaviour is given by

$$\frac{|\xi|}{\sinh(2l|\xi|)} \underset{|\xi| \to +\infty}{\sim} 2|\xi| e^{-2l|\xi|}.$$

Hence, the transmission through the pontoon behaves as a low-pass filter in the transverse direction. The mean mode in y is transmitted and gives the one-dimensional coefficient $H^{\mathrm{s}}/2l$, while oscillatory transverse modes are exponentially damped across the interior region below the structure. In particular, perturbations with a nonzero transverse average may generate a response close to a constant mode in y on the opposite side, whereas zero-mean perturbations mainly excite nonzero transverse modes and are therefore much more strongly attenuated.

6.2 Discrete formulations

In this section, we present the discrete approximation framework for the coupled fluid–structure interaction problem.

6.2.1 Approximation framework

General notations

Each subdomain $\Omega_\bullet$ with $\bullet \in \{\mathrm{f}^-, \mathrm{f}^+, \mathrm{f}, \mathrm{s}\}$ is endowed with a conforming mesh $\mathcal{T}_h^\bullet$ of $N_{\mathrm{el}}^\bullet$ open cells $\omega_c^\bullet$, $c \in [\![1, N_{\mathrm{el}}^\bullet]\!]$, such that

$$\overline{\Omega_\bullet} = \bigcup_{\omega_c^\bullet \in \mathcal{T}_h^\bullet} \overline{\omega_c^\bullet}, \qquad \omega_c^\bullet \cap \omega_v^\bullet = \emptyset, \quad \forall c \neq v.$$

For each element $\omega_c^\bullet$, we denote by $h_c^\bullet$ its diameter, and $h^\bullet = \max_c h_c^\bullet$; the barycenter is $\boldsymbol{x}_{\mathrm{b}}^{c,\bullet} = (x_{\mathrm{b}}^{c,\bullet}, y_{\mathrm{b}}^{c,\bullet})$, and the cell extents in the x- and y-directions are $h_x^{c,\bullet}$ and $h_y^{c,\bullet}$, respectively.
The definition of mesh faces plays a distinct role in each subdomain:

- On the *fluid* part(s) $\mathcal{T}_h^{\mathrm{f}}$ (or $\mathcal{T}_h^{\mathrm{f}^\pm}$), faces serve as interfaces for the construction of the numerical fluxes in the DG/FV subcell scheme.
- On the *structure* mesh $\mathcal{T}_h^{\mathrm{s}}$, the face set $\mathcal{F}_h^{\mathrm{s}}$ is central to the definition of the Hybrid High-Order (HHO) unknowns, which consist of both cells and faces polynomial degrees of freedom.

For each subdomain $\Omega_\bullet$, the set of all mesh faces is denoted by $\mathcal{F}_h^\bullet$, and is partitioned as $\mathcal{F}_h^\bullet = \mathcal{F}_h^{\bullet,0} \cup \mathcal{F}_h^{\bullet,\partial}$, where $\mathcal{F}_h^{\bullet,0}$ collects the internal faces (i.e. there exist $\omega_c^\bullet, \omega_v^\bullet \in \mathcal{T}_h^\bullet$ with $c \neq v$ such that $\Gamma \subset \partial\omega_c^\bullet \cap \partial\omega_v^\bullet$), and $\mathcal{F}_h^{\bullet,\partial}$ gathers the boundary faces (i.e. there exists $\omega_c^\bullet \in \mathcal{T}_h^\bullet$ such that $\Gamma \subset \partial\omega_c^\bullet \cap \partial\Omega_\bullet$). The length of a face $\Gamma \in \mathcal{F}_h^\bullet$ is denoted by $l_\Gamma^\bullet$, its barycenter by $\boldsymbol{x}_{\mathrm{b}}^{\Gamma,\bullet}$, and we set $h_\Gamma^\bullet := \min_{\omega_c^\bullet \supset \Gamma} h_c^\bullet$.

For each cell $\omega_c^\bullet \in \mathcal{T}_h^\bullet$, we denote by $\mathcal{F}_c^\bullet := \{\, \Gamma \in \mathcal{F}_h^\bullet \mid \Gamma \subset \partial\omega_c^\bullet \,\}$ the set of its faces. In the numerical tests, we use triangulations so that $\#\mathcal{F}_c^\bullet = 3$ for all cells, but the construction naturally extends to polygonal elements, a feature particularly relevant for the HHO discretization.

For any $\Gamma \in \mathcal{F}_c^\bullet$, we denote by $\mathbf{n}_{c\Gamma}^\bullet$ the unit normal to Γ pointing outward from $\omega_c^\bullet$. For internal faces $\Gamma \in \mathcal{F}_h^{\bullet,0}$, an arbitrary but globally consistent orientation is fixed, and we set $\mathbf{n}_\Gamma^\bullet := \mathbf{n}_{c\Gamma}^\bullet$ for the element $\omega_c^\bullet$ chosen as the reference cell. For boundary faces $\Gamma \in \mathcal{F}_h^{\bullet,\partial}$, we simply take $\mathbf{n}_\Gamma^\bullet := \mathbf{n}_{c\Gamma}^\bullet$. If there is no ambiguity, we denote by $\mathbf{n}_{cv(k)}^\bullet := \mathbf{n}_{\Gamma_{cv(k)}^\bullet}^\bullet$ the normal vector associated with the k-th face of $\omega_c^\bullet$ shared with its neighbor $\omega_{v(k)}^\bullet$, for $k \in [\![1, \#\mathcal{F}_c^\bullet]\!]$.

Remark 6.2.1 (*Orientation of face normals*). *Each subdomain $\Omega_\bullet$ possesses its own orientation map for the face normals $\mathbf{n}_\Gamma^\bullet$. This ensures that internal fluxes and reconstruction operators are defined consistently within each mesh. At the fluid–structure interface Γ_{fs}, the normals of $\mathcal{F}_h^{\mathrm{f}}$ and $\mathcal{F}_h^{\mathrm{s}}$ are of opposite signs, i.e. $\mathbf{n}_\Gamma^{\mathrm{f}} = -\mathbf{n}_\Gamma^{\mathrm{s}}$, so that coupling conditions can be written in a unified way by combining traces from both sides. This convention will be used systematically in the subsequent formulations.*

For each cell $\omega_c^\bullet \in \mathcal{T}_h^\bullet$, we denote by $\pi_c^{k,\bullet}$ the local L^2-orthogonal projector onto $\mathbb{P}^k(\omega_c^\bullet)$, and by $\pi_h^{k,\bullet}$ the global projector onto $\mathbb{P}^k(\mathcal{T}_h^\bullet)$ obtained by gathering all the local projectors $\pi_c^{k,\bullet}$. The existence and uniqueness of $\pi_c^{k,\bullet}$ follow directly from the Riesz representation theorem in $\mathbb{P}^k(\omega_c^\bullet)$ for the standard $L^2(\omega_c^\bullet)$ inner product.

For any subdomain $\Omega_\bullet$ with mesh $\mathcal{T}_h^\bullet$ and associated face set $\mathcal{F}_h^\bullet$, we introduce the following shorthands for the discrete (broken) scalar products for $(v, w) \in L^2(\mathcal{T}_h)^2$ and $(\mu, \nu) \in L^2(\partial\mathcal{T}_h)^2$:

$$
\begin{aligned}
(v, w)_{\mathcal{T}_h^\bullet} &:= \sum_{\omega_c^\bullet \in \mathcal{T}_h^\bullet} (v, w)_{\omega_c^\bullet} = \sum_{\omega_c^\bullet \in \mathcal{T}_h^\bullet} \int_{\omega_c^\bullet} v(\boldsymbol{x}, t)\, w(\boldsymbol{x}, t)\, \mathrm{d}\boldsymbol{x}, \\
\big\langle \mu, \nu \big\rangle_{\partial\mathcal{T}_h^\bullet} &:= \sum_{\omega_c^\bullet \in \mathcal{T}_h^\bullet} \big\langle \mu, \nu \big\rangle_{\partial\omega_c^\bullet} = \sum_{\omega_c^\bullet \in \mathcal{T}_h^\bullet} \int_{\partial\omega_c^\bullet} \mu(\boldsymbol{x}, t)\, \nu(\boldsymbol{x}, t)\, \mathrm{d}s, \\
\big\langle \mu, \nu \big\rangle_{\mathcal{F}_h^\bullet} &:= \sum_{\omega_c^\bullet \in \mathcal{T}_h^\bullet} \big\langle \mu, \nu \big\rangle_{\mathcal{F}_c^\bullet} = \sum_{\omega_c^\bullet \in \mathcal{T}_h^\bullet} \sum_{\Gamma \in \mathcal{F}_c^\bullet} \int_\Gamma \mu(\boldsymbol{x}, t)\, \nu(\boldsymbol{x}, t)\ \mathrm{d}s.
\end{aligned}
$$

The extension to vector-valued functions are straightforward, with componentwise or dot-product integration. At the global level, when the distinction between subdomains is unnecessary, we will simply write $(\cdot,\cdot)_{\mathcal{T}_h}$ and $\langle \cdot, \cdot \rangle_{\mathcal{F}_h}$, understanding the sums as taken over all elements and faces of the full domain Ω.

Remark 6.2.2 (*Discretization strategy and related notations*). *A HHO discretization is developed in the solid region mesh $\mathcal{T}_h^{\mathrm{s}}$, corresponding to the interior region Ω_{s}, whereas the surrounding fluid mesh $\mathcal{T}_h^{\mathrm{f}}$ (resp. meshes $\mathcal{T}_h^{\mathrm{f}^\pm}$ in the pontoon configuration) is (resp. are) handled with the monolithic DG/FV subcell method detailed in Chapter 3. Each numerical framework is presented independently in the forthcoming subsections, together with the corresponding local notations and functional settings.*

Mesh-dependent functional spaces for the DG/FV subcell method

We now adapt the notations of Chapter 3 for the discrete spaces associated with the FV–DG subcell discretization in the exterior fluid regions Ω_{f} or $\Omega_{\mathrm{f}^\pm} := \Omega_{\mathrm{f}^+} \cup \Omega_{\mathrm{f}^-}$.

For any integer $k \geq 0$, we consider the space of broken polynomials on $\mathcal{T}_h^\bullet$ (with $\bullet \in \{\mathrm{f}, \mathrm{f}^\pm\}$) that writes

$$\mathbb{P}^k(\mathcal{T}_h^\bullet) := \{\, v \in L^2(\Omega_\bullet) \mid \forall \omega_c^\bullet \in \mathcal{T}_h^\bullet,\ v_{|\omega_c^\bullet} \in \mathbb{P}^k(\omega_c^\bullet) \,\},$$

where $\mathbb{P}^k(\omega_c^\bullet)$ denotes the space of polynomials of total degree at most k defined on the cell $\omega_c^\bullet$. Its dimension is given by $\dim \mathbb{P}^k(\omega_c^\bullet) = \frac{(k+1)(k+2)}{2}$. We denote the local number of degrees of freedom by $N_k^\bullet := \dim \mathbb{P}^k(\omega_c^\bullet)$, which is independent of the specific element $\omega_c^\bullet$ for a given polynomial degree k.

Again, piecewise polynomial functions are denoted with a subscript h, and we use the usual shortcut

$$v_h^{c,\bullet} := v_{h|\omega_c^\bullet}.$$

The space of vector-valued functions is denoted $\mathbf{P}^k(\Xi) := (\mathbb{P}^k(\Xi))^3$ for any relevant domain Ξ (e.g. $\Xi = \mathcal{T}_h^\bullet$, $\mathcal{F}_h^\bullet$, or $\omega_c^\bullet$).

For any cell $\omega_c^\bullet \in \mathcal{T}_h^\bullet$, a local basis of $\mathbb{P}^k(\omega_c^\bullet)$ is written

$$\boldsymbol{\Psi}_{\omega_c^\bullet} := \{\, \psi_j^{c,\bullet} \,\}_{j \in [\![1, N_k]\!]}, \tag{6.30}$$

and we denote by $\underline{v}^{c,\bullet} := \{\, \underline{v}_j^{c,\bullet} \mid j \in [\![1, N_k]\!] \,\}$ the vector of *local degrees of freedom* (DOFs) associated with $\omega_c^\bullet$, such that

$$v_h^{c,\bullet}(\boldsymbol{x}, t) = \sum_{m=1}^{N_k} \underline{v}_m^{c,\bullet}(t)\, \psi_m^{c,\bullet}(\boldsymbol{x}) \quad \text{with} \quad \underline{v}^\bullet := \big\{\, \underline{v}^{c,\bullet} \mid c \in [\![1, N_{\mathrm{el}}^\bullet]\!] \big\},$$

the vector of *global degrees of freedom* associated with $v_h \in \mathbb{P}^k(\mathcal{T}_h^\bullet)$.

We also introduce the corresponding broken gradient operator

$$\nabla_h^{k,\bullet} : \mathbb{P}^k(\mathcal{T}_h^\bullet) \to \mathbb{P}^{k-1}(\mathcal{T}_h^\bullet),$$

defined for any $v_h \in \mathbb{P}^k(\mathcal{T}_h^\bullet)$ by

$$(\nabla_h^{k,\bullet} v_h)_{|\omega_c^\bullet} := \nabla_{\boldsymbol{x}}(v_{h|\omega_c^\bullet}), \qquad \forall \omega_c^\bullet \in \mathcal{T}_h^\bullet.$$

For any given $v_h \in \mathbf{P}^k(\mathcal{T}_h^\bullet)$ and for any interface $\Gamma_{cv(k)}^\bullet \in \mathcal{F}_h^{\bullet,0}$ shared by two neighboring cells $\omega_c^\bullet$ and $\omega_{v(k)}^\bullet$, we denote by $v_k^{-,\bullet}$ and $v_k^{+,\bullet}$ the traces of $v_{h|\omega_c^\bullet}$ and $v_{h|\omega_{v(k)}^\bullet}$ on $\Gamma_{cv(k)}^\bullet$, respectively that is the *interior* and *exterior* traces with respect to $\omega_c^\bullet$.

In the fluid domain, the discrete unknown $\boldsymbol{v}_h^{c,\bullet} := (\eta_h^{c,\bullet}, \boldsymbol{q}_h^{c,\bullet})^\top$ is expanded locally as

$$\boldsymbol{v}_h^{c,\bullet}(\boldsymbol{x}, t) = \sum_{m=1}^{N_k} \underline{\boldsymbol{v}}_m^{c,\bullet}(t)\, \psi_m^{c,\bullet}(\boldsymbol{x}), \tag{6.31}$$

where $\underline{\boldsymbol{v}}_m^{c,\bullet}(t) \in \mathbb{R}^3$ collects the polynomial coefficients.

Submesh-dependent functional spaces and operators

For a fluid mesh $\mathcal{T}_h^\bullet$ (with $\bullet \in \{\mathrm{f}, \mathrm{f}^\pm\}$), we subdivide every element $\omega_c^\bullet \in \mathcal{T}_h^\bullet$ into a finite number of subcells, $\mathcal{T}_{\omega_c^\bullet}^k$ denotes the sub-partition of $\omega_c^\bullet$ into N_s subcells:

$$\overline{\omega}_c^\bullet = \bigcup_{m \in [\![1, N_s]\!]} \overline{S}_m^{c,\bullet},$$

with volume $|S_m^{c,\bullet}|$, gathered into the diagonal matrix

$$\mathbb{D}_c^\bullet := \mathrm{diag}_{m \in [\![1, N_s]\!]}(\, |S_m^{c,\bullet}| \,) \in \mathscr{M}_{N_s}(\mathbb{R}).$$

The face between $S_m^{c,\bullet}$ and $S_p^{c,\bullet}$ is denoted by $\Gamma_{mp}^{c,\bullet}$, with length $\ell_{mp}^\bullet$. The number of faces of subcell $S_m^{c,\bullet}$ is $N_f^{m,\bullet}$, the corresponding set of faces is $\mathcal{F}_m^{c,\bullet}$, such that $\#\mathcal{F}_m^{c,\bullet} = N_f^{m,\bullet}$, and $N_f^{c,\bullet}$ is the number of subcell faces lying strictly inside $\omega_c^\bullet$, i.e. faces not belonging to $\partial\omega_c^\bullet$.

For each $S_m^{c,\bullet} \in \mathcal{T}_{\omega_c^\bullet}^k$, the set $\mathcal{V}_m^{c,\bullet}$ contains its face-neighboring subcells (with $\#\mathcal{V}_m^{c,\bullet} = N_f^{m,\bullet}$), and we distinguish the subset $\breve{\mathcal{V}}_m^{c,\bullet}$, which gathers only the neighboring subcells contained in the same parent cell $\omega_c^\bullet$. Hence, a neighboring subcell S_p can belong either to the same element ($S_p \in \mathcal{T}_{\omega_c^\bullet}^k$, then $S_p = S_p^{c,\bullet}$), or to an adjacent element ($S_p \in \mathcal{T}_{\omega_v^\bullet}^k$, then $S_p = S_p^{v,\bullet}$).

For each face $\Gamma_{mp}^{c,\bullet}$ shared by two subcells $S_m^{c,\bullet}$ and $S_p^{v,\bullet} \in \mathcal{V}_m^{c,\bullet}$, $\mathbf{n}_{mp}^\bullet$ is the unit normal vector to $\Gamma_{mp}^{c,\bullet}$, oriented outward from $S_m^{c,\bullet}$. These normals are consistently oriented through the sign function $\varepsilon_{mp}^{c,\bullet}$ defined as

$$\varepsilon_{mp}^{c,\bullet} = \begin{cases} +1, & \text{if } S_m^{c,\bullet} \text{ and } S_p^{v,\bullet} \text{ are neighbors and } m < p, \\ -1, & \text{if } S_m^{c,\bullet} \text{ and } S_p^{v,\bullet} \text{ are neighbors and } m > p, \\ 0, & \text{if } S_p^{v,\bullet} \notin \mathcal{V}_m^{c,\bullet}. \end{cases}$$

We denote by $\mathcal{P}_m^{c,\bullet}$ the set of vertices $\boldsymbol{x}_p$ of the subcell $S_m^{c,\bullet}$, and by $\mathcal{V}_p^\bullet$ the set of all subcells sharing the vertex $\boldsymbol{x}_p$. The extended neighborhood of the subcell $S_m^{c,\bullet}$ is:

$$\mathcal{N}(S_m^{c,\bullet}) := \bigcup_{\boldsymbol{x}_p \in \mathcal{P}_m^{c,\bullet}} \mathcal{V}_p^\bullet,$$

which contains $S_m^{c,\bullet}$ itself, as well as its face- and vertex-neighboring subcells $S_p^{v,\bullet}$, whether they belong to the same cell $\omega_c^\bullet$ or to one of its neighboring elements. Finally, $\mathcal{V}_{mp}^\bullet$ is the set of all subcells sharing at least one vertex lying on the face $\Gamma_{mp}^{c,\bullet}$.

For $\omega_c^\bullet \in \mathcal{T}_h^\bullet$, the space of piecewise-constant functions over its sub-partition $\mathcal{T}_{\omega_c^\bullet}^k$ is

$$\mathbb{P}^0(\mathcal{T}_{\omega_c^\bullet}^k) := \left\{ v \in L^\infty(\omega_c^\bullet) \;\middle|\; \forall S_m^{c,\bullet} \in \mathcal{T}_{\omega_c^\bullet}^k,\ v_{|S_m^{c,\bullet}} \in \mathbb{P}^0(S_m^{c,\bullet}) \right\}.$$

and $\{ \mathbb{1}_m^{c,\bullet} \mid m \in [\![1, N_s]\!] \}$ is the collection of subcell indicator functions:

$$\mathbb{1}_m^{c,\bullet}(\boldsymbol{x}) := \begin{cases} 1, & \text{if } \boldsymbol{x} \in S_m^{c,\bullet}, \\ 0, & \text{otherwise,} \end{cases}$$

leading to the *local mean-value projector* $\overline{\pi}_c^{k,\bullet} : \mathbb{P}^k(\omega_c^\bullet) \to \mathbb{P}^0(\mathcal{T}_{\omega_c^\bullet}^k)$:

$$\overline{\pi}_c^{k,\bullet} v_h^{c,\bullet} := \sum_{m=1}^{N_s} \overline{v}_m^{c,\bullet}\, \mathbb{1}_m^{c,\bullet} =: \overline{v}_h^{c,\bullet} \quad \text{with} \quad \overline{v}^{c,\bullet} := \{ \overline{v}_m^{c,\bullet} \mid m \in [\![1, N_s]\!] \},$$

where the coefficients $\overline{v}^{c,\bullet}$ represent the *subcell mean values*, defined as averages of $v_h^{c,\bullet}$ over each subcell $S_m^{c,\bullet} \in \mathcal{T}_{\omega_c^\bullet}^k$.

When there is no ambiguity, $\overline{v}^{c,\bullet}$ may be viewed either as a vector in $\mathbb{R}^{N_s}$, as the collection of N_s mean values, or as the associated piecewise-constant function $\overline{v}_h^{c,\bullet} \in \mathbb{P}^0(\mathcal{T}_{\omega_c^\bullet}^k)$.

We now assemble the local subpartitions into the global sub-mesh and the corresponding global piecewise-constant space as

$$\widetilde{\mathcal{T}}_h^\bullet := \bigcup_{\omega_c^\bullet \in \mathcal{T}_h^\bullet} \mathcal{T}_{\omega_c^\bullet}^k, \qquad \mathbb{P}^0(\widetilde{\mathcal{T}}_h^\bullet) := \bigtimes_{\omega_c^\bullet \in \mathcal{T}_h^\bullet} \mathbb{P}^0(\mathcal{T}_{\omega_c^\bullet}^k).$$

The global projection operator onto this space, obtained by assembling all the local projectors $\overline{\pi}_c^{k,\bullet}$, is denoted by $\overline{\pi}_h^{k,\bullet}$. In what follows, piecewise-constant functions belonging to $\mathbb{P}^0(\widetilde{\mathcal{T}}_h^\bullet)$ are denoted by an overline and a subscript h. For any $\overline{v}_h \in \mathbb{P}^0(\widetilde{\mathcal{T}}_h^\bullet)$ and any subcell $S_m^{c,\bullet} \in \widetilde{\mathcal{T}}_h^\bullet$, we have the natural identification

$$\overline{v}_m^{c,\bullet} = \overline{v}_{h|S_m^{c,\bullet}}.$$

HHO functional spaces and interpolators

We now introduce the discrete functional setting used on the structure domain Ω_s, endowed with the conforming mesh $\mathcal{T}_h^s$ and its skeleton $\mathcal{F}_h^s$.

For any cell $\omega_c^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}$ and face $\Gamma \in \mathcal{F}_h^{\mathrm{s}}$, we consider the polynomial spaces $\mathbb{P}^k(\omega_c^{\mathrm{s}})$ the space of polynomials of total degree at most k on ω_c^{s}, and by $\mathbb{P}^k(\Gamma)$ the space of polynomials of degree at most k on the face Γ. Their dimensions are $\dim \mathbb{P}^k(\omega_c^{\mathrm{s}}) = \frac{(k+1)(k+2)}{2}$ and $\dim \mathbb{P}^k(\Gamma) = k+1$ for straight faces. We denote by $\{\varphi_j^{c,\mathrm{s}}\}_{j\in[\![1,N_k^{\mathrm{s}}]\!]}$ a basis of $\mathbb{P}^k(\omega_c^{\mathrm{s}})$ (with $N_k^{\mathrm{s}} = \dim \mathbb{P}^k(\omega_c^{\mathrm{s}})$), and by $\{\varphi_j^{\Gamma,\mathrm{s}}\}_{j\in[\![1,M_k^{\mathrm{s}}]\!]}$ a basis of $\mathbb{P}^k(\Gamma)$ (with $M_k^{\mathrm{s}} = \dim \mathbb{P}^k(\Gamma)$).

For a given cell ω_c^{s} with face set $\mathcal{F}_c^{\mathrm{s}} := \{\,\Gamma \in \mathcal{F}_h^{\mathrm{s}} \mid \Gamma \subset \partial\omega_c^{\mathrm{s}}\,\}$, the *local HHO approximation space* is defined as

$$\underline{\mathsf{U}}_c^{k,\mathrm{s}} := \mathbb{P}^k(\omega_c^{\mathrm{s}}) \times \left\{ \bigtimes_{\Gamma\in\mathcal{F}_c^{\mathrm{s}}} \mathbb{P}^k(\Gamma) \right\}.$$

A generic local vector of DOFs is then written $\underline{u}_c^{\mathrm{s}} := (u_c^{\mathrm{s}},\ (u_\Gamma^{\mathrm{s}})_{\Gamma\in\mathcal{F}_c^{\mathrm{s}}}) \in \underline{\mathsf{U}}_c^{k,\mathrm{s}}$ with $u_c^{\mathrm{s}} \in \mathbb{P}^k(\omega_c^{\mathrm{s}})$ and $u_\Gamma^{\mathrm{s}} \in \mathbb{P}^k(\Gamma)$ for each face $\Gamma \in \mathcal{F}_c^{\mathrm{s}}$.

Denoting by $\mathcal{F}_h^{\mathrm{s},0}$ the set of interior faces of $\mathcal{T}_h^{\mathrm{s}}$ and by $\mathcal{F}_h^{\mathrm{s},\partial}$ the set of boundary faces (on $\partial\Omega_{\mathrm{s}}$) s.t. $\mathcal{F}_h^{\mathrm{s}} = \mathcal{F}_h^{\mathrm{s},0} \cup \mathcal{F}_h^{\mathrm{s},\partial}$, the *global approximation HHO space* is therefore

$$\underline{\mathsf{U}}_h^{k,\mathrm{s}} := \left\{ \bigtimes_{\omega_c^{\mathrm{s}}\in\mathcal{T}_h^{\mathrm{s}}} \mathbb{P}^k(\omega_c^{\mathrm{s}}) \right\} \times \left\{ \bigtimes_{\Gamma\in\mathcal{F}_h^{\mathrm{s}}} \mathbb{P}^k(\Gamma) \right\},$$

or equivalently

$$\underline{\mathsf{U}}_h^{k,\mathrm{s}} := \left\{ \left((u_c^{\mathrm{s}})_c,\ (u_\Gamma^{\mathrm{s}})_\Gamma\right) \;\middle|\; (u_c^{\mathrm{s}}, u_\Gamma^{\mathrm{s}}) \in \mathbb{P}^k(\omega_c^{\mathrm{s}}) \times \mathbb{P}^k(\Gamma),\quad \forall(\omega_c^{\mathrm{s}},\Gamma) \in \mathcal{T}_h^{\mathrm{s}} \times \mathcal{F}_h^{\mathrm{s}} \right\}.$$

We denote by $\underline{u}_h^{\mathrm{s}} \in \underline{\mathsf{U}}_h^{k,\mathrm{s}}$ the global vector of DOFs, obtained by stacking all local vectors $\underline{u}_c^{\mathrm{s}}$ and $\underline{u}_\Gamma^{\mathrm{s}}$.

Let us also define the local H^1-like seminorm

$$\left\|\underline{u}_c^{\mathrm{s}}\right\|_{1,c}^2 := \|\nabla_{\boldsymbol{x}} u_c^{\mathrm{s}}\|_{\omega_c^{\mathrm{s}}}^2 + \left|\underline{u}_c^{\mathrm{s}}\right|_{1,\partial\omega_c^{\mathrm{s}}}^2, \qquad \left|\underline{u}_c^{\mathrm{s}}\right|_{1,\partial\omega_c^{\mathrm{s}}}^2 := \sum_{\Gamma\in\mathcal{F}_c^{\mathrm{s}}} h_\Gamma^{-1} \|u_\Gamma^{\mathrm{s}} - u_c^{\mathrm{s}}\|_\Gamma^2.$$

Here, h_Γ denotes the face diameter and the term h_Γ^{-1} ensures the scaling between volume and face contributions in the seminorm. The global version of this seminorm is then defined as

$$\left\|\underline{u}_h^{\mathrm{s}}\right\|_{1,h}^2 := \sum_{\omega_c^{\mathrm{s}}\in\mathcal{T}_h^{\mathrm{s}}} \left\|\underline{u}_c^{\mathrm{s}}\right\|_{1,c}^2.$$

The global L^2-orthogonal projector $\pi_h^{k,\mathrm{s}} : L^2(\Omega_{\mathrm{s}}) \to \mathbb{P}^k(\mathcal{T}_h^{\mathrm{s}})$ is defined cellwise by

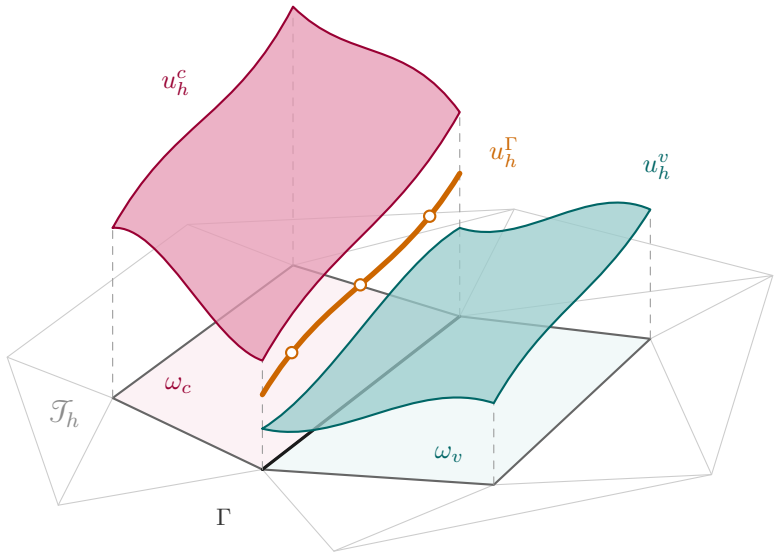


Figure 6.3: Schematic representation of an HHO solution between two adjacent mesh cells ω_c and ω_v.

$$(\pi_h^{k,\mathrm{s}} u)_{|\omega_c^{\mathrm{s}}} = \pi_c^{k,\mathrm{s}} u_c \qquad \forall \omega_c^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}.$$

The local elliptic projector $\pi_X^{1,k} : \mathsf{H}^1(X) \to \mathbb{P}^k(X)$ is classically defined as the unique polynomial $\pi_X^{1,k}u \in \mathbb{P}^k(X)$ satisfying

$$\big(\nabla_{\boldsymbol{x}}(\pi_X^{1,k}u - u), \nabla_{\boldsymbol{x}}v\big)_X = 0, \qquad \big(\pi_X^{1,k}u - u, 1\big)_X = 0, \qquad \forall v \in \mathbb{P}^k(X)$$

Taking $X = \omega_c^{\mathrm{s}}$ (resp. $X = \Gamma$) gives the corresponding projectors on a cell ω_c^{s} (resp. on a face Γ).

For a piecewise H^1 function u defined on Ω_{s}, the global elliptic projector $\pi_h^{1,k,\mathrm{s}} : \mathsf{H}^1(\Omega_{\mathrm{s}}) \to \mathbb{P}^k(\mathcal{T}_h^{\mathrm{s}})$ is defined cellwise by

$$(\pi_h^{1,k,\mathrm{s}}u)_{|\omega_c^{\mathrm{s}}} = \pi_c^{1,k,\mathrm{s}}(u_c) \qquad \forall \omega_c^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}.$$

Remark 6.2.3 (*Computation of the elliptic projector*). *The local elliptic projector $\pi_c^{1,k+1}$ can be efficiently computed without requiring the full knowledge of the continuous function u, but only its L^2 projections on the cell and its faces. Indeed, for a function $u \in \mathsf{H}^1(\omega_c^{\mathrm{s}})$ and any $v \in \mathbb{P}^{k+1}(\omega_c^{\mathrm{s}})$, integration by parts gives*

$$\big(\nabla_{\boldsymbol{x}}u, \nabla_{\boldsymbol{x}}v\big)_{\omega_c^{\mathrm{s}}} = -\big(u, \Delta_{\boldsymbol{x}}v\big)_{\omega_c^{\mathrm{s}}} + \sum_{\Gamma \in \mathcal{F}_c^{\mathrm{s}}} \big\langle u, \nabla_{\boldsymbol{x}}v \cdot \mathbf{n}_\Gamma^{\mathrm{s}}\big\rangle_\Gamma.$$

Since $\Delta_{\boldsymbol{x}}v \in \mathbb{P}^{k-1}(\omega_c^{\mathrm{s}}) \subset \mathbb{P}^k(\omega_c^{\mathrm{s}})$ and $\nabla_{\boldsymbol{x}}v \cdot \mathbf{n}_\Gamma^{\mathrm{s}} \in \mathbb{P}^k(\Gamma)$ for all $\Gamma \in \mathcal{F}_c^{\mathrm{s}}$, the right-hand side only depends on the L^2 projections of u onto $\mathbb{P}^k(\omega_c^{\mathrm{s}})$ and $\mathbb{P}^k(\Gamma)$. Replacing u by its L^2 projections in these quantities leads to the computable relation

$$\big(\nabla_{\boldsymbol{x}}\pi_c^{1,k+1,\mathrm{s}}u, \nabla_{\boldsymbol{x}}v\big)_{\omega_c^{\mathrm{s}}} = -\big(\pi_c^{k,\mathrm{s}}u, \Delta_{\boldsymbol{x}}v\big)_{\omega_c^{\mathrm{s}}} + \sum_{\Gamma \in \mathcal{F}_c^{\mathrm{s}}} \big\langle \pi_\Gamma^{k,\mathrm{s}}u, \nabla_{\boldsymbol{x}}v \cdot \mathbf{n}_\Gamma^{\mathrm{s}}\big\rangle_\Gamma, \qquad \forall v \in \mathbb{P}^{k+1,0}(\omega_c^{\mathrm{s}}),$$

supplemented with the mean-value condition

$$\big(\pi_c^{1,k+1,\mathrm{s}}u - \pi_c^{k,\mathrm{s}}u, 1\big)_{\omega_c^{\mathrm{s}}} = 0.$$

Consequently, the elliptic projector $\pi_c^{1,k+1,\mathrm{s}}u$ can be computed locally using:

- *the cellwise L^2 projection $\pi_c^{k,\mathrm{s}}u$ on $\mathbb{P}^k(\omega_c^{\mathrm{s}})$,*
- *the facewise L^2 projections $\pi_\Gamma^{k,\mathrm{s}}u$ on $\mathbb{P}^k(\Gamma)$ for all $\Gamma \in \mathcal{F}_c^{\mathrm{s}}$.*

This construction is particularly convenient for implementation, since the elliptic projector only requires known L^2 projections as input data, and its computation reduces to solving a local symmetric positive definite system of dimension $\dim(\mathbb{P}^{k+1,0}(\omega_c^{\mathrm{s}}))$.

The local interpolation operator is then defined as

$$\underline{\mathsf{I}}_c^{k,\mathrm{s}} : u \in \mathsf{H}^1(\omega_c^{\mathrm{s}}) \mapsto \underline{\mathsf{I}}_c^{k,\mathrm{s}}u := \big(\pi_c^{k,\mathrm{s}}u_c,\ (\pi_\Gamma^{k,\mathrm{s}}u_\Gamma)_{\Gamma \in \mathcal{F}_c^{\mathrm{s}}}\big) \in \underline{\mathsf{U}}_c^{k,\mathrm{s}}.$$

By gathering the face unknowns that belong to $\mathcal{F}_h^{\mathrm{s},0}$, we obtain the global interpolation operator

$$\underline{\mathsf{I}}_h^{k,\mathrm{s}} : u \in \mathsf{H}^1(\Omega_{\mathrm{s}}) \mapsto \underline{\mathsf{I}}_h^{k,\mathrm{s}}u := \big((\pi_c^{k,\mathrm{s}}u_c)_{\omega_c^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}},\ (\pi_\Gamma^{k,\mathrm{s}}u_\Gamma)_{\Gamma \in \mathcal{F}_h^{\mathrm{s}}}\big) \in \underline{\mathsf{U}}_h^{k,\mathrm{s}},$$

with single-valued face unknowns on the interior faces.

For any cell $\omega_c^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}$, we define the *local potential reconstruction* of order $k+1$ as the linear operator

$$\mathsf{p}_c^{k+1,\mathrm{s}} : \underline{u}_c^{\mathrm{s}} = (u_c^{\mathrm{s}}, (u_\Gamma^{\mathrm{s}})_{\Gamma \in \mathcal{F}_c^{\mathrm{s}}}) \in \underline{\mathsf{U}}_c^{k,\mathrm{s}} \mapsto \mathsf{p}_c^{k+1,\mathrm{s}}\underline{u}_c^{\mathrm{s}} \in \mathbb{P}^{k+1}(\omega_c^{\mathrm{s}}),$$

as, for $w \in \mathbb{P}^{k+1}(\omega_c^{\mathrm{s}})$, the unique solution of

$$\begin{cases} \big(\nabla_{\boldsymbol{x}}(\mathsf{p}_c^{k+1,\mathrm{s}}\underline{u}_c^{\mathrm{s}}), \nabla_{\boldsymbol{x}}w\big)_{\omega_c^{\mathrm{s}}} = -\big(u_c^{\mathrm{s}}, \Delta_{\boldsymbol{x}}w\big)_{\omega_c^{\mathrm{s}}} + \sum_{\Gamma \in \mathcal{F}_c^{\mathrm{s}}} \big\langle u_\Gamma^{\mathrm{s}}, \nabla_{\boldsymbol{x}}w \cdot \mathbf{n}_\Gamma^{\mathrm{s}}\big\rangle_\Gamma, \\ \big(\mathsf{p}_c^{k+1,\mathrm{s}}\underline{u}_c^{\mathrm{s}} - u_c^{\mathrm{s}}, 1\big)_{\omega_c^{\mathrm{s}}} = 0 \end{cases} \tag{6.32}$$

Remark 6.2.4 (*Computation of the potential reconstruction*). *Using integration by parts in the first term on the right-hand side of (6.32), an equivalent form reads*

$$\big(\nabla_{\boldsymbol{x}}(\mathsf{p}_c^{k+1,\mathrm{s}}\underline{u}_c^{\mathrm{s}}), \nabla_{\boldsymbol{x}}w\big)_{\omega_c^{\mathrm{s}}} = \big(\nabla_{\boldsymbol{x}}u_c^{\mathrm{s}}, \nabla_{\boldsymbol{x}}w\big)_{\omega_c^{\mathrm{s}}} + \sum_{\Gamma \in \mathcal{F}_c^{\mathrm{s}}} \big\langle u_\Gamma^{\mathrm{s}} - u_c^{\mathrm{s}}, \nabla_{\boldsymbol{x}}w \cdot \mathbf{n}_\Gamma^{\mathrm{s}}\big\rangle_\Gamma, \quad \forall w \in \mathbb{P}^{k+1}(\omega_c^{\mathrm{s}}).$$

The *global potential reconstruction* of order $k+1$ is then defined by gathering the local reconstructions, i.e. as the linear operator

$$\mathsf{p}_h^{k+1,\mathrm{s}} : \underline{\mathsf{U}}_h^{k,\mathrm{s}} \to \bigtimes_{\omega_c^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}} \mathbb{P}^{k+1}(\omega_c^{\mathrm{s}})$$

where we have elementwise $(\mathsf{p}_h^{k+1,\mathrm{s}} \underline{u}_h^{\mathrm{s}})_{|\omega_c^{\mathrm{s}}} := \mathsf{p}_c^{k+1,\mathrm{s}} \underline{u}_c^{\mathrm{s}}$.

Remark 6.2.5 (*Role of the potential reconstruction*). *The local reconstruction $\mathsf{p}_c^{k+1,\mathrm{s}} \underline{u}_c^{\mathrm{s}}$ is a polynomial on ω_c^{s} of degree $k+1$, i.e. one degree higher than the cell unknown $u_c^{\mathrm{s}} \in \mathbb{P}^k(\omega_c^{\mathrm{s}})$. Comparing the defining identities of the elliptic projector $\pi_c^{1,k+1}$ and of the reconstruction $\mathsf{p}_c^{k+1,\mathrm{s}}$ (tested with the same polynomials) shows that, for all $v \in \mathsf{H}^1(\omega_c^{\mathrm{s}})$,*

$$\mathsf{p}_c^{k+1,\mathrm{s}}\big(\underline{\mathsf{I}}_c^{k,\mathrm{s}} v\big) = \pi_c^{1,k+1,\mathrm{s}} v \tag{6.33}$$

that is, the composition of the potential reconstruction with the local interpolator returns the $(k+1)$-degree elliptic projector of v on the cell ω_c^{s}. The potential reconstruction is designed to be a local approximation of the exact solution within the element, and more specifically, to provide a high-order representation of its gradient. In essence, while the original discrete data is split between element and face contributions (a so-called hybrid representation), the reconstruction operator regathers this information into a unified, higher-order object that lives purely in the polynomial space $\mathbb{P}^{k+1}(\omega_c^{\mathrm{s}})$.

Finally, for any cell $\omega_c^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}$, we define the *local gradient reconstruction* of order k as the linear operator

$$\mathsf{G}_c^{k,\mathrm{s}} : \underline{u}_c^{\mathrm{s}} \in \underline{\mathsf{U}}_c^{k,\mathrm{s}} \mapsto \mathsf{G}_c^{k,\mathrm{s}} \underline{u}_c^{\mathrm{s}} := \nabla_{\boldsymbol{x}}\big(\mathsf{p}_c^{k+1,\mathrm{s}} \underline{u}_c^{\mathrm{s}}\big) \in \big(\mathbb{P}^k(\omega_c^{\mathrm{s}})\big)^2. \tag{6.34}$$

By construction of the potential reconstruction, this operator uniquely satisfies

$$\big(\mathsf{G}_c^{k,\mathrm{s}} \underline{u}_c^{\mathrm{s}}, \boldsymbol{w}\big)_{\omega_c^{\mathrm{s}}} = -\big(u_c^{\mathrm{s}}, \nabla_{\boldsymbol{x}} \cdot \boldsymbol{w}\big)_{\omega_c^{\mathrm{s}}} + \sum_{\Gamma \in \mathcal{F}_c^{\mathrm{s}}} \big\langle u_\Gamma^{\mathrm{s}}, \boldsymbol{w} \cdot \mathbf{n}_\Gamma^{\mathrm{s}} \big\rangle_\Gamma, \qquad \forall \boldsymbol{w} \in \nabla_{\boldsymbol{x}} \mathbb{P}^{k+1}(\omega_c^{\mathrm{s}}).$$

Topological nature of the fluid-structure contact-line

In the numerical scheme, we must account for the topological nature of the fluid-structure interface $\partial\Omega_{\mathrm{fs}}$, which depends on the chosen configuration. Let N_{fs} denotes the number of disjoint connected components defining the interface:

- In the cylinder configuration ($N_{\mathrm{fs}} = 1$), the contact-line is a single continuous boundary. Consequently, the potential trace on the interface is a scalar function ψ_h^{s}, and the associated Dirichlet-to-Neumann operator Λ_h acts as a standard scalar operator,
- In the pontoon configuration ($N_{\mathrm{fs}} = 2$), the interface consists of two disconnected segments (the left and right boundaries). As previously described, the potential trace is a vector-valued function $\boldsymbol{\psi}_h^{\mathrm{s}} = (\psi_h^{\mathrm{s},-}, \psi_h^{\mathrm{s},+})^\top$, and the DtN operator $\boldsymbol{\Lambda}_h$ naturally takes the form of a 2×2 block matrix, coupling the disjoint exterior fluid domains through the interior potential.

To maintain a unified mathematical formulation for both configurations throughout the numerical sections, we may adopt the boldface notation $\boldsymbol{\psi}_h^{\mathrm{s}}$ and $\boldsymbol{\Lambda}_h$ for the discrete interface potential and coupling operator, with the understanding that they reduce to their scalar counterparts when $N_{\mathrm{fs}} = 1$.

6.2.2 SSP-RK discretization of the ODE for the interface potential trace

The time evolution of the potential trace at the fluid–structure interface is governed by

$$\partial_t \boldsymbol{\psi}^{\mathrm{s}} = -\mathrm{g}\zeta - \tfrac{1}{2}|\boldsymbol{u}|^2 \qquad \text{on } (0,T) \times \partial\Omega_{\mathrm{fs}}. \tag{6.35}$$

This equation is integrated in time using the same SSP-RK framework introduced in Time discretization and time-marching algorithm.

Indeed, let t^n denote the current time level and Δt^n the time step. For a generic r–stage SSP-RK method, the discrete update of (6.35) reads

$$\begin{aligned}
\boldsymbol{\psi}^{\mathrm{s},(0)} &= \boldsymbol{\psi}^{\mathrm{s},n},\\
\boldsymbol{\psi}^{\mathrm{s},(i)} &= \sum_{j=0}^{i-1} \mathsf{a}_{ij}\, \boldsymbol{\psi}^{\mathrm{s},(j)} + \Delta t^n\, \mathsf{b}_i \left(-\mathrm{g}\zeta^{(i-1)} - \tfrac{1}{2}\,|\boldsymbol{u}_h^{(i-1)}|^2\right), \qquad \forall i \in [\![1, r]\!],\\
\boldsymbol{\psi}^{\mathrm{s},n+1} &= \boldsymbol{\psi}^{\mathrm{s},(r)},
\end{aligned}$$

where $\zeta^{(i-1)}$ and $\boldsymbol{u}^{(i-1)}$ denote traces at stage $i-1$, and $(\mathsf{a}_{ij}, \mathsf{b}_i)$ are the coefficients of the chosen SSP-RK scheme.

6.2.3 Fast Fourier Transform computation of the DtN operator

In the pontoon configuration, the DtN operator admits an explicit representation as a Fourier multiplier in the transverse variable y (see Proposition 6.1.1):

$$\begin{aligned}
\Lambda_-\boldsymbol{\psi}^{\mathrm{s}} &= H^{\mathrm{s}} D \coth(2lD)\psi_- - \frac{H^{\mathrm{s}} D}{\sinh(2lD)}\psi_+,\\
\Lambda_+\boldsymbol{\psi}^{\mathrm{s}} &= -\frac{H^{\mathrm{s}} D}{\sinh(2lD)}\psi_- + H^{\mathrm{s}} D \coth(2lD)\psi_+,
\end{aligned}$$

where $\boldsymbol{\psi}^{\mathrm{s}} = (\psi_-, \psi_+)^\top$ gathers the Dirichlet traces of the potential on $\partial\Omega_{\mathrm{fs}^-}$ and $\partial\Omega_{\mathrm{fs}^+}$, allowing to mimic the basic ideas of BEM.

Indeed, from a numerical viewpoint, the action of $\boldsymbol{\Lambda}$ on $\boldsymbol{\psi}^{\mathrm{s}}$ can efficiently be computed by means of the *Fast Fourier Transform* (FFT). We denote by $\mathfrak{F}_h$ and $\mathfrak{F}_h^{-1}$ the discrete Fourier transform and its inverse, acting on discrete vectors $\{f_j\}_{j\in[\![0,N_y-1]\!]}$ defined on a uniform grid. In the discrete setting, ψ_- and ψ_+ may be sampled at the same uniformly distributed points in the transverse direction y, ensuring that the degrees of freedom on both interfaces are perfectly matching.

In Fourier space, the operator $D = |D_y|$ is represented by multiplication by $|\xi|$. For a discrete representation $\boldsymbol{\psi}^{\mathrm{s}}(y_j)$ on a uniform grid $y_j = j\Delta y$, with $j \in [\![0, N_y - 1]\!]$ and $\Delta y = \frac{L_y}{N_y}$, the discrete frequencies are

$$\xi_k = \begin{cases} \dfrac{2\pi k}{L_y} & 0 \le k \le \frac{N_y}{2},\\[2mm] \dfrac{2\pi(k-N_y)}{L_y} & \frac{N_y}{2} < k < N_y, \end{cases}$$

where L_y is the total transverse length of the computational domain, with periodic boundary conditions in the y-direction. Set $\mu_k = |\xi_k|$ and define the symbols at the zero frequency by continuity,

$$a_k = \begin{cases} \mu_k \coth(2l\mu_k), & \mu_k > 0,\\ (2l)^{-1}, & \mu_k = 0, \end{cases} \qquad b_k = \begin{cases} \dfrac{\mu_k}{\sinh(2l\mu_k)}, & \mu_k > 0,\\ (2l)^{-1}, & \mu_k = 0. \end{cases}$$

For each frequency ξ_k,

$$\begin{aligned}
\widehat{\Lambda_-\boldsymbol{\psi}^{\mathrm{s}}}(\xi_k) &= H^{\mathrm{s}} a_k \widehat{\psi_-}(\xi_k) - H^{\mathrm{s}} b_k \widehat{\psi_+}(\xi_k),\\
\widehat{\Lambda_+\boldsymbol{\psi}^{\mathrm{s}}}(\xi_k) &= -H^{\mathrm{s}} b_k \widehat{\psi_-}(\xi_k) + H^{\mathrm{s}} a_k \widehat{\psi_+}(\xi_k),
\end{aligned}$$

where $\widehat{\psi_\pm}(\xi_k)$ denote the discrete Fourier transforms of $\psi_\pm$, that is

$$\widehat{\psi_\pm}(\xi_k) = \mathfrak{F}_h(\psi_\pm^j)(\xi_k), \qquad \psi_\pm^j = \mathfrak{F}_h^{-1}(\widehat{\psi_\pm})(y_j).$$

Multiplying these transforms by the corresponding Fourier symbols and applying the inverse discrete transform yields the normal discharges $H^{\mathrm{s}}\nabla_{\boldsymbol{x}}\phi^{\mathrm{s}} \cdot \mathbf{n}$ on $\partial\Omega_{\mathrm{fs}^-}$ and $\partial\Omega_{\mathrm{fs}^+}$ with spectral accuracy and computational complexity $\mathcal{O}(N_y \log N_y)$.

Remark 6.2.6 (*One-dimensional reduction*). *The FFT-based DtN operator may serve as a reference for comparison with the HHO elliptic solver, where $\boldsymbol{\Lambda}$ is approximated numerically through the interior problem. Let us note that, in the case where we consider a longitudinally invariant geometry with no variation in the y-direction for the incoming flow, the traces reduce to two scalar flux values (one on each side of the pontoon) given in (6.26).*

6.2.4 HHO formulation of the interior elliptic problem

We now apply the HHO method to the interior elliptic problem

$$\begin{cases} \nabla_{\boldsymbol{x}} \cdot \left(H^{\mathrm{s}} \nabla_{\boldsymbol{x}} \phi^{\mathrm{s}}\right) = 0 & \text{in } \Omega_{\mathrm{s}}, \\ \phi^{\mathrm{s}} = \psi^{\mathrm{s}} & \text{on } \partial\Omega_{\mathrm{fs}}. \end{cases} \tag{6.36}$$

Remark 6.2.7 (*Well-posedness of the interior elliptic problem*). *The diffusion coefficient, i.e. the water depth here, satisfies*

$$H^{\mathrm{s}} \in L^\infty(\Omega_{\mathrm{s}}), \qquad H_{\max} \geq H^{\mathrm{s}}(\boldsymbol{x}) \geq H_{\min} > 0 \quad \text{a.e. in } \Omega_{\mathrm{s}},$$

and the boundary data $\psi^{\mathrm{s}} \in \mathsf{H}^{1/2}(\partial\Omega_{\mathrm{s}})$ is prescribed by the solution of the previous forced ODE. The bilinear form

$$\mathscr{A} : \mathsf{H}^1(\Omega_{\mathrm{s}}) \times \mathsf{H}^1(\Omega_{\mathrm{s}}) \ni (\phi, w) \mapsto \left(H^{\mathrm{s}} \nabla_{\boldsymbol{x}} \phi, \nabla_{\boldsymbol{x}} w\right)_{\Omega_{\mathrm{s}}},$$

is continuous on $\mathsf{H}^1(\Omega_{\mathrm{s}}) \times \mathsf{H}^1(\Omega_{\mathrm{s}})$ since $H^{\mathrm{s}} \in L^\infty(\Omega_{\mathrm{s}})$. We also define the space of homogeneous Dirichlet test functions

$$\mathsf{H}^1_0(\Omega_{\mathrm{s}}) := \left\{ w \in \mathsf{H}^1(\Omega_{\mathrm{s}}) \mid w_{|\partial\Omega_{\mathrm{s}}} = 0 \right\}.$$

Since Ω_{s} is bounded and Lipschitz, the Poincaré inequality holds on $\mathsf{H}^1_0(\Omega_{\mathrm{s}})$, and together with $H^{\mathrm{s}} \geq H_{\min} > 0$, it ensures coercivity of $\mathscr{A}$:

$$\mathscr{A}(w, w) \geq H_{\min} \|\nabla_{\boldsymbol{x}} w\|^2_{L^2(\Omega_{\mathrm{s}})} \geq \mathsf{c}_{\mathscr{P}}^{-2} H_{\min} \|w\|^2_{\mathsf{H}^1(\Omega_{\mathrm{s}})} \qquad \forall w \in \mathsf{H}^1_0(\Omega_{\mathrm{s}}),$$

where $\mathsf{c}_{\mathscr{P}} > 0$ is the Poincaré constant related to Ω_{s}.

The objective is now to introduce a discrete counterpart for the DtN operator and compute a discrete potential $\underline{\phi}^{\mathrm{s}}_h$ in the global HHO space $\underline{\mathsf{U}}^{k,\mathrm{s}}_h$, or equivalently, a discrete normal discharge $H^{\mathrm{s}}\, \boldsymbol{u}^{\mathrm{s}} \cdot \mathbf{n}$ on $\partial\Omega_{\mathrm{fs}}$.

Remark 6.2.8 (*Nature of the Dirichlet boundary condition*). *In (6.36), ψ^{s} denotes the local scalar trace of the potential evaluated pointwise on the boundary $\partial\Omega_{\mathrm{fs}}$. In the pontoon configuration where the interface is disjoint, this boundary data corresponds piecewise to the components of the global vector $\boldsymbol{\psi}^{\mathrm{s}} = (\psi^{\mathrm{s},-}, \psi^{\mathrm{s},+})^\top$ introduced earlier, such that $\psi^{\mathrm{s}} = \psi^{\mathrm{s},-}$ on the left boundary $\partial\Omega_{\mathrm{fs}^-}$ and $\psi^{\mathrm{s}} = \psi^{\mathrm{s},+}$ on the right boundary $\partial\Omega_{\mathrm{fs}^+}$.*

For each element $\omega^{\mathrm{s}}_c \in \mathcal{T}^{\mathrm{s}}_h$, we define the local discrete bilinear form

$$\mathscr{A}^{\mathrm{s}}_c\left(\underline{\phi}^{\mathrm{s}}_c, \underline{w}^{\mathrm{s}}_c\right) := \left(H^{c,\mathrm{s}}_h\, \mathbf{G}^{k,\mathrm{s}}_c \underline{\phi}^{\mathrm{s}}_c, \mathbf{G}^{k,\mathrm{s}}_c \underline{w}^{\mathrm{s}}_c\right)_{\omega^{\mathrm{s}}_c} + \mathsf{stab}^{\mathrm{s}}_c\left(\underline{\phi}^{\mathrm{s}}_c, \underline{w}^{\mathrm{s}}_c\right), \qquad \forall \underline{w}^{\mathrm{s}}_c \in \underline{\mathsf{U}}^{k,\mathrm{s}}_c, \tag{6.37}$$

where $H^{c,\mathrm{s}}_h := \pi^{k,\mathrm{s}}_c H^{\mathrm{s}} \in \mathbb{P}^k(\omega^{\mathrm{s}}_c)$, $\mathbf{G}^{k,\mathrm{s}}_c$ denotes the local reconstructed gradient defined in (6.34), and $\mathsf{stab}^{\mathrm{s}}_c$ is a local symmetric stabilization term enforcing consistency between cell and face unknowns.

Following the standard residual-based design [81], the stabilization can be defined as a weighted least-squares penalization of cell and face residuals that vanish for interpolated polynomials of degree $\leq k+1$:

$$\mathsf{stab}^{\mathrm{s}}_c\left(\underline{\phi}^{\mathrm{s}}_c, \underline{w}^{\mathrm{s}}_c\right) := \sum_{\Gamma \in \mathcal{F}^{\mathrm{s}}_c} h_\Gamma^{-1} \left(\delta^{k,\mathrm{s}}_{c\Gamma} \underline{\phi}^{\mathrm{s}}_c - \delta^{k,\mathrm{s}}_c \underline{\phi}^{\mathrm{s}}_c, \delta^{k,\mathrm{s}}_{c\Gamma} \underline{w}^{\mathrm{s}}_c - \delta^{k,\mathrm{s}}_c \underline{w}^{\mathrm{s}}_c\right)_\Gamma, \qquad \forall \underline{w}^{\mathrm{s}}_c \in \underline{\mathsf{U}}^{k,\mathrm{s}}_c, \tag{6.38}$$

where the local residual operators $\delta^{k,\mathrm{s}}_c$ and $\delta^{k,\mathrm{s}}_{c\Gamma}$, are defined as

$$\begin{aligned} \delta^{k,\mathrm{s}}_c : \underline{w}^{\mathrm{s}}_c \in \underline{\mathsf{U}}^{k,\mathrm{s}}_c \mapsto \delta^{k,\mathrm{s}}_c \underline{w}^{\mathrm{s}}_c := \pi^{k,\mathrm{s}}_c \left(\mathsf{p}^{k+1,\mathrm{s}}_c \underline{w}^{\mathrm{s}}_c - w^{\mathrm{s}}_c\right) \in \mathbb{P}^k(\omega^{\mathrm{s}}_c), \qquad \forall \omega^{\mathrm{s}}_c \in \mathcal{T}^{\mathrm{s}}_h, \\ \delta^{k,\mathrm{s}}_{c\Gamma} : \underline{w}^{\mathrm{s}}_c \in \underline{\mathsf{U}}^{k,\mathrm{s}}_c \mapsto \delta^{k,\mathrm{s}}_{c\Gamma} \underline{w}^{\mathrm{s}}_c := \pi^{k,\mathrm{s}}_\Gamma \left(\mathsf{p}^{k+1,\mathrm{s}}_c \underline{w}^{\mathrm{s}}_c - w^{\mathrm{s}}_\Gamma\right) \in \mathbb{P}^k(\Gamma), \qquad \forall \Gamma \in \mathcal{F}^{\mathrm{s}}_c. \end{aligned}$$

Remark 6.2.9 (*Properties of the local stabilization term*). *The local stabilization bilinear form $\mathsf{stab}^{\mathrm{s}}_c : \underline{\mathsf{U}}^{k,\mathrm{s}}_c \times \underline{\mathsf{U}}^{k,\mathrm{s}}_c \to \mathbb{R}$ defined in (6.38) satisfies the following classical HHO assumptions:*

(S1) *$\mathsf{stab}^{\mathrm{s}}_c$ is symmetric and positive semidefinite,*

(S2) *there exists $\alpha \in \mathbb{R}_+$, independent of h and of the cell ω^{s}_c, such that*

$$\alpha^{-1} \|\underline{w}^{\mathrm{s}}_c\|^2_{1,c} \leq \mathscr{A}^{\mathrm{s}}_c\left(\underline{w}^{\mathrm{s}}_c, \underline{w}^{\mathrm{s}}_c\right) \leq \alpha \|\underline{w}^{\mathrm{s}}_c\|^2_{1,c}, \quad \forall \underline{w}^{\mathrm{s}}_c \in \underline{\mathsf{U}}^{k,\mathrm{s}}_c,$$

(S3) *for all $w \in \mathbb{P}^{k+1}(\omega^{\mathrm{s}}_c)$ and all $\underline{w}^{\mathrm{s}}_c \in \underline{\mathsf{U}}^{k,\mathrm{s}}_c$, it holds that $\mathsf{stab}^{\mathrm{s}}_c\left(\underline{\mathsf{I}}^{k,\mathrm{s}}_c w, \underline{w}^{\mathrm{s}}_c\right) = 0$.*

The global discrete bilinear form then reads: find $\underline{\phi}_h^{\mathrm{s}} \in \underline{\mathsf{U}}_h^{k,\mathrm{s}}$ such that, for all $\underline{w}_h^{\mathrm{s}} \in \underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}}$,

$$\mathcal{A}_h^{\mathrm{s}}(\underline{u}_h^{\mathrm{s}}, \underline{w}_h^{\mathrm{s}}) := \sum_{\omega_c^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}} \left(H^{\mathrm{s}}\, \mathsf{G}_c^{k,\mathrm{s}} \underline{\phi}_c^{\mathrm{s}}, \mathsf{G}_c^{k,\mathrm{s}} \underline{w}_c^{\mathrm{s}}\right)_{\omega_c^{\mathrm{s}}} + \sum_{\omega_c^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}} \sum_{\Gamma \in \mathcal{F}_c^{\mathrm{s}}} h_\Gamma^{-1} \left(\delta_{c\Gamma}^{k,\mathrm{s}} \underline{\phi}_c^{\mathrm{s}} - \delta_c^{k,\mathrm{s}} \underline{\phi}_c^{\mathrm{s}}, \delta_{c\Gamma}^{k,\mathrm{s}} \underline{w}_c^{\mathrm{s}} - \delta_c^{k,\mathrm{s}} \underline{w}_c^{\mathrm{s}}\right)_\Gamma \tag{6.39}$$

$$= \sum_{\omega_c^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}} \mathcal{A}_c^{\mathrm{s}}(\underline{u}_c^{\mathrm{s}}, \underline{w}_c^{\mathrm{s}}), \tag{6.40}$$

where the subspace $\underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}}$ is defined by

$$\underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}} := \left\{ \underline{w}_h^{\mathrm{s}} \in \underline{\mathsf{U}}_h^{k,\mathrm{s}} \mid w_\Gamma^{\mathrm{s}} = 0 \text{ for all } \Gamma \in \mathcal{F}_h^{\mathrm{s},\partial} \right\}.$$

The Dirichlet boundary condition $\phi^{\mathrm{s}} = \psi^{\mathrm{s}}$ is imposed strongly by prescribing the face degrees of freedom on $\partial\Omega_{\mathrm{fs}}$:

$$\phi_\Gamma^{\mathrm{s}} = \pi_\Gamma^{k,\mathrm{s}} \psi^{\mathrm{s}}, \qquad \forall \Gamma \in \mathcal{F}_h^{\mathrm{s},\partial}. \tag{6.41}$$

Once the discrete potential $\underline{\phi}_h^{\mathrm{s}}$ is obtained, the reconstructed gradient $\mathsf{G}_h^{k,\mathrm{s}} \underline{\phi}_h^{\mathrm{s}}$ provides a high-order representation of $\nabla_{\boldsymbol{x}} \phi^{\mathrm{s}}$ in each cell. The normal discharge transmitted to the fluid region is then computed as

$$H^{\mathrm{f}}\, \boldsymbol{u}^{\mathrm{f}} \cdot \mathbf{n}^{\mathrm{f}} = -H^{\mathrm{s}} \left(\mathsf{G}_h^{k,\mathrm{s}} \underline{\phi}_h^{\mathrm{s}}\right) \cdot \mathbf{n}^{\mathrm{s}} \quad \text{on } \partial\Omega_{\mathrm{fs}}. \tag{6.42}$$

Proposition 6.2.10 (*Well-posedness of the discrete interior problem*). *Assume that the coefficient H^{s} satisfies $0 < H_{\min} \le H^{\mathrm{s}}(\boldsymbol{x}) \le H_{\max} < +\infty$ a.e. in Ω_{s}, and that the stabilization $\mathsf{stab}_c^{\mathrm{s}}$ satisfies* (S1)–(S3) *for all $\omega_c^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}$. Then, for any Dirichlet datum $\psi^{\mathrm{s}} \in \mathsf{H}^{1/2}(\partial\Omega_{\mathrm{fs}})$, the discrete problem* (6.39) *admits a unique solution $\underline{\phi}_h^{\mathrm{s}} \in \underline{\mathsf{U}}_h^{k,\mathrm{s}}$ satisfying* (6.41). *Moreover, the discrete solution satisfies the stability estimate*

$$\|\underline{\phi}_h^{\mathrm{s}}\|_{1,h} \lesssim \|\psi^{\mathrm{s}}\|_{\mathsf{H}^{1/2}(\partial\Omega_{\mathrm{fs}})}.$$

Proof. See Appendix D. □

The algebraic structure of this HHO discretization may be enhanced with a static condensation procedure. The main point is that the HHO unknowns are split into cell unknowns and face unknowns. The cell unknowns are purely local and can therefore be eliminated element by element. The resulting global system is posed only on the mesh skeleton, and the interested reader is referred to Appendix E for more details.

6.2.5 Local monolithic DG/FV subcell method for the exterior region

The discretization of the NSW system in the exterior fluid domains Ω_{f} (or $\Omega_{\mathrm{f}\pm}$ in the pontoon configuration) follows the local monolithic DG/FV subcell strategy described in Chapter 3.

This method is either applied on $\mathcal{T}_h^\bullet$ with $\bullet = \mathrm{f}$ (cylinder region) or independently on each exterior mesh $\mathcal{T}_h^\bullet$ with $\bullet \in \{\mathrm{f}^-, \mathrm{f}^+\}$ (infinite pontoon). Each fluid domain is advanced in time through synchronized explicit solvers, while the coupling with the interior region Ω_{s} is enforced through the interface conditions at $\partial\Omega_{\mathrm{fs}}$. The discrete interior elliptic problem provides, at each time step, the discrete potential trace ψ_h^{s} and the corresponding normal discharge $H^{\mathrm{s}} \boldsymbol{u}_h^{\mathrm{s}} \cdot \mathbf{n} = \Lambda \psi_h^{\mathrm{s}}$, which serve as boundary data for the hyperbolic solvers on both sides of the obstacle.

For the exterior mesh(es) $\mathcal{T}_h^\bullet$, the conservative variables $\boldsymbol{v}_h^\bullet = (\eta_h^\bullet, \boldsymbol{q}_h^\bullet)$ satisfy the semi-discrete DG system

$$\left(\partial_t \boldsymbol{v}_h^\bullet, \varphi_h^\bullet\right)_{\mathcal{T}_h^\bullet} + \left(\mathcal{U}_h^\bullet \circ \boldsymbol{v}_h^\bullet, \varphi_h^\bullet\right)_{\mathcal{T}_h^\bullet} = 0,$$

where the nonlinear residual operator $\mathcal{U}_h^\bullet$ is defined as

$$\left(\mathcal{U}_h^\bullet \circ \boldsymbol{v}_h^\bullet, \varphi_h^\bullet\right)_{\mathcal{T}_h^\bullet} := -\left(\mathbb{F}\left(\boldsymbol{v}_h^\bullet\right), \varphi_h^\bullet\right)_{\mathcal{T}_h^\bullet} + \left\langle \mathbb{F}^*, \varphi_h^\bullet \right\rangle_{\partial\mathcal{T}_h^\bullet},$$

where $\mathbb{F}(\boldsymbol{v}_h^\bullet)$ denotes the physical flux of the shallow-water equations and $\mathbb{F}^*$ is a consistent numerical flux ensuring stability and conservation across mesh interfaces, typically chosen as a LF or HLL-type flux ensuring both stability and conservation (see Chapter 3 for explicit definitions).

At the subcell level, the DG formulation admits an equivalent finite-volume interpretation. As detailed in the discrete flux reconstruction framework of Section 4.3, the weak DG form can be recast into a system of FV-like subcell equations for the subcell mean values

$$\overline{\boldsymbol{v}}_m^{c,\bullet} := \frac{1}{|S_m^{c,\bullet}|} \left(\boldsymbol{v}_h^{c,\bullet}, \mathbb{1}_{S_m^{c,\bullet}}\right)_{\omega_c^\bullet} = \frac{1}{|S_m^{c,\bullet}|} \int_{S_m^{c,\bullet}} \boldsymbol{v}_h^{c,\bullet}\, \mathrm{d}\boldsymbol{x}, \qquad \forall m \in [\![1, N_s^\bullet]\!]$$

within each DG element $\omega_c^\bullet$:

$$\frac{\mathrm{d}\overline{\boldsymbol{v}}_m^{c,\bullet}}{\mathrm{d}t} = -\frac{1}{|S_m^{c,\bullet}|}\sum_{S_p^{v,\bullet}\in\mathcal{V}_m^{c,\bullet}} \ell_{mp}^\bullet \widehat{\mathbb{F}}_{mp}^\bullet, \qquad \forall m\in[\![1,N_s^\bullet]\!], \quad \forall \omega_c^\bullet\in\mathcal{T}_h^\bullet,$$

where $\widehat{\mathbb{F}}_{mp}^\bullet$ denotes the reconstructed subcell flux derived from the DG residuals:

$$\widehat{\mathbb{F}}_c^\bullet = -\mathbb{A}_c^{\bullet,\top}\mathscr{L}_c^{-1,\bullet}\left(\mathbb{D}_c^\bullet\mathbb{P}_c^\bullet\mathbb{M}_c^{-1,\bullet}\boldsymbol{\Xi}_c^\bullet + \partial\mathbb{F}_c^\bullet\right),$$

where, for each element $\omega_c^\bullet\in\mathcal{T}_h^\bullet$:

- $\mathbb{R}^{N_k}\ni\boldsymbol{\Xi}_c^\bullet$ is the vector of DG residual i.e.

$$(\boldsymbol{\Xi}_p^{c,\bullet})_{p\in[\![1,N_k^\bullet]\!]} := \sum_{\Gamma\in\mathcal{F}_c}\int_\Gamma \mathbb{F}_\Gamma^{*,\mathrm{DG}}\psi_p^c\,\mathrm{d}s - \int_{\omega_c}\mathbb{F}(\boldsymbol{v}_h^c,b_h^c)\cdot\nabla_{\boldsymbol{x}}\psi_p^c\,\mathrm{d}\boldsymbol{x},$$

- $\mathscr{M}_{N_k^\bullet}(\mathbb{R})\ni\mathbb{M}_c$ is the mass matrix:

$$(\mathbb{M}_{mp}^c)_{(m,p)\in[\![1,N_k^\bullet]\!]^2} := \int_{\omega_c}\psi_m^c\psi_p^c\,\mathrm{d}\boldsymbol{x},$$

- $\mathscr{M}_{N_f^{c,\bullet}\times N_k^\bullet}(\mathbb{R})\ni\mathbb{P}_c^\bullet$ is the projection matrix allowing to switch from the DG polynomial to the subcell mean values, i.e. for all $S_m^{c,\bullet}\in\mathcal{T}_{\omega_c^\bullet}^k$,

$$(\mathbb{P}_c)_{(m,p)\in[\![1,N_s^\bullet]\!]\times[\![1,N_k^\bullet]\!]} = \frac{1}{|S_m^c|}\int_{S_m^c}\psi_p^c(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x},$$

- $\mathbb{R}^{N_f^{c,\bullet}}\ni\mathbb{F}_c^\bullet$ collects the traces of the normal component of the interior flux on the sub-interfaces s.t. for all $S_m^{c,\bullet}\in\mathcal{T}_{\omega_c^\bullet}^k$,

$$(\ell_{mp}^\bullet\mathbb{F}_{mp}^{c,\bullet})_{p\in[\![1,N_f^{c,\bullet}]\!]} := \int_{\Gamma_{mp}^{c,\bullet}}\pi_h^{k,\bullet}\mathbb{F}(\boldsymbol{v}_h)\cdot\mathbf{n}_{mp}\,\mathrm{d}s,$$

- $\mathscr{M}_{N_k^\bullet\times N_f^{c,\bullet}}(\mathbb{R})\ni\mathbb{A}_c^\bullet$ stands for the adjacency matrix of cell ω_c, where $\left(\mathbb{A}_{mp}^{c,\bullet}\right)_{(m,p)\in[\![1,N_k^\bullet]\!]\times[\![1,N_f^{c,\bullet}]\!]} = \varepsilon_{mp}^c$,

- $\mathbb{R}^{N_s^\bullet}\ni\partial\mathbb{F}_c^\bullet$ contains the cell boundary contribution as the integral of DG fluxes over the cell boundary, i.e.

$$(\partial\mathbb{F}_m^{c,\bullet})_{m\in[\![1,N_s^\bullet]\!]} = \int_{\partial\omega_c^\bullet\cap\partial S_m^{c,\bullet}}\mathbb{F}^*\cdot\mathbf{n}_{mp}\,\mathrm{d}s,$$

- $\mathscr{M}_{N_f^{c,\bullet}\times N_s^\bullet}(\mathbb{R})\ni\mathscr{L}_c^{-1,\bullet}$ is the inverse of the Laplacian matrix of the interior subgrid graph $\mathbb{L}_c^\bullet := \mathbb{A}_c^{\bullet,\top}\mathbb{A}_c^\bullet$ on the orthogonal of its kernel. For any $\lambda\neq 0$, this generalized inverse writes

$$\mathscr{L}_c^{-1,\bullet} = \left(\mathbb{L}_c^\bullet + \frac{\lambda}{N_s^\bullet}\mathbf{1}\otimes\mathbf{1}\right)^{-1} - \frac{1}{\lambda N_s^\bullet}\mathbf{1}\otimes\mathbf{1}.$$

Remark 6.2.11 (*Reconstruction of subcell fluxes*). *In Chapter 3, we display two different strategies to obtain the equivalency between the DG and FV subcell formulations. In the present work, we rely on reconstructing the subcell fluxes from the DG residuals, which is more efficient and easier to implement.*

In order to enhance robustness in the presence of steep gradients or near wet–dry transitions, thanks to the above FV subcell reformulation, we can apply a local monolithic DG/FV subcell stabilization operator to the polynomial representation $\boldsymbol{v}_h^{n,\bullet}$ at each time step or substep:

$$\mathcal{S}_h^{n,\star,\bullet}:\mathbf{P}^k(\mathcal{T}_h^\bullet)\ni\boldsymbol{v}_h^{n,\bullet}\mapsto\mathcal{S}_h^{n,\star,\bullet}\circ\boldsymbol{v}_h^{n,\bullet} = (\overline{\pi}_h^{k,\bullet})^{-1}\circ\overline{\boldsymbol{v}}_h^{n,\star,\bullet}\in\mathbf{P}^k(\mathcal{T}_h^\bullet),$$

where $\overline{\boldsymbol{v}}_h^{n,\star,\bullet}\in\mathbf{P}^0(\widetilde{\mathcal{T}}_h^\bullet)$ denotes the limited subcell mean values obtained from the convex limiting procedure introduced in Section 3.3.

6.2.6 Semi-discrete DG–HHO formulation

We now introduce the semi-discrete formulation of the coupled fluid–structure problem (6.15)-(6.18):

Find $\left(\boldsymbol{\psi}_h^{\mathrm{s}}, \underline{\phi}_h^{\mathrm{s}}, \boldsymbol{v}_h^{\bullet}\right) \in \mathscr{C}^1\big(0,T;L^2(\partial\Omega_{\mathrm{fs}})\big) \times \underline{\mathsf{U}}_h^{k,\mathrm{s}} \times \mathscr{C}^1\big(0,T;\mathbf{P}^k(\mathcal{T}_h^{\bullet})\big)$ such that, for $\bullet \in \{\mathrm{f},\mathrm{f}^-,\mathrm{f}^+\}$, the following system holds:

$$\begin{cases} \partial_t \boldsymbol{\psi}_h^{\mathrm{s}} = -\mathrm{g}\zeta_h - \frac{1}{2}\left|\boldsymbol{u}_h\right|^2 & \text{on } (\mathcal{F}_h^{\mathrm{s},\partial} \cap \mathcal{F}_h^{\bullet,\partial}) \times (0,T), \\ H^{\bullet}\left(\nabla_{\boldsymbol{x}}\phi^{\bullet}\cdot\mathbf{n}^{\bullet}\right) = -\boldsymbol{\Lambda}_h\boldsymbol{\psi}_h^{\mathrm{s}} & \text{on } (\mathcal{F}_h^{\mathrm{s},\partial} \cap \mathcal{F}_h^{\bullet,\partial}) \times (0,T), \\ \left(\partial_t \boldsymbol{v}_h^{\bullet}, \varphi_h^{\bullet}\right)_{\mathcal{T}_h^{\bullet}} + \left(\mathcal{U}_h^{\bullet}\circ\boldsymbol{v}_h^{\bullet}, \varphi_h^{\bullet}\right)_{\mathcal{T}_h^{\bullet}} = 0 & \forall \varphi_h^{\bullet} \in \mathbb{P}^k(\mathcal{T}_h^{\bullet}), \\ \boldsymbol{v}_h^{\bullet}(0) = (\overline{\pi}_h^{k,\bullet})^{-1}\overline{\boldsymbol{v}}_h^{\bullet,0}, \qquad \boldsymbol{\psi}_h^{\mathrm{s}}(0) = \boldsymbol{\psi}_h^{\mathrm{s,in}} & \text{in } \mathcal{T}_h^{\bullet}\times\{0\} \text{ and on } \mathcal{F}_h^{\mathrm{s},\partial}\times\{0\}, \end{cases} \tag{6.43}$$

where the DtN operator $\Lambda_h : L^2(\partial\Omega_{\mathrm{fs}}) \to L^2(\partial\Omega_{\mathrm{fs}})$ is defined autonomously for every data ψ_h^{s} by

$$\boldsymbol{\Lambda}_h\boldsymbol{\psi}_h^{\mathrm{s}} := H^{\mathrm{s}}\left(\mathbf{G}_h^{k,\mathrm{s}}\underline{\phi}_h^{\mathrm{s}}\right)\cdot\mathbf{n}^{\mathrm{s}} \quad \text{on } (\mathcal{F}_h^{\mathrm{s},\partial} \cap \mathcal{F}_h^{\bullet,\partial}) \times (0,T),$$

with $\underline{\phi}_h^{\mathrm{s}} \in \underline{\mathsf{U}}_h^{k,\mathrm{s}}$ being the unique solution of the local HHO problem

$$\begin{cases} \mathscr{A}_h^{\mathrm{s}}\left(\underline{\phi}_h^{\mathrm{s}}, \underline{w}_h^{\mathrm{s}}\right) = 0 & \forall \underline{w}_h^{\mathrm{s}} \in \underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}}, \\ \phi_{\Gamma,h}^{\mathrm{s}} = \pi_{\Gamma}^{k,\mathrm{s}}\boldsymbol{\psi}_h^{\mathrm{s}} & \forall \Gamma \in \mathcal{F}_h^{\mathrm{s},\partial}. \end{cases}$$

Remark 6.2.12 (*On the discrete DtN operator*). *As highlighted in the above formulation, the discrete Dirichlet-to-Neumann operator $\boldsymbol{\Lambda}_h$ serves as a critical coupling mechanism between the interior elliptic problem and the exterior hyperbolic solvers. Note that its implementation can be adapted to the specific geometry of the pontoon configuration, as the operator $\boldsymbol{\Lambda}_h$ is explicitly known in Fourier space.*

6.2.7 Fully-discrete monolithic DG/FV–HHO–SSP-RK formulation

We now present the fully-discrete counterpart of the semi-discrete system (6.43). For clarity, we describe the first-order explicit Euler version; higher-order SSP–Runge–Kutta schemes follow by adjusting the time-stage indices accordingly.

Assume that all required quantities $\left(\boldsymbol{\psi}_h^{\mathrm{s},n}, \underline{\phi}_h^{\mathrm{s},n}, \boldsymbol{v}_h^{\bullet,n}\right)$ are known at time level t^n. The fully-discrete DG–HHO formulation then reads as follows:

Find $\left(\boldsymbol{\psi}_h^{\mathrm{s},n+1}, \underline{\phi}_h^{\mathrm{s},n+1}, \boldsymbol{v}_h^{\bullet,n+1}\right) \in L^2(\partial\Omega_{\mathrm{fs}}) \times \underline{\mathsf{U}}_h^{k,\mathrm{s}} \times \mathbf{P}^k(\mathcal{T}_h^{\bullet})$ for $\bullet \in \{\mathrm{f},\mathrm{f}^-,\mathrm{f}^+\}$, the following system holds:

$$\begin{cases} \boldsymbol{\psi}_h^{\mathrm{s},n+1} = \boldsymbol{\psi}_h^{\mathrm{s},n} + \Delta t^n\left(-\mathrm{g}\zeta_h^n - \frac{1}{2}|\boldsymbol{u}_h^n|^2\right) & \text{on } \mathcal{F}_h^{\mathrm{s},\partial} \cap \mathcal{F}_h^{\bullet,\partial}, \\ H^{\bullet}\left(\nabla_{\boldsymbol{x}}\phi^{\bullet,n+1}\cdot\mathbf{n}^{\bullet}\right) = -\boldsymbol{\Lambda}_h^{n+1}\boldsymbol{\psi}_h^{\mathrm{s},n+1} & \text{on } \mathcal{F}_h^{\mathrm{s},\partial} \cap \mathcal{F}_h^{\bullet,\partial}, \\ \boldsymbol{v}_h^{\bullet,n+1} = \mathcal{S}_h^{n,\star,\bullet}\circ\left\{\boldsymbol{v}_h^{\bullet,n} - \Delta t^n\left(\mathcal{U}_h^{\bullet}\circ\boldsymbol{v}_h^{\bullet,n}\right)\right\} & \forall \varphi_h^{\bullet} \in \mathbb{P}^k(\mathcal{T}_h^{\bullet}), \\ \boldsymbol{v}_h^{\bullet,0} = (\overline{\pi}_h^{0,k,\bullet})^{-1}\overline{\boldsymbol{v}}_h^{\bullet,0}, \qquad \boldsymbol{\psi}_h^{\mathrm{s},0} = \boldsymbol{\psi}_h^{\mathrm{s,in}} & \text{in } \mathcal{T}_h^{\bullet} \text{ and on } \mathcal{F}_h^{\mathrm{s},\partial}, \end{cases} \tag{6.44}$$

where, for every boundary datum $\boldsymbol{\psi}_h^{\mathrm{s},n+1}$, the discrete DtN operator reads

$$\boldsymbol{\Lambda}_h^{n+1}\boldsymbol{\psi}_h^{\mathrm{s},n+1} := H^{\mathrm{s}}\left(\mathbf{G}_h^{k,\mathrm{s}}\underline{\phi}_h^{\mathrm{s},n+1}\right)\cdot\mathbf{n}^{\mathrm{s}} \quad \text{on } \mathcal{F}_h^{\mathrm{s},\partial} \cap \mathcal{F}_h^{\bullet,\partial}, \tag{6.45}$$

and $\underline{\phi}_h^{\mathrm{s},n+1} \in \underline{\mathsf{U}}_h^{k,\mathrm{s}}$ is the unique solution at time t^{n+1} of the problem

$$\begin{cases} \mathscr{A}_h^{\mathrm{s}}\left(\underline{\phi}_h^{\mathrm{s},n+1}, \underline{w}_h^{\mathrm{s}}\right) = 0 & \forall \underline{w}_h^{\mathrm{s}} \in \underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}}, \\ \phi_{\Gamma,h}^{\mathrm{s},n+1} = \pi_{\Gamma}^{k,\mathrm{s}}\boldsymbol{\psi}_h^{\mathrm{s},n+1} & \forall \Gamma \in \mathcal{F}_h^{\mathrm{s},\partial}, \end{cases} \tag{6.46}$$

The coupled system (6.43) encapsulates the full dynamics of the problem, and one can describe the role of each component as follows:

(*i*) The interface ODE that governs the evolution of the local potential $\boldsymbol{\psi}_h^{\mathrm{s}}$ is solved explicitly on $\mathcal{F}_h^{\mathrm{s},\partial} \cap \mathcal{F}_h^{\bullet,\partial}$,

(*ii*) The solution of the previous interface ODE provides the necessary boundary data for the interior elliptic problem, that defines the discrete realization of the DtN operator,

(*iii*) The normal flux at the interface is computed through the DtN operator and used as boundary condition for the hyperbolic solver. Indeed, on the structural boundary $\partial\Omega_{\mathrm{fs}}$, the standard numerical flux $\mathbb{F}^*$ is substituted by the incoming coupling term $\boldsymbol{\Lambda}_h \boldsymbol{\psi}_h^{\mathrm{s}}$, effectively tying the exterior fluid motion to the interior elliptic constraint.

Remark 6.2.13 (*Coupling fluxes and DtN operator*). *At a given Runge–Kutta stage, solving the interior elliptic problem yields a discrete approximation of the Dirichlet-to-Neumann operator $\boldsymbol{\Lambda}_h$ on the interface $\partial\Omega_{\mathrm{fs}}$. Within the HHO framework, this operator leads to a polynomial reconstruction localized on each contact-line face. More precisely, for every face $\Gamma \in \mathcal{F}_h^{\mathrm{s},\partial} \cap \mathcal{F}_h^{\bullet,\partial}$, one obtains a polynomial function $\lambda_{h|\Gamma} \in \mathbb{P}^k(\Gamma)$ approximating the normal flux. This polynomial representation naturally enables a high-order coupling with the exterior hyperbolic solver.*

Remark 6.2.14 (*Numerical formulation*). *Although the continuous model is written above in terms of the variables $(\zeta^{\mathrm{f}}, \boldsymbol{u}^{\mathrm{f}})$, the numerical experiments presented in this chapter are performed with the variables $(\eta^{\mathrm{f}}, \boldsymbol{q}^{\mathrm{f}})$, with $\boldsymbol{q}^{\mathrm{f}} = H^{\mathrm{f}}\boldsymbol{u}^{\mathrm{f}}$, as in Chapter 3. This is a practical choice: we reuse the robust monolithic DG/FV subcell solver developed previously, together with its well-balanced treatment, positivity-preserving mechanism and shock capturing properties. This also means that the exterior solver does not explicitly enforce the irrotationality constraint appearing in the continuous reformulation. The resulting scheme should therefore be understood as a pragmatic conservative approximation of the coupled model, designed to test the DG/FV–HHO coupling mechanism while retaining the numerical robustness of the solver introduced in Chapter 3. A fully dedicated discretization of the conservative irrotational formulation requires to produce a new exterior hyperbolic solver (relying on the variables $(\zeta^{\mathrm{f}}, \boldsymbol{u}^{\mathrm{f}})$) and is beyond the scope of the present chapter.*

6.3 Numerical experiments

In this section, we present three numerical experiments devoted to the validation and illustration of the coupled DG/FV–HHO strategy introduced in the previous sections. First, we check that the different numerical realizations of the DtN operator give consistent results in a configuration where a reference reduced model is available. Second, we use the fully two-dimensional setting to investigate wave interactions with fixed surface obstacles of more general shape.

Compared with the numerical tests presented in Chapter 5, the present experiments are not only designed to recover a one-dimensional behavior. The model developed in this chapter is genuinely two-dimensional, since the coupling takes place along a two-dimensional interface and the interior response of the structure is obtained either from a two-dimensional elliptic problem or from its explicit counterpart in the special pontoon geometry. Nevertheless, the infinite pontoon configuration remains useful as a benchmark, because it admits a pseudo-one-dimensional reduction and then allows us to compare three different evaluations of the same coupling mechanism: the explicit one-dimensional formula, the Fourier-based DtN operator, and the HHO approximation of the interior elliptic problem.

6.3.1 Geometry of the structures

We consider two types of partially immersed fixed obstacles: the first one is the infinite pontoon of Section 5.4.1: when the solution does not depend on y, the two-dimensional pontoon model reduces to the one-dimensional wave-structure interactions model studied in Chapter 5. In this regime, the interior discharge can be computed explicitly from the two traces ψ_-^{s} and ψ_+^{s} as detailed in Section 6.1.4. This reduced formula provides a direct reference for the FFT-based and HHO-based evaluations of the DtN operator.

The second is a cylindrical obstacle: in the horizontal plane, the wetted region below the obstacle is an open ball

$$\Omega_{\mathrm{s}} = \left\{ \boldsymbol{x} \in \mathbb{R}^2 \mid |\boldsymbol{x} - \boldsymbol{x}_{\mathrm{c}}| < R \right\} =: B(\boldsymbol{x}_{\mathrm{c}}, R),$$

where $\boldsymbol{x}_{\mathrm{c}} = \boldsymbol{0}$ is the center of the obstacle and $R > 0$ is its radius. In the numerical tests below, we take $R = 5$. The exterior fluid region is defined as $\Omega_{\mathrm{f}} = \Omega \setminus \overline{\Omega_{\mathrm{s}}}$ and the fluid-structure interface is the closed curve,

$$\partial\Omega_{\mathrm{fs}} = \left\{ \boldsymbol{x} \in \mathbb{R}^2 \mid |\boldsymbol{x} - \boldsymbol{x}_{\mathrm{c}}| = R \right\} =: D(\boldsymbol{x}_{\mathrm{c}}, R),$$

and the normal vector $\mathbf{n}^{\mathrm{s}}$ is oriented from the obstacle region towards the exterior fluid region.

The immersed underside of the cylinder is defined through the equilibrium water-height as follows:

$$H^{\mathrm{s}}(\boldsymbol{x}) = H_0^{\mathrm{s}} + \kappa_{\mathrm{b}} \left(R^2 - |\boldsymbol{x} - \boldsymbol{x}_{\mathrm{c}}|^2 \right),$$

for $|\boldsymbol{x} - \boldsymbol{x}_{\mathrm{c}}| \leq R$ and we set $H_0^{\mathrm{s}} = 5$ m and $\kappa_{\mathrm{b}} = 1$ m.

6.3.2 Wave-pontoon interactions and DtN solver comparison

PARAMETERS

Polynomial degree: $k = 3$
Time-marching order: $p = 3$
Final time: $T_{\max} = 6$ sec
Boundary conditions: homogeneous Neumann

Computational domain:
$\Omega_{\mathrm{f}-} = [-40, -5] \times [-20, 20]$ ($N_{\mathrm{el}} = 894$)
$\Omega_{\mathrm{s}} = [-5, 5] \times [-20, 20]$ ($N_{\mathrm{el}} = 368$)
$\Omega_{\mathrm{f}+} = [5, 40] \times [-20, 20]$ ($N_{\mathrm{el}} = 892$)

We consider the interaction between a single incoming wave and a fixed pontoon. This test is performed in the infinite pontoon configuration introduced in Section 6.3.1.

First, we check that the coupled DG/FV–HHO strategy reproduces the expected qualitative behaviour: an incoming wave reaches the pontoon, part of the wave is reflected back to the left exterior domain, while another part is transmitted to the right exterior domain. Second, in the infinite pontoon geometry, the coupling can be evaluated in three different ways: (*i*) the explicit reduced formula obtained in the one-dimensional limit, (*ii*) the Fourier-based DtN operator, (*iii*) the HHO approximation of the interior elliptic problem.

The initial condition is the right-going single-wave of Section 5.4.4. In the exterior fluid domains, we prescribe

$$\eta_0(\boldsymbol{x}) = H_0 + A_{\mathrm{w}} \operatorname{sech}^2\left(\gamma(x - x_{\mathrm{c}})\right), \qquad \gamma = \sqrt{\frac{3A_{\mathrm{w}}}{4H_0}}, \quad H_0 = 10, \quad A_{\mathrm{w}} = 4.35, \quad x_{\mathrm{c}} = -30.$$

and the corresponding initial discharge:

$$\boldsymbol{q}_0(\boldsymbol{x}) = \left(\sqrt{\frac{\mathrm{g}}{H_0}}\,(\eta_0(\boldsymbol{x}) - H_0)\,\eta_0(\boldsymbol{x}), 0\right)^{\top}.$$

This single-wave is initially located in the left exterior domain and propagates towards the pontoon. Since the object is fixed, the response is entirely produced by the constraint imposed below the pontoon: the incoming wave reaches the left contact line of the pontoon, a transmitted wave propagates in the right exterior domain and a reflected wave travels back leftward. This behaviour is shown in the two-dimensional snapshots of Figure 6.5 and in the one-dimensional profiles of Figure 6.6, both obtained using the HHO method for the elliptic part.

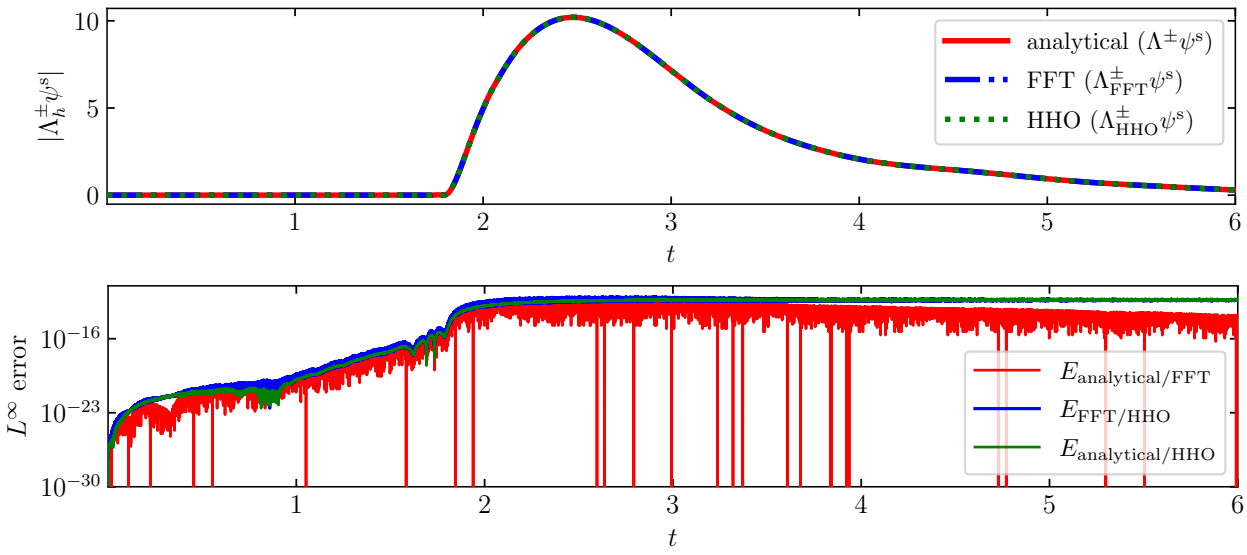


Figure 6.4: Wave-pontoon interactions – Comparison of the transmitted discharge computed with the analytical formula, the FFT-based DtN operator, and the HHO approximation, together with the pairwise L^{∞} errors between these three evaluations.

We observe that the solver reproduces the expected interactions without producing any transverse artefacts, and the reflected and transmitted components are well resolved. This shows that the DG/FV subcell correction handles the steep fronts generated by the interaction in a stable way. If one wants to see the long-term evolution of the reflected and transmitted waves, the simulation can be continued for a longer time, and can expect the same result as Figure 5.14 from Chapter 5. Readers can find an animation of this benchmark in the following link.

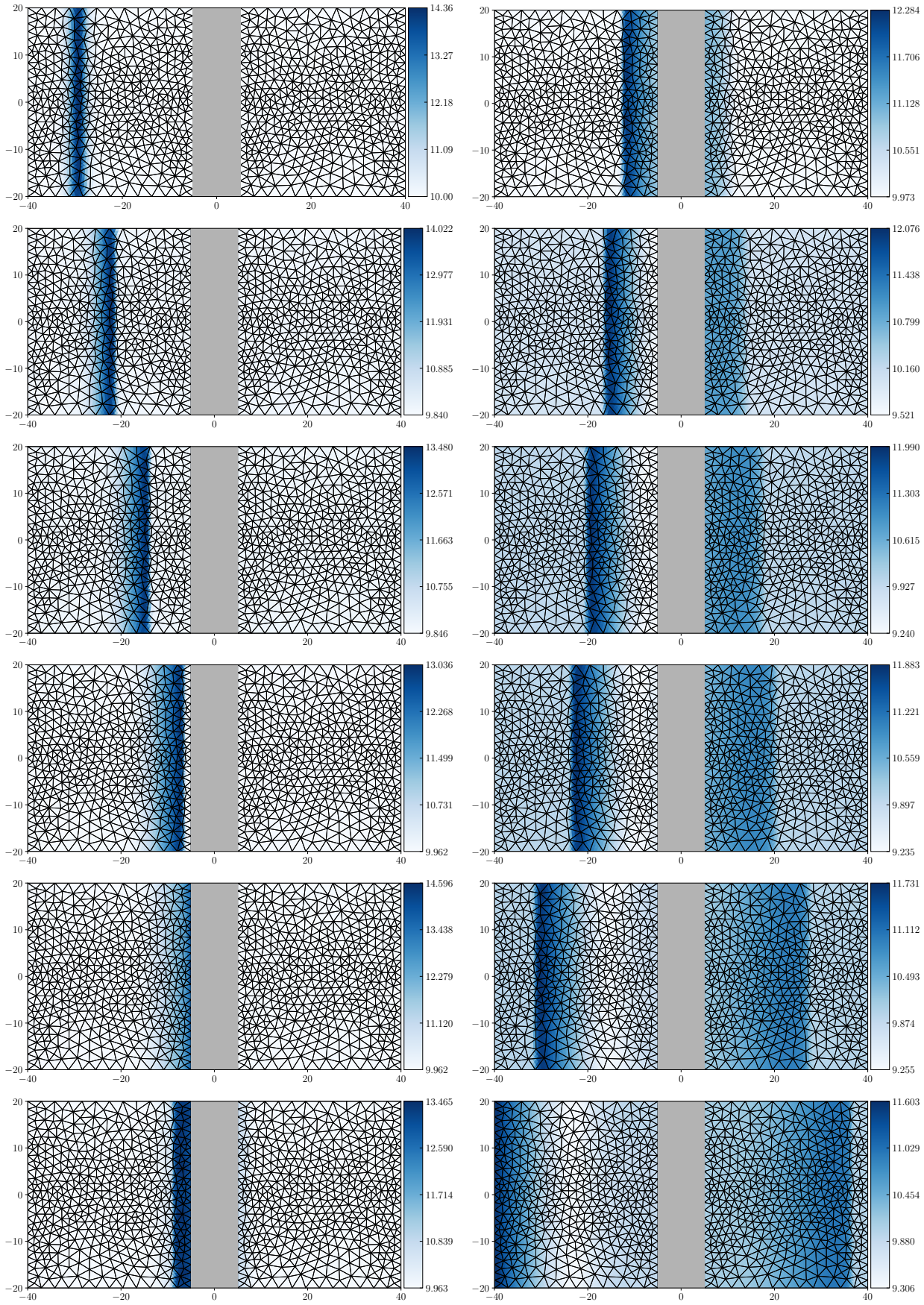


Figure 6.5: Wave-pontoon interactions – Time evolution of the free-surface elevation η during the period $(0, 5]$ sec, ordered columnwise from top to bottom (animation).

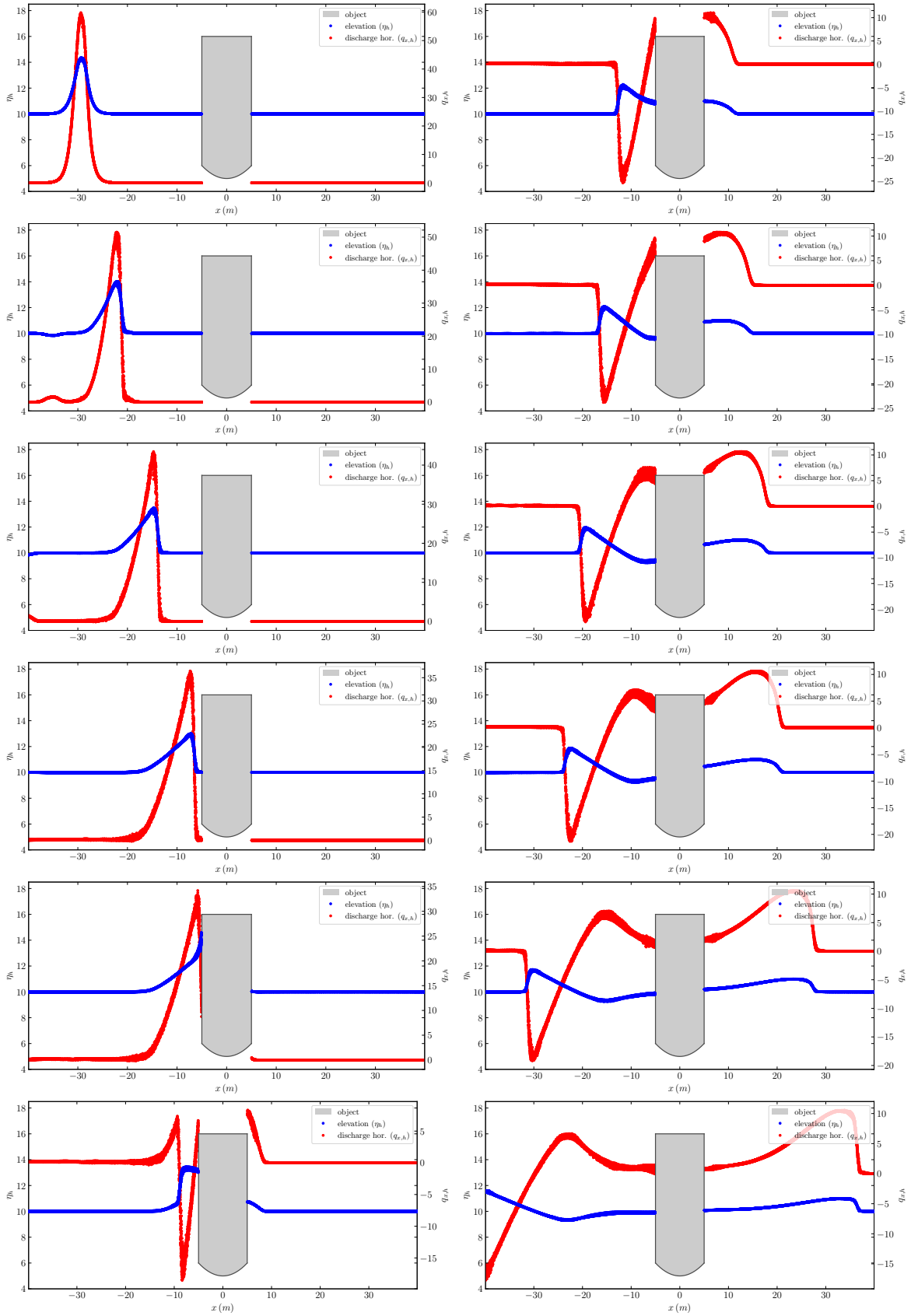


Figure 6.6: Wave-pontoon interactions – Columnwise, from top to bottom: time evolution of free-surface elevation η and horizontal discharge q_x at several times in the period $(0, 5]$ sec (animation).

We also compare the three numerical evaluations of the coupling operator. At each time, the normal discharge is evaluated on the two contact sides of the pontoon and we measure the discrepancy between two evaluations $\mathfrak{A}$ and $\mathfrak{B}$ of the DtN response through the discrete L^∞ error

$$E_{\mathfrak{A}/\mathfrak{B}}(t) = \max_{\bullet\in\{-,+\}} \max_{x\in\Omega_{\mathrm{f}\bullet}} \left| q_x^{\mathfrak{A}}(t,\bullet l,y) - q_x^{\mathfrak{B}}(t,\bullet l,y) \right|.$$

In particular, we consider the three pairwise comparisons

$$(\mathfrak{A},\mathfrak{B}) \in \{(\text{analytical},\text{FFT}),(\text{analytical},\text{HHO}),(\text{FFT},\text{HHO})\}.$$

The corresponding discharge and error curves are shown in Figure 6.4, where we observe that the three transmitted discharges are almost superposed during the whole simulation. The analytical and Fourier evaluations are almost indistinguishable. Their error starts around 10^{-28} and remains at machine precision, while the largest error observed in the three comparisons is about 9×10^{-12}.

Time series are shown on Figure 6.7, confirming that the three error curves have very similar variations in time. The small oscillatory patterns visible in this curve can occur at extremely small amplitudes and are most likely due to round-off effects and the finite FFT representation.
The HHO-based comparisons are also very accurate, and the error remains very small compared with the magnitude of the transmitted discharge. The agreement displayed in Figures 6.4 and 6.7 confirms that the HHO approximation provides an accurate approximation of the DtN operator.

6.3.3 Incident wave interactions with a fixed cylindrical obstacle

We now consider the interactions between an incoming wave and a fixed cylindrical obstacle. This test is genuinely two-dimensional, as in contrast with the pontoon case, the interface is a closed curve, and the incoming wave interactions may generate complex patterns. The flow is expected to be diffracted and reflected by the cylinder, while producing direct transmitted perturbations downstream.

The aim of this test is to check that the coupled DG/FV–HHO method remains stable and gives a coherent wave-structure description with this new 2D geometry.

Flat-bottom case

PARAMETERS

Polynomial degree: $k=3$
Time-marching order: $p=3$
Final time: $T_{\max}=10$ sec
Boundary conditions: homogeneous Neumann

Computational domain:
$\Omega_{\mathrm{f}} = ([-40,40]\times[-20,20]) \setminus B(\mathbf{0},5)$ ($N_{\mathrm{el}}=2340$)
$\Omega_{\mathrm{s}} = B(\mathbf{0},5)$ ($N_{\mathrm{el}}=80$)

We first consider the flat-bottom case i.e. we still consider the topography $b\equiv0$. The initial condition is the same right-going single-wave as in the previous test. The wave is initially located in the left part of the exterior fluid domain and propagates towards the cylinder.
Figure 6.9 shows the time evolution of the free-surface elevation together with the corresponding blending coefficients. The computation is performed on the unstructured meshes displayed in Figure 6.8, where the exterior fluid mesh surrounds the circular solid region.

This test is physically more demanding than the previous one: the incident wave first reaches the upstream part of the circular interface and a reflected component is generated in front of the obstacle, while the flow is redirected along the cylinder interface. A transmitted perturbation emerges from below and interacts with the surface diffraction patterns. The method clearly reproduces these genuinely two-dimensional wave field patterns. One can observe the formation of a cylindrical shock wave produced near the obstacle and propagating towards the right part of the computational domain. This phenomenon is similar to the wave patterns already observed in benchmarks from Chapter 3, where comparable two-dimensional effects were produced by topographic features emerging from the water. This sequence is visible in the two-dimensional snapshots of Figure 6.9, and appears even more clearly in the associated animation.

In addition to the results of Chapter 3, we observe that the blending coefficient is often strongly reduced near the cylindrical interface; the relaxation of the blending coefficient near the obstacle can be interpreted as a stabilizing response of the scheme in a region where the flux is externally imposed by the coupled solver (and where no regularity is expected).

This flat-topography cylinder test confirms that the coupled DG/FV–HHO method may successfully handle problems involving closed fluid–solid interfaces, even with high-order polynomials approximations and coarse meshes, which is promising for practical applications.

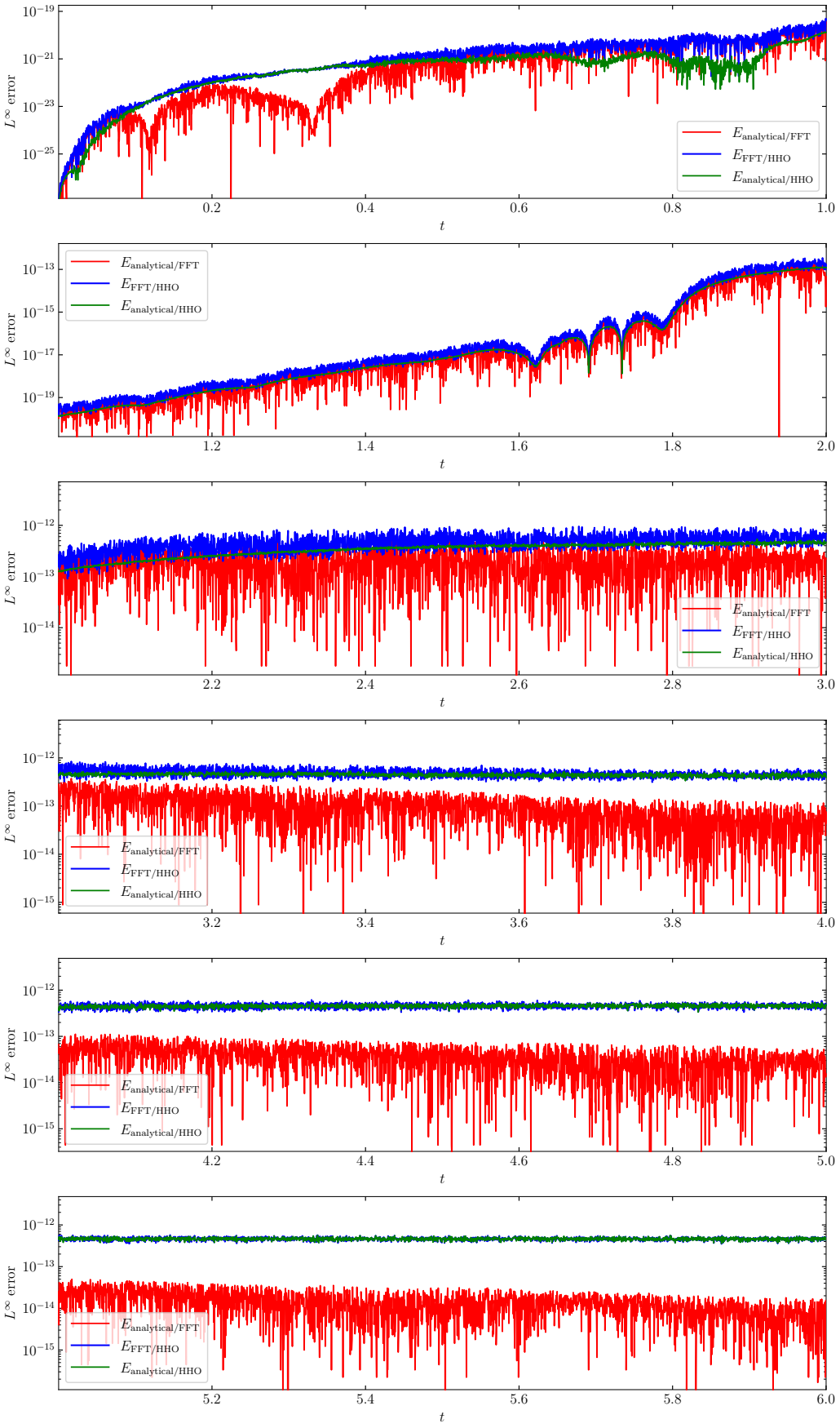


Figure 6.7: Wave-pontoon interactions – Time series of the pairwise L^∞ errors between the analytical formula, the FFT-based DtN operator and the HHO approximation.

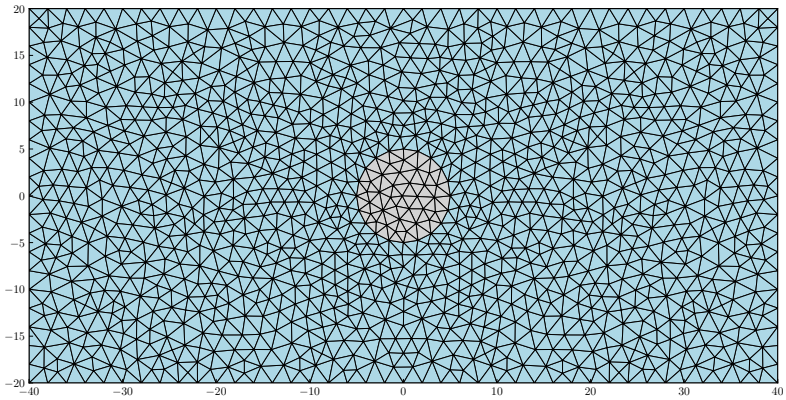


Figure 6.8: Unstructured meshes for Ω_f ($N_\mathrm{el} = 2340$) and Ω_s ($N_\mathrm{el} = 80$).

Non-flat bottom case

PARAMETERS

Polynomial degree: $k = 1$
Time-marching order: $p = 2$
Final time: $T_\mathrm{max} = 10$ sec
Boundary conditions: homogeneous Neumann

Computational domain:
$\Omega_\mathrm{f} = ([-40, 40] \times [-20, 20]) \setminus B(\mathbf{0}, 5)$ ($N_\mathrm{el} = 5060$)
$\Omega_\mathrm{s} = B(\mathbf{0}, 5)$ ($N_\mathrm{el} = 162$)

In this second variation, we consider the same wave-cylinder interactions problem, but with a non-flat topography, meaning we reintroduce the geometric topography source term, defined explicitly in Nonlinear shallow-water model in two space dimension.

The purpose of this second configuration is to test the coupled DG/FV–HHO method in a more demanding situation, where the wave patterns are affected both by the fixed cylindrical obstacle and by the topography. We consider a sloping beach on the right side of the cylinder, defined by

$$b(\boldsymbol{x}) = \max\big(0, x - 15\big).$$

The incident wave is still the same but the amplitude is now set to $A_\mathrm{w} = 8.35$ m, leading to higher nonlinear interactions.

This test is also particularly relevant to check the robustness of the method; indeed, the right part of the domain has a dry region (sloping beach). The interaction with the obstacle first generates a genuinely two-dimensional cylindrical shock wave, which then propagates rightwards and reaches the shoreline. This test also validates the compatibility between the well-balancing property, the positivity-preserving DG/FV subcell mechanism detailed in Chapter 3, and the HHO approximation of the interior elliptic problem. Several one-dimensional slices (at $y = 0$), together with the blending coefficients, are shown in Figure 6.10, while the corresponding free-surface profiles are shown in Figure 6.12.
These numerical results show that the cylindrical shock wave reaches the sloping beach with a run-up, and is then reflected leftward. This run-up and backwash motions are clearly visible in the numerical solution, as the computation captures this behaviour without visible spurious oscillations near the dry area. The two-dimensional evolutions of the free-surface elevation and of the discharge norm are displayed in Figure 6.11.
This test gives promising results: it shows that the coupled DG/FV–HHO method is robust for the two-dimensional wave-structure interactions problem. An animation of this benchmark is available here.

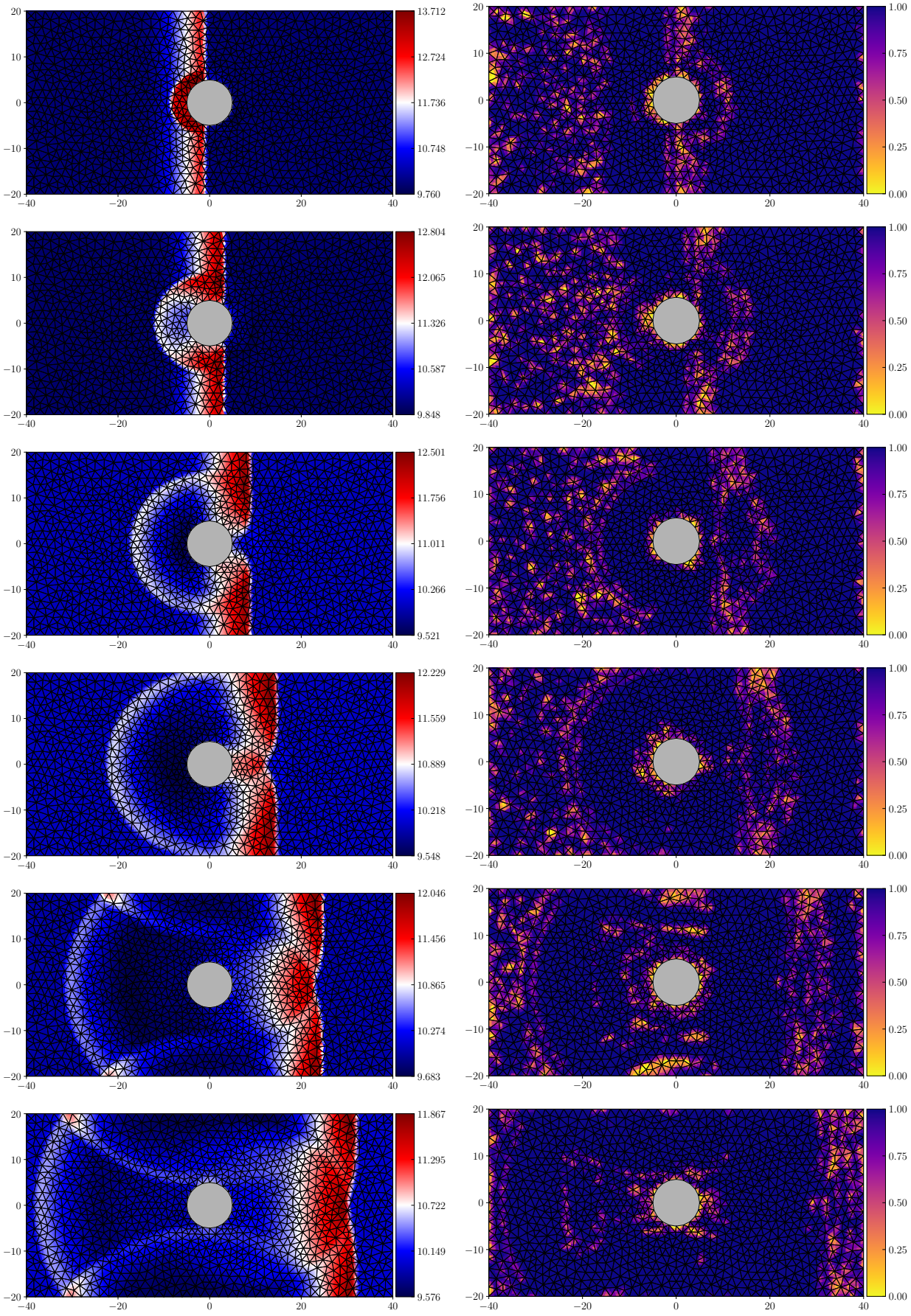


Figure 6.9: Single-wave interactions with a fixed cylinder − Free-surface elevation and blending coefficients at several times during the period $[2.15, 5.45]$ sec, ordered columnwise from top to bottom (animation).

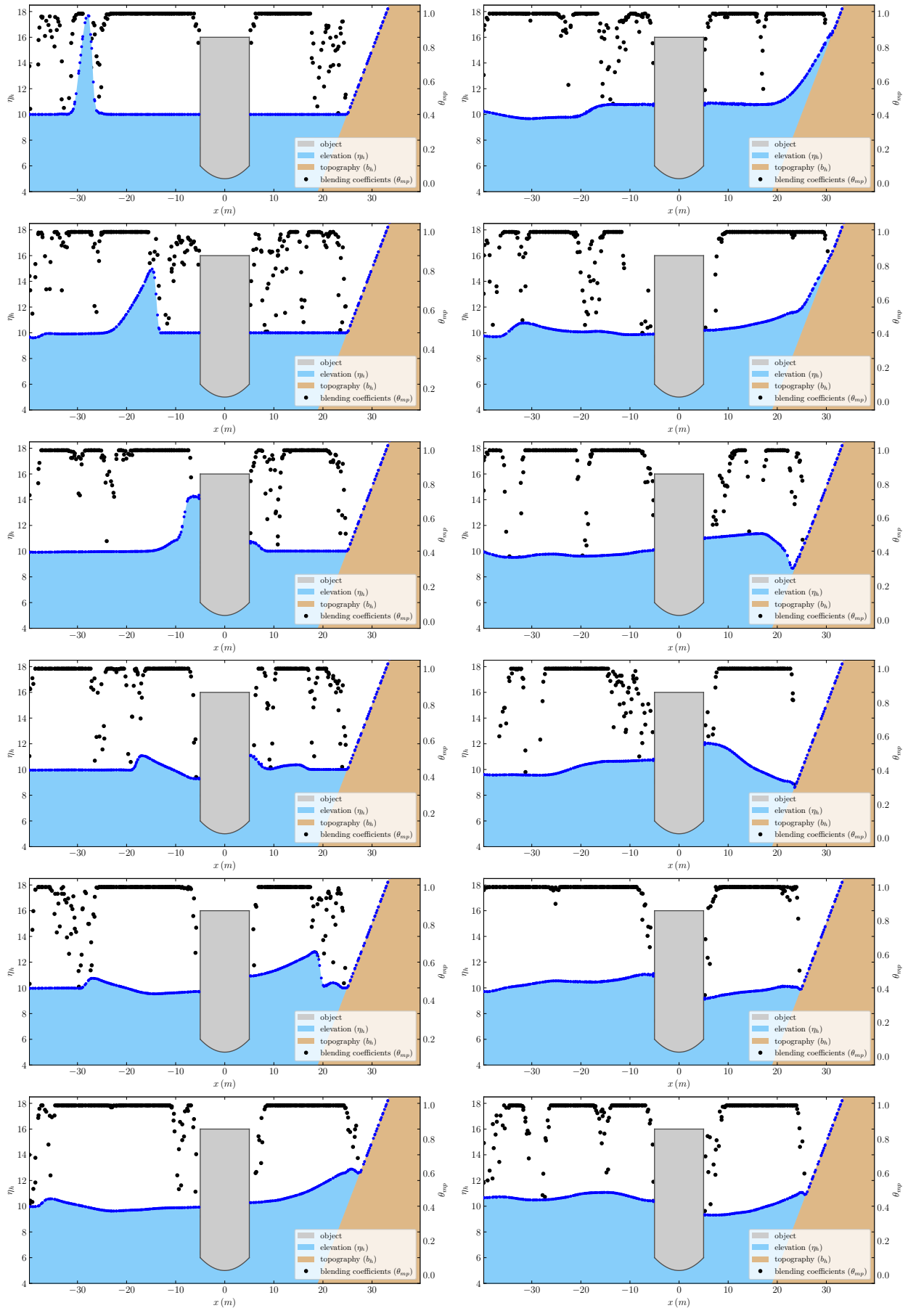


Figure 6.10: Single-wave interactions with a fixed cylinder over a sloping beach − Free surface elevation and blending coefficients during the period $[0, 10]$ sec on slice $y = 0$ (animation).

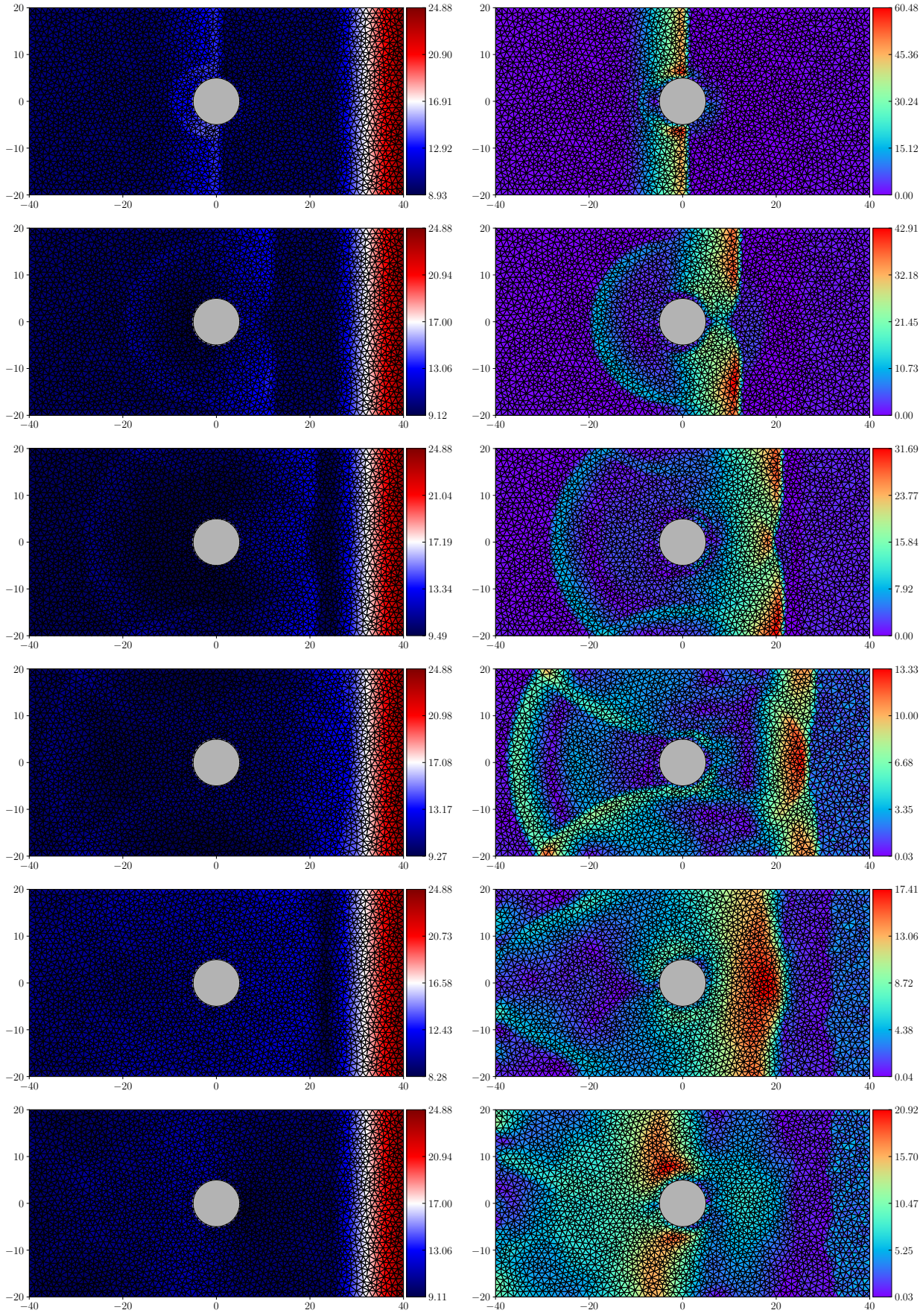


Figure 6.11: Single-wave interactions with a fixed cylinder over a sloping beach − Time evolution during the period $[2, 9]$ sec of the free-surface elevation (left) and discharge norm (right) (animation).

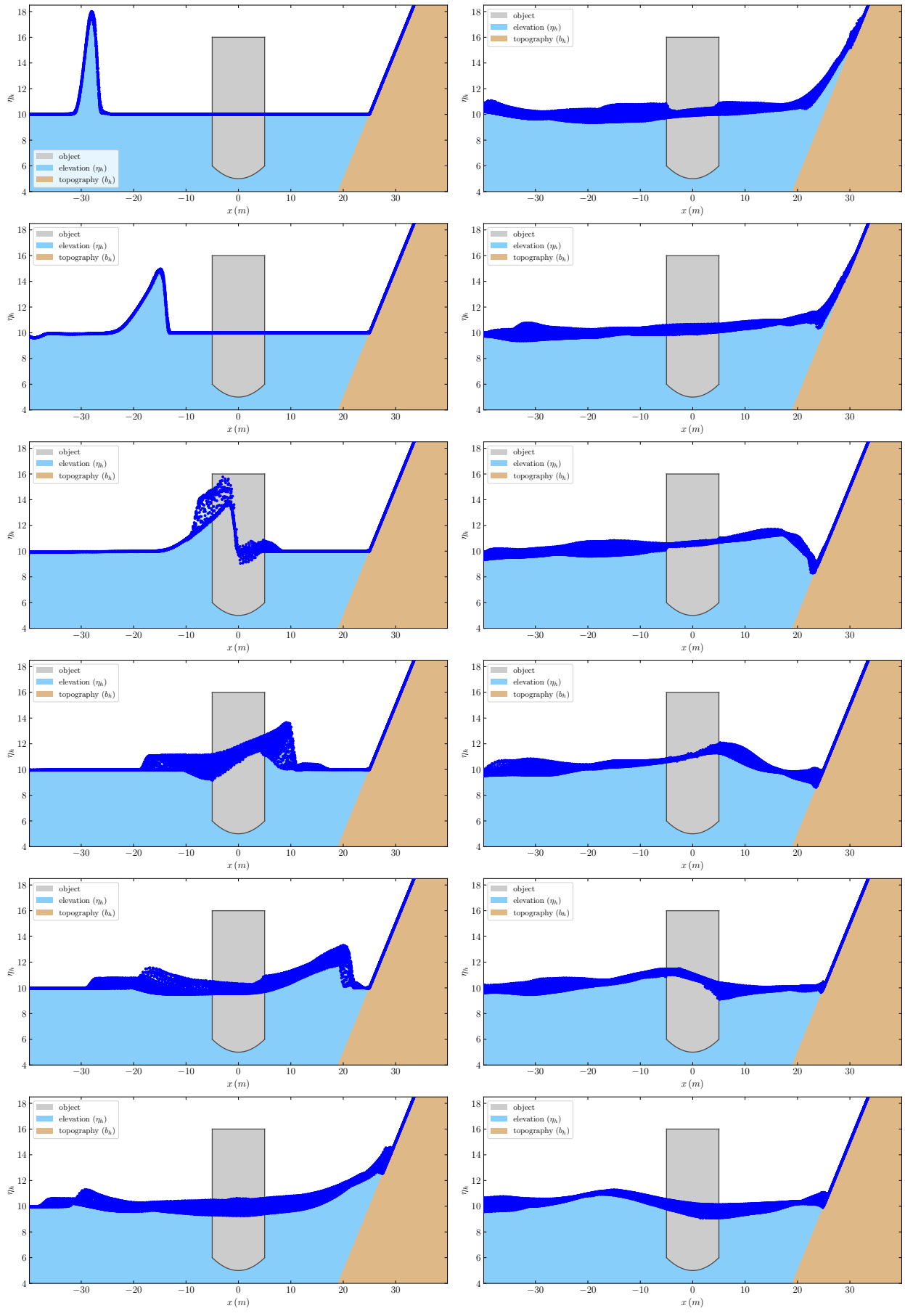


Figure 6.12: Single-wave interactions with a fixed cylinder over a sloping beach – Free surface elevation density profiles during the period $[0, 10]$ sec (animation).

Conclusion and perspectives

This thesis is devoted to the construction and the study of high-order numerical methods for shallow-water type models, with a special attention to stability, robustness, handling moving meshes, and wave-structure interactions. The main idea followed throughout the manuscript is to keep the accuracy of DG methods in smooth regions, while using lower-order FV methods where robustness is necessary. This leads to monolithic DG/FV schemes in which the high-order and low-order parts are combined inside an *a priori* update.

The starting point of this work is the construction of a monolithic DG/FV subcell method for the one-dimensional NSW. In Chapter 1, a local blending strategy between a high-order DG discretization and a first-order FV subcell scheme is introduced. Particular attention is paid to the treatment of bathymetry source term, positivity preservation, and the well-balancing property. This also leads to the introduction of a *two-scale hydrostatic reconstruction* method, both at the cell and subcell levels, allowing the method to preserve lake-at-rest states. Several numerical experiments show that this method is able to accurately capture smooth solutions while remaining robust near singularities.

Chapter 2 considers the one-dimensional SGN equations. These equations incorporate dispersive effects and provide a more accurate description of wave propagation in shallow-water. They can be written as a system of hyperbolic equations coupled with an additional dispersive source term arising from an auxiliary elliptic problem. The proposed strategy combines the monolithic DG/FV discretization of the hyperbolic part with a SWIP-DG discretization of the elliptic correction. This provides a first coupled hyperbolic–elliptic solver and shows that the monolithic framework can naturally be coupled with other approximation strategies for the study of dispersive wave equations.

This monolithic DG/FV framework is also developed for the two-dimensional NSW equations on unstructured meshes in Chapter 3, the method is generalized to simplicial meshes while preserving the main admissibility properties of the model. The resulting scheme is validated on a wide range of demanding benchmark problems, highlighting both the robustness and the accuracy of the approach, even considering high-order approximations on coarse meshes.

Once the fixed-grid framework had been established, Chapter 4 introduced an Arbitrary Lagrangian–Eulerian adaptation of the 2D method. The objective is to construct a conservative discreet formulation on moving meshes while preserving the structure of the NSW equations. Special attention is paid to the compatibility between mesh motions, which requires the study of auxiliary elliptic problems. This work paves the way for simulations involving free boundaries and moving floating structures.

The second part focuses on wave-structure interactions problems. In Chapter 5, a first model describing the interaction between shallow-water flows and a partially immersed floating structure in a pseudo-one-dimensional setting is considered. The coupled system combines NSW equations around the structure, an elliptic problem in the interior region below the object and a system of ODEs for the structure dynamics. A numerical method based on the monolithic DG/FV framework and a SWIP-DG discretization of the elliptic problem is developed and validated on several representative test cases involving a vertical motion (heaving) of the object.

Finally, Chapter 6 introduced a new genuinely two-dimensional wave-structure interactions model, involving a fixed cylindrical obstacle. The interior elliptic problem is discretized using a HHO method, which is particularly well suited to this type of problem, as it allows a naturel description of the DtN operator. The resulting DG/FV–HHO coupling is successfully validated on flat and non-flat bottom configurations. The numerical results show that the method is able to handle strong shocks, wave reflection, diffraction, refraction and transmission under the object, with wet-dry transitions, and complex interactions between the flow and the structure within a fully two-dimensional setting.

Beyond the contributions presented in the main body of the manuscript, very recent work has made it possible to assemble several of its principal components into functional prototypes for more freely moving surface obstacles. Although these latest developments are not presented as a dedicated contribution in the thesis, they build directly on the models, discretizations and software components developed throughout the manuscript.

A first prototype concerns the cylindrical obstacle studied in Chapter 6. The fixed-object formulation has been extended with a vertical degree of freedom, following the model described in Appendix C. The horizontal projection of the cylinder and the surrounding planar meshes remain fixed, while the wetted surface moves vertically. The resulting implementation combines the exterior monolithic DG/FV solver, an HHO treatment of the constrained flow and Bernoulli pressure below the object, and Newton's equation for the heave motion. A prescribed vertical oscillation was first used to verify the transfer of the body motion to the fluid and the generation of outgoing cylindrical waves. The complete coupling was then exercised by releasing the cylinder away from its equilibrium position and by sending a single incident wave towards a freely heaving cylinder. These computations produce the expected return-to-equilibrium dynamics, wave radiation, reflection, transmission and diffraction while preserving the symmetry of the axisymmetric configurations. In the freely moving cases, the displacement is no longer prescribed: the pressure below the object and the rigid-body motion are determined together during the computation.

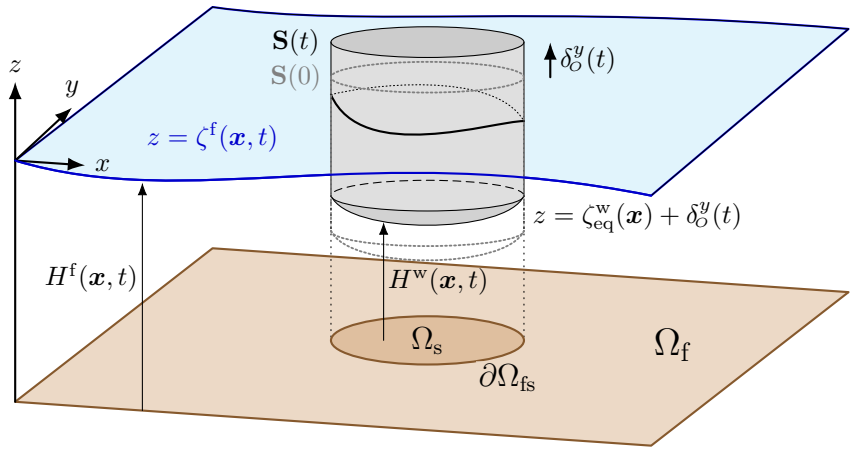


Figure 6.13: Geometry of the vertically moving cylindrical surface obstacle. The dotted contour represents the initial position $\mathbf{S}(0)$, while the solid contour represents $\mathbf{S}(t)$ after the vertical translation $\delta_O^y(t)$. The horizontal regions Ω_{f} and Ω_{s} and the contact-line projection $\partial\Omega_{\mathrm{fs}}$ remain fixed.

The corresponding simulations can be found online: cylindrical wave generator, return to equilibrium & freely heaving cylinder.

A second prototype combines the pseudo-one-dimensional floating-pontoon model of Chapter 5 with the moving-mesh framework of Chapter 4. In addition to the vertical heave motion, the pontoon is given a horizontal surge degree of freedom. Its two contact lines consequently move with the structure, causing the left and right exterior meshes to stretch and compress in opposite directions. The prototype retains the complete coupling architecture: the exterior ALE shallow-water solver supplies the interface data, the interior HHO problem determines the Bernoulli pressure and constrained discharge, the vertical pressure resultant drives the heave motion, and the imbalance of the lateral hydrostatic forces drives the horizontal motion. The resulting rigid-body velocities are then returned to both the fluid transmission conditions and the ALE mesh update. A further extension also allows the pontoon to pitch about its centre of mass, so that its orientation, as well as its horizontal and vertical position, evolves with the flow.

Several complementary configurations have been computed with this second prototype. Prescribing only a horizontal oscillation of the pontoon isolates the moving-interface mechanism and shows that the simultaneous compression and extension of the two meshes can be handled without introducing visible disturbances at the remote boundaries. More complete experiments remove this prescription and activate the coupled rigid-body dynamics. A single wave impacting a pontoon initially at equilibrium generates both surge and heave, together with reflected and transmitted waves and a visible deformation of the exterior meshes. In another demanding demonstrator, the pontoon is released above its hydrostatic equilibrium with non-zero initial vertical and horizontal velocities. The body then falls, oscillates, radiates waves and travels horizontally while the two ALE meshes undergo a large deformation. This last configuration is particularly useful because no component of the subsequent motion is prescribed: the exterior flow, the interior pressure and discharge, both rigid-body degrees of freedom and the mesh motion all remain active in the same computation. In a further single-wave experiment, pitch is activated alongside surge and heave. The corresponding three-body configuration is a first step towards simulating farms of wave energy converters (WECs): waves reflected or radiated by one pontoon alter the forces on its neighbours [13]. Estimating energy production would also require a power-take-off model.

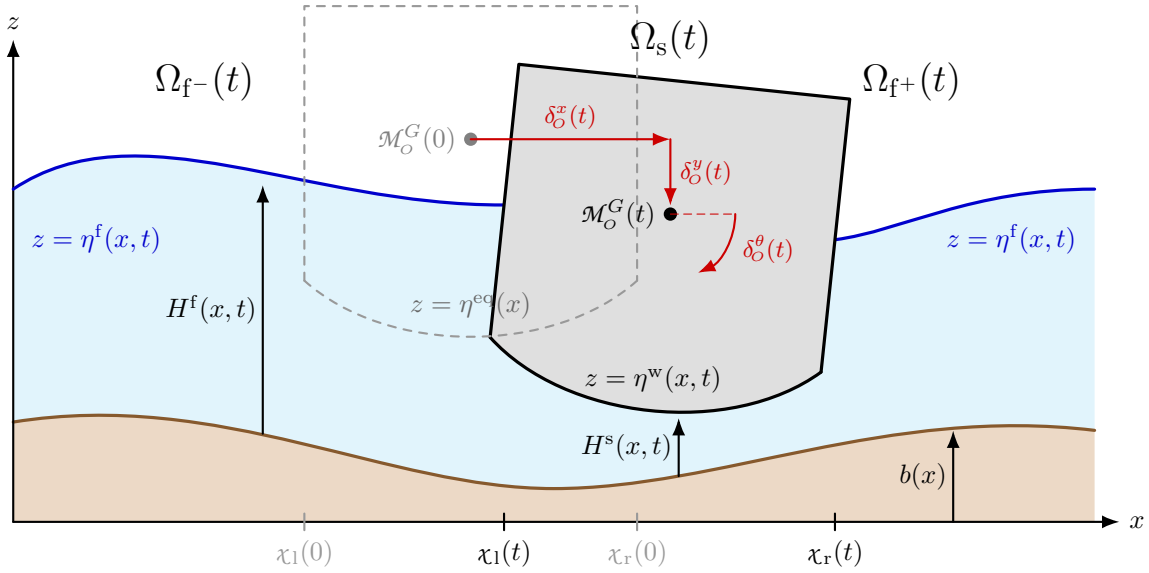


Figure 6.14: Pseudo-one-dimensional wave–structure interaction problem with surge, heave and pitch of the floating pontoon.

The corresponding simulations are also available online: prescribed horizontal motion, single wave impact & pontoon released with an initial velocity. A separate single-wave simulation of one pontoon with surge, heave and pitch is also available. More generally, the same formulation extends straightforwardly from one to K floating objects by coupling K interior HHO–rigid-body subsystems through $K+1$ ALE fluid meshes, as illustrated by this single-wave simulation involving three freely moving pontoons and a three-pontoon simulation including pitch.

These prototypes obviously do not replace the systematic consistency, convergence, robustness and stability studies that would be required for a complete presentation of the corresponding models: their main significance at this stage is instead integrative. They show that the apparently separate ingredients developed in the thesis (i.e. high-order monolithic DG/FV approximations, HHO elliptic solvers, pressure–structure coupling, Newton dynamics, ALE fluxes and conservative moving meshes) can already operate together in functional wave–structure interaction computations. They therefore constitute a direct starting point for the continuation of this work and for a future contribution devoted specifically to freely moving surface obstacles.

These various contributions and recent developments form a coherent progression from robust NSW solvers to complex wave-structure interactions problems. Several directions remain open after this work:

- A first natural perspective is to consolidate the freely moving-object prototypes described above. This requires a complete analysis of the coupled formulations, systematic convergence and robustness studies, and a more general treatment of horizontal hydrodynamic forces. The ALE strategy could then be extended to objects with non-straight boundaries and moving curved contact lines, building in particular on the fixed-domain model of [152]. Larger motions would also motivate more advanced mesh-motion strategies and, eventually, local remeshing.

- A second perspective is to extend the strategy of Chapter 2 to two space dimensions on SGN and *Isobe–Kakinuma* type models. The objective is to use the monolithic DG/FV method for the hyperbolic part and an HHO method for the elliptic correction. This would give a new fully two-dimensional high-order framework for dispersive shallow-water models, as in [211].

- Another possible direction concerns the extension of the monolithic subcell idea to other systems. In particular, such methods may be useful for compressible flows and for Navier–Stokes type equations, for instance in the context of the ANR project *HYPERSONICS* led by F. Vilar. The idea would be to use this *a priori* framework to control the hyperbolic part of the Navier–Stokes equations, and maybe adapt it to use implicit or semi-implicit time integration.

- Another perspective concerns the use of Godunov–Peshkov–Romenski type models for wave-structure interactions problems. These models provide a very general hyperbolic framework for continuum mechanics, and can describe fluids, solids and more complex media within a unified system of first-order equations.

It would therefore be interesting to investigate whether such formulations could be used to model the interaction between waves and floating structures, while keeping a fully hyperbolic structure compatible with robust high-order schemes.

♯ A further direction, in collaboration with Ali Haidar, is the use of physics-informed neural networks for NSW type equations, to bring tools from numerical analysis into their construction. In particular, one could try to design neural solvers which respect the main mathematical properties of the equations, such as conservation, stability, and admissibility of the water-height.

♯ Another direction concerns the design of mesh-motion strategies preserving differential constraints. This work has already been initiated with Walter Boscheri, Mattia Luppi, Lidia Gude Vila and Elena Bernardelli, and aims at understanding whether ALE methods can be constructed so that the mesh motion itself helps preserving structural constraints of the continuous system. This perspective is directly in line with the structure-preserving philosophy followed throughout the thesis, but applied here to differential constraints.

♯ A last perspective concerns conservation laws and NSW equations on surfaces. This direction is motivated by ongoing work with Marien Hanot. For large-scale simulations, it is natural to work directly on curved surfaces. However, this raises several geometric difficulties; many methods either project the problem into the ambient space, or mesh the surface with a geometry error that must then be controlled. A possible way forward is to use the spaces developed in [95] for building a discrete de Rham scheme on manifolds.

Several of these perspectives are already being pursued, and preliminary results can be found at sachacardonna.github.io/research/ and sachacardonna.github.io/codes/.

'Cause baby I'm just a scared and lonely rider, but I gotta know how it feels
I want to know if love is wild, babe, I want to know if love is real
FROM: ***"Born to Run"***, Bruce Springsteen – *Born to Run* (1993)

Appendices

This final part gathers additional developments that are considered useful for the understanding of the manuscript, but that would have overloaded the main body of the thesis. These appendices contain complementary technical material, implementation details and specific analytical results that support the methodology developed in the previous parts. It is organized as follows:

- **Appendix A.** This appendix is devoted to practical aspects of the implementation. Starting from a functional discontinuous Galerkin code, we describe how to build, in a systematic way, a finite-volume subcell solver consistent with the monolithic DG/FV schemes introduced in the manuscript.
- **Appendix B.** We prove that the intermediate states arising in the first-order finite-volume Riemann intermediate states preserve the positivity of the water-height. This result is a key ingredient for establishing the stability of the monolithic schemes.
- **Appendix C.** We propose an extension of the model from Chapter 6 to the case of a floating structure with heaving motion.
- **Appendix D.** This appendix addresses the well-posedness of the HHO scheme used to solve the elliptic problem arising in the wave-structure interactions model (6.14).
- **Appendix E.** Some details are given about the algebraic formulation of the HHO problem (6.39) and the static condensation procedure, allowing for an efficient implementation.
- **Appendix F.** We provide expressions of the blending coefficients used in the monolithic schemes in order to ensure the preservation of discrete entropy inequalities for the 1D NSW equations.
- **Appendix G.** This introduces the functional framework used in the manuscript. The relevant function spaces, continuous operators and notations are defined in a consistent way in order to avoid any ambiguities in the subsequent chapters.

APPENDIX A

Implementation aspects

All the numerical results presented in this thesis are obtained from developments carried out within WaveBox, the multi-model C++ platform initiated by F. Marche for the simulation of water-wave models.

This appendix presents the main principles for implementing DG methods for hyperbolic systems in two dimensions, with a design that naturally extends to DG/FV subcell schemes. The emphasis is principally on essential algorithmic choices and data structures that ensure consistency, robustness, and a smooth transition between DG and subcell FV discretizations, rather than on exhaustive theoretical or coding details. The framework also incorporates an Arbitrary Lagrangian–Eulerian (ALE) extension, allowing it to handle moving mesh configurations.

A.1 Semi-discrete DG formulation of hyperbolic systems

We consider a generic system of conservation laws in two space dimensions:

$$\partial_t \boldsymbol{u}(\boldsymbol{x},t) + \nabla_{\boldsymbol{x}} \cdot \mathbb{F}(\boldsymbol{u}(\boldsymbol{x},t)) = \boldsymbol{S}(\boldsymbol{u}(\boldsymbol{x},t)), \qquad \boldsymbol{x} = (x,y) \in \Omega \subset \mathbb{R}^2,\ t > 0, \tag{A.1}$$

with $\boldsymbol{u} : \Omega \times [0,T] \to \mathbb{R}^m$, $\mathbb{F} : \mathbb{R}^m \to \mathbb{R}^{m\times 2}$ and $\boldsymbol{S} : \Omega \times [0,T] \to \mathbb{R}^m$. The system is closed with appropriate boundary conditions on $\partial\Omega$:

$$\boldsymbol{u}(\boldsymbol{x},0) = \boldsymbol{u}_0(\boldsymbol{x}), \qquad \boldsymbol{x} \in \Omega,$$

and we assume that the Jacobian matrices of the flux function $A(\boldsymbol{u}) = \nabla_{\boldsymbol{u}} \mathbb{F}_i(\boldsymbol{u})$ admit real eigenvalues and a complete set of eigenvectors.

A semi-discrete global DG formulation for (A.1) may be written as: find $\boldsymbol{u}_h \in \mathbf{P}^k(\mathcal{T}_h) := (\mathbb{P}^k(\mathcal{T}_h))^m$ such that, for all $\varphi_h \in \mathbb{P}^k(\mathcal{T}_h)$,

$$\big(\partial_t \boldsymbol{u}_h, \varphi_h\big)_{\mathcal{T}_h} - \big(\mathbb{F}(\boldsymbol{u}_h), \nabla_{\boldsymbol{x}} \varphi_h\big)_{\mathcal{T}_h} + \big\langle \mathbb{F}_\Gamma^{*,\mathrm{DG}} \varphi_h \big\rangle_{\mathcal{T}_h} = \big(\boldsymbol{S}(\boldsymbol{u}_h), \varphi_h\big)_{\mathcal{T}_h}, \tag{A.2}$$

equivalently expressed in a local form as: for each $\omega_c \in \mathcal{T}_h$, find $\boldsymbol{u}_h^c \in \mathbf{P}^k(\omega_c)$ such that, for all $\varphi_h^c \in \mathbb{P}^k(\omega_c)$,

$$\int_{\omega_c} \partial_t \boldsymbol{u}_h^c \varphi_h^c \,\mathrm{d}\boldsymbol{x} - \int_{\omega_c} \mathbb{F}(\boldsymbol{u}_h^c) \cdot \nabla_{\boldsymbol{x}} \varphi_h^c \,\mathrm{d}\boldsymbol{x} + \sum_{\Gamma \in \mathcal{F}_c} \int_\Gamma \mathbb{F}_\Gamma^{*,\mathrm{DG}} \varphi_h^c \,\mathrm{d}s = \int_{\omega_c} \boldsymbol{S}(\boldsymbol{u}_h^c) \varphi_h^c \,\mathrm{d}\boldsymbol{x}. \tag{A.3}$$

Replacing $\boldsymbol{u}_h^c$ by its polynomial expansion, we obtain the following semi-discrete formulation for the polynomial moments $\underline{\boldsymbol{u}}^c = \{\underline{\boldsymbol{u}}_m^c \mid m \in [\![1, N_k]\!]\}$ for all $\varphi_p^c \in \mathbb{P}^k(\omega_c)$, $p \in [\![1, N_k]\!]$:

$$\sum_{m=1}^{N_k} \frac{\mathrm{d}\underline{\boldsymbol{u}}_m^c}{\mathrm{d}t} \int_{\omega_c} \psi_m^c \psi_p^c \,\mathrm{d}\boldsymbol{x} + \sum_{\Gamma \in \mathcal{F}_c} \int_\Gamma \mathbb{F}_\Gamma^{*,\mathrm{DG}} \psi_p^c \,\mathrm{d}s - \int_{\omega_c} \mathbb{F}(\boldsymbol{u}_h^c) \cdot \nabla_{\boldsymbol{x}} \psi_p^c \,\mathrm{d}\boldsymbol{x} = \int_{\omega_c} \boldsymbol{S}(\boldsymbol{u}_h^c) \psi_p^c \,\mathrm{d}\boldsymbol{x}. \tag{A.4}$$

This may now be also written in a compact matrix-vector form, similarly as in (3.8) from Remark 3.2.1. This form is of particular interest from an implementation perspective: once the appropriate data structures are in place, the semi-discrete scheme essentially reduces to successive matrix–vector products.

Remark A.1.1 (*Use of linear algebra libraries*). *From a practical point of view, we use the dedicated linear algebra library* **`Eigen`** *[130] in C++, that greatly facilitates the management of vectors and matrices arising in the code.*

A.2 Data structures and design choices

The implementation of a DG solver requires careful consideration of data structures and design choices to ensure efficiency, clarity, and maintainability. Below, we outline the key components and their roles in the overall architecture of the solver.

A.2.1 The (unstructured) mesh object

The mesh object must provide both the raw geometrical data and the topological relations required by a DG solver and any limiting extension. We denote by N_V the number of vertices, N_E the number of elements, N_F the number of faces, and N_{BF} the number of boundary faces, and we list the main components of a mesh object in Figure A.1.

```
unstructured_mesh:

  counts:
    num_vertices (N_V)         # total number of mesh vertices
    num_elements (N_E)         # number of triangular elements
    num_faces (N_F)            # number of faces
    num_boundary_faces (N_BF)  # number of boundary faces
    num_dirichlet_faces, num_neumann_faces, ... # per type

  extents:
    x_min, x_max, y_min, y_max # bounding box
    domain_length_scale        # characteristic length
    domain_area                # total area of the mesh

  input_geometry:
    vertex_coordinates [N_V,2]        # coordinates of mesh vertices
    element_to_vertices [N_E,N_F]       # connectivity: vertices of each element
    boundaryface_to_vertices [N_BF,2] # vertices of boundary faces

  derived_topology:
    vertex_to_elements [N_V, N_E] (sparse) # incidence: vertex to elements
    vertex_to_faces    [N_V, N_F] (sparse) # incidence: vertex to faces
    face_to_vertices   [N_F,2]             # vertices of each face
    face_to_elements   [N_F,2]             # adjacent elements of each face
    element_to_faces   [N_E,N_F]             # faces of each element
    element_to_elements [N_E,N_F]            # neighbors per element
    boundaryface_to_vertices_by_type:      # connectivity per BC type
        dirichlet [N_D,2], neumann [N_N,2], ...
    face_boundary_id [N_BF]                # boundary condition tag per face

  derived_geometry:
    face_areas [N_F]               # length of faces
    face_normals [N_F,2]           # outward unit normals
    element_areas [N_E]            # area of each element
    element_barycenters [N_E,2]    # element barycenters
    element_scales [N_E,2]         # scaling per direction
    element_size [N_E]             # circumscribed radius
    affine_maps                    # reference-to-physical transformations
    jacobians                      # Jacobian determinants
    metric_factors (rx, ry, sx, sy) # geometric factors for derivatives
```

Figure A.1: The `unstructured_mesh` object and its topological and geometrical data.

A.2.2 Quadrature rules

A central ingredient of any high-order discretization method is the numerical approximation of integrals. Given a domain $\omega_c \subset \mathbb{R}^d$ and a sufficiently smooth function $f : \omega_c \to \mathbb{R}$, one seeks an approximation

$$\int_{\omega_c} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \;\approx\; \sum_{i=1}^{N_q} w_i\, f(\boldsymbol{x}_i),$$

where N_q denotes the number of quadrature points, $\{\boldsymbol{x}_i\}_{i\in[\![1,N_q]\!]} \subset \omega_c$ are the quadrature points and $\{w_i\}_{i\in[\![1,N_q]\!]}$ the associated weights. The *degree of exactness* of a quadrature rule is defined as the largest integer r such that the above formula is exact for all polynomial functions f of total degree at most r.

Remark A.2.1 (*Exact vs. quadrature-based integration*). *For complex systems of conservation laws with non-polynomial fluxes or nonlinear terms, exact integration is generally not feasible. Instead, one resorts to numerical quadrature rules or equivalent collocation strategies for the flux. To ensure the expected accuracy, quadrature rules must be chosen with sufficient level of approximation quality. In what follows, the notation $\oint$ will be used to indicate quadrature-based integration, while $\int$ denotes exact integration. For polynomials of degree at most $2k$ (volume) or $2k+1$ (surface), the two notations coincide since the quadrature is then exact.*

On structured meshes (e.g. Cartesian grids), quadrature rules are tensor-product based; on unstructured meshes, a common strategy is to define quadrature rules on a *reference element*, typically the unit simplex

$$\omega_{\mathrm{ref}} := \{(r,s) \in \mathbb{R}^2 \mid r \geq 0,\ s \geq 0,\ r+s \leq 1\},$$

and to map quadrature points to each physical element $\omega_c \subset \Omega$ via an element-wise mapping $\Phi : \omega_{\mathrm{ref}} \to \omega_c$ (which is an affine diffeomorphism in the case of straight-edged triangles), see Figure A.2.

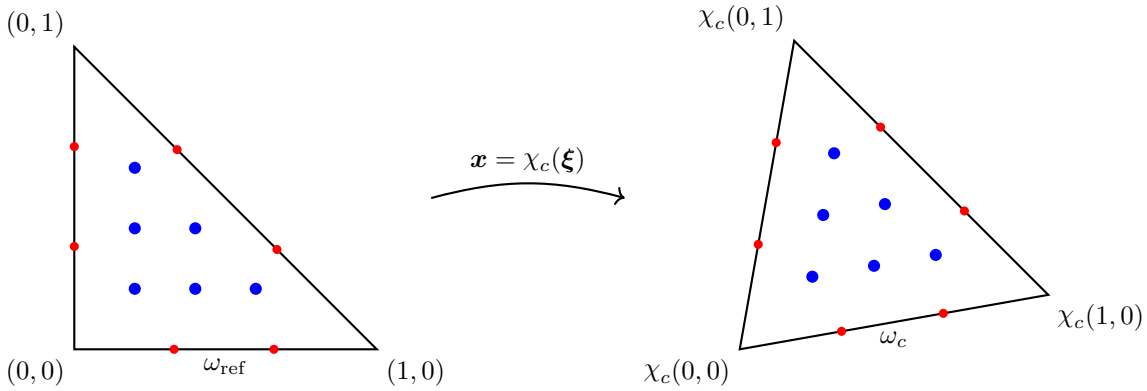


Figure A.2: Reference quadrature nodes (left) mapped to a physical element (right) through the affine mapping $\boldsymbol{x} = \chi_c(\boldsymbol{\xi})$. Edge nodes are shown in red and interior nodes in blue.

The integral over a physical element can then be approximated as

$$\oint_{\omega_c} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \sum_{i=1}^{N_q} w_i\,(f \circ \chi_c)(\boldsymbol{\xi}_i)\,|\det \nabla_{\boldsymbol{\xi}} \chi_c(\boldsymbol{\xi}_i)|,$$

where $(\boldsymbol{\xi}_i, w_i)$ are the quadrature points and weights on the reference triangle.

Remark A.2.2 (*Simplex case*). *In the case of straight-edged affine triangles, the mapping $\chi_c : \omega_{\mathrm{ref}} \to \omega_c$ is affine and its Jacobian determinant is constant over the element. In this setting, the quadrature formula simplifies to*

$$\oint_{\omega_c} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \frac{|\omega_c|}{|\omega_{\mathrm{ref}}|} \sum_{i=1}^{N_q} w_i\,(f \circ \chi_c)(\boldsymbol{\xi}_i),$$

where $|\omega_c|$ denotes the volume of the physical triangle and $|\omega_{\mathrm{ref}}|$ the volume of the reference element.

```
quadrature:

  quad_elements:
    num_quad_points_per_elem (NqE)
    ref_coords_r [NqE]           # x-coordinates on reference triangle
    ref_coords_s [NqE]           # y-coordinates on reference triangle
    phys_coords_xy [NqE, N_E, 2] # mapped coords in physical elements
    weights [NqE]                # quad. weights (reference)

  quad_faces:
    num_quad_points_per_face (NqF)
    ref_coords_1d [NqF]             # reference coords on [-1,1]
    phys_coords_xy [NqF, N_F, 2]    # mapped coords in physical faces
    weights [NqF]                   # quad. weights
    normals_x [NqF, N_F]            # x-component of outgoing normal at nodes
    normals_y [NqF, N_F]            # y-component of outgoing normal at nodes
```

Figure A.3: The `quadrature` object, split into element and face data.

A.2.3 Discrete basis functions

On the physical element ω_c, the discrete solution is expanded in a polynomial space $\mathbf{P}^k(\omega_c)$ of dimension N_k^3, with $N_k = (k+1)(k+2)/2$ being the number of degrees of freedom related to the polynomial degree k. A family of basis functions $\{\varphi_m^c\}_{m=1}^{N_k}$ is chosen, and the approximation is written as

$$\boldsymbol{u}_h^c(\boldsymbol{x},t) = \sum_{m=1}^{N_k} \underline{\boldsymbol{u}}_m^c(t)\varphi_m^c(\boldsymbol{x}).$$

In the same way, any higher-order derivative with respect to $\boldsymbol{x}$ are obtained by differentiating the basis functions, so that for instance

$$\nabla_{\boldsymbol{x}}^{(k)}\boldsymbol{u}_h^c(\boldsymbol{x},t) = \sum_{m=1}^{N_k} \underline{\boldsymbol{u}}_m^c(t)\nabla_{\boldsymbol{x}}^{(k)}\varphi_m^c(\boldsymbol{x}).$$

From a theoretical point of view, the choice of the basis functions is not supposed to alter the numerical solution: any complete basis of $\mathbb{P}^k(\omega_c)$ yields the same approximation space. In practice, however, the computational properties may differ substantially. For instance, orthogonal modal bases may produce diagonal (or block diagonal) mass matrices, while monomial bases are notoriously known to produce ill-conditioned mass or stiffness matrices for high-order approximations. In Figure A.4, we propose a simple design for a `basis_functions` object that gathers the main ingredients related to the basis functions and their evaluation at quadrature nodes.

A.2.4 Modal and nodal representations objects

In polynomial methods, the solutions are naturally described through two complementary frameworks:

- *Modal representation.* Each component of the variables vector is expanded in a polynomial basis. The unknowns are the polynomial moments, referred to as the fields, and represent the actual degrees of freedom of the method.
- *Nodal representation.* This gathers the pointwise values of the numerical solution at quadrature nodes inside elements and faces. These values are essential for approximating integrals via quadrature rules

A bridge between the two representations may be provided by the L^2-projection, for any function $f \in L^2(\omega_c)$ and defined as the unique polynomial $\pi_c^k f$ such that

$$\int_{\omega_c} (f - \pi_c^k f)\,\varphi_m \,\mathrm{d}\boldsymbol{x} = 0, \qquad m \in [\![1, N_k]\!].$$

Expanding $\pi_c^k f$ in the basis, we get $\pi_c^k f = \sum_{p=1}^{N_k} \underline{f}_p\,\varphi_p$, the coefficients $\underline{f}_p$ being identified as the *modes* of f. These are collected in the coefficient vector $\underline{\boldsymbol{F}} = (\underline{f}_p)_{p\in[\![1,N_k]\!]}$ that is obtained by solving the mass-matrix system

$$\mathbb{M}_c\,\underline{\boldsymbol{F}} = \boldsymbol{R}_c,$$

```
basis_functions:

  polynomial_degree: k
  num_nodes_per_element: N_q^E
  num_nodes_per_face: N_q^F

  element_values:
    val [N_q^E, N_E]      # basis functions at volume quad. points
    val_dx [N_q^E, N_E]   # derivatives wrt x
    val_dy [N_q^E, N_E]   # derivatives wrt y

  face_values:
    val_on_faces [N_q^F, N_F, N_E]
    val_dx_on_faces [N_q^F, N_F, N_E]
    val_dy_on_faces [N_q^F, N_F, N_E]

  interpolation_support:
    lagrange_nodes_ref [N_k,2]   # interpolation nodes on reference element
    mass_mat [N_q^F, N_q^F]          # mass matrix
    inv_mass_mat [N_q^F, N_q^F]      # inverse mass matrix
```

Figure A.4: The `basis_functions` object: storage of basis functions and derivatives at quadrature nodes.

where the right-hand side integrals $\boldsymbol{R}_c := (\oint_{\omega_c} f\varphi_m \, \mathrm{d}\boldsymbol{x})_{m\in[\![1,N_k]\!]}$ are approximated by quadrature. Again, considering affine triangles, $|\det \nabla_{\boldsymbol{x}} \chi_c|$ is constant and one obtains

$$\mathbb{M}_c = \frac{|\omega_c|}{|\omega_{\mathrm{ref}}|}\mathbb{V}^\top \mathbb{W}\mathbb{V}, \qquad \boldsymbol{R}_c = \frac{|\omega_c|}{|\omega_{\mathrm{ref}}|}\mathbb{V}^\top \mathbb{W}\boldsymbol{F}^\star,$$

where $\mathbb{V} = (\varphi_m(\boldsymbol{\xi}_i))_{(i,m)\in[\![1,N_q]\!]\times[\![1,N_k]\!]}$, $\mathbb{W} = \mathrm{diag}(w_1,\dots,w_{N_q})$, and $\boldsymbol{F}^\star = (f(\boldsymbol{x}_i))_{i\in[\![1,N_q]\!]}$. Hence

$$\underline{\boldsymbol{F}} = (\mathbb{V}^\top \mathbb{W}\mathbb{V})^{-1}\mathbb{V}^\top \mathbb{W}\boldsymbol{F}^\star.$$

Once the modal moments $\underline{\boldsymbol{F}}$ are known, nodal values at quadrature nodes follow by evaluation, i.e.

$$(\pi_c^k f(\boldsymbol{x}_i))_{i\in[\![1,N_q]\!]} = \mathbb{V}\underline{\boldsymbol{F}}.$$

Remark A.2.3 (*Orthonormal bases*). *When the basis is orthonormal and the quadrature is exact for products up to degree $2k$, the matrix $\mathbb{V}^\top\mathbb{W}\mathbb{V}$ reduces to the identity and the coefficients are simply $\underline{\boldsymbol{F}} = \mathbb{V}^\top\mathbb{W}\,\boldsymbol{F}^\star$.*

A.3 Monolithic DG/FV subcell framework

The extension from a standard DG solver to a monolithic DG/FV subcell framework requires a significant amount of additional implementation: subdivision of elements into subcells, associated topology and new connectivity relations. Once these structures are in place, however, the transition itself is remarkably efficient. A carefully designed DG solver can be adapted with relatively minor changes, since the algebraic formulation of the subcell method relies on the same fundamental building blocks.

This approach also provides a flexible computational framework; by enforcing the blending parameter Θ at the subcell level, one can seamlessly recover different schemes within the same code: a pure FV method when $\Theta = 0$, a standard DG method when $\Theta = 1$ (on which *classical* limiters can then be applied), and full monolithic DG/FV subcell schemes for intermediate values. Hence, the monolithic formulation does not only extend DG solvers: it unifies FV and DG methods within a single consistent framework.

A.3.1 Subcell semi-discrete reformulation

Let us recall the equivalency theorem (Theorem 3.2.5 from Chapter 3) adapted in the context of hyperbolic systems of conservation laws:

Theorem A.3.1 (*Equivalence between DG and FV-like subcell schemes*). *The classical unlimited Discontinuous Galerkin scheme expressed in compact form through residuals can be recast into N_s FV-like subcell schemes as*

$$\frac{\mathrm{d}\overline{\boldsymbol{u}}^c}{\mathrm{d}t} = -\mathbb{D}_c^{-1}\left(\mathbb{A}_c\widehat{\mathbb{F}}_c + \partial\mathbb{F}_c\right) + \overline{\boldsymbol{S}}_c \tag{A.5}$$

where $\overline{\boldsymbol{S}}_c := \mathbb{P}_c\mathbb{M}_c^{-1}\boldsymbol{S}_c$ is a vector containing the submean values of source term projection, i.e. $(\overline{\boldsymbol{S}}_c)_m = \overline{\boldsymbol{S}}_m^c$. The reconstructed flux $\widehat{\mathbb{F}}_c$, that contains all the interior faces reconstructed fluxes weighted by the face length, i.e. $\ell_{mp}\widehat{\mathbb{F}}_{mp}$ writes as

$$\widehat{\mathbb{F}}_c = -\mathbb{A}_c^\top\mathscr{L}_c^{-1}\left(\mathbb{D}_c\mathbb{P}_c\mathbb{M}_c^{-1}\boldsymbol{\Xi}_c + \partial\mathbb{F}_c\right), \tag{A.6}$$

where $\partial\mathbb{F}_c \in \mathbb{R}^{N_s}$ carries the cell boundary contribution as the integral of DG fluxes over the cell boundary, i.e.

$$(\partial\mathbb{F}_m^c)_{m\in[\![1,N_s]\!]} = \int_{\partial\omega_c\cap\partial S_m^c} \mathbb{F}_n^{*,\mathrm{DG}}\,\mathrm{d}\boldsymbol{x}.$$

Remark A.3.2 (*Computational aspects*). *The only terms depending on time are $\boldsymbol{\Xi}_c$ (the DG residual) and the boundary contribution $\partial\mathbb{F}_c$. All the other terms can be precomputed for all ω_c, by means of a mapping to a reference element. Choosing the same set of basis functions across all elements, as Legendre functions for example, will lead to an unique definition for $\mathbb{A}_c$, $\mathbb{P}_c$ and $\mathscr{L}_c^{-1}$.*

The equivalency theorem stated above holds for the unlimited DG formulation, and thus provides only a formal correspondence between the two methods. In practice, however, the purpose of the subcell framework is to incorporate limiting strategies in order to guarantee stability and robustness near discontinuities. To that end, one must go beyond the bare equivalence and design the numerical scheme in a way that combines two key ingredients:

- the reconstruction of DG polynomials into FV-like subcell averages, as given by the theorem,
- and the use of genuine finite-volume fluxes at the subcell level, on which the limiting procedure will rely.

We recall a possible and practical formulation of the semi-discrete monolithic DG/FV subcell scheme, which reads as follows:

Definition A.3.3 (*Monolithic DG/FV subcell scheme*). *The subcell monolithic DG/FV convex property preserving semi-discrete scheme supplemented reads as*

$$\frac{\mathrm{d}\overline{\boldsymbol{u}}_m^c}{\mathrm{d}t} = -\frac{1}{|S_m^c|}\sum_{S_p^v\in\mathscr{V}_m^c}\ell_{mp}\widetilde{\mathbb{F}}_{mp} + \overline{\boldsymbol{S}}_m^c, \tag{A.7}$$

where $\widetilde{\mathbb{F}}_{mp} = \mathbb{F}_{mp}^{*,\mathrm{FV}} + \Theta_{mp}\left(\widehat{\mathbb{F}}_{mp} - \mathbb{F}_{mp}^{*,\mathrm{FV}}\right)$, with the reconstructed high-order flux recalled in (A.5) and the classical FV flux at the subcell level.

This shift from a purely DG formulation to a monolithic DG/FV subcell scheme implies some modification of the solver's architecture but most of the new computational objects are designed as extensions (*mix-ins*) or unherited classes from the original DG ones, ensuring consistency while enabling the additional operations required by the subcell method.

A.3.2 Subcell geometry and topology

Let summarize the main notations introduced in Chapter 3 related to the sub-mesh and connectivity. These structures provide the additional geometric and topological informations required when extending the classical mesh objects to the subcell level. For each element ω_c, we introduce the notations detailed in Table A.1.
The corresponding objects are naturally implemented by extending the classical mesh data structures so that element-wise and subcell-wise quantities can be accessed in a consistent way. The submesh is constructed once at the level of the reference element, and then mapped to each physical element through the usual affine (or isoparametric) transformations. This approach avoids recomputing the subdivision for every cell and ensures consistency across the mesh.

Notation	Description
$\mathcal{T}^k_{\omega_c}$	Sub-partition of element ω_c into N_s subcells S^c_m, with associated volume matrix $\mathbb{D}_c = \mathrm{diag}(\lvert S^c_m\rvert)$.
Γ^c_{mp}	Face shared by subcells S^c_m and S_p, with length ℓ_{mp}. The set of all such faces is $\mathcal{F}^c_m$, with $\#\mathcal{F}^c_m = n^m_f$.
$\mathcal{V}^c_m$	Set of face-neighboring subcells of S^c_m. The restricted set $\breve{\mathcal{V}}^c_m$ contains only those neighbors lying entirely inside ω_c.
$\mathbf{n}_{mp}$	Normal vector associated to face Γ^c_{mp}, oriented by the sign $\{-1, 1\}$.
$\mathcal{P}^c_m$	Set of vertices of subcell S^c_m. For each vertex $\boldsymbol{x}_p$, $\mathcal{V}_p$ denotes the set of subcells sharing this vertex.
$\mathcal{N}(S^c_m)$	Neighborhood of S^c_m including itself, its face-neighbors, and node-neighbors, i.e., $\mathcal{N}(S^c_m) = \bigcup_{\boldsymbol{x}_p \in \mathcal{P}^c_m} \mathcal{V}_p$.
$\mathcal{V}_{mp}$	Set of subcells sharing at least one vertex with the face Γ^c_{mp}.

Table A.1: Summary of geometrical and topological notations for the FV-subcell framework on an element ω_c.

For instance, given a polynomial order k, we divide ω_{ref} into N_s disjoint subcells. For the fully triangular case considered in this work, this is achieved by introducing $k+1$ subdivisions along each edge, and generating the interior nodes via barycentric coordinates

$$\boldsymbol{\xi}^{\mathrm{sub}}_{i,j} = \left(\tfrac{i}{k+1}, \tfrac{j}{k+1}\right), \qquad i,j \geq 0, \quad i+j \leq k+1.$$

Here, the $^{\mathrm{sub}}$ supercript emphasizes that these points are the subdivision nodes of the reference element, used specifically for constructing subcells, and should not be confused with the generic reference coordinate $\boldsymbol{\xi}$ employed in the mapping $\Phi : \omega_{\mathrm{ref}} \to \omega_c$.

Each subcell S^{ref}_m is then formed by connecting adjacent subdivision nodes, so that

$$\overline{\omega_{\mathrm{ref}}} = \bigcup_{m=1}^{N_s} \overline{S^{\mathrm{ref}}_m}, \qquad N_s = \min\{(d+1)^2 \mid (d+1)^2 \geq \dim \mathbb{P}^k(\omega_{\mathrm{ref}}),\, d \in \mathbb{N}\}.$$

In practice, the subcells are stored once in reference coordinates, with their node coordinates, face lengths, barycenter, and normals. The physical subcells are then obtained by applying the affine map $\Psi_c : \omega_{\mathrm{ref}} \to \omega_c$ s.t.

$$\boldsymbol{x} = \Psi_c(\boldsymbol{\xi}^{\mathrm{sub}}) = \boldsymbol{x}^c_{\mathrm{b}} + \nabla_{\boldsymbol{\xi}^{\mathrm{sub}}} \Psi_c \, \boldsymbol{\xi}^{\mathrm{sub}},$$

so that volumes and barycenters scale as

$$\lvert S^c_m\rvert = \lvert \det(\nabla_{\boldsymbol{\xi}^{\mathrm{sub}}} \Psi_c)\rvert \, \lvert S^{\mathrm{ref}}_m\rvert, \qquad \boldsymbol{x}^c_{\mathrm{b}} = \chi_c(\boldsymbol{\xi}^{\mathrm{sub}}_{\mathrm{b}}).$$

In contrast with the global `unstructured_mesh` object, the submesh relies on lighter container without member functions, designed to group together related data. This choice is computationally efficient, since most operations at the subcell level involve iterating over large arrays of sub-elements or sub-faces. For this reason, we define two dedicated structures: `subElement`, which stores the geometry and connectivity of a single subcell, and `subFace`, which provides the corresponding information for subfaces, see Figure A.5 for more details. Together, these structures support efficient access to both element-based and face-based loops, a key ingredient in finite volume subcell methods.

Remark A.3.4 (*Subcell generation algorithm*). *Some details about the subcell generation procedure:*

- *The algorithm systematically computes face lengths, oriented normals, barycenters, and volumes at the reference level. These are then pushed forward to the physical mesh via the Jacobian* $\nabla_{\boldsymbol{\xi}^{\mathrm{sub}}} \Psi_c$,
- *Both boundary subcells (those touching* $\partial\omega_{\mathrm{ref}}$*) and interior subcells are identified, together with the corresponding boundary normal IDs and orientation signs,*

– *For quadrangular subdivision, a similar procedure is used, leading to* $N_s = \dim \mathbb{P}^k(\omega_{\text{ref}})$ *subcells (some triangular, some quadrangular depending on the pattern),*

– *The subdivision process also initializes connectivity data: neighbors through faces, vertex-based neighbors (via a* `node_to_subcells` *map), face-to-normal correspondences, and boundary tags. The additional* `node_to_subcells` *relation is particularly useful to enlarge stencils and to access diagonal neighbors that cannot be reached by face adjacency alone,*

– *Initialization routines such as* `SubFace` *construction, boundary condition propagation, neighbor-face indices and many others complete the setup of the submesh.*

A.3.3 Inherited DG components and additional objects

All notions related to numerical quadrature can be directly inherited and applied at the subcell scale. Once the subdivision of each physical element ω_c into N_s subcells $\{S_m^c\}_{m\in[\![1,N_s]\!]}$ has been performed, each subcell can be treated as a miniature DG element endowed with its own set of quadrature points and weights.

The construction of these quadrature rules follows the same principles as in the classical DG setting: interior quadrature rules are defined by mapping Gauss-type nodes from the reference subcell S_m^{ref} to the corresponding physical subcell S_m^c, while boundary quadrature rules are defined along each subcell face. Thus, both volume and face integrals in the monolithic DG/FV scheme rely on the same family of quadrature rules as in the parent DG method, but evaluated locally within each subcell.

From an implementation point of view, we introduce an additional quadrature object which extends the existing `quadrature` class. This object stores, for each subcell, the local quadrature nodes in both reference and physical coordinates, their associated weights, and the quadrature nodes lying on subcell faces. The latter are used to compute numerical fluxes between adjacent subcells and to enforce the internal coupling that characterizes the monolithic DG/FV formulation. The same quadrature order as in the parent DG element is typically retained for each subcell, which guarantees consistency and preserves the overall accuracy of the scheme.

Remark A.3.5 (*Quadrature design choices*). *It is worth mentioning that depending on the chosen subdivision type, some arbitrary but necessary design decisions must be made: for instance, in a mixed triangular–quadrangular subdivision, quadrature rules must be defined consistently on both geometries. A simple option consists in splitting each quadrangle into two triangles, at the expense of increasing the number of quadrature points, while the alternative we chose relies on tensor-product 1D quadrature rules defined directly on the quadrangular subcells, see Figure A.6.*

A.4 SSP-RK–DG/FV-subcells solver workflow

The execution of the FV-subcells stabilized DG solver extends the classical DG workflow by introducing an additional subcell layer, local reconstruction operators, and a convex blending between FV and reconstructed DG fluxes. After a preprocessing phase, the solver advances the solution in time by applying an SSP Runge–Kutta scheme to the FV-subcell mean values, while the DG polynomial representation is reconstructed from these subcells at every stage.

Initialization and preprocessing. This stage is once again performed once and for all before the time loop and includes:

– defining the physical and numerical parameters of the test case,

– loading the unstructured mesh and assembling parent-element connectivity,

– constructing the reference submesh and its mapping to all physical elements,

– building quadrature rules on elements, faces, subcells, and subfaces,

– evaluating the reference basis functions $\{\varphi_m^c\}_{m\in[\![1,N_k]\!]}$ and their derivatives at all DG and FV-subcell quadrature nodes,

– assembling the element-wise mass matrices $\{\mathbb{M}_c\}_{c\in\mathcal{T}_h}$ and their inverses,

```
submesh:

  counts:
    num_subcells              # total number of subcells
    num_subcells_per_element  # average subdivision
    num_faces_per_subcell     # typical subfaces per subcell

    area_subcells           # vector of subcell areas
    scaling_subfaces        # geometric scaling factors per subface

    node_to_subcells        # map: node id -> set of incident subcells

  structures:

    subElement:
      id                # global identifier of the subcell
      local_id          # local index within parent element
      parent_element    # id of the parent mesh element
      type              # subcell type (triangle, quad, ...)

      nodes                       # list of vertexes
      nodes_ids                   # indices of subcell vertices
      volume                      # subcell volume
      barycenter                  # geometric barycenter
      amplitude_x, amplitude_y    # characteristic scales

      n_faces           # number of faces
      n_bdr_faces       # number of boundary faces
      faces_id          # ids of subfaces
      faces_length      # length of each subface
      normals           # outward normals

      neighbours              # ids of face neighboring subcells
      neighbours_exterior     # ids of vertex neighboring subcells
      neighbour_face          # corresponding subface id for each neighbor

      normals_id           # ids of normals
      normals_sgn          # orientation of each normal
      bdr_normals_id       # ids of boundary normals
      boundary_condition   # boundary tag for each boundary face

    subFace:
      id                # global identifier of the subface
      subcells          # adjacent subcells
      subcells_faces    # corresponding local face indices in subcells
      nodes             # vertices of the subface
      length            # subface length
      normal_id         # id of the associated normal
      normal_sgn        # orientation with respect to subcell
      is_boundary       # boolean flag (boundary or internal)

  matrices:
    adjacency_matrix A_c             # subgrid adjacency
    volume_matrices D_c              # subcell volumes
    laplacian_matrix                 # discrete Laplacian on subgrid
    subgrid_graph L_c                # L_c = A_c * A_c^T
    inverse_graph L_c^-1             # inverse of L_c on the orthogonal of its kernel
    projection_matrices P_c, P_c^-1  # modal-to-nodal mappings
```

Figure A.5: The `submesh` object extends the unstructured mesh with dedicated structures (`subElement`, `subFace`), connectivity operators, and global data.

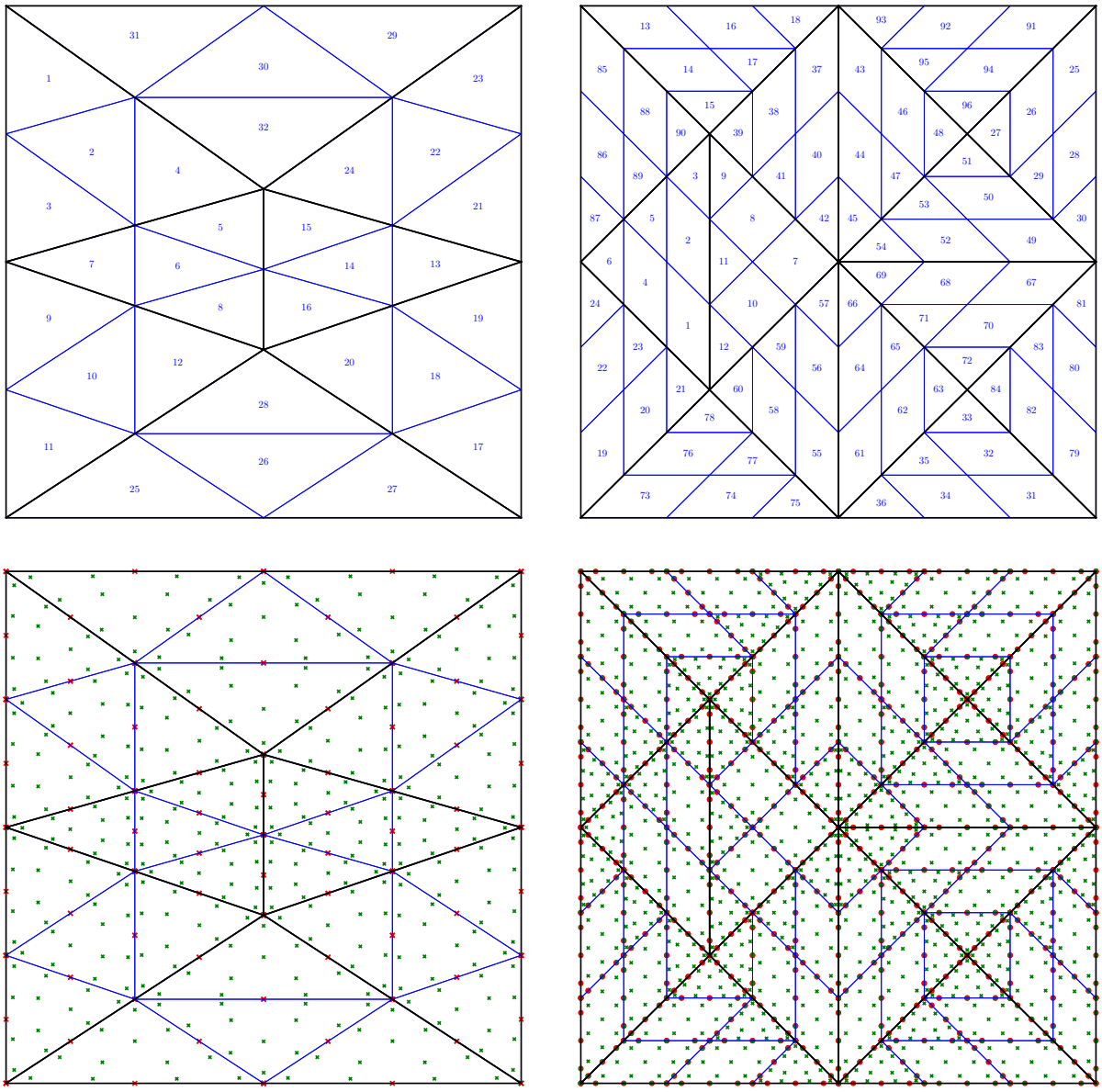


Figure A.6: Examples of reference-element subdivisions used in the FV-subcell framework. Left: division into triangular subcells in the case $N_s \geq N_k$, with the corresponding quadrature points for subcell interiors and faces. Right: division into both triangular and quadrangular subcells in the case $N_s = N_k$, again with the associated subcell quadrature points. One of my thesis advisors should be quite happy to see that I managed to make good use of these figures in my thesis.

– assembling the subcell matrices $\{\mathbb{A}_c, \mathbb{D}_c, \mathscr{L}_c, \mathbb{P}_c\}_{c\in\mathcal{T}_h}$ required for flux reconstruction.

The semi-discrete Eulerian formulation combining DG and subcell finite volumes is solved using the fully explicit workflow detailed in Algorithm 1.

A.5 Extension to Arbitrary Lagrangian-Eulerian (ALE) framework

The transition from a fixed-grid Eulerian solver to an Arbitrary Lagrangian-Eulerian (ALE) framework introduces a fundamental paradigm shift in the implementation: the computational mesh $\mathcal{T}_h(t)$ is no longer static. Consequently, all geometrical quantities, quadrature mappings, and metric terms become time-dependent and must be dynamically updated during the simulation. Furthermore, the numerical fluxes must be modified to account for the grid velocity $\boldsymbol{w}^g$, and the mesh motion itself must be computed by solving an auxiliary elliptic problem. All the notations and definitions used in this section come from Chapter 4.

A.5.1 Dynamic mesh and time-dependent geometry

In the static DG solver, geometrical data (element areas, subcell volumes, normals, face lengths, and Jacobians) were computed once and for all during the initialization phase. In the ALE framework, the `unstructured_mesh` and `submesh` objects must be extended to support dynamic coordinate updates $\boldsymbol{x}_p(t)$ for all $p \in \mathcal{V}_h(t)$, see Figure A.7.

From an object-oriented perspective, it is efficient to separate the constant topological connectivity (which remains invariant since the mesh does not undergo topological changes or remeshing in our setting) from the time-dependent geometrical metrics. We introduce a `dynamic_mesh` class deriving from `unstructured_mesh`, containing an `update_geometry()` routine that is called whenever the mesh vertices are displaced. This routine recalculates:

- Physical coordinates of the mesh vertices $\boldsymbol{x}_p(t)$,
- Element areas $|\omega_c(t)|$, barycenters $\boldsymbol{x}_{\mathrm{b}}^c(t)$, and characteristic lengths $h_c(t)$,
- Face lengths $\ell_\Gamma(t)$ and oriented normals $\mathbf{n}_{cv}(t)$ for both DG elements $\omega_c(t)$, $c \in [\![1, n_{\mathrm{el}}]\!]$, and its subcells $S_m^c(t)$, $m \in [\![1, N_s]\!]$,
- Affine mappings, Jacobians $\mathbb{J}_c(t)$, and metric factors from the reference element to the current physical configurations.

```
dynamic_mesh (extends unstructured_mesh):

  counts:
    // Topology remains constant (NV, NE, NF, NBF)

  dynamic_geometry:
    grid_velocity      [NV, 2]     // Nodal grid velocity wg
    vertex_coords_old  [NV, 2]     // Coordinates at t^n
    vertex_coords_curr [NV, 2]     // Coordinates at current RK stage

  methods:
    update_mesh_position(dt)       // Moves vertices according to wg
    update_metrics()               // Recomputes areas, normals, Jacobians, etc
```

Figure A.7: Extension of the mesh object to support dynamic ALE updates.

A.5.2 Grid velocity computation and $\mathbb{P}^1$ finite element solvers

Before advancing the mesh geometry at each Runge-Kutta stage or time step, the grid velocity field $\boldsymbol{w}^g$ must be determined over the entire computational domain. As introduced in Chapter 4, this is achieved by solving an elliptic problem subject to boundary conditions dictated by the moving geometries $\partial\Omega_{\mathrm{mov}}(t)$.

Input: FV-subcell mean values $\overline{\boldsymbol{u}}^{c,n}$ at time t^n
Output: FV-subcell mean values $\overline{\boldsymbol{u}}^{c,n+1}$ at time t^{n+1}

`// 1. Time step evaluation`
Compute Δt^n from a CFL condition evaluated on the FV-subcell states

`// 2. Runge-Kutta stages`
for $j = 0, \dots, s-1$ **do**

`// High-order DG step`
for *each element* $\omega_c \in \mathcal{T}_h$ **do**
Evaluate $\boldsymbol{u}_h^{c,(j)}$ and physical fluxes at element quadrature nodes
Build face traces $\boldsymbol{u}_h^{c,\pm,(j)}$, apply BCs, and compute DG numerical fluxes
Assemble elemental DG residual: $\mathscr{R}_{\mathrm{DG}}^{c,(j)} = \mathbb{M}_c^{-1}(\boldsymbol{\Xi}_c^{(j)} + \boldsymbol{S}_c^{(j)})$
Compute subcell source terms: $\overline{\boldsymbol{S}}_m^{c,(j)} = \mathbb{P}_c\mathbb{M}_c^{-1}\Big(\int_{\omega_c} \boldsymbol{S}_h^{(j)}\varphi_h \,\mathrm{d}\boldsymbol{x}\Big)$
end

`// First-order FV step`
for *each subface* $\Gamma_{mp} \in \mathcal{F}_m^c$ **do**
Build int./ext. FV states using subcell connectivity
Impose boundary conditions when $\Gamma_{mp} \in \mathcal{F}_h^\partial$
Compute first-order FV flux $\mathbb{F}_{mp}^{*,\mathrm{FV},(j)}$
end

`// Subcell flux blending and residuals`
for *each element* $\omega_c \in \mathcal{T}_h$ **do**
Reconstruct DG flux on internal subfaces: $\widehat{\mathbb{F}}_c^{(j)} = -\mathbb{A}_c^\top \mathscr{L}_c^{-1}\Big(\mathbb{D}_c\mathbb{P}_c\mathbb{M}_c^{-1}\boldsymbol{\Xi}_c^{(j)} + \partial\mathbb{F}_c^{(j)}\Big)$
end
for *each subface* $\Gamma_{mp} \in \mathcal{F}_m^c$ **do**
Compute blending coefficients $\Theta_{mp}^{(j)} \in [0,1]$
Form the blended interface flux: $\widetilde{\mathbb{F}}_{mp}^{(j)} = \mathbb{F}_{mp}^{*,\mathrm{FV},(j)} + \Theta_{mp}^{(j)}\big(\widehat{\mathbb{F}}_{mp}^{(j)} - \mathbb{F}_{mp}^{*,\mathrm{FV},(j)}\big)$
Assemble the final FV-subcell residual: $\mathscr{R}_{\boldsymbol{u},m}^{c,(j)} = -\frac{1}{|S_m^c|}\sum_{S_p^v \in \mathcal{V}_m^c} \ell_{mp}\widetilde{\mathbb{F}}_{mp}^{(j)} + \overline{\boldsymbol{S}}_m^{c,(j)}$
end

`// Stage update & reconstruction`
for *each element* $\omega_c \in \mathcal{T}_h$ **do**
Form stage combination: $\Sigma_{\boldsymbol{u}}^{c,(j)} := \sum_{i=0}^{s-1} \mathtt{a}_{ij}\,\mathscr{R}_{\boldsymbol{u}}^{c,(i)}$
Update FV-subcell mean values: $\overline{\boldsymbol{u}}^{c,(j)} = \overline{\boldsymbol{u}}^{c,n} + \Delta t^n \Sigma_{\boldsymbol{u}}^{c,(j)}$
Reconstruct DG coefficients: $\underline{\boldsymbol{u}}^{c,(j)} = \mathbb{R}_c\,\overline{\boldsymbol{u}}^{c,(j)}$ (using $\mathbb{R}_c = \mathbb{P}_c^{-1}$ or pseudo-inverse)
end
end

`// 3. Final solution update`
Update final FV-subcell mean values for the next time step:

$$\overline{\boldsymbol{u}}^{c,n+1} = \overline{\boldsymbol{u}}^{c,n} + \Delta t^n \sum_{i=0}^{s-1} \mathtt{b}_i\,\mathscr{R}_{\boldsymbol{u}}^{c,(i)}$$

Algorithm 1: SSP-RK–DG/FV-subcells time step.

Since the mesh update relies strictly on vertex displacements, the grid velocity is approximated using a continuous $\mathbb{P}^1$ Galerkin finite element method defined on the current triangulation $\mathcal{T}_h(t)$. To support arbitrary mesh deformations and prevent element inversion, the DG code embeds two distinct elliptic $\mathbb{P}^1$ solvers, differing primarily in their assembly and coupling strategies:

1. *Assembly of the global stiffness matrix.* Depending on the chosen physical analogy, the local element contributions (integrated using the gradients of standard $\mathbb{P}^1$ hat functions) are assembled into a global sparse matrix:
   - *Harmonic extension*: the x and y components of the grid velocity are assumed to be independent. The solver assembles a single scalar stiffness matrix $\mathbb{K}_\Delta \in \mathscr{M}_{N_v}(\mathbb{R})$,
   - *Pseudo-elasticity*: the mesh is treated as a deformable solid. The x and y displacements are inherently coupled through the Lamé parameters λ and μ, so the solver assembles a larger block-matrix $\mathbb{K}_{\text{elas}} \in \mathscr{M}_{2N_v}(\mathbb{R})$.

   In both cases, a variable stiffness scaling (e.g., the Winslow strategy) can be applied locally during the assembly to stiffen the smallest elements and preserve mesh grading.
2. *Boundary conditions.* The prescribed structural velocity $\partial_t \boldsymbol{x}_{\text{mov}}(t)$ is strongly enforced on the moving boundaries $\partial\Omega_{\text{mov}}(t)$. On the external boundary $\partial\Omega_{\text{fix}}$, depending on the geometric configuration, the solver enforces either a strict no-slip condition ($\boldsymbol{w}^g = \mathbf{0}$) or a slip condition ($\boldsymbol{w}^g \cdot \mathbf{n} = 0$). In our implementation, both Dirichlet and slip conditions are robustly imposed at the algebraic level using a large penalty parameter method.
3. *Linear solve.* Once assembled and penalized, the linear systems are solved to extract the nodal grid velocities:
   - For the Laplacian model, two decoupled linear systems $\mathbb{K}_\Delta \underline{\boldsymbol{w}}^g_\bullet = \underline{\boldsymbol{f}}^g_\bullet$ (with $\bullet \in \{x, y\}$) are solved successively.
   - For the pseudo-elasticity model, a single global system $\mathbb{K}_{\text{elas}} \underline{\boldsymbol{w}}^g = \underline{\boldsymbol{f}}^g$ is solved simultaneously.

   All system resolutions are performed using the highly optimized `SparseLU` direct factorization provided by the `Eigen` C++ library.

A.5.3 Interpolation at Gauss points

Once the nodal velocities $\boldsymbol{w}^g_p$ are computed, they must be accurately evaluated at the quadrature nodes inside the elements and on the faces to compute the modified ALE fluxes. Because we chose a $\mathbb{P}^1$ approximation for the grid velocity, the interpolation at any Gauss point $\boldsymbol{x}_q(t)$ inside an element $\omega_c(t)$ reduces to a straightforward barycentric combination.

Let $\lambda_1, \lambda_2, \lambda_3$ be the barycentric coordinates of the quadrature point $\boldsymbol{x}_q(t)$ with respect to the three vertices $\boldsymbol{x}_A(t), \boldsymbol{x}_B(t), \boldsymbol{x}_C(t)$ of the triangle $\omega_c(t)$. By definition of the affine mapping from the reference element ω_0 (where the quadrature points are defined by their reference coordinates (r, s) as previously shown in Figure A.2), the barycentric coordinates are invariant under the ALE mapping:

$$\lambda_1 = 1 - r - s, \quad \lambda_2 = r, \quad \lambda_3 = s.$$

Thus, the grid velocity at the quadrature point is simply obtained by:

$$\boldsymbol{w}^g_h(\boldsymbol{x}_q(t), t) = \lambda_1 \boldsymbol{w}^g_h(\boldsymbol{x}_A(t), t) + \lambda_2 \boldsymbol{w}^g_h(\boldsymbol{x}_B(t), t) + \lambda_3 \boldsymbol{w}^g_h(\boldsymbol{x}_C(t), t).$$

This linear interpolation is computationally inexpensive, only requiring the fixed reference coordinates (r, s) already stored in the `quadrature` object. The same logic applies to face quadrature nodes, using a 1D linear interpolation between the two nodes defining the face segment.

A.5.4 Time-dependent quadrature and basis functions

As introduced in Chapter 4, the ALE framework relies on a time-dependent mapping

$$\boldsymbol{x}_c(\cdot, t) : \omega_c(0) \ni \boldsymbol{\xi} \mapsto \boldsymbol{x}_c(\boldsymbol{\xi}, t) \in \omega_c(t),$$

together with its inverse

$$\boldsymbol{\xi}_c(\cdot, t) : \omega_c(t) \to \omega_c(0).$$

The physical basis functions used on the moving element $\omega_c(t)$ are therefore obtained by transporting fixed reference basis functions defined on $\omega_c(0)$. More precisely, for any reference basis function $\widetilde{\Psi}_c^{(i,j)} \in \mathbb{P}^k(\omega_c(0))$, we define

$$\Psi_c^{(i,j)}(\boldsymbol{x}, t) := \widetilde{\Psi}_c^{(i,j)}\big(\boldsymbol{\xi}_c(\boldsymbol{x}, t)\big), \qquad \boldsymbol{x} \in \omega_c(t).$$

Hence, along the ALE trajectory $\boldsymbol{x} = \boldsymbol{x}_c(\boldsymbol{\xi}, t)$, one has

$$\Psi_c^{(i,j)}\big(\boldsymbol{x}_c(\boldsymbol{\xi}, t), t\big) = \widetilde{\Psi}_c^{(i,j)}(\boldsymbol{\xi}), \qquad \forall \boldsymbol{\xi} \in \omega_c(0).$$

This is the invariance property stated in Proposition 4.2.3. In practice, it means that the values of the basis functions at the reference quadrature nodes are time-independent. Only the physical positions of these quadrature nodes move with the mesh. This separation is important from an implementation point of view, since the polynomial representation remains attached to the reference configuration, while the geometry is updated through the ALE mapping.

Let $\boldsymbol{\xi}_q$ be a quadrature node on the reference element $\omega_c(0)$. Its physical position at time t is given by

$$\boldsymbol{x}_q(t) = \boldsymbol{x}_c(\boldsymbol{\xi}_q, t).$$

The value of the basis function at this moving quadrature node is therefore simply

$$\Psi_c^{(i,j)}(\boldsymbol{x}_q(t), t) = \widetilde{\Psi}_c^{(i,j)}(\boldsymbol{\xi}_q),$$

and does not need to be recomputed when the mesh moves. The same remark applies to quadrature nodes on faces and subcells, provided that they are transported by the same elementwise ALE mapping.

The situation is different for spatial derivatives. Indeed, the gradients with respect to the physical coordinates depend on the current deformation of the element. Denoting by

$$\mathbb{J}_c(\boldsymbol{\xi}, t) := \nabla_{\boldsymbol{\xi}} \boldsymbol{x}_c(\boldsymbol{\xi}, t)$$

the Jacobian matrix of the ALE mapping, the chain rule gives

$$\nabla_{\boldsymbol{x}} \Psi_c^{(i,j)}\big(\boldsymbol{x}_c(\boldsymbol{\xi}, t), t\big) = \mathbb{J}_c^{-\top}(\boldsymbol{\xi}, t)\, \nabla_{\boldsymbol{\xi}} \widetilde{\Psi}_c^{(i,j)}(\boldsymbol{\xi}).$$

For affine simplicial elements, $\mathbb{J}_c(t)$ is constant on $\omega_c(0)$, so this update is performed once per element and per Runge–Kutta stage. For more general mappings, the same formula holds pointwise at each quadrature node.

Consequently, the implementation distinguishes between two types of data:

- the reference quantities, such as $\widetilde{\Psi}_c^{(i,j)}(\boldsymbol{\xi}_q)$ and $\nabla_{\boldsymbol{\xi}} \widetilde{\Psi}_c^{(i,j)}(\boldsymbol{\xi}_q)$, which are precomputed once and for all,
- the physical geometric quantities, such as $\boldsymbol{x}_q(t)$, $\mathbb{J}_c(t)$, $\mathrm{J}_c(t)$, face lengths, normals and subcell volumes, which must be updated whenever the mesh position is modified.

Thus, at each Runge–Kutta stage, the `basis_functions` object does not need to rebuild the basis itself. It only has to refresh the physical derivatives through

$$\nabla_{\boldsymbol{x}} \Psi_c^{(i,j)} = \mathbb{J}_c^{-\top}(t)\, \nabla_{\boldsymbol{\xi}} \widetilde{\Psi}_c^{(i,j)},$$

together with the associated quadrature metrics.

A.6 SSP-RK–ALE–DG/FV-subcells solver workflow

Here we detail the execution flow of the FV-subcells stabilized DG-ALE solver, highlighting the modifications induced by the dynamic mesh framework and the time-dependence of geometrical quantities. The overall structure of the time loop remains similar to the Eulerian case, but with additional steps for mesh updates, grid velocity computation, and metric refreshment at each Runge-Kutta (RK) stage to ensure consistency with the ALE formulation.

Initialization and preprocessing. Before entering the time-stepping loop, all static variables and reference operators are evaluated. This includes, as usual, the definition of physical parameters, loading the unstructured mesh, and assembling parent-element connectivity. The $\mathbb{P}^1$ finite element solvers for the grid velocity are initialized. Furthermore, the reference submesh, quadrature rules, reference basis functions $\{\varphi_m\}_{m=1}^{N_k}$, and their derivatives are precomputed. Finally, the element-wise mass matrices $\{\mathbb{M}_c\}_{c\in\mathcal{T}_h}$ and the time-invariant reference subcell matrices $\{\mathbb{A}_c, \mathbb{D}_c, \mathbb{S}_c, \mathbb{P}_c\}_{c\in\mathcal{T}_h}$ required for flux reconstruction are assembled.

The fully coupled, moving-mesh discrete workflow is summarized in Algorithm 2.

Input: Solution $\overline{\boldsymbol{u}}^{c,n}$ and mesh coordinates $\boldsymbol{x}_i^n$ at time t^n
Output: Solution $\overline{\boldsymbol{u}}^{c,n+1}$ and mesh coordinates $\boldsymbol{x}_i^{n+1}$ at time t^{n+1}

`// 1. Grid velocity computation at` t^n
if *Harmonic Extension Model* **then**
  Assemble $N_v \times N_v$ scalar stiffness matrix $\mathbb{K}_\Delta$ via $\mathbb{P}^1$ FE
  Solve independently for W_x and W_y using `Eigen::SparseLU`
else
  `// Pseudo-Elasticity Model`
  Assemble $2N_v \times 2N_v$ coupled stiffness matrix $\mathbb{K}_{\text{elas}}$
  Solve for interleaved components (W_x, W_y) using `Eigen::SparseLU`
end
Apply penalty method ($\lambda \approx 10^{12}$) to enforce Dirichlet/Slip boundary conditions

`// 2. Time step evaluation`
Compute Δt^n via CFL condition using FV-subcell states shifted by $\sigma^{\text{ALE}} = |\sigma - \boldsymbol{w}^g \cdot \mathbf{n}_{cv}|$

`// 3. Runge-Kutta stages`
for $j = 0, \dots, s-1$ **do**
  Advance mesh nodes to intermediate stage time $t^{(j)}$ using $\boldsymbol{w}^g$
  Call `update_geometry()` to refresh metrics: $|\omega_c(t)|$, $|S_m^c(t)|$, $\ell_{mp}(t)$, $\mathbf{n}_{cv}(t)$, ...

  `// High-order DG step`
  for *each element* $\omega_c(t) \in \mathcal{T}_h(t)$ **do**
    Evaluate $\boldsymbol{u}_h^c$ and physical ALE fluxes at element quadrature nodes
    Build face traces $\boldsymbol{u}_h^{c,\pm}$, apply BCs, and compute DG-ALE numerical fluxes $\mathbb{G}_\Gamma^{*,\text{DG}}$
    Assemble elemental DG-ALE residual: $\mathscr{R}_{\text{DG}}^{c,(j)} = \mathbb{M}_c^{-1}(\boldsymbol{\Xi}_c^{\text{ALE}} + \boldsymbol{S}_c)$
    Compute subcell source terms: $\overline{\boldsymbol{S}}_m^{c,(j)} = \mathbb{P}_c \mathbb{M}_c^{-1}\Big(\int_{\omega_c(t)} \boldsymbol{S}_h \varphi_h \, \mathrm{d}\boldsymbol{x}\Big)$
  end

  `// First-order FV step`
  for *each subface* $\Gamma_{mp}(t) \in \mathcal{F}_m^c(t)$ **do**
    Build int./ext. FV states, impose BCs, and compute first-order FV-ALE flux $\mathbb{G}_{mp}^{*,\text{FV}}$
  end

  `// Subcell flux blending and residuals`
  for *each element* $\omega_c(t) \in \mathcal{T}_h(t)$ **do**
    Reconstruct DG-ALE flux on subfaces: $\widehat{\mathbb{G}}_c^{(j)} = -\mathbb{A}_c^\top \mathscr{L}_c^{-1}\big(\mathbb{D}_c \mathbb{P}_c \mathbb{M}_c^{-1} \boldsymbol{\Xi}_c^{\text{ALE}} + \partial \mathbb{G}_c\big)$
  end
  for *each subface* $\Gamma_{mp}(t) \in \mathcal{F}_m^c(t)$ **do**
    Compute blending coefficients $\Theta_{mp} \in [0,1]$
    Form blended ALE flux: $\widetilde{\mathbb{G}}_{mp}^{(j)} = \mathbb{G}_{mp}^{*,\text{FV}} + \Theta_{mp}\big(\widehat{\mathbb{G}}_{mp}^{(j)} - \mathbb{G}_{mp}^{*,\text{FV}}\big)$
    Assemble FV-subcell residual: $|S_m^c(t^{(j)})|\mathscr{R}_{\boldsymbol{u},m}^{c,(j)} = -\sum_p \ell_{mp}(t^{(j)}) \widetilde{\mathbb{G}}_{mp}^{(j)} + |S_m^c(t^{(j)})| \overline{\boldsymbol{S}}_m^{c,(j)}$
  end

  `// Stage update & reconstruction`
  for *each element* $\omega_c(t) \in \mathcal{T}_h(t)$ **do**
    Form stage combination: $\Sigma_{\boldsymbol{u}}^{c,(j)} := \sum_{i=0}^{s-1} \mathsf{a}_{ij} \, |S_m^c(t^{(i)})| \mathscr{R}_{\boldsymbol{u}}^{c,(i)}$
    Update FV-subcell means: $|S_m^c(t^{(j)})| \overline{\boldsymbol{u}}^{c,(j)} = |S_m^c(t^n)| \overline{\boldsymbol{u}}^{c,n} + \Delta t^n \Sigma_{\boldsymbol{u}}^{c,(j)}$
    Reconstruct DG coefficients: $\underline{\boldsymbol{u}}^{c,(j)} = \mathbb{R}_c \, \overline{\boldsymbol{u}}^{c,(j)}$ (using $\mathbb{R}_c = \mathbb{P}_c^{-1}$ or pseudo-inverse)
  end
end

`// 4. Final solution update`
Update final FV-subcell mean values for the next time step:

$$|S_m^{c,n+1}| \overline{\boldsymbol{u}}_m^{c,n+1} = |S_m^{c,n}| \overline{\boldsymbol{u}}_m^{c,n} + \Delta t^n \sum_{i=0}^{s-1} \mathsf{b}_i \, |S_m^c(t^{(i)})| \mathscr{R}_{\boldsymbol{u},m}^{c,(i)}$$

Algorithm 2: SSP-RK–ALE–DG/FV-subcells time step.

APPENDIX B

Positivity-preserving first-order Riemann intermediate states

In order to construct high-order schemes that remain both accurate and stable, it is essential to rely on robust first-order building blocks. In particular, when dealing with the NSW equations, positivity of the water-height must be guaranteed at the subcell level. This property is inherited from the underlying first-order FV scheme, through the intermediate Riemann states it produces. For completeness, we provide in this appendix a proof of positivity preservation in the two-dimensional case, which can be easily adapted to the one-dimensional setting. We will rely here on the notations and framework introduced in Chapter 3.

Proposition. *The water-height positivity is ensured on first-order FV Riemann intermediate states for the subcell FV fluxes* (3.29) *and global LF flux* (3.27).

Proof. Here, since we are interested only in conservation of water-height positivity, we just need to ensure that for all $\omega_c \in \mathcal{T}_h$, for all $S_m^c \in \mathcal{T}_{\omega_c}$, $\overline{\eta}_m^{c,n} \geq \overline{b}_m^{c,n}$. Dropping the superscript n and using the notation $\mathbb{F}_{mp}^{*,\eta}$ for the first component of the numerical FV flux $\mathbb{F}_{mp}^{*,\mathrm{FV}}$, we only need to verify is that for any $S_m^c \in \mathcal{T}_{\omega_c}$, we have

$$\eta_{mp}^* = \overline{\eta}_m^c - \frac{\mathbb{F}_{mp}^{*,\eta} - \overline{\boldsymbol{q}}_m^c \cdot \mathbf{n}_{mp}}{\sigma} \geq \overline{b}_m^c, \quad \text{and} \quad \eta_{pm}^* = \overline{\eta}_p^v - \frac{\mathbb{F}_{pm}^{*,\eta} - \overline{\boldsymbol{q}}_p^v \cdot \mathbf{n}_{pm}}{\sigma} \geq \overline{b}_p^v.$$

Noticing that $\check{\overline{\eta}}_p^v - \check{\overline{\eta}}_m^c = \check{\overline{H}}_p^v - \check{\overline{H}}_m^c + \check{\overline{b}}_{mp} - \check{\overline{b}}_{mp} = \check{\overline{H}}_p^v - \check{\overline{H}}_m^c$, and using the velocity notation $\overline{\boldsymbol{u}}_m^c := \frac{\overline{\boldsymbol{q}}_m^c}{\overline{H}_m^c}$, we have

$$\begin{aligned}
\eta_{mp}^* &= \overline{\eta}_m^c - \frac{\mathbb{F}_{mp}^{*,\eta} - \overline{\boldsymbol{q}}_m^c \cdot \mathbf{n}_{mp}}{\sigma} = \overline{H}_m^c + \overline{b}_m^c - \frac{\mathbb{F}_{mp}^{*,\eta} - \overline{\boldsymbol{q}}_m^c \cdot \mathbf{n}_{mp}}{\sigma} \\
&= \frac{1}{2}\left(2\overline{H}_m^c + \check{\overline{H}}_p^v - \check{\overline{H}}_m^c\right) - \frac{1}{2\sigma}\left(\check{\overline{\boldsymbol{q}}}_p^v - \check{\overline{\boldsymbol{q}}}_m^c - 2\overline{\boldsymbol{q}}_m^c\right) \cdot \mathbf{n}_{mp} + \overline{b}_m^c \\
&= \frac{1}{2}\left(2\overline{H}_m^c + \check{\overline{H}}_p^v - \check{\overline{H}}_m^c\right) - \frac{1}{2\sigma}\left(\check{\overline{H}}_p^v \overline{\boldsymbol{u}}_m^c - \check{\overline{H}}_m^c \overline{\boldsymbol{u}}_m^c - 2\overline{H}_m^c \overline{\boldsymbol{u}}_m^c\right) \cdot \mathbf{n}_{mp} + \overline{b}_m^c.
\end{aligned}$$

Writing it this way demonstrate that we just need to have

$$\frac{1}{2}\left(2\overline{H}_m^c + \check{\overline{H}}_p^v - \check{\overline{H}}_m^c\right) - \frac{1}{2\sigma}\left(\check{\overline{H}}_p^v \overline{\boldsymbol{u}}_m^c - \check{\overline{H}}_m^c \overline{\boldsymbol{u}}_m^c - 2\overline{H}_m^c \overline{\boldsymbol{u}}_m^c\right) \cdot \mathbf{n}_{mp} \geq 0.$$

By construction, we easily have $0 \leq \check{\overline{H}}_m^c \leq \overline{H}_m^c$ and $0 \leq \check{\overline{H}}_p^v \leq \overline{H}_p^v$, so we easily get $2\overline{H}_m^c + \check{\overline{H}}_p^v - \check{\overline{H}}_m^c \geq 2\overline{H}_m^c$. For the other term, let us rearrange it into

$$\left(\check{\overline{H}}_p^v \overline{\boldsymbol{u}}_m^c - \check{\overline{H}}_m^c \overline{\boldsymbol{u}}_m^c - 2\overline{H}_m^c \overline{\boldsymbol{u}}_m^c\right) \cdot \mathbf{n}_{mp} = \left(\left(2\overline{H}_m^c - \check{\overline{H}}_m^c\right)\frac{1 + \overline{\boldsymbol{u}}_m^c \sigma^{-1}}{2} + \check{\overline{H}}_p^v \frac{1 + \overline{\boldsymbol{u}}_p^v \sigma^{-1}}{2}\right) \cdot \mathbf{n}_{mp},$$

and according to CFL (3.4) and hydrostatic reconstructions, we easily get:

$$2\overline{H}_m^c - \check{\overline{H}}_m^c \geq \overline{H}_m^c, \qquad \frac{1 + \overline{\boldsymbol{u}}_m^c \sigma^{-1}}{2} \geq \mathbf{0}, \qquad \check{\overline{H}}_p^v \geq 0, \qquad \frac{1 + \overline{\boldsymbol{u}}_p^v \sigma^{-1}}{2} \geq \mathbf{0}.$$

Gathering all the results give us that $\eta_{mp}^* \geq \overline{b}_m^c$, what we wanted. A similar proof goes for $\eta_{mp}^* \geq \overline{b}_p^v$. □

APPENDIX C

Extension of model (6.15)-(6.18) towards heaving floating structures

The formulation presented in Section 6.1.1 assumes that the object is fixed, allowing for a rigorous analysis of the corresponding IBVP. In order to study 2D simulations and compare them to the results of Chapter 5, we describe in this appendix a formal derivation of the model which allows for heaving motions, in the spirit of [34,35]. Parametrizing the free-surface under the structure as

$$\zeta^{\mathrm{w}}(\boldsymbol{x},t) = \zeta^{\mathrm{w}}_{\mathrm{eq}}(\boldsymbol{x}) + \delta_{\mathcal{O}}(t), \qquad \boldsymbol{x} \in \Omega_{\mathrm{s}}. \tag{C.1}$$

the equilibrium configuration is described by $\zeta^{\mathrm{w}}_{\mathrm{eq}}$, and the rigid motion assumption implies that the horizontal geometry of the wetted region does not change in time, and as $\zeta^{\mathrm{w}}_{\mathrm{eq}}$ is not assumed to be constant, we have:

$$H^{\mathrm{w}}(\boldsymbol{x},t) = H_0 + \zeta^{\mathrm{w}}_{\mathrm{eq}}(\boldsymbol{x}) + \delta_{\mathcal{O}}(t)$$

As in the one-dimensional case, the vertical motion induces a prescribed divergence of the interior discharge, while the horizontal redistribution of mass is governed by a solenoidal component.

Interior kinematic constraint and representation of the discharge

Inside Ω_{s}, the free-surface satisfies (C.1), so the mass conservation equation yields

$$\nabla_{\boldsymbol{x}} \cdot (H^{\mathrm{w}}\boldsymbol{u}^{\mathrm{s}}) = \nabla_{\boldsymbol{x}} \cdot \boldsymbol{q}^{\mathrm{s}} = -\dot{\delta}_{\mathcal{O}}(t), \qquad \boldsymbol{x} \in \Omega_{\mathrm{s}}, \tag{C.2}$$

where $\boldsymbol{q}^{\mathrm{s}} = H^{\mathrm{w}}\boldsymbol{u}^{\mathrm{s}}$ is the discharge under the structure. To isolate the effects of the rigid-body vertical motion from the local fluid circulation, and following a *Helmholtz–Hodge* decomposition, we consider the discharge is split into two distinct parts: an irrotational accounting for the vertical motion and a divergence-free component:

$$\boldsymbol{q}^{\mathrm{s}}(\boldsymbol{x},t) = \boldsymbol{q}^{\mathrm{s}}_{\mathrm{part}}(\boldsymbol{x},t) + \boldsymbol{q}^{\mathrm{s}}_{\mathrm{sol}}(\boldsymbol{x},t).$$

Let us remark that unlike the 1D case or infinite domains, there is no particular solution that produces no normal flux through the boundary $\partial\Omega_{\mathrm{fs}}$. Indeed, integrating (C.2) over Ω_{s} and using the Gauss-Ostrogradski divergence theorem yields

$$\int_{\partial\Omega_{\mathrm{fs}}} \boldsymbol{q}^{\mathrm{s}} \cdot \mathbf{n}\, \mathrm{d}s = -|\Omega_{\mathrm{s}}|\dot{\delta}_{\mathcal{O}}(t),$$

i.e. the vertical motion of the object inherently displaces fluid across the contact line.

We now choose the particular solution

$$\boldsymbol{q}^{\mathrm{s}}_{\mathrm{part}}(\boldsymbol{x},t) = -\dot{\delta}_{\mathcal{O}}(t)\, \nabla_{\boldsymbol{x}} \Upsilon(\boldsymbol{x}),$$

where Υ solves the well-posed Dirichlet problem

$$\begin{cases} \Delta_{\boldsymbol{x}} \Upsilon = 1 & \text{in } \Omega_{\mathrm{s}}, \\ \Upsilon = 0 & \text{on } \partial\Omega_{\mathrm{fs}}, \end{cases}$$

ensuring that $\nabla_{\boldsymbol{x}} \cdot \boldsymbol{q}^{\mathrm{s}}_{\mathrm{part}} = -\dot{\delta}_{\mathcal{O}}$. The divergence–free component $\boldsymbol{q}^{\mathrm{s}}_{\mathrm{sol}}$ describes the horizontal redistribution of mass below the structure and we retain the simplest choice, namely a spatially uniform field $\overline{\boldsymbol{q}}(t)$, so that

$$\boldsymbol{q}^{\mathrm{s}}(\boldsymbol{x},t) = \overline{\boldsymbol{q}}(t) - \dot{\delta}_{\mathcal{O}}(t)\, \nabla_{\boldsymbol{x}} \Upsilon(\boldsymbol{x}). \tag{C.3}$$

The spatially uniform ansatz for $\boldsymbol{q}^{\mathrm{s}}_{\mathrm{sol}}$ is a reduced finite-dimensional closure; it is not the general solenoidal component of a Helmholtz–Hodge decomposition. Note that $\overline{\boldsymbol{q}}(t)$ is related to, but generally differs from, the exact spatial mean of $\boldsymbol{q}^{\mathrm{s}}$, since $\nabla_{\boldsymbol{x}}\Upsilon$ may have a non-zero integral over Ω_{s}.

Interior momentum equation and elliptic problem for the Bernoulli pressure

The interior momentum equation, written in terms of the Bernoulli pressure, reads

$$\partial_t \boldsymbol{u}^{\mathrm{s}} + \boldsymbol{u}^{\mathrm{s}} \cdot \nabla_{\boldsymbol{x}} \boldsymbol{u}^{\mathrm{s}} - \frac{1}{2}\nabla_{\boldsymbol{x}}|\boldsymbol{u}^{\mathrm{s}}|^2 = -\frac{1}{\rho}\nabla_{\boldsymbol{x}}\Pi^{\mathrm{s}}, \qquad \Pi^{\mathrm{s}} = \underline{\mathrm{p}}^{\mathrm{s}} - \mathrm{p}_{\mathrm{atm}} + \rho \mathrm{g}\zeta^{\mathrm{s}} + \tfrac{1}{2}\rho|\boldsymbol{u}^{\mathrm{s}}|^2,$$

with $\boldsymbol{u}^{\mathrm{s}} = \boldsymbol{q}^{\mathrm{s}}/H^{\mathrm{w}}$ and $H^{\mathrm{w}}(\boldsymbol{x},t) = H_0 + \zeta^{\mathrm{w}}_{\mathrm{eq}}(\boldsymbol{x}) + \delta_{\mathrm{o}}(t)$. Since only $\delta_{\mathrm{o}}(t)$ depends on time, we have $\partial_t H^{\mathrm{w}} = \dot{\delta_{\mathrm{o}}}(t)$. Let us compute the time derivative:

$$\partial_t \boldsymbol{u}^{\mathrm{s}} = \partial_t\left(\frac{\boldsymbol{q}^{\mathrm{s}}}{H^{\mathrm{w}}}\right) = \frac{(\partial_t \boldsymbol{q}^{\mathrm{s}})H^{\mathrm{w}} - \boldsymbol{q}^{\mathrm{s}}(\partial_t H^{\mathrm{w}})}{(H^{\mathrm{w}})^2} = \frac{\partial_t \boldsymbol{q}^{\mathrm{s}}}{H^{\mathrm{w}}} - \frac{\partial_t H^{\mathrm{w}}}{(H^{\mathrm{w}})^2}\boldsymbol{q}^{\mathrm{s}} = \frac{\partial_t \boldsymbol{q}^{\mathrm{s}}}{H^{\mathrm{w}}} - \frac{\dot{\delta_{\mathrm{o}}}}{(H^{\mathrm{w}})^2}\boldsymbol{q}^{\mathrm{s}}.$$

By (C.3), the discharge $\boldsymbol{q}^{\mathrm{s}}$ is irrotational since it is the gradient of $\overline{\boldsymbol{q}}(t)\cdot\boldsymbol{x} - \dot{\delta_{\mathrm{o}}}(t)\Upsilon(\boldsymbol{x})$. Therefore, $\boldsymbol{q}^{\mathrm{s}} \cdot \nabla_{\boldsymbol{x}} \boldsymbol{q}^{\mathrm{s}} = \frac{1}{2}\nabla_{\boldsymbol{x}}|\boldsymbol{q}^{\mathrm{s}}|^2$.

The nonlinear correction in the Bernoulli formulation then expands as

$$\boldsymbol{u}^{\mathrm{s}} \cdot \nabla_{\boldsymbol{x}} \boldsymbol{u}^{\mathrm{s}} - \frac{1}{2}\nabla_{\boldsymbol{x}}|\boldsymbol{u}^{\mathrm{s}}|^2 = \frac{|\boldsymbol{q}^{\mathrm{s}}|^2\nabla_{\boldsymbol{x}}H^{\mathrm{w}} - \boldsymbol{q}^{\mathrm{s}}\big(\boldsymbol{q}^{\mathrm{s}}\cdot\nabla_{\boldsymbol{x}}H^{\mathrm{w}}\big)}{(H^{\mathrm{w}})^3}.$$

Inserting these expressions into the momentum equation and multiplying by H^{w} yields

$$\partial_t \boldsymbol{q}^{\mathrm{s}} - \frac{\dot{\delta_{\mathrm{o}}}}{H^{\mathrm{w}}}\boldsymbol{q}^{\mathrm{s}} + \frac{|\boldsymbol{q}^{\mathrm{s}}|^2\nabla_{\boldsymbol{x}}H^{\mathrm{w}} - \boldsymbol{q}^{\mathrm{s}}\big(\boldsymbol{q}^{\mathrm{s}}\cdot\nabla_{\boldsymbol{x}}H^{\mathrm{w}}\big)}{(H^{\mathrm{w}})^2} = -\frac{H^{\mathrm{w}}}{\rho}\nabla_{\boldsymbol{x}}\Pi^{\mathrm{s}}. \tag{C.4}$$

Taking the divergence of (C.4) gives

$$\nabla_{\boldsymbol{x}}\cdot(\partial_t \boldsymbol{q}^{\mathrm{s}}) - \nabla_{\boldsymbol{x}}\cdot\left(\frac{\dot{\delta_{\mathrm{o}}}}{H^{\mathrm{w}}}\boldsymbol{q}^{\mathrm{s}}\right) + \nabla_{\boldsymbol{x}}\cdot\left(\frac{|\boldsymbol{q}^{\mathrm{s}}|^2\nabla_{\boldsymbol{x}}H^{\mathrm{w}} - \boldsymbol{q}^{\mathrm{s}}\big(\boldsymbol{q}^{\mathrm{s}}\cdot\nabla_{\boldsymbol{x}}H^{\mathrm{w}}\big)}{(H^{\mathrm{w}})^2}\right) = -\nabla_{\boldsymbol{x}}\cdot\left(\frac{H^{\mathrm{w}}}{\rho}\nabla_{\boldsymbol{x}}\Pi^{\mathrm{s}}\right),$$

and using the divergence constraint (C.2), we simply have:

$$\nabla_{\boldsymbol{x}}\cdot(\partial_t \boldsymbol{q}^{\mathrm{s}}) = \partial_t(\nabla_{\boldsymbol{x}}\cdot\boldsymbol{q}^{\mathrm{s}}) = \partial_t(-\dot{\delta_{\mathrm{o}}}(t)) = -\ddot{\delta_{\mathrm{o}}}(t).$$

Inserting this result and isolating the pressure term, we obtain the elliptic equation for the interior Bernoulli pressure:

$$-\nabla_{\boldsymbol{x}}\cdot\left(\frac{H^{\mathrm{w}}}{\rho}\nabla_{\boldsymbol{x}}\Pi^{\mathrm{s}}\right) = -\ddot{\delta_{\mathrm{o}}}(t) - \nabla_{\boldsymbol{x}}\cdot\left(\frac{\dot{\delta_{\mathrm{o}}}}{H^{\mathrm{w}}}\boldsymbol{q}^{\mathrm{s}}\right) + \mathfrak{R}(\dot{\delta_{\mathrm{o}}}, \overline{\boldsymbol{q}}), \tag{C.5}$$

where the nonlinear remainder $\mathfrak{R}$ gathers the purely spatial and quadratic terms:

$$\mathfrak{R}(\dot{\delta_{\mathrm{o}}}, \overline{\boldsymbol{q}}) = \nabla_{\boldsymbol{x}}\cdot\left(\frac{|\boldsymbol{q}^{\mathrm{s}}|^2\nabla_{\boldsymbol{x}}H^{\mathrm{w}} - \boldsymbol{q}^{\mathrm{s}}\big(\boldsymbol{q}^{\mathrm{s}}\cdot\nabla_{\boldsymbol{x}}H^{\mathrm{w}}\big)}{(H^{\mathrm{w}})^2}\right).$$

Evolution equation for the mean discharge

Integrating the interior momentum equation $\partial_t \boldsymbol{u}^{\mathrm{s}} + \boldsymbol{u}^{\mathrm{s}}\cdot\nabla_{\boldsymbol{x}}\boldsymbol{u}^{\mathrm{s}} - \frac{1}{2}\nabla_{\boldsymbol{x}}|\boldsymbol{u}^{\mathrm{s}}|^2 = -\frac{1}{\rho}\nabla_{\boldsymbol{x}}\Pi^{\mathrm{s}}$ over Ω_{s} and applying the divergence theorem yields an evolution equation for the macroscopic flow. Indeed, the time derivative integral becomes

$$\int_{\Omega_{\mathrm{s}}} \partial_t \boldsymbol{u}^{\mathrm{s}}\,\mathrm{d}\boldsymbol{x} = \frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega_{\mathrm{s}}}\frac{\boldsymbol{q}^{\mathrm{s}}}{H^{\mathrm{w}}}\,\mathrm{d}\boldsymbol{x} = \frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega_{\mathrm{s}}}\left(\frac{\overline{\boldsymbol{q}}(t)}{H^{\mathrm{w}}} - \dot{\delta_{\mathrm{o}}}(t)\frac{\nabla_{\boldsymbol{x}}\Upsilon}{H^{\mathrm{w}}}\right)\,\mathrm{d}\boldsymbol{x},$$

and injecting the decomposition (C.3) leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega_{\mathrm{s}}}\frac{\boldsymbol{q}^{\mathrm{s}}}{H^{\mathrm{w}}}\,\mathrm{d}\boldsymbol{x} = \frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega_{\mathrm{s}}}\frac{\overline{\boldsymbol{q}}(t) - \dot{\delta_{\mathrm{o}}}(t)\nabla_{\boldsymbol{x}}\Upsilon(\boldsymbol{x})}{H^{\mathrm{w}}(\boldsymbol{x},\delta_{\mathrm{o}})}\,\mathrm{d}\boldsymbol{x} = \frac{\mathrm{d}}{\mathrm{d}t}\left\{\overline{\boldsymbol{q}}(t)\int_{\Omega_{\mathrm{s}}}\frac{\mathrm{d}\boldsymbol{x}}{H^{\mathrm{w}}(\boldsymbol{x},\delta_{\mathrm{o}})} - \dot{\delta_{\mathrm{o}}}(t)\int_{\Omega_{\mathrm{s}}}\frac{\nabla_{\boldsymbol{x}}\Upsilon(\boldsymbol{x})}{H^{\mathrm{w}}(\boldsymbol{x},\delta_{\mathrm{o}})}\,\mathrm{d}\boldsymbol{x}\right\}.$$

As in [34], the last term is identified as an added-mass effect, allowing to rewrite this system compactly:

$$|\Omega_{\mathrm{s}}|\,\frac{\mathrm{d}}{\mathrm{d}t}\big(\alpha(\delta_{\mathrm{o}})\overline{\boldsymbol{q}} - \dot{\delta_{\mathrm{o}}}\boldsymbol{\beta}(\delta_{\mathrm{o}})\big) + \int_{\Omega_{\mathrm{s}}}\left(\boldsymbol{u}^{\mathrm{s}}\cdot\nabla_{\boldsymbol{x}}\boldsymbol{u}^{\mathrm{s}} - \frac{1}{2}\nabla_{\boldsymbol{x}}|\boldsymbol{u}^{\mathrm{s}}|^2\right)\,\mathrm{d}\boldsymbol{x} = -\frac{1}{\rho}\int_{\partial\Omega_{\mathrm{fs}}}\Pi^{\mathrm{s}}\,\mathbf{n}\,\mathrm{d}s,$$

with $\alpha(\delta_o)$ and $\boldsymbol{\beta}(\delta_o)$ defined as

$$\alpha(\delta_o) := \frac{1}{|\Omega_s|}\int_{\Omega_s} \frac{\mathrm{d}\boldsymbol{x}}{H^w(\boldsymbol{x},\delta_o)}, \qquad \boldsymbol{\beta}(\delta_o) := \frac{1}{|\Omega_s|}\int_{\Omega_s} \frac{\nabla_{\boldsymbol{x}}\Upsilon(\boldsymbol{x})}{H^w(\boldsymbol{x},\delta_o)}\,\mathrm{d}\boldsymbol{x}.$$

Using again (C.3), the transmission conditions on the contact line naturally write

$$H\boldsymbol{u}\cdot\mathbf{n} = \big(-\dot{\delta_o}\,\nabla_{\boldsymbol{x}}\Upsilon(\boldsymbol{x}) + \overline{\boldsymbol{q}}\big)\cdot\mathbf{n} \quad \text{and} \quad \Pi = \Pi^s \qquad \text{on } \partial\Omega_{fs},$$

and the equations governing the full wave-structure interactions problem for a heaving object can be summarized as

$$\begin{cases}
\partial_t\zeta^f + \nabla_{\boldsymbol{x}}\cdot(H^f\boldsymbol{u}^f) = 0 & \text{in } \Omega_f\times(0,T),\\
\partial_t\boldsymbol{u}^f + \boldsymbol{u}^f\cdot\nabla_{\boldsymbol{x}}\boldsymbol{u}^f + g\nabla_{\boldsymbol{x}}\zeta^f = \mathbf{0} & \text{in } \Omega_f\times(0,T),\\
\boldsymbol{q}^s = \overline{\boldsymbol{q}} - w_o\nabla_{\boldsymbol{x}}\Upsilon & \text{in } \Omega_s\times(0,T),\\
-\nabla_{\boldsymbol{x}}\cdot\Big(\frac{H^w}{\rho}\nabla_{\boldsymbol{x}}\Pi^s\Big) = -\dot{w_o} - \nabla_{\boldsymbol{x}}\cdot\big(\frac{w_o}{H^w}\boldsymbol{q}^s\big) + \mathfrak{R}(\boldsymbol{x},w_o,\overline{\boldsymbol{q}}) & \text{in } \Omega_s\times(0,T),\\
|\Omega_s|\,\frac{\mathrm{d}}{\mathrm{d}t}\big(\alpha(\delta_o)\overline{\boldsymbol{q}} - w_o\boldsymbol{\beta}(\delta_o)\big) + \int_{\Omega_s}\big(\boldsymbol{u}^s\cdot\nabla_{\boldsymbol{x}}\boldsymbol{u}^s - \frac{1}{2}\nabla_{\boldsymbol{x}}|\boldsymbol{u}^s|^2\big)\,\mathrm{d}\boldsymbol{x} = -\frac{1}{\rho}\int_{\partial\Omega_{fs}}\Pi^s\,\mathbf{n}\,\mathrm{d}s & \text{in } (0,T),\\
\dot{\delta_o} = w_o & \text{in } (0,T),\\
\dot{w_o} = \tau_{\text{buoy}}^{-2}\Big[\frac{1}{\rho g|\Omega_s|}\int_{\Omega_s}\big(\Pi^s - \frac{1}{2}\rho|\boldsymbol{u}^s|^2\big)\,\mathrm{d}\boldsymbol{x} - \delta_o\Big] & \text{in } (0,T),\\
\Pi^f = \rho g\zeta^f + \frac{1}{2}\rho|\boldsymbol{u}^f|^2 & \text{in } \Omega_f\times(0,T),\\
\Pi^s = \underline{p}^s - p_{\text{atm}} + \rho g\zeta^s + \frac{1}{2}\rho|\boldsymbol{u}^s|^2 & \text{in } \Omega_s\times(0,T),\\
H^f\boldsymbol{u}^f\cdot\mathbf{n} = \big(\overline{\boldsymbol{q}} - w_o\nabla_{\boldsymbol{x}}\Upsilon\big)\cdot\mathbf{n} & \text{on } \partial\Omega_{fs}\times(0,T),\\
\Pi^f = \Pi^s & \text{on } \partial\Omega_{fs}\times(0,T),\\
\boldsymbol{u}^f\cdot\mathbf{n}^\perp = \frac{\boldsymbol{q}^s}{H^w}\cdot\mathbf{n}^\perp & \text{on } \partial\Omega_{fs}\times(0,T).
\end{cases} \tag{C.6}$$

where the different equations in system (C.6) encapsulate the fully coupled dynamics:

(*i*) The first two equations are the standard nonlinear shallow-water equations in the exterior fluid domain Ω_f, governing the free-surface elevation ζ and the horizontal velocity $\boldsymbol{u}$,

(*ii*) The third and fourth equations describe the interior region beneath the structure, they decompose the discharge $\boldsymbol{q}^s$ into a mean component $\overline{\boldsymbol{q}}$ and a geometric correction driven by the vertical velocity $w_o := \dot{\delta_o}$, and they define the elliptic problem satisfied by the interior Bernoulli pressure Π^s,

(*iii*) The fifth equation governs the evolution of the macroscopic discharge $\overline{\boldsymbol{q}}(t)$, compared with the fixed-structure case, the vertical motion introduces added-mass effects through the geometric coefficients $\alpha(\delta_o)$ and $\boldsymbol{\beta}(\delta_o)$,

(*iv*) The sixth and seventh equations govern the rigid-body motion of the structure, as in the 1D model of [19], the vertical displacement satisfies a forced oscillator equation as described in (5.14), with a hydrostatic restoring force and a forcing term coming from the averaged nonlinear pressure under the hull,

(*v*) The remaining equations define the Bernoulli pressures Π and Π^s, and enforce the transmission conditions on the contact line $\partial\Omega_{fs}$, in particular, the normal flux condition is modified by the vertical motion through the term $\big(\overline{\boldsymbol{q}} - w_o\nabla_{\boldsymbol{x}}\Upsilon\big)\cdot\mathbf{n}$.

APPENDIX D

Well-posedness of the HHO scheme for the wave-structure model

In this appendix, we provide a proof of the well-posedness of the Hybrid High-Order (HHO) scheme introduced in Section 6.2.4 for the wave-structure interactions model. The discrete elliptic operator plays a central role in the definition of the DtN map at the fluid–structure interface, and its stability and uniqueness properties are essential for ensuring the robustness of the fully coupled numerical scheme. Under standard assumptions on the interior water-height and on the stabilization operator, we prove that the discrete HHO problem admits a unique solution and satisfies a stability estimate consistent with the continuous theory.

Proposition. *Assume that the coefficient H^{s} satisfies $0 < H_{\min} \le H^{\mathrm{s}}(\boldsymbol{x}) \le H_{\max} < +\infty$ a.e. in Ω_{s}, and that the stabilization $\mathsf{stab}_c^{\mathrm{s}}$ satisfies* (S1)–(S3) *for all $\omega_c^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}$. Then, for any Dirichlet datum $\psi^{\mathrm{s}} \in H^{1/2}(\partial\Omega_{\mathrm{fs}})$, the discrete problem* (6.39) *admits a unique solution $\underline{\phi}_h^{\mathrm{s}} \in \underline{\mathsf{U}}_h^{k,\mathrm{s}}$ satisfying* (6.41). *Moreover, the discrete solution satisfies the stability estimate*

$$\|\underline{\phi}_h^{\mathrm{s}}\|_{1,h} \lesssim \|\psi^{\mathrm{s}}\|_{H^{1/2}(\partial\Omega_{\mathrm{fs}})}.$$

Proof. We divide the proof into several steps that will be detailed below.

(*i*) *Coercivity and continuity.* For any cell $\omega_c^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}$ and any $\underline{w}_c^{\mathrm{s}} \in \underline{\mathsf{U}}_c^{k,\mathrm{s}}$, recall from (6.37) that

$$\mathcal{A}_c^{\mathrm{s}}(\underline{w}_c^{\mathrm{s}}, \underline{w}_c^{\mathrm{s}}) = \left(H^{\mathrm{s}}\, \mathbf{G}_c^{k,\mathrm{s}} \underline{w}_c^{\mathrm{s}}, \mathbf{G}_c^{k,\mathrm{s}} \underline{w}_c^{\mathrm{s}}\right)_{\omega_c^{\mathrm{s}}} + \mathsf{stab}_c^{\mathrm{s}}(\underline{w}_c^{\mathrm{s}}, \underline{w}_c^{\mathrm{s}}).$$

Using the uniform bounds $0 < H_{\min} \le H^{\mathrm{s}}(\boldsymbol{x}) \le H_{\max} < +\infty$, we first obtain

$$\left(H^{\mathrm{s}}\, \mathbf{G}_c^{k,\mathrm{s}} \underline{w}_c^{\mathrm{s}}, \mathbf{G}_c^{k,\mathrm{s}} \underline{w}_c^{\mathrm{s}}\right)_{\omega_c^{\mathrm{s}}} \ge H_{\min} \left\|\mathbf{G}_c^{k,\mathrm{s}} \underline{w}_c^{\mathrm{s}}\right\|_{\omega_c^{\mathrm{s}}}^2.$$

Therefore,

$$\begin{aligned}
\mathcal{A}_c^{\mathrm{s}}(\underline{w}_c^{\mathrm{s}}, \underline{w}_c^{\mathrm{s}}) &\ge H_{\min} \left\|\mathbf{G}_c^{k,\mathrm{s}} \underline{w}_c^{\mathrm{s}}\right\|_{\omega_c^{\mathrm{s}}}^2 + \mathsf{stab}_c^{\mathrm{s}}(\underline{w}_c^{\mathrm{s}}, \underline{w}_c^{\mathrm{s}}) \\
&\ge \min(1, H_{\min}) \left(\left\|\mathbf{G}_c^{k,\mathrm{s}} \underline{w}_c^{\mathrm{s}}\right\|_{\omega_c^{\mathrm{s}}}^2 + \mathsf{stab}_c^{\mathrm{s}}(\underline{w}_c^{\mathrm{s}}, \underline{w}_c^{\mathrm{s}})\right).
\end{aligned}$$

Similarly, the upper bound on H^{s} gives

$$\begin{aligned}
\mathcal{A}_c^{\mathrm{s}}(\underline{w}_c^{\mathrm{s}}, \underline{w}_c^{\mathrm{s}}) &\le H_{\max} \left\|\mathbf{G}_c^{k,\mathrm{s}} \underline{w}_c^{\mathrm{s}}\right\|_{\omega_c^{\mathrm{s}}}^2 + \mathsf{stab}_c^{\mathrm{s}}(\underline{w}_c^{\mathrm{s}}, \underline{w}_c^{\mathrm{s}}) \\
&\le \max(1, H_{\max}) \left(\left\|\mathbf{G}_c^{k,\mathrm{s}} \underline{w}_c^{\mathrm{s}}\right\|_{\omega_c^{\mathrm{s}}}^2 + \mathsf{stab}_c^{\mathrm{s}}(\underline{w}_c^{\mathrm{s}}, \underline{w}_c^{\mathrm{s}})\right).
\end{aligned}$$

On the other hand, the local stability property (S2) of the HHO stabilization gives

$$\alpha^{-1} \|\underline{w}_c^{\mathrm{s}}\|_{1,c}^2 \le \left\|\mathbf{G}_c^{k,\mathrm{s}} \underline{w}_c^{\mathrm{s}}\right\|_{\omega_c^{\mathrm{s}}}^2 + \mathsf{stab}_c^{\mathrm{s}}(\underline{w}_c^{\mathrm{s}}, \underline{w}_c^{\mathrm{s}}) \le \alpha \|\underline{w}_c^{\mathrm{s}}\|_{1,c}^2,$$

where $\alpha > 0$ is independent of h and of the cell geometry. Combining the previous estimates yields the local norm equivalence

$$\alpha^{-1} \min(1, H_{\min}) \|\underline{w}_c^{\mathrm{s}}\|_{1,c}^2 \le \mathcal{A}_c^{\mathrm{s}}(\underline{w}_c^{\mathrm{s}}, \underline{w}_c^{\mathrm{s}}) \le \alpha \max(1, H_{\max}) \|\underline{w}_c^{\mathrm{s}}\|_{1,c}^2. \tag{D.1}$$

Summing (D.1) over all cells of the mesh gives, for all $\underline{w}_h^{\mathrm{s}} \in \underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}}$,

$$\alpha^{-1} \min(1, H_{\min}) \|\underline{w}_h^{\mathrm{s}}\|_{1,h}^2 \le \mathcal{A}_h^{\mathrm{s}}(\underline{w}_h^{\mathrm{s}}, \underline{w}_h^{\mathrm{s}}) \le \alpha \max(1, H_{\max}) \|\underline{w}_h^{\mathrm{s}}\|_{1,h}^2.$$

The lower bound gives coercivity. The upper bound gives the diagonal boundedness. Since $\mathcal{A}_h^{\mathrm{s}}$ is symmetric and non-negative, or equivalently by applying the Cauchy–Schwarz inequality to the gradient and stabilization terms, this boundedness implies the full continuity estimate

$$|\mathcal{A}_h^{\mathrm{s}}(\underline{u}_h^{\mathrm{s}}, \underline{w}_h^{\mathrm{s}})| \lesssim \|\underline{u}_h^{\mathrm{s}}\|_{1,h} \|\underline{w}_h^{\mathrm{s}}\|_{1,h}.$$

Thus, the global bilinear form $\mathcal{A}_h^{\mathrm{s}}$ is coercive and continuous on $\underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}}$ with respect to the discrete energy norm $\|\cdot\|_{1,h}$. Since the face components on the Dirichlet boundary vanish in $\underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}}$, this semi-norm is a genuine norm on this space.

(*ii*) *Uniqueness of the homogeneous problem.* In the homogeneous case $\psi^{\mathrm{s}} = 0$, the discrete space is $\underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}}$ and the problem (6.39) reduces to

$$\text{Find } \underline{\phi}_{h,0}^{\mathrm{s}} \in \underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}} \quad \text{such that} \quad \mathcal{A}_h^{\mathrm{s}}\big(\underline{\phi}_{h,0}^{\mathrm{s}}, \underline{w}_h^{\mathrm{s}}\big) = 0, \quad \forall \underline{w}_h^{\mathrm{s}} \in \underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}}.$$

Choosing the test function $\underline{w}_h^{\mathrm{s}} = \underline{\phi}_{h,0}^{\mathrm{s}}$ in the above variational formulation, we obtain

$$\mathcal{A}_h^{\mathrm{s}}\big(\underline{\phi}_{h,0}^{\mathrm{s}}, \underline{\phi}_{h,0}^{\mathrm{s}}\big) = 0.$$

From Step 1, the global bilinear form $\mathcal{A}_h^{\mathrm{s}}$ is coercive on $\underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}}$ with respect to the norm $\|\cdot\|_{1,h}$:

$$\mathcal{A}_h^{\mathrm{s}}\big(\underline{w}_h^{\mathrm{s}}, \underline{w}_h^{\mathrm{s}}\big) \gtrsim \|\underline{w}_h^{\mathrm{s}}\|_{1,h}^2, \qquad \forall \underline{w}_h^{\mathrm{s}} \in \underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}}.$$

Applying this inequality to $\underline{w}_h^{\mathrm{s}} = \underline{\phi}_{h,0}^{\mathrm{s}}$ gives

$$0 = \mathcal{A}_h^{\mathrm{s}}\big(\underline{\phi}_{h,0}^{\mathrm{s}}, \underline{\phi}_{h,0}^{\mathrm{s}}\big) \gtrsim \|\underline{\phi}_{h,0}^{\mathrm{s}}\|_{1,h}^2,$$

so that $\|\underline{\phi}_{h,0}^{\mathrm{s}}\|_{1,h} = 0$ and therefore $\underline{\phi}_{h,0}^{\mathrm{s}} = \underline{0}$ in $\underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}}$.

(*iii*) *Non-homogeneous Dirichlet data.* Let $\phi^{\mathrm{s,ext}} \in \mathsf{H}^1(\Omega_{\mathrm{s}})$ be an H^1-extension[1] of ψ^{s} inside Ω_{s}, that is,

$$\phi^{\mathrm{s,ext}}|_{\partial\Omega_{\mathrm{fs}}} = \psi^{\mathrm{s}} \quad \text{and} \quad \|\phi^{\mathrm{s,ext}}\|_{H^1(\Omega_{\mathrm{s}})} \lesssim \|\psi^{\mathrm{s}}\|_{H^{1/2}(\partial\Omega_{\mathrm{fs}})}.$$

We define the discrete lifting $\underline{\phi}_{h,\psi}^{\mathrm{s}} \in \underline{\mathsf{U}}_h^{k,\mathrm{s}}$ by

$$\underline{\phi}_{h,\psi}^{\mathrm{s}} := \mathsf{I}_h^{k,\mathrm{s}} \phi^{\mathrm{s,ext}},$$

where $\mathsf{I}_h^{k,\mathrm{s}}$ denotes the global HHO interpolation operator acting cellwise and facewise by local L^2-projections. By construction of $\mathsf{I}_h^{k,\mathrm{s}}$, the face component on a boundary face $\Gamma \in \mathcal{F}_h^{\mathrm{s},\partial}$ is given by

$$\phi_{\Gamma,\psi}^{\mathrm{s}} = \pi_{\Gamma}^{k,\mathrm{s}}\big(\phi^{\mathrm{s,ext}}|_{\Gamma}\big) = \pi_{\Gamma}^{k,\mathrm{s}} \psi^{\mathrm{s}}, \qquad \forall \Gamma \in \mathcal{F}_h^{\mathrm{s},\partial},$$

since $\phi^{\mathrm{s,ext}}$ coincides with ψ^{s} on $\partial\Omega_{\mathrm{fs}}$. Therefore, $\underline{\phi}_{h,\psi}^{\mathrm{s}}$ enforces the discrete boundary condition (6.41). Moreover, the stability of the interpolation operator $\mathsf{I}_h^{k,\mathrm{s}}$ in the discrete norm $\|\cdot\|_{1,h}$ yields

$$\|\underline{\phi}_{h,\psi}^{\mathrm{s}}\|_{1,h} = \|\mathsf{I}_h^{k,\mathrm{s}} \phi^{\mathrm{s,ext}}\|_{1,h} \lesssim \|\phi^{\mathrm{s,ext}}\|_{H^1(\Omega_{\mathrm{s}})} \lesssim \|\psi^{\mathrm{s}}\|_{H^{1/2}(\partial\Omega_{\mathrm{fs}})}.$$

We then seek the discrete solution as the sum of a homogeneous part and the lifting,

$$\underline{\phi}_h^{\mathrm{s}} = \underline{\phi}_{h,0}^{\mathrm{s}} + \underline{\phi}_{h,\psi}^{\mathrm{s}}, \qquad \underline{\phi}_{h,0}^{\mathrm{s}} \in \underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}}.$$

Substituting this decomposition into the discrete problem (6.39) and using that $\underline{\phi}_{h,0}^{\mathrm{s}}$ has vanishing face components on $\partial\Omega_{\mathrm{fs}}$, we obtain the equivalent problem

$$\text{Find } \underline{\phi}_{h,0}^{\mathrm{s}} \in \underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}} \quad \text{such that} \quad \mathcal{A}_h^{\mathrm{s}}\big(\underline{\phi}_{h,0}^{\mathrm{s}}, \underline{w}_h^{\mathrm{s}}\big) = -\mathcal{A}_h^{\mathrm{s}}\big(\underline{\phi}_{h,\psi}^{\mathrm{s}}, \underline{w}_h^{\mathrm{s}}\big), \quad \forall \underline{w}_h^{\mathrm{s}} \in \underline{\mathsf{U}}_{h,0}^{k,\mathrm{s}}.$$

[1] The existence of such an extension follows from the continuity and surjectivity of the trace operator $\mathsf{H}^1(\Omega_{\mathrm{s}}) \to \mathsf{H}^{1/2}(\partial\Omega_{\mathrm{fs}})$ and a continuous right-inverse.

We then view the right-hand side as the following functional

$$\ell[\underline{\phi}^{\mathrm{s}}_{h,\psi}](\underline{w}^{\mathrm{s}}_h) := -\mathscr{A}^{\mathrm{s}}_h(\underline{\phi}^{\mathrm{s}}_{h,\psi}, \underline{w}^{\mathrm{s}}_h), \qquad \underline{w}^{\mathrm{s}}_h \in \underline{\mathsf{U}}^{k,\mathrm{s}}_{h,0}.$$

The continuity of $\mathscr{A}^{\mathrm{s}}_h$ with respect to $\|\cdot\|_{1,h}$ implies

$$\big|\ell[\underline{\phi}^{\mathrm{s}}_{h,\psi}](\underline{w}^{\mathrm{s}}_h)\big| = \big|\mathscr{A}^{\mathrm{s}}_h(\underline{\phi}^{\mathrm{s}}_{h,\psi}, \underline{w}^{\mathrm{s}}_h)\big| \lesssim \|\underline{\phi}^{\mathrm{s}}_{h,\psi}\|_{1,h}\, \|\underline{w}^{\mathrm{s}}_h\|_{1,h}, \qquad \forall \underline{w}^{\mathrm{s}}_h \in \underline{\mathsf{U}}^{k,\mathrm{s}}_{h,0},$$

so that $\ell[\underline{\phi}^{\mathrm{s}}_{h,\psi}]$ is a continuous linear functional on $\underline{\mathsf{U}}^{k,\mathrm{s}}_{h,0}$.
On the other hand, Step 1 has shown that $\mathscr{A}^{\mathrm{s}}_h$ is coercive and continuous on $\underline{\mathsf{U}}^{k,\mathrm{s}}_{h,0}$ with respect to $\|\cdot\|_{1,h}$. Therefore, the Lax–Milgram lemma applies to the variational problem on $\underline{\mathsf{U}}^{k,\mathrm{s}}_{h,0}$ and guarantees the existence and uniqueness of $\underline{\phi}^{\mathrm{s}}_{h,0}$. The discrete solution

$$\underline{\phi}^{\mathrm{s}}_h = \underline{\phi}^{\mathrm{s}}_{h,0} + \underline{\phi}^{\mathrm{s}}_{h,\psi}$$

is thus uniquely defined and satisfies the boundary condition (6.41).

(*iv*) *Stability estimate.* We now derive a bound on the discrete solution in terms of the boundary data. Starting from the variational formulation obtained in Step 3,

$$\mathscr{A}^{\mathrm{s}}_h(\underline{\phi}^{\mathrm{s}}_{h,0}, \underline{w}^{\mathrm{s}}_h) = -\mathscr{A}^{\mathrm{s}}_h(\underline{\phi}^{\mathrm{s}}_{h,\psi}, \underline{w}^{\mathrm{s}}_h), \qquad \forall\, \underline{w}^{\mathrm{s}}_h \in \underline{\mathsf{U}}^{k,\mathrm{s}}_{h,0},$$

we choose the test function $\underline{w}^{\mathrm{s}}_h = \underline{\phi}^{\mathrm{s}}_{h,0}$. We then obtain the identity

$$\mathscr{A}^{\mathrm{s}}_h(\underline{\phi}^{\mathrm{s}}_{h,0}, \underline{\phi}^{\mathrm{s}}_{h,0}) = -\mathscr{A}^{\mathrm{s}}_h(\underline{\phi}^{\mathrm{s}}_{h,\psi}, \underline{\phi}^{\mathrm{s}}_{h,0}).$$

The left-hand side can be estimated from below using the coercivity property established in Step 1, namely

$$\mathscr{A}^{\mathrm{s}}_h(\underline{\phi}^{\mathrm{s}}_{h,0}, \underline{\phi}^{\mathrm{s}}_{h,0}) \gtrsim \|\underline{\phi}^{\mathrm{s}}_{h,0}\|^2_{1,h}.$$

On the other hand, the right-hand side can be bounded from above by applying the Cauchy–Schwarz inequality and the continuity of $\mathscr{A}^{\mathrm{s}}_h$ with respect to $\|\cdot\|_{1,h}$:

$$\big|\mathscr{A}^{\mathrm{s}}_h(\underline{\phi}^{\mathrm{s}}_{h,\psi}, \underline{\phi}^{\mathrm{s}}_{h,0})\big| \lesssim \|\underline{\phi}^{\mathrm{s}}_{h,\psi}\|_{1,h}\, \|\underline{\phi}^{\mathrm{s}}_{h,0}\|_{1,h}.$$

Combining these two inequalities gives

$$\|\underline{\phi}^{\mathrm{s}}_{h,0}\|^2_{1,h} \lesssim \|\underline{\phi}^{\mathrm{s}}_{h,\psi}\|_{1,h}\, \|\underline{\phi}^{\mathrm{s}}_{h,0}\|_{1,h},$$

then assuming $\underline{\phi}^{\mathrm{s}}_{h,0} \neq \underline{0}$ and dividing both sides by $\|\underline{\phi}^{\mathrm{s}}_{h,0}\|_{1,h}$ lead to

$$\|\underline{\phi}^{\mathrm{s}}_{h,0}\|_{1,h} \lesssim \|\underline{\phi}^{\mathrm{s}}_{h,\psi}\|_{1,h}.$$

Finally, using the interpolation stability estimate established in Step 3,

$$\|\underline{\phi}^{\mathrm{s}}_{h,\psi}\|_{1,h} \lesssim \|\psi^{\mathrm{s}}\|_{H^{1/2}(\partial\Omega_{\mathrm{fs}})},$$

and applying the triangle inequality to $\underline{\phi}^{\mathrm{s}}_h = \underline{\phi}^{\mathrm{s}}_{h,0} + \underline{\phi}^{\mathrm{s}}_{h,\psi}$, we get

$$\|\underline{\phi}^{\mathrm{s}}_h\|_{1,h} \le \|\underline{\phi}^{\mathrm{s}}_{h,0}\|_{1,h} + \|\underline{\phi}^{\mathrm{s}}_{h,\psi}\|_{1,h} \lesssim \|\psi^{\mathrm{s}}\|_{H^{1/2}(\partial\Omega_{\mathrm{fs}})}.$$

This shows that the discrete solution satisfies the desired stability bound

$$\|\underline{\phi}^{\mathrm{s}}_h\|_{1,h} \lesssim \|\psi^{\mathrm{s}}\|_{H^{1/2}(\partial\Omega_{\mathrm{fs}})},$$

and the proof is complete. □

APPENDIX E

HHO algebraic formulation and static condensation

As mentioned in Chapter 6, one advantage of the HHO method is that static condensation can be used to reduce the linear system to a problem involving only the degrees of freedom located on the skeleton of the mesh. We now detail this procedure and provide some implementation remarks on the method. Readers interested in this method are referred to [86] for a comprehensive overview of this class of schemes.

Let $\omega_c^{\mathrm{s}} \in \mathcal{T}_h^{\mathrm{s}}$ be a fixed cell. We recall that a local HHO unknown is a tuple

$$\underline{u}_c^{\mathrm{s}} = \big(u_c^{\mathrm{s}}, (u_\Gamma^{\mathrm{s}})_{\Gamma\in\mathcal{F}_c^{\mathrm{s}}}\big) \in \underline{\mathrm{U}}_c^{k,\mathrm{s}},$$

where $u_c^{\mathrm{s}} \in \mathbb{P}^k(\omega_c^{\mathrm{s}})$ is the cell polynomial and $u_\Gamma^{\mathrm{s}} \in \mathbb{P}^k(\Gamma)$ are the face polynomials. We still denote by $\{\varphi_i^{c,\mathrm{s}}\}_{i\in[\![1,N_k^{\mathrm{s}}]\!]}$ and $\{\varphi_j^{\Gamma,\mathrm{s}}\}_{j\in[\![1,M_k^{\mathrm{s}}]\!]}$ the chosen polynomial bases of $\mathbb{P}^k(\omega_c^{\mathrm{s}})$ and $\mathbb{P}^k(\Gamma)$, respectively. The corresponding local coefficient vectors are denoted by $\mathbf{U}_c^{\mathrm{s}} \in \mathbb{R}^{N_k^{\mathrm{s}}}$, $\mathbf{U}_\Gamma^{\mathrm{s}} \in \mathbb{R}^{M_k^{\mathrm{s}}}$, and $\mathbf{U}_{\mathcal{F}_c}^{\mathrm{s}} := (\mathbf{U}_\Gamma^{\mathrm{s}})_{\Gamma\in\mathcal{F}_c^{\mathrm{s}}} \in \mathbb{R}^{M_k^{\mathrm{s}}\#\mathcal{F}_c^{\mathrm{s}}}$. Thus the total number of local degrees of freedom is

$$N_c^{\mathrm{s}} = N_k^{\mathrm{s}} + M_k^{\mathrm{s}}\#\mathcal{F}_c^{\mathrm{s}}.$$

For later use, we introduce the canonical local HHO basis functions. The cell basis functions are

$$\underline{\varphi}_{c,i} := \big(\varphi_i^{c,\mathrm{s}}, (0)_{\Gamma\in\mathcal{F}_c^{\mathrm{s}}}\big), \qquad i \in [\![1, N_k^{\mathrm{s}}]\!],$$

and the face basis functions are

$$\underline{\varphi}_{\Gamma,j} := \big(0, (0, \dots, \varphi_j^{\Gamma,\mathrm{s}}, \dots, 0)\big), \qquad j \in [\![1, M_k^{\mathrm{s}}]\!], \qquad \Gamma \in \mathcal{F}_c^{\mathrm{s}}.$$

In the second definition, the only non-zero face component is the one associated with the face Γ.

The local bilinear form $\mathcal{A}_c^{\mathrm{s}}$ defines a local symmetric matrix $\mathbb{K}^c \in \mathcal{M}_{N_c^{\mathrm{s}}\times N_c^{\mathrm{s}}}(\mathbb{R})$. After ordering first the cell DOFs and then the face DOFs, the local algebraic system reads

$$\begin{bmatrix} \mathbb{K}_{cc}^c & \mathbb{K}_{c\mathcal{F}_c}^c \\ \mathbb{K}_{\mathcal{F}_c c}^c & \mathbb{K}_{\mathcal{F}_c\mathcal{F}_c}^c \end{bmatrix} \begin{bmatrix} \mathbf{U}_c^{\mathrm{s}} \\ \mathbf{U}_{\mathcal{F}_c}^{\mathrm{s}} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_c^c \\ \mathbf{b}_{\mathcal{F}_c}^c \end{bmatrix}. \tag{E.1}$$

The entries are given by the local bilinear form evaluated on the local basis functions, for faces $\Gamma, \Gamma' \in \mathcal{F}_c^{\mathrm{s}}$:

$$\begin{aligned} (\mathbb{K}_{cc}^c)_{(i,j)\in[\![1,N_k^{\mathrm{s}}]\!]^2} &= \mathcal{A}_c^{\mathrm{s}}(\underline{\varphi}_{c,j}, \underline{\varphi}_{c,i}), \\ (\mathbb{K}_{c\Gamma}^c)_{(i,j)\in[\![1,N_k^{\mathrm{s}}]\!]\times[\![1,M_k^{\mathrm{s}}]\!]} &= \mathcal{A}_c^{\mathrm{s}}(\underline{\varphi}_{\Gamma,j}, \underline{\varphi}_{c,i}), \\ (\mathbb{K}_{\Gamma\Gamma'}^c)_{(i,j)\in[\![1,M_k^{\mathrm{s}}]\!]^2} &= \mathcal{A}_c^{\mathrm{s}}(\underline{\varphi}_{\Gamma',j}, \underline{\varphi}_{\Gamma,i}). \end{aligned}$$

The block $\mathbb{K}_{c\mathcal{F}_c}^c$ is obtained by concatenating the matrices $\mathbb{K}_{c\Gamma}^c$, i.e., if $\mathcal{F}_c^{\mathrm{s}} = \{\Gamma_i\}_{i\in[\![1,\#\mathcal{F}_c^{\mathrm{s}}]\!]}$, then

$$\mathbb{K}_{c\mathcal{F}_c}^c = \begin{bmatrix} \mathbb{K}_{c\Gamma_1}^c & \mathbb{K}_{c\Gamma_2}^c & \cdots & \mathbb{K}_{c\Gamma_{\#\mathcal{F}_c^{\mathrm{s}}}}^c \end{bmatrix} \in \mathcal{M}_{N_k^{\mathrm{s}}\times M_k^{\mathrm{s}}\#\mathcal{F}_c^{\mathrm{s}}}(\mathbb{R}),$$

where each block $\mathbb{K}_{c\Gamma}^c \in \mathcal{M}_{N_k^{\mathrm{s}}\times M_k^{\mathrm{s}}}(\mathbb{R})$ collects the couplings between the N_k^{s} cell DOFs and the M_k^{s} DOFs attached to the face Γ. Similarly, the blocks $\mathbb{K}_{\mathcal{F}_c c}^c$ and $\mathbb{K}_{\mathcal{F}_c\mathcal{F}_c}^c$ are assembled as

$$\mathbb{K}_{\mathcal{F}_c c}^c = \begin{bmatrix} (\mathbb{K}_{c\Gamma_1}^c)^\top \\ (\mathbb{K}_{c\Gamma_2}^c)^\top \\ \vdots \\ (\mathbb{K}_{c\Gamma_{\#\mathcal{F}_c^{\mathrm{s}}}}^c)^\top \end{bmatrix} \in \mathcal{M}_{M_k^{\mathrm{s}}\#\mathcal{F}_c^{\mathrm{s}}\times N_k^{\mathrm{s}}}(\mathbb{R}),$$

and

$$\mathbb{K}^c_{\mathcal{F}_c\mathcal{F}_c} = \begin{bmatrix} \mathbb{K}^c_{\Gamma_1\Gamma_1} & \mathbb{K}^c_{\Gamma_1\Gamma_2} & \cdots & \mathbb{K}^c_{\Gamma_1\Gamma_{\#\mathcal{F}^{\mathrm{s}}_c}} \\ \mathbb{K}^c_{\Gamma_2\Gamma_1} & \mathbb{K}^c_{\Gamma_2\Gamma_2} & \cdots & \mathbb{K}^c_{\Gamma_2\Gamma_{\#\mathcal{F}^{\mathrm{s}}_c}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{K}^c_{\Gamma_{\#\mathcal{F}^{\mathrm{s}}_c}\Gamma_1} & \mathbb{K}^c_{\Gamma_{\#\mathcal{F}^{\mathrm{s}}_c}\Gamma_2} & \cdots & \mathbb{K}^c_{\Gamma_{\#\mathcal{F}^{\mathrm{s}}_c}\Gamma_{\#\mathcal{F}^{\mathrm{s}}_c}} \end{bmatrix} \in \mathscr{M}_{M^{\mathrm{s}}_k \#\mathcal{F}^{\mathrm{s}}_c}(\mathbb{R}),$$

where each sub-block $\mathbb{K}^c_{\Gamma\Gamma'} \in \mathscr{M}_{M^{\mathrm{s}}_k}(\mathbb{R})$ encodes the interaction between the DOFs attached to faces Γ and Γ', and the symmetry of $\mathscr{A}^{\mathrm{s}}_c(\cdot,\cdot)$ implies $\mathbb{K}^c_{\Gamma\Gamma'} = (\mathbb{K}^c_{\Gamma'\Gamma})^\top$ and $\mathbb{K}^c_{\mathcal{F}_c c} = (\mathbb{K}^c_{c\mathcal{F}_c})^\top$. Let now $\mathbf{U}^{\mathrm{s}}_{\mathcal{T}_h}$ and $\mathbf{U}^{\mathrm{s}}_{\mathcal{F}_h}$ denote the global vectors collecting all cell DOFs and all face DOFs, respectively. After assembling all local contributions over the cells $\omega^{\mathrm{s}}_c \in \mathcal{T}^{\mathrm{s}}_h$, the global system has the block form

$$\begin{bmatrix} \mathbb{K}_{\mathcal{T}^{\mathrm{s}}_h\mathcal{T}^{\mathrm{s}}_h} & \mathbb{K}_{\mathcal{T}^{\mathrm{s}}_h\mathcal{F}^{\mathrm{s}}_h} \\ \mathbb{K}_{\mathcal{F}^{\mathrm{s}}_h\mathcal{T}^{\mathrm{s}}_h} & \mathbb{K}_{\mathcal{F}^{\mathrm{s}}_h\mathcal{F}^{\mathrm{s}}_h} \end{bmatrix} \begin{bmatrix} \mathbf{U}^{\mathrm{s}}_{\mathcal{T}_h} \\ \mathbf{U}^{\mathrm{s}}_{\mathcal{F}_h} \end{bmatrix} = \begin{bmatrix} \mathbf{b}^{\mathrm{s}}_{\mathcal{T}_h} \\ \mathbf{b}^{\mathrm{s}}_{\mathcal{F}_h} \end{bmatrix}.$$

The essential HHO property is that the block $\mathbb{K}_{\mathcal{T}^{\mathrm{s}}_h\mathcal{T}^{\mathrm{s}}_h}$ is block diagonal, with one independent block $\mathbb{K}^c_{cc}$ per cell. This follows from the fact that cell unknowns do not interact directly with cell unknowns from neighbouring cells; the coupling between cells is only mediated by face unknowns. Moreover, each block $\mathbb{K}^c_{cc}$ is symmetric positive definite.

From the first row of the local system (E.1), the cell DOFs satisfy

$$\mathbb{K}^c_{cc}\mathbf{U}^{\mathrm{s}}_c + \mathbb{K}^c_{c\mathcal{F}_c}\mathbf{U}^{\mathrm{s}}_{\mathcal{F}_c} = \mathbf{b}^c_c.$$

Since $\mathbb{K}^c_{cc}$ is invertible, we obtain the local elimination formula

$$\mathbf{U}^{\mathrm{s}}_c = (\mathbb{K}^c_{cc})^{-1}\left(\mathbf{b}^c_c - \mathbb{K}^c_{c\mathcal{F}_c}\mathbf{U}^{\mathrm{s}}_{\mathcal{F}_c}\right). \tag{E.2}$$

Substituting (E.2) into the second row gives the local condensed system

$$\widehat{\mathbb{K}}^c_{\mathcal{F}_c\mathcal{F}_c}\mathbf{U}^{\mathrm{s}}_{\mathcal{F}_c} = \widehat{\mathbf{b}}^c_{\mathcal{F}_c},$$

where the matrix $\widehat{\mathbb{K}}^c_{\mathcal{F}_c\mathcal{F}_c}$ is the local Schur complement associated with the elimination of the cell unknowns:

$$\widehat{\mathbb{K}}^c_{\mathcal{F}_c\mathcal{F}_c} := \mathbb{K}^c_{\mathcal{F}_c\mathcal{F}_c} - \mathbb{K}^c_{\mathcal{F}_c c}(\mathbb{K}^c_{cc})^{-1}\mathbb{K}^c_{c\mathcal{F}_c}$$

and

$$\widehat{\mathbf{b}}^c_{\mathcal{F}_c} := \mathbf{b}^c_{\mathcal{F}_c} - \mathbb{K}^c_{\mathcal{F}_c c}(\mathbb{K}^c_{cc})^{-1}\mathbf{b}^c_c.$$

After this operation has been performed on every cell, only the face unknowns remain globally coupled. The condensed global problem is therefore

$$\widehat{\mathbb{K}}_{\mathcal{F}^{\mathrm{s}}_h\mathcal{F}^{\mathrm{s}}_h}\mathbf{U}^{\mathrm{s}}_{\mathcal{F}_h} = \widehat{\mathbf{B}}^{\mathrm{s}}_{\mathcal{F}_h}. \tag{E.3}$$

The global condensed matrix is obtained by assembling the local Schur complements $\widehat{\mathbb{K}}^c_{\mathcal{F}_c\mathcal{F}_c}$ over all cells $\omega^{\mathrm{s}}_c \in \mathcal{T}^{\mathrm{s}}_h$. Equivalently, it is the Schur complement of the global cell-cell block $\mathbb{K}_{\mathcal{T}^{\mathrm{s}}_h\mathcal{T}^{\mathrm{s}}_h}$ in the full HHO matrix. The right-hand side $\widehat{\mathbf{B}}^{\mathrm{s}}_{\mathcal{F}_h}$ is assembled from the local vectors $\widehat{\mathbf{b}}^c_{\mathcal{F}_c}$. In the present interior elliptic problem, there is no volume source term. Therefore, for cells that are not affected by the non-homogeneous boundary data, the corresponding local source contribution vanishes. Non-zero contributions only arise from the weak enforcement or lifting of the Dirichlet data on $\partial\Omega_{\mathrm{fs}}$.

In practice, the face unknowns located on the Dirichlet boundary $\partial\Omega_{\mathrm{fs}}$ are prescribed by the trace variable ψ^{s}. They are therefore not solved for in the condensed system. If we split the face DOFs into free skeletal DOFs and Dirichlet skeletal DOFs, then

$$\mathbf{U}^{\mathrm{s}}_{\mathcal{F}_h} = \begin{bmatrix} \mathbf{U}^{\mathrm{s}}_{\mathcal{F}^{\mathrm{s}}_{h,0}} \\ \mathbf{U}^{\mathrm{s}}_{\mathcal{F}^{\mathrm{s}}_{h,D}} \end{bmatrix},$$

where $\mathbf{U}^{\mathrm{s}}_{\mathcal{F}^{\mathrm{s}}_{h,D}}$ is known. The known Dirichlet contribution is moved to the right-hand side. Thus the practical condensed system is solved only for the free skeletal unknowns:

$$\widehat{\mathbb{K}}_{00}\mathbf{U}^{\mathrm{s}}_{\mathcal{F}^{\mathrm{s}}_{h,0}} = \widehat{\mathbf{B}}_0 - \widehat{\mathbb{K}}_{0D}\mathbf{U}^{\mathrm{s}}_{\mathcal{F}^{\mathrm{s}}_{h,D}}.$$

Once the free face unknowns have been computed, the complete face vector $\mathbf{U}^{\mathrm{s}}_{\mathcal{F}_h}$ is known.

The cell unknowns are then recovered cell by cell using (E.2). Finally, the reconstructed potential $\mathsf{p}_c^{k+1,\mathrm{s}}\underline{u}_c^{\mathrm{s}}$ and its gradient $\mathbf{G}_c^{k,\mathrm{s}}\underline{u}_c^{\mathrm{s}}$ can be evaluated locally from the complete vector $(\mathbf{U}_c^{\mathrm{s}}, \mathbf{U}_{\mathcal{F}_c}^{\mathrm{s}})$.

The algorithm then follows the cell-by-cell structure below:

1. Construct the local potential reconstruction operator $\mathsf{p}_c^{k+1,\mathrm{s}}$, then assemble the local consistency and stabilization matrices, hence the full matrix $\mathbb{K}^c$,
2. Split $\mathbb{K}^c$ into the blocks $\mathbb{K}^c_{cc}$, $\mathbb{K}^c_{c\mathcal{F}_c}$, $\mathbb{K}^c_{\mathcal{F}_c c}$, and $\mathbb{K}^c_{\mathcal{F}_c\mathcal{F}_c}$,
3. Invert the local cell block $\mathbb{K}^c_{cc}$, compute the local Schur complement $\widehat{\mathbb{K}}^c_{\mathcal{F}_c\mathcal{F}_c}$ and the condensed right-hand side $\widehat{\mathbf{b}}^c_{\mathcal{F}_c}$,
4. Assemble all condensed local contributions into the global skeletal system and impose the prescribed Dirichlet face DOFs on $\partial\Omega_{\mathrm{fs}}$ and solve the condensed system for the remaining face unknowns,
5. Reconstruct the eliminated cell unknowns locally by (E.2).

The static condensation procedure is exact: it does not modify the discrete HHO solution and does not introduce any additional approximation. It is only an algebraic elimination of the element-based unknowns, made possible by the local structure of the HHO method. Its advantage is that the globally coupled system is posed only on the mesh skeleton, with a number of unknowns proportional to the number of faces rather than to the total number of cell and face unknowns.

APPENDIX F

Entropy considerations

This appendix relies on the notations and framework introduced in Chapter 1, and a full exhaustive study of entropy stability can be found in [265].

In the analysis of nonlinear conservation laws and related PDE systems, entropy plays a central role. Classical weak solutions are often not unique, and additional admissibility criteria are required to select the physically relevant solution. Entropy provides one of the most robust and widely used criteria. For a system of conservation laws such as the 1D homogeneous shallow-water equations

$$\partial_t \boldsymbol{v} + \partial_x \boldsymbol{F}(\boldsymbol{v}) = 0,$$

an entropy–entropy flux pair $(\mathsf{S}, \Phi_{\mathsf{S}})$ is defined as a scalar function $\mathsf{S}(\boldsymbol{v}) = (\mathsf{S}^\eta(\boldsymbol{v}), \mathsf{S}^q(\boldsymbol{v}))$ together with its associated flux $\Phi_{\mathsf{S}}(\boldsymbol{v})$ satisfying the compatibility condition

$$\nabla_{\boldsymbol{v}} \mathsf{S}(\boldsymbol{v}) \cdot \nabla_{\boldsymbol{v}} \boldsymbol{F}(\boldsymbol{v}) = \nabla_{\boldsymbol{v}} \Phi_{\mathsf{S}}(\boldsymbol{v}).$$

This condition ensures that smooth solutions satisfy the additional conservation law

$$\partial_t \mathsf{S}(\boldsymbol{v}) + \partial_x \Phi_{\mathsf{S}}(\boldsymbol{v}) = 0.$$

At the discrete level, and for weak (possibly discontinuous) solutions, this conservation law is replaced by the inequality

$$\partial_t \mathsf{S}(\boldsymbol{v}) + \partial_x \Phi_{\mathsf{S}}(\boldsymbol{v}) \leq 0,$$

which expresses dissipation of entropy across shocks. Numerical schemes respecting this property are called *entropy-stable*, and their construction is essential for robustness, particularly for systems where nonlinearity and non-convexity may lead to spurious oscillations or even instabilities.

A key object associated with $(\mathsf{S}, \Phi_{\mathsf{S}})$ is the *entropy variable* $\mathbb{U} = (\boldsymbol{U}^\eta, \boldsymbol{U}^q)$ such that

$$\mathbb{U}(\boldsymbol{v}) := \nabla_{\boldsymbol{v}} \mathsf{S}(\boldsymbol{v}),$$

which induces a symmetrization of the system. Another important quantity is the *entropy potential flux*

$$\Psi(\boldsymbol{v}) := \mathbb{U}(\boldsymbol{v}) \cdot \boldsymbol{F}(\boldsymbol{v}, b) - \Phi_{\mathsf{S}}(\boldsymbol{v}).$$

From a numerical point of view, enforcing entropy stability usually amounts to controlling the amount of artificial dissipation in the numerical flux. Many high-order methods, including DG and FV schemes on subcells, rely on a decomposition of the numerical flux into a central (entropy-conservative) part and a dissipative part. The dissipative term must be strong enough to guarantee the correct entropy inequality, but not excessive, in order to preserve the accuracy of the scheme. This balance motivates the introduction of *blending coefficients*, which interpolate between high-order fluxes and more dissipative first-order fluxes.

In the present appendix, we study entropy inequalities at both subcell and cell levels for the DG/FV blended numerical fluxes. We consider three types of inequalities:

1. A fully discrete subcell entropy stability inequality, valid for any entropy pair $(\mathsf{S}, \Phi_{\mathsf{S}})$.
2. A semi-discrete subcell entropy conservation/dissipation inequality, valid for a fixed entropy pair $(\mathsf{S}, \Phi_{\mathsf{S}})$.
3. A semi-discrete cell entropy conservation/dissipation inequality, obtained after projection in the DG polynomial space and involving cellwise reconstructions of the entropy variables.

In each case, blending coefficients are designed so that the dissipative contribution of the numerical flux satisfies the required entropy inequality. This leads to explicit expressions depending on entropy variables, entropy potentials, and variations of the numerical flux across interfaces. These constructions follow the general philosophy of entropy-stable schemes introduced by Tadmor and developed in various DG and FV contexts.

E.1 Subcell discrete inequality

Here we design a blending coefficient $\Theta^{\mathbf{S},(\mathrm{I})}_{m+\frac{1}{2}}$ such that the fully discrete scheme satisfies the following subcell entropy stability inequality for any entropy pair $(\mathbf{S}, \Phi_{\mathbf{S}})$:

Proposition. *The blending coefficient assuring subcell entropy stability at the discrete level for all* $(\mathbf{S}, \Phi_{\mathbf{S}})$ *is defined by*

$$\Theta^{\mathbf{S},(\mathrm{I})}_{m+\frac{1}{2}} = \min\left(1,\ \frac{\left(\gamma_{m+\frac{1}{2}} - \gamma^{God}\right)\left(\overline{\eta}^n_{m+1} - \overline{\eta}^n_m\right)}{2\Delta\boldsymbol{\mathcal{F}}^{\eta}_{m+\frac{1}{2}}}\right), \qquad \textit{if } (\overline{\eta}^n_{m+1} - \overline{\eta}^n_m)\Delta\boldsymbol{\mathcal{F}}^{\eta}_{m+\frac{1}{2}} > 0,$$

and by $\Theta^{\mathbf{S},(\mathrm{I})}_{m+\frac{1}{2}} = 1$ *otherwise.*

Let us give some sketch of the proof. We dropped superscript i and also the $\pm$ indices in fluxes for clarity. The goal here is to write

$$\widetilde{\boldsymbol{\mathcal{F}}}_{m+\frac{1}{2}} = \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} + \Theta^{\mathbf{S},(\mathrm{I})}_{m+\frac{1}{2}}\left(\widehat{\boldsymbol{\mathcal{F}}}_{m+\frac{1}{2}} - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}}\right),$$

in the following viscous form

$$\widetilde{\boldsymbol{\mathcal{F}}}_{m+\frac{1}{2}} = \frac{\boldsymbol{\mathcal{F}}(\overline{\boldsymbol{v}}^n_m) + \boldsymbol{\mathcal{F}}(\overline{\boldsymbol{v}}^n_{m+1})}{2} - \frac{\widetilde{\gamma}(\overline{\boldsymbol{v}}^n_m, \overline{\boldsymbol{v}}^n_{m+1})}{2}(\overline{\boldsymbol{v}}^n_{m+1} - \overline{\boldsymbol{v}}^n_m).$$

Using this

$$\boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} + \Theta_{m+\frac{1}{2}}\left(\widehat{\boldsymbol{\mathcal{F}}}_{m+\frac{1}{2}} - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}}\right) = \frac{\boldsymbol{\mathcal{F}}(\overline{\boldsymbol{v}}^n_m) + \boldsymbol{\mathcal{F}}(\overline{\boldsymbol{v}}^n_{m+1})}{2} - \frac{\widetilde{\gamma}(\overline{\boldsymbol{v}}^n_m, \overline{\boldsymbol{v}}^n_{m+1})}{2}(\overline{\boldsymbol{v}}^n_{m+1} - \overline{\boldsymbol{v}}^n_m),$$

gives us

$$\widetilde{\gamma}(\overline{\boldsymbol{v}}^n_m, \overline{\boldsymbol{v}}^n_{m+1}) = \gamma_{m+\frac{1}{2}} - 2\Theta^{\mathbf{S},(\mathrm{I})}_{m+\frac{1}{2}}\frac{\widehat{\boldsymbol{\mathcal{F}}}_{m+\frac{1}{2}} - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}}}{\overline{\boldsymbol{v}}^n_{m+1} - \overline{\boldsymbol{v}}^n_m}.$$

We now want to ensure that $\widetilde{\gamma}(\overline{\boldsymbol{v}}^n_m, \overline{\boldsymbol{v}}^n_{m+1}) \geq \gamma^{\mathrm{God}}$ with γ^{God} computed as:

$$\gamma^{\mathrm{God}} = \frac{\boldsymbol{\mathcal{F}}(\overline{\boldsymbol{v}}^n_m) + \boldsymbol{\mathcal{F}}(\overline{\boldsymbol{v}}^n_{m+1}) - 2\boldsymbol{\mathcal{F}}\left(\boldsymbol{\mathcal{W}}(0; \overline{\boldsymbol{v}}^n_m, \overline{\boldsymbol{v}}^n_{m+1})\right)}{\overline{\boldsymbol{v}}^n_{m+1} - \overline{\boldsymbol{v}}^n_m},$$

with $\boldsymbol{\mathcal{W}}(0; \overline{\boldsymbol{v}}^n_m, \overline{\boldsymbol{v}}^n_{m+1})$ being the unique entropic solution of Riemann problem. Knowing that $\gamma_{m+\frac{1}{2}} \geq \gamma^{\mathrm{God}}$, we then have $\widetilde{\gamma}(\overline{\boldsymbol{v}}^n_m, \overline{\boldsymbol{v}}^n_{m+1}) \geq \gamma^{\mathrm{God}}$ such that

$$\gamma_{m+\frac{1}{2}} - \gamma^{\mathrm{God}} \geq 2\Theta^{\mathbf{S},(\mathrm{I})}_{m+\frac{1}{2}}\frac{\widehat{\boldsymbol{\mathcal{F}}}_{m+\frac{1}{2}} - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}}}{\overline{\boldsymbol{v}}^n_{m+1} - \overline{\boldsymbol{v}}^n_m}.$$

We finally get

$$\Theta^{\mathbf{S},(\mathrm{I})}_{m+\frac{1}{2}} = \begin{cases} \min\left(1,\ \dfrac{\left(\gamma_{m+\frac{1}{2}} - \gamma^{\mathrm{God}}\right)\left(\overline{\eta}^n_{m+1} - \overline{\eta}^n_m\right)}{2\Delta\boldsymbol{\mathcal{F}}^{\eta}_{m+\frac{1}{2}}}\right) & \text{if } (\overline{\eta}^n_{m+1} - \overline{\eta}^n_m)\Delta\boldsymbol{\mathcal{F}}^{\eta}_{m+\frac{1}{2}} > 0, \\ 1 & \text{else.} \end{cases}$$

It is worth noting that the coefficient $\Theta^{\mathbf{S},(\mathrm{I})}_{m+\frac{1}{2}}$ obtained above enforces the desired subcell entropy stability for any entropy pair $(\mathbf{S}, \Phi_{\mathbf{S}})$, but it does so by reverting locally to the amount of dissipation associated with a first-order scheme. As a consequence, this construction ensures robustness but limits the accuracy near interfaces where the entropy constraint becomes active. For this reason, the fully discrete entropy-stable form presented here is mainly of theoretical interest and is not used directly in the high-order version of the method.

E.2 Subcell semi-discrete inequality

We now design a blending coefficient $\Theta^{\mathsf{S},(\mathrm{II})}_{m+\frac{1}{2}}$ such that the semi-discrete scheme satisfies the following subcell entropy conservation/dissipation inequality for a fixed entropy pair $(\mathsf{S}, \Phi_{\mathsf{S}})$:

Proposition. *The blending coefficient assuring subcell semi-discrete entropy conservation/dissipation inequality for a given* $(\mathsf{S}, \Phi_{\mathsf{S}})$ *reads*

$$\Theta^{\mathsf{S},(\mathrm{II})}_{m+\frac{1}{2}} = \frac{\Psi(\overline{\boldsymbol{v}}_{m+1}) - \Psi(\overline{\boldsymbol{v}}_m) - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot (\mathbb{U}(\overline{\boldsymbol{v}}_{m+1}) - \mathbb{U}(\overline{\boldsymbol{v}}_m))}{\Delta\boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot (\mathbb{U}(\overline{\boldsymbol{v}}_{m+1}) - \mathbb{U}(\overline{\boldsymbol{v}}_m))} \quad \textit{if } \Delta\boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot (\mathbb{U}(\overline{\boldsymbol{v}}_{m+1}) - \mathbb{U}(\overline{\boldsymbol{v}}_m)) > 0,$$

and $\Theta^{\mathsf{S},(\mathrm{II})}_{m+\frac{1}{2}} = 1$ *otherwise.*

To this end, considering the semi-discrete scheme

$$\partial_t \overline{\boldsymbol{v}}_m = -\frac{1}{|S_m|}\left(\widetilde{\boldsymbol{\mathcal{F}}}_{m+\frac{1}{2}} - \widetilde{\boldsymbol{\mathcal{F}}}_{m-\frac{1}{2}}\right) + \overline{\boldsymbol{B}}_m, \qquad \forall m \in [\![1, k+1]\!],$$

we want to obtain the following inequality:

$$\left(\mathbb{U}(\overline{\boldsymbol{v}}_{m+1}) - \mathbb{U}(\overline{\boldsymbol{v}}_m)\right) \cdot \widetilde{\boldsymbol{\mathcal{F}}}_{m+\frac{1}{2}} \leq \Psi(\overline{\boldsymbol{v}}_{m+1}) - \Psi(\overline{\boldsymbol{v}}_m),$$

with

$$\widetilde{\boldsymbol{\mathcal{F}}}_{m+\frac{1}{2}} = \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} + \Theta^{\mathsf{S},(\mathrm{II})}_{m+\frac{1}{2}}\left(\widehat{\boldsymbol{\mathcal{F}}}_{m+\frac{1}{2}} - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}}\right).$$

For the sake of simplicity let's denote $\Delta\overline{\mathbb{U}}_m := \mathbb{U}(\overline{\boldsymbol{v}}_{m+1}) - \mathbb{U}(\overline{\boldsymbol{v}}_m)$ and $\Delta\overline{\Psi}_m := \Psi(\overline{\boldsymbol{v}}_{m+1}) - \Psi(\overline{\boldsymbol{v}}_m)$. We can then obtain $\Delta\overline{\mathbb{U}}_m \cdot \widetilde{\boldsymbol{\mathcal{F}}}_{m+\frac{1}{2}} \leq \Delta\overline{\Psi}_m$ if and only if we have the two conditions

$$\begin{aligned}
\Theta^{\mathsf{S},(\mathrm{II})}_{m+\frac{1}{2}} \cdot (\Delta\boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot \Delta\overline{\mathbb{U}}_m) &\leq \Delta\overline{\Psi}_m - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot \Delta\overline{\mathbb{U}}_m, \\
\Delta\overline{\Psi}_m - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot \Delta\overline{\mathbb{U}}_m &\geq 0,
\end{aligned}$$

and these easily get us

$$\Theta^{\mathsf{S},(\mathrm{II})}_{m+\frac{1}{2}} = \begin{cases} \dfrac{\Delta\overline{\Psi}_m - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot \Delta\overline{\mathbb{U}}_m}{\Delta\boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot \Delta\overline{\mathbb{U}}_m} & \text{if } \Delta\overline{\mathbb{U}}_m \cdot \Delta\boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} > 0, \\ 1 & \text{else.} \end{cases}$$

In contrast with the fully discrete coefficient of type (I), the semi-discrete construction above does not automatically reduce the numerical flux to its first-order form. The coefficient $\Theta^{\mathsf{S},(\mathrm{II})}_{m+\frac{1}{2}}$ only modifies the flux when the entropy inequality is at risk of being violated, which means that the high-order part of the scheme is preserved wherever the solution remains sufficiently smooth. As a result, this entropy-based correction is second-order accurate in regular regions while still providing the required entropy dissipation near shocks. This makes the semi-discrete formulation more suitable for practical use than its fully discrete counterpart.

E.3 Cell semi-discrete inequality

Finally, we design a blending coefficient $\Theta^{\mathsf{S},(\mathrm{III})}_{m+\frac{1}{2}}$ such that the semi-discrete scheme satisfies the following cell entropy conservation/dissipation inequality for a fixed entropy pair $(\mathsf{S}, \Phi_{\mathsf{S}})$:

Proposition. *The blending coefficient assuring cell semi-discrete entropy conservation/dissipation inequality for a given* $(\mathsf{S}, \Phi_{\mathsf{S}})$ *is defined by*

$$\Theta^{\mathsf{S},(\mathrm{III})}_{m+\frac{1}{2}} = \frac{\Psi\left(\boldsymbol{v}(\underline{\mathbb{U}}_{m+1})\right) - \Psi\left(\boldsymbol{v}(\underline{\mathbb{U}}_{m+1})\right) - \boldsymbol{\mathcal{F}}^{\mathsf{S}}_{m+\frac{1}{2}} \cdot \Delta\underline{\boldsymbol{v}}(\mathbb{U}_m)}{\Delta\boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot \Delta\underline{\boldsymbol{v}}(\mathbb{U}_m)} \quad \textit{if } \Delta\boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot \Delta\underline{\boldsymbol{v}}(\mathbb{U}_m) > 0,$$

or is set equal to 1 if $\Delta\boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot \Delta\underline{\boldsymbol{v}}(\mathbb{U}_m) \leq 0$*, using the notation* $\Delta\underline{\boldsymbol{v}}(\mathbb{U}_m) := \boldsymbol{v}(\underline{\mathbb{U}}_{m+1}) - \boldsymbol{v}(\underline{\mathbb{U}}_{m+1})$ *together with the following:*

- $\boldsymbol{\mathcal{F}}^{\mathsf{S}}_{m+\frac{1}{2}} := \boldsymbol{\mathcal{F}}\left(\boldsymbol{v}(\underline{\mathbb{U}}_m), \boldsymbol{v}(\underline{\mathbb{U}}_{m+1})\right)$ *is the modified FV numerical flux,*
- $\left\{\phi^i_m\right\}_{m\in[\![1,k+1]\!]} = \left\{\pi^k_i \mathbb{1}^i_m\right\}_{m\in[\![1,k+1]\!]}$ *are the* L^2*-projection of* $\left\{\mathbb{1}^i_m\right\}_{m\in[\![1,k+1]\!]}$ *onto* $\mathbb{P}^k(\omega_i)$*;*
- $\mathbb{U}^i_h = \sum_{m=1}^{k+1} \mathbb{U}(\underline{\boldsymbol{v}}^i_m)\phi^i_m \in \mathbb{P}^k(\omega_i)^2$ *is the* L^2*-projection of* $\mathbb{U}(\boldsymbol{v}^i_h)$ *onto* $\mathbb{P}^k(\omega_i)^2$*,*
- *the cell boundaries fluxes defined as*

$$\begin{cases} \widehat{\boldsymbol{\mathcal{F}}}^i_{\frac{1}{2}} = \boldsymbol{\mathcal{F}}^{\mathrm{DG}}_{i-\frac{1}{2}} := \boldsymbol{\mathcal{F}}\left(\boldsymbol{v}(\underline{\mathbb{U}}_m), \boldsymbol{v}(\underline{\mathbb{U}}_{m+1})\right), \\ \widehat{\boldsymbol{\mathcal{F}}}^i_{k+\frac{3}{2}} = \boldsymbol{\mathcal{F}}^{\mathrm{DG}}_{i+\frac{1}{2}} := \boldsymbol{\mathcal{F}}\left(\boldsymbol{v}(\underline{\mathbb{U}}_m), \boldsymbol{v}(\underline{\mathbb{U}}_{m+1})\right). \end{cases}$$

To provide again an idea of the obtention of this blending coefficient, we write

$$\int_{\omega_i} \partial_t \mathbf{S}(\boldsymbol{v}^i_h)\,\mathrm{d}x = \int_{\omega_i} \nabla_{\boldsymbol{v}}\mathbf{S}(\boldsymbol{v}^i_h)\cdot\partial_t\boldsymbol{v}^i_h\,\mathrm{d}x = \int_{\omega_i} \mathbb{U}(\boldsymbol{v}^i_h)\cdot\partial_t\boldsymbol{v}^i_h\,\mathrm{d}x,$$

we can notice that $\partial_t\boldsymbol{v}^i_h \in \mathbb{P}^k(\omega_i)^2$, and we now denote by $\mathbb{U}^i_h$ the L^2-projection of $p^k_i(\boldsymbol{v}^i_h)$. Then,

$$\int_{\omega_i} \mathbb{U}(\boldsymbol{v}^i_h)\cdot\partial_t\boldsymbol{v}^i_h\,\mathrm{d}x = \int_{\omega_i} \mathbb{U}^i_h\cdot\partial_t\boldsymbol{v}^i_h\,\mathrm{d}x.$$

Reminding that $\mathbb{U}^i_h$ is the L^2-projection of $\mathbb{U}(\boldsymbol{v}^i_h)$ onto $\mathbb{P}^k(\omega_i)^2$, we get:

$$\begin{aligned}\int_{\omega_i} \mathbb{U}^i_h\cdot\partial_t\boldsymbol{v}^i_h\,\mathrm{d}x &= \int_{\omega_i}\left(\sum_{m=1}^{k+1}\mathbb{U}(\underline{\boldsymbol{v}}^i_m)\phi^i_m\right)\cdot\partial_t\boldsymbol{v}^i_h\,\mathrm{d}x = \sum_{m=1}^{k+1}\mathbb{U}(\underline{\boldsymbol{v}}^i_m)\int_{\omega_i} p^k_i(\mathbb{1}^i_m)\partial_t\boldsymbol{v}^i_h\,\mathrm{d}x \\ &= \sum_{m=1}^{k+1}\mathbb{U}(\underline{\boldsymbol{v}}^i_m)\int_{S^i_m}\partial_t\boldsymbol{v}^i_h\,\mathrm{d}x = \sum_{m=1}^{k+1}\mathbb{U}(\underline{\boldsymbol{v}}^i_m)|S^i_m|\partial_t\overline{\boldsymbol{v}}^i_m.\end{aligned}$$

Let us now consider everything on the first equation. We get then:

$$\sum_{m=1}^{k+1}\boldsymbol{U}^\eta(\underline{\eta}^i_m)|S^i_m|\partial_t\overline{\eta}^i_m = -\sum_{m=1}^{k+1}\boldsymbol{U}^\eta(\underline{\eta}^i_m)\left(\widetilde{\boldsymbol{\mathcal{F}}}^{\eta,i}_{m+\frac{1}{2}} - \widetilde{\boldsymbol{\mathcal{F}}}^{\eta,i}_{m-\frac{1}{2}}\right).$$

We are going to add and substract the following two terms to the last sum:

$$\begin{aligned}\sum_{m=1}^{k+1}\boldsymbol{U}^\eta(\underline{\eta}^i_m)|S^i_m|\partial_t\overline{\eta}^i_m = -\sum_{m=1}^{k+1}\boldsymbol{U}^\eta(\underline{\eta}^i_m)\left(\widetilde{\boldsymbol{\mathcal{F}}}^{\eta,i}_{m+\frac{1}{2}} - \widetilde{\boldsymbol{\mathcal{F}}}^{\eta,i}_{m-\frac{1}{2}}\right) &\pm \boldsymbol{U}^\eta\left(\eta^i_h(x_{i-\frac{1}{2}})\right)\Theta^i_{\frac{1}{2}}\boldsymbol{\mathcal{F}}^{\eta,\mathrm{DG}}_{i-\frac{1}{2}} \\ &\pm \boldsymbol{U}^\eta\left(\eta^i_h(x_{i+\frac{1}{2}})\right)\Theta^i_{k+\frac{3}{2}}\boldsymbol{\mathcal{F}}^{\eta,\mathrm{DG}}_{i+\frac{1}{2}},\end{aligned}$$

and we also have

$$\begin{aligned}-\sum_{m=1}^{k+1}\boldsymbol{U}^\eta(\underline{\eta}^i_m)\left(\widetilde{\boldsymbol{\mathcal{F}}}^{\eta,i}_{m+\frac{1}{2}} - \widetilde{\boldsymbol{\mathcal{F}}}^{\eta,i}_{m-\frac{1}{2}}\right) = \sum_{m=1}^{k}\left(\boldsymbol{U}^\eta(\underline{\eta}^i_{m+1}) - \boldsymbol{U}^\eta(\underline{\eta}^i_m)\right)\widetilde{\boldsymbol{\mathcal{F}}}^{\eta,i}_{m+\frac{1}{2}} &+ \boldsymbol{U}^\eta(\underline{\eta}^i_1)\widetilde{\boldsymbol{\mathcal{F}}}^{\eta,i}_{\frac{1}{2}} \\ &- \boldsymbol{U}^\eta(\underline{\eta}^i_{k+1})\widetilde{\boldsymbol{\mathcal{F}}}^{\eta,i}_{k+\frac{3}{2}}\end{aligned}$$

with

$$\begin{aligned}\widetilde{\boldsymbol{\mathcal{F}}}^{\eta,i}_{k+\frac{3}{2}} &= \left(1-\Theta^i_{k+\frac{3}{2}}\right)\boldsymbol{\mathcal{F}}^{\eta,\mathsf{S}}_{k+\frac{3}{2}} + \Theta^i_{k+\frac{3}{2}}\boldsymbol{\mathcal{F}}^{\eta,\mathrm{DG}}_{i+\frac{1}{2}}, \\ \widetilde{\boldsymbol{\mathcal{F}}}^{\eta,i}_{\frac{1}{2}} &= \left(1-\Theta^i_{\frac{1}{2}}\right)\boldsymbol{\mathcal{F}}^{\eta,\mathsf{S}}_{\frac{1}{2}} + \Theta^i_{\frac{1}{2}}\boldsymbol{\mathcal{F}}^{\eta,\mathrm{DG}}_{i-\frac{1}{2}}.\end{aligned}$$

For the sake of simplicity, let us introduce some local entropy notations for clarity:

$$\begin{cases} \boldsymbol{U}^{\eta,i}_m := \boldsymbol{U}^\eta(\underline{\eta}^i_m), \\ \boldsymbol{U}^{\eta,i}_h(x_{i\pm\frac{1}{2}}) := \boldsymbol{U}^\eta\left(\eta^i_h(x_{i\pm\frac{1}{2}})\right), \\ \Psi^{\eta,i}_h(x_{i\pm\frac{1}{2}}) := \Psi\left(\boldsymbol{U}^{\eta,i}_h(x_{i\pm\frac{1}{2}})\right). \end{cases}$$

Then we can write

$$\begin{aligned}\sum_{m=1}^{k+1} \boldsymbol{U}_m^{\eta,i}|S_m^i|\partial_t\overline{\eta}_m^i = \sum_{m=1}^{k}&\left(\boldsymbol{U}_{m+1}^{\eta,i}-\boldsymbol{U}_m^{\eta,i}\right)\widetilde{\boldsymbol{\mathcal{F}}}_{m+\frac{1}{2}}^{\eta,i}+\Theta_{\frac{1}{2}}^i\left(\boldsymbol{U}_1^{\eta,i}-\boldsymbol{U}_h^{\eta,i}(x_{i-\frac{1}{2}})\right)\boldsymbol{\mathcal{F}}_{i-\frac{1}{2}}^{\eta,\mathrm{DG}}\\&+\Theta_{k+\frac{3}{2}}^i\left(\boldsymbol{U}_h^{\eta,i}(x_{i+\frac{1}{2}})-\boldsymbol{U}_{k+1}^{\eta,i}\right)\boldsymbol{\mathcal{F}}_{i+\frac{1}{2}}^{\eta,\mathrm{DG}}-\left(1-\Theta_{\frac{1}{2}}^i\right)\boldsymbol{U}_1^{\eta,i}\boldsymbol{\mathcal{F}}_{i-\frac{1}{2}}^{\eta,\mathsf{S}}\\&+\Theta_{\frac{1}{2}}^i\boldsymbol{U}_h^{\eta,i}(x_{i-\frac{1}{2}})\boldsymbol{\mathcal{F}}_{i-\frac{1}{2}}^{\eta,\mathrm{DG}}-\left(1-\Theta_{k+\frac{3}{2}}^i\right)\boldsymbol{U}_{k+1}^{\eta,i}\boldsymbol{\mathcal{F}}_{i+\frac{1}{2}}^{\eta,\mathsf{S}}\\&+\Theta_{k+\frac{3}{2}}^i\boldsymbol{U}_h^{\eta,i}(x_{i-\frac{1}{2}})\boldsymbol{\mathcal{F}}_{i+\frac{1}{2}}^{\eta,\mathrm{DG}}.\end{aligned}$$

If we have the following condition:

$$\begin{aligned}&\sum_{m=1}^{k}\left(\boldsymbol{U}_{m+1}^{\eta,i}-\boldsymbol{U}_m^{\eta,i}\right)\widetilde{\boldsymbol{\mathcal{F}}}_{m+\frac{1}{2}}^{\eta,i}+\Theta_{\frac{1}{2}}^i\left(\boldsymbol{U}_1^{\eta,i}-\boldsymbol{U}_h^{\eta,i}(x_{i-\frac{1}{2}})\right)\boldsymbol{\mathcal{F}}_{i-\frac{1}{2}}^{\eta,\mathrm{DG}}+\Theta_{k+\frac{3}{2}}^i\left(\boldsymbol{U}_h^{\eta,i}(x_{i+\frac{1}{2}})-\boldsymbol{U}_{k+1}^{\eta,i}\right)\boldsymbol{\mathcal{F}}_{i+\frac{1}{2}}^{\eta,\mathrm{DG}}\\&\leq\Psi\left(\boldsymbol{U}_{k+1}^{\eta,i}\right)-\Psi\left(\boldsymbol{U}_1^{\eta,i}\right)+\Theta_{\frac{1}{2}}^i\left(\Psi\left(\boldsymbol{U}_1^{\eta,i}\right)-\Psi_h^i(x_{i-\frac{1}{2}})\right)+\Theta_{k+\frac{3}{2}}^i\left(\Psi_h^{\eta,i}(x_{i+\frac{1}{2}})-\Psi\left(\boldsymbol{U}_{k+1}^{\eta,i}\right)\right),\end{aligned}$$

then we obtain

$$\begin{aligned}\sum_{m=1}^{k+1}\boldsymbol{U}_m^{\eta,i}|S_m^i|\partial_t\overline{\eta}_m^i\leq&\left(1-\Theta_{\frac{1}{2}}^i\right)\left(\boldsymbol{U}_1^{\eta,i}\boldsymbol{\mathcal{F}}_{i-\frac{1}{2}}^{\eta,\mathsf{S}}-\Psi\left(\boldsymbol{U}_1^{\eta,i}\right)\right)\\&-\left(1-\Theta_{k+\frac{3}{2}}^i\right)\left(\boldsymbol{U}_{k+1}^{\eta,i}\boldsymbol{\mathcal{F}}_{i+\frac{1}{2}}^{\eta,\mathsf{S}}-\Psi\left(\boldsymbol{U}_{k+1}^{\eta,i}\right)\right)\\&+\Theta_{\frac{1}{2}}^i\left(\boldsymbol{U}_h^{\eta,i}(x_{i-\frac{1}{2}})\boldsymbol{\mathcal{F}}_{i-\frac{1}{2}}^{\eta,\mathrm{DG}}-\Psi_h^i(x_{i-\frac{1}{2}})\right)\\&-\Theta_{k+\frac{3}{2}}^i\left(\boldsymbol{U}_h^{\eta,i}(x_{i+\frac{1}{2}})\boldsymbol{\mathcal{F}}_{i+\frac{1}{2}}^{\eta,\mathrm{DG}}-\Psi_h^{\eta,i}(x_{i+\frac{1}{2}})\right)\end{aligned}$$

We get

$$\begin{aligned}\int_{\omega_i}\partial_t\mathsf{S}^\eta(\eta_h^i)\,\mathrm{d}x\leq&\left(1-\theta_{i-\frac{1}{2}}\right)\left\{U_1^{\eta,i}\boldsymbol{\mathcal{F}}^\eta\left(\boldsymbol{U}_{k+1}^{\eta,i-1},\boldsymbol{U}_1^{\eta,i}\right)-\Psi\left(\boldsymbol{U}_1^{\eta,i}\right)\right\}\\&-\left(1-\theta_{i+\frac{1}{2}}\right)\left\{U_{k+1}^{\eta,i}\boldsymbol{\mathcal{F}}^\eta\left(\boldsymbol{U}_{k+1}^{\eta,i},\boldsymbol{U}_1^{\eta,i+1}\right)-\Psi\left(\boldsymbol{U}_{k+1}^{\eta,i}\right)\right\}\\&+\theta_{i-\frac{1}{2}}\left\{U_h^{\eta,i}(x_{i-\frac{1}{2}})\boldsymbol{\mathcal{F}}^\eta\left(\boldsymbol{U}_h^{\eta,i-1}(x_{i-\frac{1}{2}}),\boldsymbol{U}_h^{\eta,i}(x_{i-\frac{1}{2}})\right)-\Psi_h^i(x_{i-\frac{1}{2}})\right\}\\&-\theta_{i+\frac{1}{2}}\left\{U_h^{\eta,i}(x_{i+\frac{1}{2}})\boldsymbol{\mathcal{F}}^\eta\left(\boldsymbol{U}_h^{\eta,i}(x_{i+\frac{1}{2}}),\boldsymbol{U}_h^{\eta,i+1}(x_{i+\frac{1}{2}})\right)-\Psi_h^{\eta,i}(x_{i+\frac{1}{2}})\right\},\end{aligned}$$

and calculating the following quantity: Combining the previous inequality written on the two neighbouring cells ω_i and ω_{i+1}, the contributions associated with the common interface $x_{i+\frac{1}{2}}$ give

$$\begin{aligned}\int_{\omega_i\cup\omega_{i+1}}\partial_t\mathsf{S}^\eta(\eta_h)\,\mathrm{d}x\leq&\left(1-\theta_{i-\frac{1}{2}}\right)\left\{\boldsymbol{U}_1^{\eta,i}\boldsymbol{\mathcal{F}}^\eta\left(\boldsymbol{U}_{k+1}^{\eta,i-1},\boldsymbol{U}_1^{\eta,i}\right)-\Psi\left(\boldsymbol{U}_1^{\eta,i}\right)\right\}\\&-\left(1-\theta_{i+\frac{3}{2}}\right)\left\{\boldsymbol{U}_{k+1}^{\eta,i+1}\boldsymbol{\mathcal{F}}^\eta\left(\boldsymbol{U}_{k+1}^{\eta,i+1},\boldsymbol{U}_1^{\eta,i+2}\right)-\Psi\left(\boldsymbol{U}_{k+1}^{\eta,i+1}\right)\right\}\\&+\theta_{i-\frac{1}{2}}\left\{\boldsymbol{U}_h^{\eta,i}(x_{i-\frac{1}{2}})\boldsymbol{\mathcal{F}}^\eta\left(\boldsymbol{U}_h^{\eta,i-1}(x_{i-\frac{1}{2}}),\boldsymbol{U}_h^{\eta,i}(x_{i-\frac{1}{2}})\right)-\Psi_h^{\eta,i}(x_{i-\frac{1}{2}})\right\}\\&-\theta_{i+\frac{3}{2}}\left\{\boldsymbol{U}_h^{\eta,i+1}(x_{i+\frac{3}{2}})\boldsymbol{\mathcal{F}}^\eta\left(\boldsymbol{U}_h^{\eta,i+1}(x_{i+\frac{3}{2}}),\boldsymbol{U}_h^{\eta,i+2}(x_{i+\frac{3}{2}})\right)-\Psi_h^{\eta,i+1}(x_{i+\frac{3}{2}})\right\}\\&+\left(1-\theta_{i+\frac{1}{2}}\right)\left\{\left(\boldsymbol{U}_1^{\eta,i+1}-\boldsymbol{U}_{k+1}^{\eta,i}\right)\boldsymbol{\mathcal{F}}^\eta\left(\boldsymbol{U}_{k+1}^{\eta,i},\boldsymbol{U}_1^{\eta,i+1}\right)-\left(\Psi\left(\boldsymbol{U}_1^{\eta,i+1}\right)-\Psi\left(\boldsymbol{U}_{k+1}^{\eta,i}\right)\right)\right\}\\+\theta_{i+\frac{1}{2}}&\left\{\left(\boldsymbol{U}_h^{\eta,i+1}(x_{i+\frac{1}{2}})-\boldsymbol{U}_h^{\eta,i}(x_{i+\frac{1}{2}})\right)\boldsymbol{\mathcal{F}}^\eta\left(\boldsymbol{U}_h^{\eta,i}(x_{i+\frac{1}{2}}),\boldsymbol{U}_h^{\eta,i+1}(x_{i+\frac{1}{2}})\right)-\left(\Psi_h^{\eta,i+1}(x_{i+\frac{1}{2}})-\Psi_h^{\eta,i}(x_{i+\frac{1}{2}})\right)\right\}.\end{aligned}$$

demonstrate we have a convex combination of order 1 dissipation and high-order dissipation on boundaries. We now need to ensure that

$$\begin{aligned}
&\Theta^i_{\frac{1}{2}}\left\{\left(\boldsymbol{U}_1^{\eta,i}-\boldsymbol{U}_h^{\eta,i}(x_{i-\frac{1}{2}})\right)\boldsymbol{\mathcal{F}}^{\eta,\mathrm{DG}}_{i-\frac{1}{2}}-\left(\Psi\left(\boldsymbol{U}_1^{\eta,i}\right)-\Psi_h^i(x_{i-\frac{1}{2}})\right)\right\}\\
&+\sum_{m=1}^{k}\left(\boldsymbol{U}_{m+1}^{\eta,i}-\boldsymbol{U}_m^{\eta,i}\right)\Theta^i_{m+\frac{1}{2}}\left(\widehat{\boldsymbol{F}}^{\eta}_{m+\frac{1}{2}}-\boldsymbol{\mathcal{F}}^{\eta,\mathsf{S}}_{m+\frac{1}{2}}\right)\\
&+\Theta^i_{k+\frac{3}{2}}\left\{\left(\boldsymbol{U}_h^{\eta,i}(x_{i+\frac{1}{2}})-\boldsymbol{U}_{k+1}^{\eta,i}\right)\boldsymbol{\mathcal{F}}^{\eta,\mathrm{DG}}_{i+\frac{1}{2}}-\left(\Psi_h^{\eta,i}(x_{i+\frac{1}{2}})-\Psi\left(\boldsymbol{U}_{k+1}^{\eta,i}\right)\right)\right\}\\
&\leq\Psi\left(\boldsymbol{U}_{k+1}^{\eta,i}\right)-\Psi\left(\boldsymbol{U}_1^{\eta,i}\right)-\sum_{m=1}^{k}\left(\boldsymbol{U}_{m+1}^{\eta,i}-\boldsymbol{U}_m^{\eta,i}\right)\boldsymbol{\mathcal{F}}^{\eta,\mathsf{S}}_{m+\frac{1}{2}}.
\end{aligned}$$

This can be seen as the following optimization problem:

$$\max_{\boldsymbol{\Theta}\in\mathbb{R}^{k+2}}\sum_{m=0}^{k+1}\Theta^i_{m+\frac{1}{2}},\qquad\text{with }\boldsymbol{A}\cdot\boldsymbol{\Theta}\leq b,\text{ and }\mathbf{0}\leq\boldsymbol{\Theta}\leq\boldsymbol{\Theta}_c,$$

with $\boldsymbol{\Theta}:=\left(\Theta^i_{\frac{1}{2}},\Theta^i_{\frac{3}{2}},\dots,\Theta^i_{k+\frac{3}{2}}\right)^\top$ and $\boldsymbol{A}:=\left(a_{\frac{1}{2}},a_{\frac{3}{2}},\dots,a_{k+\frac{3}{2}}\right)^\top$ being filled with

$$\begin{cases}
a_{\frac{1}{2}}:=\left(\boldsymbol{U}_1^{\eta,i}-\boldsymbol{U}_h^{\eta,i}(x_{i-\frac{1}{2}})\right)\boldsymbol{\mathcal{F}}^{\eta,\mathrm{DG}}_{i-\frac{1}{2}}-\left(\Psi\left(\boldsymbol{U}_1^{\eta,i}\right)-\Psi_h^i(x_{i-\frac{1}{2}})\right),\\
a_{m+\frac{1}{2}}:=\left(\boldsymbol{U}_{m+1}^{\eta,i}-\boldsymbol{U}_m^{\eta,i}\right)\left(\widehat{\boldsymbol{F}}^{\eta}_{m+\frac{1}{2}}-\boldsymbol{\mathcal{F}}^{\eta,\mathsf{S}}_{m+\frac{1}{2}}\right), & \forall m\in[\![1,k]\!]\\
a_{k+\frac{3}{2}}:=\left(\boldsymbol{U}_h^{\eta,i}(x_{i+\frac{1}{2}})-\boldsymbol{U}_{k+1}^{\eta,i}\right)\boldsymbol{\mathcal{F}}^{\eta,\mathrm{DG}}_{i+\frac{1}{2}}-\left(\Psi_h^{\eta,i}(x_{i+\frac{1}{2}})-\Psi\left(\boldsymbol{U}_{k+1}^{\eta,i}\right)\right).
\end{cases}$$

and

$$b:=\Psi\left(\boldsymbol{U}_{k+1}^{\eta,i}\right)-\Psi\left(\boldsymbol{U}_1^{\eta,i}\right)-\sum_{m=1}^{k}\left(\boldsymbol{U}_{m+1}^{\eta,i}-\boldsymbol{U}_m^{\eta,i}\right)\mathscr{F}^{\eta,\mathsf{S}}_{m+\frac{1}{2}}.$$

To solve this problem, we can use the following *Greedy Algorithm*:

Data: $b\in\mathbb{R}_+$ positive scalar, $(\boldsymbol{A},\boldsymbol{\Theta}_c)\in(\mathbb{R}_+)^{k+2}$ positives vectors.
Result: $\boldsymbol{\Theta}\in(\mathbb{R}_+)^{k+2}$ minimizing the sum $\sum_{m=0}^{k+1}\Theta^i_{m+\frac{1}{2}}$ such that $\boldsymbol{A}\cdot\boldsymbol{\Theta}\leq b$ and $\mathbf{0}\leq\boldsymbol{\Theta}\leq\boldsymbol{\Theta}_c$.
Begin;
$\boldsymbol{\Theta}\leftarrow\boldsymbol{\Theta}_c$;
$\mathcal{N}=[\![0,k+1]\!]$;
while $\boldsymbol{A}\cdot\boldsymbol{\Theta}>b$ **do**
  $p=\mathrm{argmax}_{m\in\mathcal{N}}$;
  $\mathcal{N}\leftarrow\mathcal{N}\setminus\{p\}$;
  $d=\sum_{m\in\mathcal{N}}a_{m+\frac{1}{2}}\Theta^i_{m+\frac{1}{2}}$;
  if $d>b$ **then**
    $\Theta^i_{p+\frac{1}{2}}=0$;
  else
    $\Theta^i_{p+\frac{1}{2}}=\frac{b-d}{a_{p+\frac{1}{2}}}$;
  end
end

Algorithm 3: Overview of Greedy Algorithm.

Let us notice that since

$$b=\Psi\left(\boldsymbol{U}_{k+1}^{\eta,i}\right)-\Psi\left(\boldsymbol{U}_1^{\eta,i}\right)-\sum_{m=1}^{k}\left(\boldsymbol{U}_{m+1}^{\eta,i}-\boldsymbol{U}_m^{\eta,i}\right)\boldsymbol{\mathcal{F}}^{\eta}\left(\boldsymbol{U}_m^{\eta,i},\boldsymbol{U}_{m+1}^{\eta,i}\right)$$

and $(v-u)\boldsymbol{\mathcal{F}}(u,v)\leq\Psi(v)-\Psi(u)$, we have

$$b\geq\Psi\left(\boldsymbol{U}_{k+1}^{\eta,i}\right)-\Psi\left(\boldsymbol{U}_1^{\eta,i}\right)-\sum_{m=1}^{k}\left(\Psi\left(\boldsymbol{U}_{m+1}^{\eta,i}\right)-\Psi\left(\boldsymbol{U}_m^{\eta,i}\right)\right)=0,$$

so if $\boldsymbol{\Theta} = \mathbf{0}$ we can still use this algorithm on every cell after having calculated a similar way as before, i.e. by denoting $\Delta\underline{\boldsymbol{v}}(\mathbb{U}_m) := \boldsymbol{v}(\underline{\mathbb{U}}_{m+1}) - \boldsymbol{v}(\underline{\mathbb{U}}_{m+1})$ and $\Delta\underline{\Psi}_m := \Psi\left(\boldsymbol{v}(\underline{\mathbb{U}}_{m+1})\right) - \Psi\left(\boldsymbol{v}(\underline{\mathbb{U}}_{m+1})\right)$ we have

$$\Delta\underline{\boldsymbol{v}}(\mathbb{U}_m) \cdot \widetilde{\boldsymbol{\mathcal{F}}}_{m+\frac{1}{2}} \leq \Delta\underline{\Psi}_m \Longleftrightarrow \begin{cases} \Theta^{\mathbf{S},(\mathrm{III})}_{m+\frac{1}{2}} \cdot (\Delta\boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot \Delta\underline{\boldsymbol{v}}(\mathbb{U}_m)) \leq \Delta\underline{\Psi}_m - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot \Delta\underline{\boldsymbol{v}}(\mathbb{U}_m), \\ \Delta\underline{\Psi}_m - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot \Delta\underline{\boldsymbol{v}}(\mathbb{U}_m) \geq 0, \end{cases}$$

and we then get

$$\Theta^{\mathbf{S},(\mathrm{III})}_{m+\frac{1}{2}} = \begin{cases} \dfrac{\Delta\underline{\Psi}_m - \boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot \Delta\underline{\boldsymbol{v}}(\mathbb{U}_m)}{\Delta\boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} \cdot \Delta\underline{\boldsymbol{v}}(\mathbb{U}_m)} & \text{if } \Delta\underline{\boldsymbol{v}}(\mathbb{U}_m) \cdot \Delta\boldsymbol{\mathcal{F}}_{m+\frac{1}{2}} > 0, \\ 1 & \text{else.} \end{cases}$$

The cell-based coefficient $\Theta^{\mathbf{S},(\mathrm{III})}_{m+\frac{1}{2}}$ is designed to enforce a semi-discrete entropy conservation/dissipation inequality at the level of the DG cells. In contrast with the fully discrete construction and even with the subcell semi-discrete case, this formulation is in principle compatible with an arbitrarily high order of accuracy, since the entropy correction acts on the projected DG fields and does not intrinsically restrict the polynomial degree. However, the computation of $\Theta^{\mathbf{S},(\mathrm{III})}_{m+\frac{1}{2}}$ requires a nontrivial reformulation of the underlying first-order FV fluxes, so that the relevance and consistency of the previously introduced blending coefficients are no longer guaranteed. This makes the practical use of the full cell-level entropy limiter more delicate, and motivates keeping this construction mainly at a theoretical level in the present work.

APPENDIX G

Continuous settings and functional spaces

In this appendix, we collect some basic functional analysis tools and notations thet are used throughout the manuscript.

G.1 Smooth function spaces

Let $X \subset \mathbb{R}^n$, with $n \geq 1$, denote a measurable open bounded set. In practice, the dimension n will usually be equal to d when X is a domain or a mesh element, or equal to $d-1$ when X represents a mesh face.

We begin by recalling the classical spaces of smooth functions that serve as a foundation for the definition of distributional derivatives and Sobolev spaces. The space of infinitely differentiable (real-valued) functions on X is denoted by

$$\mathscr{C}^\infty(X) := \{v : X \to \mathbb{R} \mid \partial^{\boldsymbol{\alpha}} v \text{ exists and is continuous for all } \boldsymbol{\alpha} \in \mathbb{N}^n\}.$$

A function $v \in \mathscr{C}^\infty(X)$ is said to have *compact support* in X if there exists a compact subset $K \Subset X$ such that $v(x) = 0$ for all $x \notin K$. The set of such functions is denoted by

$$\mathscr{C}_c^\infty(X) := \{v \in \mathscr{C}^\infty(X) \mid \operatorname{supp}(v) \Subset X\}.$$

This space plays a central role in the definition of weak derivatives and serves as the standard set of test functions in distribution theory.

For any integer $k \geq 0$, we also introduce

$$\mathscr{C}^k(X) := \Big\{ v : X \to \mathbb{R} \;\Big|\; \partial^{\boldsymbol{\alpha}} v \text{ is continuous for } \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n) \text{ s.t. } \sum_{i=1}^n \alpha_i \leq k \Big\}.$$

and its subspace of compactly supported functions

$$\mathscr{C}_c^k(X) := \mathscr{C}^k(X) \cap \mathscr{C}_c^\infty(X).$$

The spaces $\mathscr{C}^k(X)$ and $\mathscr{C}_c^k(X)$ will be used mainly for regularity statements and approximation arguments.

G.2 Classical differential operators

We recall here the standard differential operators used throughout the manuscript in both the continuous and the discrete formulations.

We denote by

$$\nabla := (\partial_1, \dots, \partial_n)^\top$$

the classical nabla operator, understood as the vector collecting all partial derivatives with respect to the underlying coordinate variables. In situations where several independent variables can coexist (for instance physical coordinates $\boldsymbol{x}$, reference coordinates $\boldsymbol{\xi}$, etc.), we specify the differentiation variable by writing explicitly the subscript:

$$\nabla_{\boldsymbol{x}} := (\partial_{x_1}, \dots, \partial_{x_n})^\top,$$

whenever the derivatives are taken with respect to a vector of variables $\boldsymbol{x} = (x_1, \dots, x_n)$. Throughout the manuscript, the subscript may be omitted whenever there is no ambiguity.

The basic operators associated with ∇ are defined as follows.

♯ *Gradient.* For a scalar function $v : X \to \mathbb{R}$, the gradient of v is the vector field

$$\mathbf{grad}\, v := \nabla v = (\partial_1 v, \dots, \partial_n v)^\top.$$

♯ *Divergence & jacobian.* For a vector field $\boldsymbol{v} = (v_1, \dots, v_n)^\top : X \to \mathbb{R}^n$, the Jacobian matrix is defined by

$$\mathbb{J}_{\boldsymbol{v}} := \nabla \boldsymbol{v} = (\partial_j v_i)_{(i,j)\in[\![1,n]\!]^2},$$

while the divergence is the trace of this matrix:

$$\operatorname{div} \boldsymbol{v} := \nabla \cdot \boldsymbol{v} = \sum_{i=1}^{n} \partial_i v_i.$$

♯ *Laplacian.* For a scalar function v, the Laplace operator is defined by

$$\Delta v := \operatorname{div} \circ \mathbf{grad}\, v = \nabla \cdot (\nabla v) = \sum_{i=1}^{n} \partial_{ii}^2 v.$$

When acting on a vector field $\boldsymbol{v} = (v_1, \dots, v_n)^\top$, it is applied componentwise:

$$\Delta \boldsymbol{v} := (\Delta v_1, \dots, \Delta v_n)^\top.$$

♯ *Curl (rotational).* The curl operator is defined in two or three spatial dimensions. In two dimensions, for $\boldsymbol{v} = (v_1, v_2)$,

$$\operatorname{curl} \boldsymbol{v} := \nabla \times \boldsymbol{v} = \partial_1 v_2 - \partial_2 v_1.$$

In three dimensions, for $\boldsymbol{v} = (v_1, v_2, v_3)$,

$$\mathbf{curl}\, \boldsymbol{v} := \nabla \times \boldsymbol{v} = (\partial_2 v_3 - \partial_3 v_2,\ \partial_3 v_1 - \partial_1 v_3,\ \partial_1 v_2 - \partial_2 v_1)^\top.$$

Remark (*Tensor-valued functions*). *For tensor-valued functions $\mathbb{T} : X \to \mathbb{R}^{n\times m}$, all the previous operators are defined componentwise. We shall then freely write $\nabla\mathbb{T}$, $\nabla \cdot \mathbb{T}$, or $\Delta\mathbb{T}$ whenever the context is unambiguous.*

G.3 Lebesgue spaces

We now consider functions $v : X \to \mathbb{R}$ that are Lebesgue measurable and we denote by $\int_X v(x)\,\mathrm{d}x$ the Lebesgue integral of v over X. Whenever no ambiguity may arise, we may omit both the dependence on x and the measure, and simply write $\int_X v$.

For $p \in [1, +\infty]$, we define the L^p-norm of a (real-valued) measurable function v on X as

$$\|v\|_{L^p(X)} := \begin{cases} \left(\displaystyle\int_X |v|^p \,\mathrm{d}x\right)^{\frac{1}{p}}, & \text{if } 1 \le p < +\infty, \\ \inf\{M \in \mathbb{R} \mid |v(x)| \le M \text{ for a.e. } x \in X\}, & \text{if } p = +\infty. \end{cases}$$

The Lebesgue space $L^p(X)$ is then defined as

$$L^p(X) := \{v \text{ measurable on } X \mid \|v\|_{L^p(X)} < +\infty\}.$$

Equipped with the norm $\|\cdot\|_{L^p(X)}$, the space $L^p(X)$ is a Banach space. Moreover, when $p < +\infty$, the space $\mathscr{C}_c^\infty(X)$ of infinitely differentiable functions compactly supported in X is dense in $L^p(X)$.

We will denote by $(\cdot,\cdot)_X$ the standard L^2 inner product on X, defined for all functions $v, w \in L^2(X)$ by

$$(v, w)_X := \int_X vw \,\mathrm{d}x,$$

and extended componentwise when v and w are vector or tensor-valued. Similarly, for any subset $\Sigma \subset \partial X$ endowed with the Lebesgue surface measure $\mathrm{d}s$, the L^2 inner product on Σ is denoted by $\langle \mu, \nu \rangle_\Sigma$ and defined for all $\mu, \nu \in L^2(\Sigma)$ as

$$\langle \mu, \nu \rangle_\Sigma := \int_\Sigma \mu\nu \,\mathrm{d}s,$$

again extended componentwise in the vector and tensor cases.

Remark (*Case $p = 2$*). *$L^2(X)$ is a real Hilbert space equipped with the scalar product $(\cdot \mid \cdot)_X$ and the associated norm $\|\cdot\|_X$. In what follows, we will often omit the index X when the domain is clear from the context.*

For the case $p = +\infty$, we recall that the corresponding Lebesgue space is defined by

$$L^\infty(X) := \{v \text{ measurable on } X \mid \operatorname*{ess\,sup}_{x \in X} |v(x)| < +\infty\}.$$

In contrast with the case $p < +\infty$, the finiteness of this norm does not require X to be bounded. The space $L^\infty(X)$ consists of functions that remain essentially bounded on X, even when X is unbounded.

Remark (*Equivalence classes in L^p spaces*). *Strictly speaking, the spaces $L^p(X)$ as defined above do not consist of functions in the classical sense, but of equivalence classes of measurable functions. Two measurable functions v and w are identified whenever they coincide almost everywhere on X, so we may therefore regard $L^p(X)$ as the quotient space*

$$L^p(X) = \mathscr{L}^p(X) / \sim,$$

where $\mathscr{L}^p(X)$ denotes the set of measurable functions v such that $\int_X |v|^p \, \mathrm{d}x < +\infty$ (or $\operatorname{ess\,sup}_{x \in X} |v(x)| < +\infty$ when $p = +\infty$), and $\sim$ is the equivalence relation defined by equality almost everywhere. In practice, it is customary to work with representatives of these classes and to refer to them simply as "functions" whenever the context is clear. This slight abuse of notation will be adopted throughout the manuscript.

G.4 Distributions and weak derivatives

We briefly recall the distributional framework underlying Sobolev spaces and weak formulations. Let $\mathscr{D}(X) := \mathscr{C}_c^\infty(X)$ denote the space of smooth compactly supported test functions on X. Its dual space $\mathscr{D}'(X)$, the space of distributions on X, consists of all continuous linear forms on $\mathscr{D}(X)$.

For any locally integrable function $v \in L^1_{\mathrm{loc}}(X)$, we define its associated distribution (still denoted v) by

$$\langle\!\langle v, \varphi \rangle\!\rangle := \big(v, \varphi\big)_X, \qquad \forall \varphi \in \mathscr{D}(X).$$

This identification is injective modulo equality almost everywhere, and provides the canonical embedding $L^1_{\mathrm{loc}}(X) \hookrightarrow \mathscr{D}'(X)$.

One may then define weak derivatives within this framework. Given a multi-index $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, the *distributional derivative* of $T \in \mathscr{D}'(X)$ is the distribution $\partial^\alpha T$ defined by

$$\langle\!\langle \partial^\alpha T, \varphi \rangle\!\rangle := (-1)^{\|\alpha\|_1} \, \langle\!\langle T, \partial^\alpha \varphi \rangle\!\rangle, \qquad \forall \varphi \in \mathscr{D}(X),$$

where $\|\alpha\|_1 := \sum_{i=1}^n \alpha_i$. If T is induced by a function $v \in L^1_{\mathrm{loc}}(X)$ and if $\partial^\alpha v$ exists in the classical sense, then the above definition is fully consistent:

$$\langle\!\langle \partial^\alpha v, \varphi \rangle\!\rangle = \big(\partial^\alpha v, \varphi\big)_X, \qquad \forall \varphi \in \mathscr{D}(X).$$

This distributional viewpoint is the basis for the definition of Sobolev spaces, which require the existence of weak derivatives $\partial^\alpha v$ belonging to $L^p(X)$ for all multi-indices of order up to s. It also underlies the construction of weak formulations, where derivatives are most of the time transferred onto test functions by means of integration by parts.

G.5 Sobolev spaces

Let $X \subset \mathbb{R}^n$ be as above. Denote by ∂_i the partial derivative with respect to the i-th Cartesian coordinate. For any multi-index $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, we introduce the distributional derivative

$$\partial^{\boldsymbol{\alpha}} v := \partial_1^{\alpha_1} \cdots \partial_n^{\alpha_n} v,$$

with the convention that $\partial^{\mathbf{0}} v := v$. For any integer $s \geq 0$, we define the set of multi-indices of order less than or equal to s as

$$A_n^s := \{\boldsymbol{\alpha} \in \mathbb{N}^n \mid \|\boldsymbol{\alpha}\|_1 \leq s\}.$$

For $p \in [1, +\infty]$, the Sobolev space $\mathsf{W}^{s,p}(X)$ is defined by

$$\mathsf{W}^{s,p}(X) := \big\{v \in L^p(X) \mid \forall \boldsymbol{\alpha} \in A_n^s, \, \partial^{\boldsymbol{\alpha}} v \in L^p(X)\big\},$$

equipped with the Sobolev norm

$$\|v\|_{\mathsf{W}^{s,p}(X)} := \sum_{\boldsymbol{\alpha}\in A_n^s} \|\partial^{\boldsymbol{\alpha}} v\|_{L^p(X)}.$$

Equipped with such a norm, $\mathsf{W}^{s,p}(X)$ is a Banach space. The space $\mathsf{W}_0^{s,p}(X)$ denotes the closure of $\mathscr{C}_c^\infty(X)$ in $\mathsf{W}^{s,p}(X)$.

If we consider again the case $p = 2$, we adopt the classical notation

$$\mathsf{H}^s(X) := \mathsf{W}^{s,2}(X), \qquad \mathsf{H}_0^s(X) := \mathsf{W}_0^{s,2}(X),$$

and $\mathsf{H}^s(X)$ is a Hilbert space when endowed with the scalar product

$$\big(v, w\big)_{\mathsf{H}^s(X)} := \sum_{\boldsymbol{\alpha}\in A_n^s} \big(\partial^{\boldsymbol{\alpha}} v, \partial^{\boldsymbol{\alpha}} w\big)_X.$$

When the boundary ∂X of the domain is sufficiently smooth, the *trace operator*

$$\gamma : \mathsf{H}^1(X) \to \mathsf{H}^{1/2}(\partial X)$$

is continuous and surjective. The trace space $\mathsf{H}^{1/2}(\partial X)$ consists of the restrictions to ∂X of $\mathsf{H}^1(X)$ functions and is endowed with the norm

$$\|\mu\|_{\mathsf{H}^{1/2}(\partial X)} := \inf\big\{ \|v\|_{\mathsf{H}^1(X)} \mid v \in \mathsf{H}^1(X),\ \gamma(v) = \mu \big\}.$$

Its dual, denoted by $\mathsf{H}^{-1/2}(\partial X)$, is defined through the pivot space $L^2(\partial X)$ and will be used to describe natural (Neumann-type) boundary conditions. These spaces are essential for the formulation of variational problems involving fluxes or boundary stresses.

Even though a significant part of this thesis is devoted to hyperbolic systems and their numerical approximation, elliptic and diffusive problems also appear in various components of the analysis, in particular in the treatment of the interior fluid region under floating structures. This motivates the introduction of function spaces that naturally accommodate differential operators acting on vector fields.

Let $\mathscr{D}$ denote a linear differential operator acting on vector fields $\boldsymbol{\tau} : X \to \mathbb{R}^d$, such as, for instance $\mathscr{D} \in \{\mathrm{div}, \mathbf{curl}\}$. We define, for $p \in [1, +\infty]$,

$$\mathbf{W}^p(\mathscr{D}; X) := \big\{ \boldsymbol{\tau} \in L^p(X)^d \;\big|\; \mathscr{D}\boldsymbol{\tau} \in L^p(X)^{m_{\mathscr{D}}} \big\},$$

where $m_{\mathscr{D}}$ is the number of components of $\mathscr{D}\boldsymbol{\tau}$. This space is endowed with the graph norm

$$\|\boldsymbol{\tau}\|_{W^p(\mathscr{D};X)} := \Big(\|\boldsymbol{\tau}\|^p_{L^p(X)^d} + \|\mathscr{D}\boldsymbol{\tau}\|^p_{L^p(X)^{m_{\mathscr{D}}}} \Big)^{1/p}.$$

For $p = 2$, we adopt the Hilbertian notation $\mathbf{H}(\mathscr{D}; X) := \mathbf{W}^2(\mathscr{D}; X)$, so that, for instance,

$$\mathbf{H}(\mathrm{div}; X) := \big\{ \boldsymbol{\tau} \in L^2(X)^d \mid \nabla\cdot\boldsymbol{\tau} \in L^2(X) \big\}, \qquad \mathbf{H}(\mathbf{curl}; X) := \big\{ \boldsymbol{\tau} \in L^2(X)^3 \mid \nabla\times\boldsymbol{\tau} \in L^2(X)^3 \big\}.$$

The corresponding norms read

$$\|\boldsymbol{\tau}\|_{\mathbf{H}(\mathrm{div};X)} = \big(\|\boldsymbol{\tau}\|^2_{L^2(X)^d} + \|\nabla\cdot\boldsymbol{\tau}\|^2_{L^2(X)} \big)^{1/2}, \qquad \|\boldsymbol{\tau}\|_{\mathbf{H}(\mathbf{curl};X)} = \big(\|\boldsymbol{\tau}\|^2_{L^2(X)^3} + \|\nabla\times\boldsymbol{\tau}\|^2_{L^2(X)^3} \big)^{1/2}.$$

When ∂X is sufficiently regular, the trace of the normal component of a field in $\mathbf{H}(\mathrm{div}; X)$ and that of the tangential component of a field in $\mathbf{H}(\mathbf{curl}; X)$ are both well defined in $\mathsf{H}^{-1/2}(\partial X)$.

G.6 Fourier transform

For a function $f \in L^1(\mathbb{R}^d)$, the (continuous) Fourier transform and its inverse are defined by

$$(\mathfrak{F} f)(\xi) = \int_{\mathbb{R}^d} e^{-\mathrm{i}\, x\cdot\xi}\, f(x)\, \mathrm{d}x, \qquad (\mathfrak{F}^{-1} g)(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{\mathrm{i}\, x\cdot\xi}\, g(\xi)\, \mathrm{d}\xi.$$

By density, these formulas extend to $f \in L^2(\mathbb{R}^d)$, in which case the Fourier transform defines a unitary operator on $L^2(\mathbb{R}^d)$:

$$\|\mathfrak{F} f\|_{L^2(\mathbb{R}^d)} = (2\pi)^{d/2}\, \|f\|_{L^2(\mathbb{R}^d)}.$$

We also recall the definition of the Schwartz class. The space $\mathcal{S}(\mathbb{R}^d)$ is defined as the space of all functions $f \in \mathscr{C}^\infty(\mathbb{R}^d)$ whose derivatives decay faster than any polynomial at infinity. More precisely,

$$\mathcal{S}(\mathbb{R}^d) := \left\{ f \in \mathscr{C}^\infty(\mathbb{R}^d) \;\middle|\; \sup_{x\in\mathbb{R}^d} \left| x^{\boldsymbol{\alpha}} \partial^{\boldsymbol{\beta}} f(x) \right| < +\infty, \quad \forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^d \right\}.$$

This space is stable under differentiation, multiplication by polynomials, and Fourier transform. In particular, if $f \in \mathcal{S}(\mathbb{R}^d)$, then $\mathfrak{F} f \in \mathcal{S}(\mathbb{R}^d)$.

A key property of the Fourier transform is that differentiation in physical space corresponds to multiplication in Fourier space:

$$\mathfrak{F}(\partial^{\boldsymbol{\alpha}} f)(\xi) = (\mathrm{i}\xi)^{\boldsymbol{\alpha}}\, \mathfrak{F} f(\xi),$$

for any multi-index $\boldsymbol{\alpha} \in \mathbb{N}^d$. This correspondence allows one to represent differential or pseudo-differential operators as Fourier multipliers.

Given a sufficiently regular function $m : \mathbb{R}^d \to \mathbb{C}$, called a *symbol*, one defines the Fourier multiplier operator $m(D)$ acting on f as

$$(m(D)f)(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{\mathrm{i}\, x\cdot\xi}\, m(\xi)\, (\mathfrak{F} f)(\xi)\, \mathrm{d}\xi \;=\; \mathfrak{F}^{-1}\big(m(\xi)\, \mathfrak{F} f(\xi)\big)(x).$$

The operator $m(D)$ therefore acts as multiplication by $m(\xi)$ in the frequency domain, and is linear whenever m is fixed.

G.7 Some fundamental results

We recall here a few fundamental theorems that will serve as a theoretical foundation for the variational, projection, and existence results appearing throughout the manuscript.

Hölder's inequality. *Let $X \subset \mathbb{R}^n$ be a bounded domain. For any conjugate exponents $(p, q) \in [1, +\infty]^2$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$, it holds that $vw \in L^1(X)$ for all $v \in L^p(X)$ and $w \in L^q(X)$, with*

$$\int_X |vw| \, \mathrm{d}x \leq \|v\|_{L^p(X)} \|w\|_{L^q(X)}.$$

The special case $p = q = 2$ corresponds to the *Cauchy-Schwarz inequality*, which will be used in the forthcoming analysis.

Gauss-Ostrogradski divergence theorem. *Let $X \subset \mathbb{R}^n$ be a bounded domain with Lipschitz boundary ∂X and outward unit normal vector $\mathbf{n}$. If $\boldsymbol{v} \in \mathscr{C}^1(\overline{X})^n$ and $w \in \mathscr{C}^1(\overline{X})$, the classical integration by parts identity holds:*

$$\int_X w\, (\nabla \cdot \boldsymbol{v}) \; \mathrm{d}x = -\int_X \nabla w \cdot \boldsymbol{v} \; \mathrm{d}x + \int_{\partial X} w\, (\boldsymbol{v} \cdot \mathbf{n}) \; \mathrm{d}s.$$

By density and continuity of the trace operator on $\mathsf{H}(\mathrm{div}; X)$, this formula extends to all $\boldsymbol{v} \in \mathsf{H}(\mathrm{div}; X)$ and $w \in \mathsf{H}^1(X)$ in the weak sense, where $\boldsymbol{v} \cdot \mathbf{n} \in \mathsf{H}^{-1/2}(\partial X)$ denotes the normal trace of $\boldsymbol{v}$.

This theorem provides the analytical backbone of variational formulations and finite element methods, as it allows to transfer derivatives from one function to another. It will be repeatedly used in the weak formulations throughout this study.

Poincaré inequality. *Let $X \subset \mathbb{R}^n$ be a bounded Lipschitz domain and consider the Sobolev space $\mathsf{H}^1_0(X)$ defined as the closure of $\mathscr{C}^\infty_c(X)$ in $\mathsf{H}^1(X)$. There exists a constant $\mathsf{c}_{\mathscr{P}} = \mathsf{c}_{\mathscr{P}}(X) > 0$ such that*

$$\|v\|_{L^2(X)} \leq \mathsf{c}_{\mathscr{P}} \|\nabla v\|_{L^2(X)^n} \qquad \forall v \in \mathsf{H}^1_0(X).$$

In particular, this inequality implies the equivalence of norms on $\mathsf{H}^1_0(X)$:

$$\|v\|_{\mathsf{H}^1(X)} \sim \|\nabla v\|_{L^2(X)^n} \qquad \forall v \in \mathsf{H}^1_0(X).$$

The previous result is essential for establishing stability and coercivity properties of elliptic operators.

Lax-Milgram theorem. *Let V be a real Hilbert space with scalar product $(\cdot,\cdot)_V$ and norm $\|\cdot\|_V$. Let $\mathscr{A} : V \times V \to \mathbb{R}$ be a bilinear form that is continuous, that is, there exists $\mathsf{c} > 0$ such that*

$$|\mathscr{A}(u,v)| \leq \mathsf{c}\|u\|_V\|v\|_V \qquad \forall u, v \in V,$$

and coercive i.e. there exists $\alpha > 0$ such that

$$\mathscr{A}(v,v) \geq \alpha\|v\|_V^2 \qquad \forall v \in V.$$

Let $\ell : V \to \mathbb{R}$ be a continuous linear form. Then, there exists a unique $u \in V$ such that

$$\mathscr{A}(u,v) = \ell(v) \qquad \forall v \in V.$$

This theorem will play a central role whenever elliptic problems are formulated in a weak form, in particular for the DtN operator induced by a floating structure in a shallow-water governed flow.

Riesz representation theorem. *Let V be a real Hilbert space equipped with $(\cdot,\cdot)_V$ some scalar product and its associated norm $\|\cdot\|_V$. For every continuous linear form $\ell \in V'$ (the dual space of V), there exists a unique element $u_\ell \in V$ such that*

$$\ell(v) = (u_\ell, v)_V \qquad \forall v \in V,$$

and moreover,

$$\|\ell\|_{V'} = \|u_\ell\|_V.$$

This establishes an isometric isomorphism between V' and V, providing a canonical identification of a continuous linear form with a unique element of the underlying Hilbert space. In particular, it ensures the existence and uniqueness of L^2-orthogonal projections onto closed subspaces of $L^2(X)$ or polynomial subspaces, a property that will be used in the construction of some discrete projectors.

List of Figures

1 Illustration of the different coastal wave regimes over a varying bathymetry. 6
2 Schematic representation of a standard FV method on an unstructured mesh. 7
3 Schematic representation of DG approximation u_h across two neighboring control volumes ω_c and ω_v. 8
4 Schematic representation of the Arbitrary Lagrangian–Eulerian (ALE) mapping on a simplicial unstructured mesh. The fixed reference domain $\Omega(0)$ (left) with coordinates $\boldsymbol{\xi}$ is mapped continuously in time onto the moving physical domain $\Omega(t)$ (right) with coordinates $\boldsymbol{x}(\boldsymbol{\xi}, t)$. The mapping purely displaces the nodes while preserving straight edges, generating a deformed triangular element (red) whose trajectory is tracked by the grid velocity $\boldsymbol{w}^g$. 14

1.1 Main notations for prebalanced shallow-water equations. The free-surface elevation is denoted by η, the bathymetry by b and the water-height by H. The horizontal velocity is denoted by u and the horizontal discharge by q. 20
1.2 Subcell decomposition of ω_i through $k+2$ flux points. 22
1.3 Piecewise polynomial function v_h^i and associated sub-mean values. 23
1.4 Periodic sinusoidal solution − Free surface elevation computed at $T_{\max} = 0.6$ for $k = 3$ and $N_{\text{el}} = 50$ 37
1.5 Sinusoidal solution on dry bed − Free surface elevation computed at $T_{\max} = 0.09$ sec for $k = 3$ and $n_{\text{el}} = 50$, with a zoom on dry-bed. 38
1.6 Sinusoidal solution on dry bed − Free surface elevation computed at $T_{\max} = 0.09\,s$ for $k = 8$ and $n_{\text{el}} = 10$ (left), with a zoom on the wet-dry transition (right). 39
1.7 Dam-break on a wet bed − Free surface elevation computed at $T_{\max} = 0.075$ sec for $k = 3$ and $N_{\text{el}} = 50$. 39
1.8 Dam-break − Free surface elevation computed at $T_{\max} = 0.075$ sec for $k = 3$ and $n_{\text{el}} = 50$, with a zoom on both rarefaction and shock waves. 40
1.9 Dam-break on a wet bed − Free surface elevation computed at $T_{\max} = 0.075$ sec for $k = 9$ and $n_{\text{el}} = 10$. 40
1.10 Dam-break on a dry bed − Free surface elevation computed in the time interval $[0.005, 0.05]$, for $k = 3$ and $N_{\text{el}} = 50$. 41
1.11 Dam-break on a dry bed − Zoom on top front and dry front, computed in the time interval $[0.005, 0.05]$, for $k = 3$ and $N_{\text{el}} = 50$. 41
1.12 Preservation of a motionless steady state (first case) − Free surface elevation computed at $T_{\max} =$ 500 sec, for $k = 3$ and $N_{\text{el}} = 100$. 42
1.13 Preservation of a motionless steady state (second case) − Free surface elevation computed at $T_{\max} = 500$ sec, for $k = 3$ and $N_{\text{el}} = 100$, with a zoom on dry structure. 43
1.14 Preservation of a motionless steady state (third case) − Free surface elevation computed at $T_{\max} = 500$ sec, for $k = 3$ and $N_{\text{el}} = 100$, with a zoom on dry structure. 43
1.15 Transcritical flow over a bump without shock − Surface elevation (left) and discharge (right) computed at several times, for $k = 3$ and $N_{\text{el}} = 100$. 45
1.16 Carrier & Greenspan's transient solution − Surface elevation computed at several times between 0.0 sec and 100.0 sec − Comparison between $k = 3$, $N_{\text{el}} = 50$ and $k = 8$, $N_{\text{el}} = 10$ setups. 46
1.17 Carrier & Greenspan's transient solution − Surface elevation computed at several times between 0 and 100 sec for $k = 3$ and $N_{\text{el}} = 50$ (zoom on dry bed). 46
1.18 Carrier & Greenspan's transient solution − Surface elevation computed at several times between 0 and 100 sec for $k = 8$ and $N_{\text{el}} = 10$ (full domain). 47
1.19 Carrier & Greenspan's periodic solution − Surface elevation computed in the time interval $[20.25T, 20.50T]$ for $k = 3$ and $N_{\text{el}} = 50$ (zoom on dry bed). 48
1.20 Carrier & Greenspan's periodic solution − Surface elevation computed in the time interval $[20.25T, 20.50T]$ for $k = 8$ and $N_{\text{el}} = 10$ (full domain). 48

1.21 Periodic planar surface in a parabolic domain. Surface elevation computed for $k = 3$ and $N_{\text{el}} = 50$ over the time intervals $[4T, 4.5T]$ (left) and $[4.5T, 5T]$ (right). 49
1.22 Periodic planar surface in a parabola − Surface elevation computed in the time interval $[4, 4.5T]$ for $k = 6$ and $N_{\text{el}} = 30$. 50
1.23 Periodic planar surface in a parabola − L^2-errors between numerical and analytical solutions and convergence rates for η at time $t = 0.6$ sec. 50
1.24 Vacuum occurrence wave over a step − Surface elevation computed during the time interval $[0.00, 8.00]$ for $k = 3$ and $N_{\text{el}} = 50$. 51
1.25 Vacuum occurrence wave over a step − Surface elevation computed from $t = 0$ sec to $t = T_{\max}$, for $k = 4$ and $N_{\text{el}} = 100$. 52

2.1 Main notations for Serre-Green-Naghdi equations. The free-surface elevation is denoted by η, ζ denoting the deviation from the rest state, the bathymetry is b and the water-height by H. The horizontal velocity is again denoted by u and the horizontal discharge by q. 54
2.2 Oscillating dam-break − Surface elevation computed for $k = 2$ and $N_{\text{el}} = 800$ at time $T_{\max} = 47.5$ sec (animation). 62
2.3 Solitary-waves over a piecewise-linear bottom − Time-series of the free-surface elevation at wave gauges locations, for the 4 experiments with increasing wave amplitudes. 63

3.1 Schematic representation of the 2D shallow-water flow configuration, illustrating the free-surface elevation $\eta(\boldsymbol{x}, t)$, the bathymetry $b(\boldsymbol{x})$, and the water-height $H(\boldsymbol{x}, t)$. 67
3.2 Two cases: subneighbor S_p inside cell ω_c (left), and subneighbor S_p inside neighbor cell ω_v (right). 69
3.3 Relation between the projectors π_h^k, $\overline{\pi}_h^k$ and its extension $\widetilde{\pi}_h^k$. 70
3.4 Partition of a mesh element ω_c into $N_s = N_k$ subcells for $\mathbb{P}^1$ (left), $\mathbb{P}^2$ (center) and $\mathbb{P}^3$ (right) cases. 71
3.5 Partition of a mesh element ω_c into $N_s > N_k$ subcells for $\mathbb{P}^1$ (left), $\mathbb{P}^2$ (center) and $\mathbb{P}^3$ (right) cases. 71
3.6 Representation of the two fluxes computed on each subinterface between two neighboring subcells S_m^c and S_p^v. The blending coefficient Θ_{mp} then allows to combine those two fluxes in a convex way to get the final flux $\widetilde{\mathbb{F}}_{mp}$ used in the monolithic scheme. 79
3.7 Steady vortex − Exact (left) and $\mathbb{P}^3$ numerical (right) height at final time $t = 1$ sec on 800 cells. 86
3.8 Steady vortex − Exact (left) and $\mathbb{P}^3$ numerical (right) discharge norm at final time $t = 1$ sec on 800 cells. 86
3.9 Steady vortex − L^2-errors between numerical and analytical solutions and convergence rates for η at time $t = 1$ sec. 87
3.10 Well-balancing − $\mathbb{P}^4$ solution at final time $t = 20$ sec − 2D free-surface elevation (left) and discharge norm (right). 87
3.11 Well-balancing − At $t = 20$ sec, from left to right and top to bottom: free-surface elevation, horizontal discharge, blending coefficient and vertical discharge density profiles. 88
3.12 Well-balancing with dry area − $\mathbb{P}^4$ solution on a domain slice ($y = 0.7$), located on the dry area. 88
3.13 Well-balancing with dry area − At $t = 20$ sec, from top to bottom: topography, free-surface elevation, discharge norm, and map of blending coefficient means per subcell. 89
3.14 Dam break on a wet bed − $\mathbb{P}^4$ free-surface elevation and horizontal discharge between $t = 0$ and $t = 32$ sec on the slice $y = 100$ (animation). 90
3.15 Dam break on a wet bed − At $t = 18$ sec, from top to bottom: exact elevation, unlimited DG scheme, monolithic DG/FV scheme and map of blending coefficient means per subcell (animation). 91
3.16 Dam break on a wet bed − $\mathbb{P}^4$ free-surface elevation and horizontal discharge density profiles at $t = 32$ sec (animation). 91
3.17 Dam break on a wet bed with friction − $\mathbb{P}^4$ free-surface elevation and blending density profiles at $t = 24$ sec for $n_f = 0.5$ (left), 2 (center) and 10 (right). 92
3.18 Dam break on a dry bed − Density profile of $\mathbb{P}^3$ solution on unstructured (left) and structured (right) meshes. 92
3.19 Dam break on a dry bed − At $t = 20$ sec, $\mathbb{P}^3$ free-surface elevation and blending map on unstructured mesh. 93
3.20 Dam break on a dry bed − At $t = 20$ sec, $\mathbb{P}^3$ free-surface elevation and blending map on structured mesh. 93
3.21 Dam break on a dry bed with friction − Snapshots of $\mathbb{P}^3$ free-surface elevation and blending density profiles for $t \in [10, 60]$ sec for $k_f = 0.5$. 94
3.22 Dam break on a dry bed with friction − At $t = 40$ sec, $\mathbb{P}^3$ free-surface elevation (top) and discharge norm (bottom). 94

3.23 Circular dam break on a wet bed – $\mathbb{P}^4$ free-surface elevation (top) and discharge norm (bottom) at $t = \frac{T_{\max}}{2}$ (left) and at $t = T_{\max}$ (right) (animation). 96
3.24 Circular dam-break on a wet-bed – $\mathbb{P}^4$ free-surface elevation at $t = T_{\max}$ on slice $y = 0$ (resp. $x = 0$) and horizontal (resp. vertical) discharge (animation). 97
3.25 Circular dam-break on a wet-bed – $\mathbb{P}^4$ free-surface elevation at $t = \frac{T_{\max}}{2}$ (left) and at $t = T_{\max}$ (right) on $\{(x, y) \in \Omega \mid y = \frac{\pi x}{3}$ and $x, y \geq 0\}$ (animation). 97
3.26 Circular dam break on a dry bed – $\mathbb{P}^4$ free-surface elevation (left) and blending coefficients (right) at $t = T_{\max}$ (right). 98
3.27 Circular dam break on a dry bed – $\mathbb{P}^4$ free-surface elevation (left) and blending coefficients (right) at $t = T_{\max}$ on slice $y = 0$. 98
3.28 Carrier & Greenspan transient solution – Snapshots of $\mathbb{P}^2$ elevation density profile on full domain (left) and zoom on dry bed (right) for $t \in [1, 21]$ sec. 100
3.29 Carrier & Greenspan transient solution – At $t = 10$ sec, exact elevation (top) and $\mathbb{P}^2$ numerical elevation (bottom). 100
3.30 Carrier & Greenspan transient solution – Snapshots of $\mathbb{P}^2$ horizontal discharge density profile for $t \in [1, 21]$ sec (left) and elevation density profile at final time (right). 100
3.31 Carrier & Greenspan periodic solution – Snapshots of $\mathbb{P}^2$ elevation density profile for $t \in [80, 86]$ sec (left) and for $t \in [154, 160]$ sec (right) (animation). 101
3.32 Carrier & Greenspan periodic solution – Snapshots of $\mathbb{P}^2$ free-surface elevation for $t \in [154, 160]$ sec (animation). 102
3.33 Single-wave on a conical island – Experimental setup : computational domain with island geometry and gauge locations. 102
3.34 Single-wave on a conical island – Numerical/experimental comparisons of the time-series of the surface-elevation at the different gauges for $t \in [2, 10]$ sec. 103
3.35 Tidal-wave overtopping the conical island – Snapshots of $\mathbb{P}^2$ free-surface elevation on slice $y = 12.5$ for $t \in [0, 12]$ sec. 104
3.36 Tidal-wave overtopping the conical island – Snapshots of $\mathbb{P}^2$ free-surface elevation (1^{st} and 3^{rd} line) and their corresponding blending map (2^{nd} and 4^{th} line) for $t \in [1, 12]$ sec (animation). 106
3.37 Single wave-rock interaction – Snapshots of $\mathbb{P}^6$ free-surface elevation on slice $x = 15$ for $t \in [0, 5]$ sec (animation). 107
3.38 Single wave-rock interaction – Snapshots of $\mathbb{P}^6$ free-surface elevation for $t \in [0, 5]$ sec (animation). 109
3.39 Tsunami with conical obstacles – Columnwise, from top to bottom: snapshots of $\mathbb{P}^2$ free-surface elevation on slice $y = 12.5$ for $t \in [0, 2]$ sec (animation). 110
3.40 Tsunami with conical obstacles – Snapshots of $\mathbb{P}^2$ water-height for $t \in [0, 2]$ sec, columnwise, from left to right (animation). 111

4.1 Evolution of the ALE mesh/submesh around an oscillating cylinder at five consecutive time steps. Left: standard Laplacian smoothing (animation). Right: pseudo-elasticity approach (animation). 127
4.2 Typical continuous piecewise-linear ($\mathbb{P}^1$) Lagrange basis function $\varphi_i(\boldsymbol{x})$ associated with the degree of freedom $\boldsymbol{x}_i$ on a two-dimensional triangular mesh $\mathcal{T}_h$. The function takes the value 1 at its associated node and vanishes at all other vertices, ensuring $\mathsf{W}_h \subset \mathsf{H}^1(\Omega)$. 128
4.3 Snapshots of the ALE mesh and submesh under the pseudo-one-dimensional explicit mapping over a complete deformation cycle. 131
4.4 ALE well-balancing – $\mathbb{P}^3$ solution at $t = 20$ sec, from left to right and top to bottom: free-surface elevation, horizontal discharge, blending coefficient and vertical discharge density profiles on the deformed mesh (animation). 141
4.5 Oscillating piston test case – At $t = 7.6$ sec, from left to right: free-surface elevation and horizontal discharge q_x (animation). 143
4.6 Oscillating piston test case – One-dimensional snapshots of the free-surface elevation (left column) and horizontal discharge (right column) during [2.5, 4] sec (animation). 144
4.7 Stopped piston with sloping beach – Snapshots of the ALE mesh deformation (left column) and free-surface elevation (right column) during the piston motion, for $t \in [0, 1.25]$ sec (animation). 145
4.8 Stopped piston with sloping beach – Two-dimensional snapshots of the wave propagation and run-up over the sloping beach, with map of blending coefficients, for $t \in [2.5, 7.5]$ sec (animation). 146
4.9 Stopped piston with sloping beach – Columnwise snapshots of the free-surface elevation density profile for $t \in [1.6, 7.6]$ sec (animation). 148
4.10 Oscillating cylindrical wavemaker (horizontal motion) – At final time: free-surface elevation (left) and discharge norm (right). 149

4.11 Oscillating cylindrical wavemaker (horizontal motion) − Two-dimensional snapshots of the free-surface elevation during the period $[0.5, 2.5]$ sec, ordered columnwise from top to bottom. 149
4.12 Oscillating cylindrical wavemaker (circular motion) − At final time, from left to right: free-surface elevation and blending map (animation). 150
4.13 Oscillating cylindrical wavemaker (circular motion) − Columnwise from top to bottom: two-dimensional snapshots of the free-surface elevation during the period $[0.5, 7.5]$ sec (animation). 151

5.1 Schematic representation of the one-dimensional wave-structure interactions problem. 156
5.2 Wave interacting with a fixed object and topography − Initial state (left) and final state at $t = 100$ sec (animation). 169
5.3 Wave interacting with a fixed object and topography − Time evolution of the free-surface elevation η during the period [0,10] sec, ordered columnwise from top to bottom (animation). 170
5.4 Unstructured meshes for $\Omega_{\mathrm{f}-}$ (left), Ω_{s} (center) and $\Omega_{\mathrm{f}+}$ (right) used during the simulation. 171
5.5 Wave generator (1D & 2D) − At time $t = 5$ s: comparison of the free-surface elevation η between the one-dimensional solver (left) and the two-dimensional solver (right). The 2D result is displayed as a density profile, gathering all the numerical values along the longitudinal direction (animation 1D & animation 2D). 171
5.6 Wave generator (1D) − Time evolution of the free-surface elevation η and the horizontal discharge q during the period $[2, 4]$ sec, ordered columnwise from top to bottom (animation). 172
5.7 Wave generator (2D) − Time evolution of the free-surface elevation η during the period $(2, 4)$ sec, ordered columnwise from top to bottom. The label displayed at the center of each panel indicates the vertical coordinate $(\mathcal{M}_O^G)_z$ of the center of mass of the floating object (animation 2D). 173
5.8 Wave generator (2D) − Columnwise from top to bottom: time evolution of the discharge norm during the period $(0, 2)$ sec. The label displayed at the center of each panel indicates the vertical coordinate $(\mathcal{M}_O^G)_z$ of the center of mass of the floating object (animation 2D). 174
5.9 Shock-wave interactions with a vertically freely heaving floating object − Elevation and discharge density profiles at the final time $t = 100$ sec. 175
5.10 Shock-wave interactions with a vertically freely heaving floating object − Time evolution of the free-surface elevation η during the period $[1.2, 5]$ sec, ordered columnwise from top to bottom (animation). 176
5.11 Shock-wave interactions with a vertically freely heaving floating object − Columnwise snapshots of the free-surface elevation density profile during the period $[0.1, 9]$ sec, ordered from top to bottom (animation). 177
5.12 Single-wave interactions with a freely heaving floating object − Unstructured meshes for $\Omega_{\mathrm{f}-}$ (left), Ω_{s} (center) and $\Omega_{\mathrm{f}+}$ (right) used during the simulation. 178
5.13 Single-wave interactions with a freely heaving floating object − Time evolution of the free-surface elevation η and the horizontal discharge q_x during the period $[0.4, 5.4]$ sec, ordered columnwise from top to bottom (animation). 179
5.14 Single-wave interactions with a freely heaving floating object − Elevation and discharge profiles at the final time $t = 100$ sec. 180
5.15 Single-wave interactions with a freely heaving floating object − Snapshots of the free-surface elevation (left column) and the discharge norm (right column) during the period $[0, 4.3]$ sec (animation). 181
5.16 Return to equilibrium (1D) − Comparison of the reduced analytical solution and the elliptic-Newton computation for the vertical displacement δ_O and velocity $\dot{\delta_O}$. 182
5.17 Return to equilibrium (1D) − Errors between the reduced analytical solution and the elliptic-Newton computation for δ_O and $\dot{\delta_O}$. 183
5.18 Return to equilibrium (2D) − Free-surface elevation profile η during the period $(0, 10]$ sec. Left column: initial upward displacement ${\delta_O}^0 = 3.5$ (upward animation). Right column: initial downward displacement ${\delta_O}^0 = -3.5$ (downward animation). 184

6.1 Geometry of a partially immersed cylindrical surface obstacle in a shallow-water flow. The free-surface is given by $z = \zeta^{\mathrm{f}}(\boldsymbol{x}, t)$, the wetted surface of the structure by $z = \zeta^{\mathrm{w}}(\boldsymbol{x})$, and the computational domain is split into an exterior region Ω_{f} and an interior region Ω_{s} beneath the structure. 186
6.2 Geometry of a partially immersed cylindrical floating structure in a shallow-water flow. The structure is an infinite pontoon with rectangular cross section, and the fluid domain is the union of two half-planes. 190
6.3 Schematic representation of an HHO solution between two adjacent mesh cells ω_c and ω_v. 198

6.4 Wave-pontoon interactions − Comparison of the transmitted discharge computed with the analytical formula, the FFT-based DtN operator, and the HHO approximation, together with the pairwise L^∞ errors between these three evaluations. 207
6.5 Wave-pontoon interactions − Time evolution of the free-surface elevation η during the period $(0, 5]$ sec, ordered columnwise from top to bottom (animation). 208
6.6 Wave-pontoon interactions − Columnwise, from top to bottom: time evolution of free-surface elevation η and horizontal discharge q_x at several times in the period $(0, 5]$ sec (animation). 209
6.7 Wave-pontoon interactions − Time series of the pairwise L^∞ errors between the analytical formula, the FFT-based DtN operator and the HHO approximation. 211
6.8 Unstructured meshes for Ω_f ($N_\mathrm{el} = 2340$) and Ω_s ($N_\mathrm{el} = 80$). 212
6.9 Single-wave interactions with a fixed cylinder − Free-surface elevation and blending coefficients at several times during the period $[2.15, 5.45]$ sec, ordered columnwise from top to bottom (animation). 213
6.10 Single-wave interactions with a fixed cylinder over a sloping beach − Free surface elevation and blending coefficients during the period $[0, 10]$ sec on slice $y = 0$ (animation). 214
6.11 Single-wave interactions with a fixed cylinder over a sloping beach − Time evolution during the period $[2, 9]$ sec of the free-surface elevation (left) and discharge norm (right) (animation). 215
6.12 Single-wave interactions with a fixed cylinder over a sloping beach − Free surface elevation density profiles during the period $[0, 10]$ sec (animation). 216
6.13 Geometry of the vertically moving cylindrical surface obstacle. The dotted contour represents the initial position $\mathbf{S}(0)$, while the solid contour represents $\mathbf{S}(t)$ after the vertical translation $\delta_O^y(t)$. The horizontal regions Ω_f and Ω_s and the contact-line projection $\partial\Omega_\mathrm{fs}$ remain fixed. 218
6.14 Pseudo-one-dimensional wave–structure interaction problem with surge, heave and pitch of the floating pontoon. 219

A.1 The `unstructured_mesh` object and its topological and geometrical data. 226
A.2 Reference quadrature nodes (left) mapped to a physical element (right) through the affine mapping $\boldsymbol{x} = \chi_c(\boldsymbol{\xi})$. Edge nodes are shown in red and interior nodes in blue. 227
A.3 The `quadrature` object, split into element and face data. 228
A.4 The `basis_functions` object: storage of basis functions and derivatives at quadrature nodes. 229
A.5 The `submesh` object extends the unstructured mesh with dedicated structures (`subElement`, `subFace`), connectivity operators, and global data. 233
A.6 Examples of reference-element subdivisions used in the FV-subcell framework. Left: division into triangular subcells in the case $N_s \geq N_k$, with the corresponding quadrature points for subcell interiors and faces. Right: division into both triangular and quadrangular subcells in the case $N_s = N_k$, again with the associated subcell quadrature points. One of my thesis advisors should be quite happy to see that I managed to make good use of these figures in my thesis. 234
A.7 Extension of the mesh object to support dynamic ALE updates. 235

Bibliography

[1] MB Abbott, HM Petersen, and O Skovgaard, *On the numerical modelling of short waves in shallow water*, Journal of Hydraulic Research **16** (1978), no. 3, 173–204.

[2] Rémi Abgrall, *Some remarks about conservation for residual distribution schemes*, Computational Methods in Applied Mathematics **18** (2018), no. 3, 327–351.

[3] V. Aizinger and C. Dawson, *A discontinuous Galerkin method for two-dimensional flow and transport in shallow water*, Advances in Water Resources **25** (2002), 67–84.

[4] F. Alcrudo and P. Garcia-Navarro, *A high-resolution godunov-type scheme in finite volumes for the 2d shallow-water equations*, Internat. J. Numer. Methods Fluids (1993).

[5] Y. Allaneau and A. Jameson, *Connections between the filtered discontinuous Galerkin method and the flux reconstruction approach to high order discretizations*, Comput. Meth. Appl. Mech. Engrg. **200** (2011), 3628–3636.

[6] V.R. Ambati and O. Bokhove, *Space–time discontinuous galerkin finite element method for shallow water flows*, Journal of Computational and Applied Mathematics **204** (2007), 452–462.

[7] C.J. Amick, L.E. Fraenkel, and J.F. Toland, *On the stokes conjecture for the wave of extreme form*, Acta Mathematica **148** (1982), 193–214.

[8] K. Anastasiou and C.T. Chan, *Solution of the 2d shallow water equations using the finite volume method on unstructured triangular meshes*, Int J Numer Methods Fluids **24** (1997), 1225–1245.

[9] Douglas N. Arnold, *An interior penalty finite element method with discontinuous elements*, SIAM Journal on Numerical Analysis **19** (1982), no. 4, 742–760.

[10] L. Arpia and M. Ricchiuto, *Well balanced residual distribution for the ALE spherical shallow water equations on moving adaptive meshes*, J. Comput. Phys. **405** (2019), 109–173.

[11] E Audusse, F Bouchut, M.-O Bristeau, R Klein, and B Perthame, *A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows*, SIAM Journal on Scientific Computing **25** (2004), no. 6, 2050–2065.

[12] E. Audusse, F. Bouchut, M.-O. Bristeau, R. Klein, and B. Perthame, *A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows*, SIAM J. Sci. Comput. **25** (2004), no. 6, 2050–2065.

[13] A. Babarit, *On the park effect in arrays of oscillating wave energy converters*, Renewable Energy **58** (2013), 68–78.

[14] Ivo Babuška, *The finite element method with penalty*, Mathematics of Computation **27** (1973), no. 122, 221–228.

[15] Ivo Babuška and Miloš Zlámal, *Nonconforming elements in the finite element method with penalty*, SIAM Journal on Numerical Analysis **10** (1973), no. 5, 863–875.

[16] Garth A. Baker, *Finite element methods for elliptic equations using nonconforming elements*, Mathematics of Computation **31** (1977), no. 137, 45–59.

[17] D. Balsara, C. Altmann, C.D. Munz, and M. Dumbser, *A sub-cell based indicator for troubled zones in RKDG schemes and a novel class of hybrid RKDG+HWENO schemes*, J. Comp. Phys. **226** (2007), 586–620.

[18] S.R.M. Barros and J.W. Cardenas, *A nonlinear galerkin method for the shallow-water equations on periodic domains*, J. Comput. Phys. **172** (2001), 592–608.

[19] G. Beck and D. Lannes, *Freely floating objects on a fluid governed by the boussinesq equations*, Annales de l'Institut Henri Poincaré C **39** (2022), no. 3, 575–646.

[20] Geoffrey Beck, David Lannes, and Lisl Weynans, *A numerical method for wave-structure interactions in the Boussinesq regime*, ESAIM: Mathematical Modelling and Numerical Analysis **59** (2025), no. 6, 2895–2931.

[21] C. Beels, P. Troch, K. De Visch, J.P. Kofoed, and G De Backer, *Application of the time-dependent mild-slope equations for the simulation of wake effects in the lee of a farm of wave dragon wave energy converters*, Renewable Energ **35** (2010), no. 8, 1644–1661.

[22] A. Bermudez, A. Dervieux, J.-A. Desideri, and M.E. Vazquez, *Upwind schemes for the two-dimensional shallow water equations with variable depth using unstructured meshes*, Comput. Methods Appl. Mech. Engrg. **155** (1998), 49–72.

[23] Alfredo Bermudez and Ma Elena Vazquez, *Upwind methods for hyperbolic conservation laws with source terms*, Computers & Fluids **23** (1994), no. 8, 1049–1071.

[24] P.-E. Bernard, J.-F. Remacle, R. Comblen, V. Legat, and K. Hillewaert, *High-order discontinuous galerkin schemes on general 2d manifolds applied to the shallow water equations*, J. Comput. Phys. **228** (2009), 6514–6535.

[25] Markus Berndt, Jérôme Breil, Stéphane Galera, Milan Kucharik, Pierre-Henri Maire, and Mikhail Shashkov, *Two-step hybrid conservative remapping for multimaterial arbitrary lagrangian–eulerian methods*, Journal of Computational Physics **230** (2011), no. 17, 6664–6687.

[26] C. Berthon, *Robustness of MUSCL schemes for 2D unstructured meshes*, J. Comput. Phys. **218** (2006), no. 2, 495–509.

[27] C Berthon, *Why the muscl-hancock scheme is l1-stable*, Numer. Math. **104** (2006), no. 1, 27–46.

[28] C. Berthon and V. Desveaux, *An entropy preserving mood scheme for the euler equations*, to appear in J. Finite Volumes (2014).

[29] C. Berthon and F. Marche, *A positive preserving high order VFRoe scheme for shallow water equations: a class of relaxation schemes*, SIAM J. Sci. Comput. **30** (2008), no. 5, 2587–2612.

[30] C. Berthon, F. Marche, and R. Turpault, *An efficient scheme on wet/dry transitions for shallow water equations with friction*, Comput. Fluids **48** (2011), no. 1, 192 – 201.

[31] C. Berthon, M. M'Baye, M.H. Le, and D. Seck, *A well-defined moving steady states capturing godunov-type scheme for shallow-water model*, Int. J. Finite Volume **15** (2020).

[32] Christophe Berthon, *Stability of the muscl schemes for the euler equations*, Commun. Math. Sci. **3** (2005), no. 2, 133–157.

[33] Rupak Biswas, Karen D. Devine, and Joseph E. Flaherty, *Parallel, adaptive finite element methods for conservation laws*, Applied Numerical Mathematics **14** (1994), 255 – 283.

[34] E Bocchi, *Floating structures in shallow water: local well-posedness in the axisymmetric case*, SIAM Journal on Mathematical Analysis **52** (2020), no. 1, 306–339.

[35] E. Bocchi, J. He, and G. Vergara-Hermosilla, *Modelling and simulation of a wave energy convertor*, A venir, 2020.

[36] P. Bonneton, E. Barthélemy, F. Chazel, R. Cienfuegos, D. Lannes, F. Marche, and M. Tissier, *Recent advances in serre–green naghdi modelling for wave transformation, breaking and runup processes*, European Journal of Mechanics-B/Fluids **30** (2011), no. 6, 589–597.

[37] P. Bonneton, F. Chazel, D. Lannes, F. Marche, and M. Tissier, *A splitting approach for the fully nonlinear and weakly dispersive Green-Naghdi model*, Journal of Computational Physics **230** (2011), no. 4, 1479–1498.

[38] Jay P. Boris and David L. Book, *Flux-Corrected Transport. i. SHASTA, a fluid transport algorithm that works*, Journal of Computational Physics **11** (1973), no. 1, 38–69.

[39] W. Boscheri, R. Loubere, and M. Dumbser, *Direct arbitrary-lagrangian–eulerian ader-mood finite volume schemes for multidimensional hyperbolic conservation laws*, Journal of Computational Physics **292** (2015), 56–87.

[40] Walter Boscheri, Matteo Semplice, and Michael Dumbser, *Central WENO subcell finite volume limiters for ADER discontinuous galerkin schemes on fixed and moving unstructured meshes*, Communications in Computational Physics **25** (2019), 311–346.

[41] U Bosi, A.-P Engsig-Karup, C Eskilsson, and M Ricchiuto, *A spectral/hp element depth-integrated model for nonlinear wave–body interaction*, Comput. Methods Appl. Mech. Engrg **348** (2019), 222–249.

[42] F. Bouchut, J. Le Sommer, 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), 35–63.

[43] J. Boussinesq, *Théorie des ondes et des remous qui se propagent le long d'un canal rectangulaire horizontal, en communiquant au liquide contenu dans ce canal des vitesses sensiblement pareilles de la surface au fond.*, Journal de Mathématiques Pures et Appliquées (1872), 55–108 (fre).

[44] K Budal, *Theory for absorption of wave power by a system of interacting bodies*, Journal of Ship Research **21** (1977), no. 4.

[45] A. Burbeau, P. Sagaut, and C.-H. Bruneau, *A problem-independent limiter for high-order Runge-Kutta discontinuous Galerkin methods*, J. Comput. Phys. **169** (2001), no. 1, 111 – 150.

[46] S Busto, M Dumbser, C Escalante, N Favrie, and S Gavrilyuk, *On high order ader discontinuous galerkin schemes for first order hyperbolic reformulations of nonlinear dispersive systems*, Journal of Scientific Computing **87** (2021), no. 2, 1–47.

[47] S Busto, M Dumbser, S Gavrilyuk, and K Ivanova, *On thermodynamically compatible finite volume methods and path-conservative ader discontinuous galerkin schemes for turbulent shallow water flows*, Journal of Scientific Computing **88** (2021), no. 1, 1–45.

[48] S. Camarri, M.V. Salvetti, B. Koobus, and A. Dervieux, *A low-diffusion MUSCL scheme for LES on unstructured grids*, Computers and Fluids **33** (2004), no. 9, 1101–1129.

[49] A. Canestrelli, A. Siviglia, M. Dumbser, and E.F. Toro, *Well-balanced high-order centred schemes for non-conservative hyperbolic systems. applications to shallow water equations with fixed and mobile bed*, Advances in Water Resources **32** (2009), no. 6, 634–644.

[50] Alberto Canestrelli, Michael Dumbser, Annunziato Siviglia, and Eleuterio F. Toro, *Well-balanced high-order centered schemes on unstructured meshes for shallow water equations with fixed and mobile bed*, Advances in Water Resources **33** (2010), no. 3, 291–303.

[51] S. Cardonna, A. Haidar, F. Marche, and F. Vilar, *Local subcell monolithic dg/fv methods for nonlinear shallow-water models with source terms*, (submitted) (2026).

[52] S. Cardonna, F. Marche, and F. Vilar, *A high-order robust subcell monolithic dg/fv formulation for nonlinear shallow-water equations on unstructured grids*, (submitted) (2026).

[53] Mark H. Carpenter, Travis C. Fisher, Eric J. Nielsen, and Steven H. Frankel, *Entropy stable spectral collocation schemes for the Navier–Stokes equations: Discontinuous interfaces*, SIAM Journal on Scientific Computing **36** (2014), no. 5, B835–B867.

[54] G. Carrier and H. Greenspan, *Water waves of finite amplitude on a sloping beach*, Journal of Fluid Mechanics **2** (1958), 97—109.

[55] E. Casoni, J. Peraire, and A. Huerta, *One-dimensional shock-capturing for high-order discontinuous Galerkin methods*, Int. J. Numer. Meth. Fluids **71** (2013), 737–755.

[56] A. Castro and D. Lannes, *Fully nonlinear long-wave models in the presence of vorticity*, Journal of Fluid Mechanics **759** (2014), 642–675.

[57] ———, *Fully nonlinear long-waves models in presence of vorticity*, J. Fluid Mech. **759** (2014), 642–675.

[58] Manuel J. Castro Díaz, Enrique D. Fernández-Nieto, and Ana M. Ferreiro, *Sediment transport models in shallow water equations and numerical approach by high order finite volume methods*, Computers & Fluids **37** (2008), no. 3, 299–316.

[59] M.J. Castro Diaz, J.A. Lopez-Garcia, and C. Parès, *High order exactly well-balanced numerical methods for shallow water systems*, J. Comput. Phys. **246** (2013), 242–264.

[60] Vincenzo Casulli, *Semi-implicit finite difference methods for the two-dimensional shallow water equations*, Journal of Computational Physics **86** (1990), no. 1, 56–74.

[61] P. Caussignac and R. Touzani, *Solution of three-dimensional boundary layer equations by a discontinuous finite element method, part i: Numerical analysis of a linear model problem*, Computer methods in applied mechanics and engineering **78** (1990), no. 3, 249–271.

[62] ———, *Solution of three-dimensional boundary layer equations by a discontinuous finite element method, part ii: Implementation and numerical results*, Computer methods in applied mechanics and engineering **79** (1990), no. 1, 1–20.

[63] L. Cea and M.E. Vasquez-Cendon, *Unstructured finite volume discretisation of bed friction and convective flux in solute transport models linked to the shallow water equations*, J. Comput. Phys. (2012).

[64] C. Chalons, F. Coquel, E. Godlewski, P.A. Raviart, and N. Seguin, *Godunov-type schemes for hyperbolic systems with parameter dependent source. The case of Euler system with friction.*, Math. Models and Methods in Appl. Sci. **20** (2010), no. 11, 2109–2166.

[65] G. Chavent and B. Cockburn, *The local projection-discontinuous-galerkin finite element method for scalar conservation laws*, ESAIM: Mathematical Modelling and Numerical Analysis **23** (1989), no. 4, 565–592.

[66] X.N. Chen and S.D. Sharma, *A slender ship moving at a near-critical speed in a shallow channel*, J. Fluid Mech. **291** (1995), 263–285.

[67] A. Chertock, S. Cui, A. Kurganov, and T. Wu, *Well-balanced positivity preserving central-upwind scheme for the shallow water system with friction terms*, Int J Numer Methods Fluids **78** (2015), 355–383.

[68] P.H. Chiu, L. Lee, and T.W.H. Sheu, *A dispersion-relation-preserving algorithm for a nonlinear shallow-water wave equation*, J. Comput. Phys. **228** (2009), 8034–8052.

[69] Mirco Ciallella, Lorenzo Micalizzi, Philipp Öffner, and Davide Torlo, *An arbitrary high order and positivity preserving method for the shallow water equations*, Computers & Fluids **247** (2022), 105630.

[70] S. Clain, S. Diot, and R. Loubère, *A high-order finite volume method for systems of conservation laws—multi-dimensional optimal order detection (mood)*, J. Comput. Phys. **230** (2011), 4028–4050.

[71] B. Cockburn, S. Hou, and C.-W. Shu, *The Runge-Kutta Discontinuous Galerkin Method for Conservation Laws V: Multidimensional Systems*, J. Comp. Phys. **141** (1998), 199–224.

[72] B. Cockburn and C.-W. Shu, *Tvb runge-kutta local projection discontinuous galerkin finite element method for conservation laws. ii. general framework*, Mathematics of computation **52** (1989), no. 186, 411–435.

[73] ———, *The runge-kutta local projection-discontinuous-galerkin finite element method for scalar conservation laws*, ESAIM: Mathematical Modelling and Numerical Analysis **25** (1991), no. 3, 337–361.

[74] Stéphane Cordier, Minh Hoang Le, and Tomás Morales de Luna, *Bedload transport in shallow water models: Why splitting (may) fail, how hyperbolicity (can) help*, Advances in Water Resources **34** (2011), no. 8, 980–989.

[75] M. de la Asuncion and M.J. Castro, *Simulation of tsunamis generated by landslides using adaptive mesh refinement on GPU*, J. Comput. Phys. **345** (2017), 91–110.

[76] J. N. de la Rosa and C. D. Munz, *Hybrid DG/FV schemes for magnetohydrodynamics and relativistic hydrodynamics*, Comp. Phys. Commun. **222** (2018), 113–135.

[77] A.J.-C. de Saint-Venant, *Théorie du mouvement non-permanent des eaux, avec application aux crues des rivières et à l'introduction des marées dans leur lit*, C.R. Acad. Sci. Paris, Section Mécanique **73** (1871), 147–154.

[78] O. Delestre, C. Lucas, P.-A. Ksinant, F. Darboux, C. Laguerre, T.-N. Tuoi Vo, F. James, and S. Cordier, *SWASHES: a compilation of shallow water analytic solutions for hydraulic and environmental studies*, Internat. J. Numer. Methods Fluids **72** (2012), no. 3, 269–300.

[79] O Delestre and F Marche, *A numerical scheme for a viscous shallow water model with friction*, Journal of Scientific Computing **48** (2011), no. 1, 41–51.

[80] O. Delestre and F. Marche, *A numerical scheme for a viscous shallow water model with friction*, J. Sci. Comput. **48** (2011), no. 1-3, 41–51.

[81] D. Di Pietro and A. Ern, *Hybrid high-order methods for variable-diffusion problems on general meshes*, C.R. Acad. Sci. Paris, Ser. I **353** (2015), 31–34.

[82] D. Di Pietro, A. Ern, and J.-L. Guermond, *Discontinuous Galerkin methods for anisotropic semidefinite diffusion with advection*, SIAM J. Numer. Anal. **46** (2008), no. 2, 805–831.

[83] D. A. Di Pietro and A. Ern, *Discrete functional analysis tools for discontinuous Galerkin methods with application to the incompressible Navier-Stokes equations*, Math. Comp. **79** (2010), no. 271, 1303–1330.

[84] ———, *Mathematical aspects of discontinuous galerkin methods*, Mathématiques and Applications, vol. 69, Springer, 2012.

[85] D. A. Di Pietro and F. Marche, *Weighted interior penalty discretization of fully nonlinear and weakly dispersive free surface shallow water flows*, J. Comput. Phys. **355** (2018), 285–309.

[86] Daniele A. Di Pietro and Jérôme Droniou, *The Hybrid High-Order method for polytopal meshes: Design, analysis, and applications*, Modeling, Simulation and Applications, vol. 19, Springer International Publishing, 2020.

[87] A.S. Dimakopoulos, A. Guercio, and G. Cuomo, *Advanced numerical modelling of tsunami wave propagation, transformation and run-up*, preprint (2019).

[88] S. Diot, S. Clain, and R. Loubère, *Improved detection criteria for the multi-dimensional optimal order detection (MOOD) on unstructured meshes with very high-order polynomials*, Computers and Fluids **64** (2012), 43–63.

[89] S. Diot, R. Loubère, and S. Clain, *The MOOD method in the three-dimensional case: very-high-order finite volume method for hyperbolic systems*, Int. J. Numer. Meth. Fluids **73** (2013), 362–392.

[90] D.Kuzmin, *Slope limiting for discontinuous galerkin approximations with a possibly non-orthogonal taylor basis*, Int J Numer Methods Fluids **71** (2013), no. 9, 1178–1190.

[91] J. Donea, S. Giuliani, and J. P. Halleux, *An arbitrary lagrangian-eulerian finite element method for transient dynamic fluid-structure interactions*, Computer Methods in Applied Mechanics and Engineering **33** (1982), no. 1–3, 689–723.

[92] J. Donea, A. Huerta, J.-P. Ponthot, and A. Rodríguez-Ferran, *Arbitrary lagrangian-eulerian methods*, Encyclopedia of Computational Mechanics (E. Stein, R. de Borst, and T. J. R. Hughes, eds.), John Wiley & Sons, 2004, pp. 413–437.

[93] J. Donea, A. Huerta, J.-Ph. Ponthot, and A. Rodríguez-Ferran, *Arbitrary lagrangian–eulerian methods, the encyclopedia of computational mechanics*, pp. 413–437, Wiley, 2004.

[94] Jim Douglas and Todd Dupont, *Interior penalty procedures for elliptic and parabolic galerkin methods*, Computing Methods in Applied Sciences, Lecture Notes in Physics, vol. 58, Springer, Berlin, Heidelberg, 1976, pp. 207–216.

[95] Jérôme Droniou, Marien Hanot, and Todd Oliynyk, *A polytopal discrete de Rham complex on manifolds, with application to the Maxwell equations*, Journal of Computational Physics **527** (2025), 113886.

[96] M. Dryja, *On discontinuous Galerkin methods for elliptic problems with discontinuous coefficients*, Comput. Methods Appl. Math. **3** (2003), no. 1, 76–85.

[97] J. Du, C.-W. Shu, and M. Zhang, *A simple weighted essentially non-oscillatory limiter for the correction procedure via reconstruction (CPR) framework*, Appl. Num. Math. **95** (2015), 173–198.

[98] F. Duarte, R. Gormaz, and S. Natesan, *Arbitrary lagrangian-eulerian method for navier-stokes equations with moving boundaries*, Computer Methods in Applied Mechanics and Engineering **193** (2004), no. 45–47, 4819–4836.

[99] V. Duchêne, S. Israwi, and R. Talhouk, *A new green-naghdi model in the camassa-holm regime and full justification of asymptotic models for the propagation of internal waves*, arXiv preprint arXiv:1304.4554 (2013).

[100] Vincent Duchêne, Samer Israwi, and Raafat Talhouk, *A new class of two-layer green–naghdi systems with improved frequency dispersion*, Studies in Applied Mathematics **137** (2016), no. 3, 356–415.

[101] M. Dumbser and R. Loubère, *A simple robust and accurate a posteriorisub-cell finite volume limiter for the discontinuousGalerkin method on unstructured meshes*, J. Comp. Phys. **319** (2016), 163–199.

[102] M. Dumbser, A. Uuriintsetseg, and O. Zanotti, *On arbitrary-lagrangian-eulerian one-step weno schemes for stiff hyperbolic balance laws*, Communications in Computational Physics **14** (2013), no. 2, 301–327.

[103] M. Dumbser, O. Zanotti, R. Loubère, and S. Diot, *A posteriori subcell limiting of the discontinuous Galerkin finite element method for hyperbolic conservation laws*, J. Comput. Phys. **278** (2014), 47–75.

[104] Michael Dumbser, Dinshaw S. Balsara, Eleuterio F. Toro, and Claus-Dieter Munz, *A unified framework for the construction of one-step finite volume and discontinuous galerkin schemes on unstructured meshes*, Journal of Computational Physics **227** (2008), no. 18, 8209–8253.

[105] Michael Dumbser and Raphaël Loubère, *A simple robust and accurate a posteriori sub-cell finite volume limiter for the discontinuous galerkin method on unstructured meshes*, Journal of Computational Physics **319** (2016), 163–199.

[106] A. Duran, Q. Liang, and F. Marche, *On the well-balanced numerical discretization of shallow water equations on unstructured meshes*, J. Comput. Phys. **235** (2013), 565–586.

[107] A. Duran and F. Marche, *Recent advances on the discontinuous Galerkin method for shallow water equations with topography source terms*, Comput. Fluids **101** (2014), 88–104.

[108] A. Duran, F. Marche, C. Berthon, and R. Turpault, *Numerical discretizations for shallow water equations with source terms on unstructured meshes*, Proceedings of the Fourtheenth International Conference on Hyperbolic Problems: Theory Numerics and Applications - HYP 2012, Padova (AIMS Series on Applied Mathematics, ed.), vol. 8, 2014, pp. 541–549.

[109] ———, *Asymptotic preserving scheme for the shallow water equations with source terms on unstructured meshes*, J. Comput. Phys. **287** (2015), no. 184–206.

[110] G.A. El, R.H.J. Grimshaw, and N.F. Smyth, *Unsteady undular bores in fully nonlinear shallow-water theory,*, Phys. Fluids **18** (2006).

[111] K.S. Erduran, V. Kutija, and C.J.M. Hewett, *Performance of finite volume solutions to the shallow water equations with shock-capturing schemes*, Int J Numer Methods Fluids **40** (2002), 1237–1273.

[112] C. Eskilsson and S.J.Sherwin, *Discontinuous galerkin spectral/hp element modelling of dispersive shallow water systems*, J. Sci. Comput. **22** (2005), 269–288.

[113] C Farhat, P Geuzaine, and C Grandmont, *The discrete geometric conservation law and the nonlinear stability of ale schemes for the solution of flow problems on moving grids*, J. Comput. Phys. **174** (2001), no. 2, 669–694.

[114] M. Feistauer, V. Dolejší, and V. Kučera, *On the discontinuous Galerkin method for the simulation of compressible flow with wide range of Mach numbers*, Computing and Visualization in Science **10** (2007), no. 1, 17–27.

[115] Travis C. Fisher and Mark H. Carpenter, *High-order entropy stable finite difference schemes for nonlinear conservation laws: Finite domains*, Journal of Computational Physics **252** (2013), 518–557.

[116] M. Folley, A. Babarit, B. Child, D. Forehand, L. O'Boyle, K Silverthorne, J. Spinneken, V. Stratigaki, and P. Troch, *A review of numerical modelling of wave energy converter arrays.*, Proc. Of the ASME 2012 31th International Conference on Ocean, Offshore and Arctic Engineering, 2012.

[117] L. Fraccarollo and E.F. Toro, *Experimental and numerical assessment of the shallow water model for two-dimensional dam-break type problems*, J. Hydraulic Res. **33** (1995), no. 6, 843–863.

[118] Elena Gaburro, Walter Boscheri, Simone Chiocchetti, and Mario Ricchiuto, *Discontinuous galerkin schemes for hyperbolic systems in non-conservative variables: Quasi-conservative formulation with subcell finite volume corrections*, Computer Methods in Applied Mechanics and Engineering **431** (2024), 117311.

[119] J.M. Gallardo, C. Parés, and M Castro, *On a well-balanced high-order finite volume scheme for shallow water equations with topography and dry areas*, J. Comput. Phys. **227** (2007), no. 1, 574–601.

[120] T. Gallouët, J.M. Hérard, and N. Seguin, *Some approximate godunov schemes to compute shallow-water equations with topography*, Computers and Fluids (2003).

[121] H. Gao and Z. J. Wang, *A conservative correction procedure via reconstruction formulation with the Chain-Rule divergence evaluation*, J. Comp. Phys. **232** (2013), 7–13.

[122] S. Gavrilyuk, B. Nkonga, K.-M. Shyue, and L. Truskinovsky, *Generalized Riemann problem for dispersive equations*, preprint - HAL Id : hal-01958328 (2019).

[123] F.X. Giraldo, J.S. Hesthaven, and T. Warburton, *Nodal high-order discontinuous galerkin methods for the spherical shallow water equations*, J. Comput. Phys. **181** (2002), no. 2, 499 – 525.

[124] S. Giuliani, *An algorithm for continuous rezoning of the hydrodynamic grid in arbitrary lagrangian-eulerian computer codes*, Nuclear Engineering and Design **72** (1982), no. 2, 205–212.

[125] E. Godlewski, M. Parisot, J. Sainte-Marie, and F. Wahl, *Congested shallow water model: on floating body*, SMAI Journal of Computational Mathematics **in press** (2022).

[126] S. Gottlieb, C.-W. Shu, and Tadmor E., *Strong stability preserving high order time discretization methods*, SIAM Review **43** (2001), 89–112.

[127] N Goutal, *Proceedings of the 2nd workshop on dam-break wave simulation*, Department Laboratoire National d'Hydraulique, Groupe Hydraulique Fluviale, 1997.

[128] A.E Green and P.M. Naghdi, *A derivation of equations for wave propagation in water of variable depth*, Journal of Fluid Mechanics **78** (1976), no. 2, 237–246.

[129] J. Grue, E.N. Pelinovsky, D. Fructus, T. Talipova, and C. Kharif, *Formation of undular bores and solitary waves in the strait of malacca caused by the 26 december 2004 indian ocean tsunami*, J. Geophys. Res. **113** (2008).

[130] G. Guennebaud and B. Jacob, *Eigen v3*, http://eigen.tuxfamily.org, 2010.

[131] J.-L. Guermond, R. Pasquetti, and B. Popov, *Entropy viscosity method for nonlinear conservation laws*, J. Comput. Phys. **230** (2011), 4248–4267.

[132] J.-L. Guermond, B. Popov, E. Tovar, and C. Kees, *Robust explicit relaxation technique for solving the Green-Naghdi equations*, J. Comput. Phys. (2019).

[133] Jean-Luc Guermond, Bojan Popov, and Ignacio Tomas, *Invariant domain preserving discretization-independent schemes and convex limiting for hyperbolic systems*, Computer Methods in Applied Mechanics and Engineering **347** (2019), 143–175.

[134] M. Guerra, *Numerical and experimental modeling of extreme flood waves and inundation zone*, Ph.D. thesis, Pontificia Universidad Catolica de Chile, 2010.

[135] S. Guibourg, *Modélisation numérique et expérimentale des houles bidimensionnelles en zone cotière,*, PhD Thesis, Université Joseph Fourier-Grenoble I, 1994.

[136] H. Guillard and C. Farhat, *On the significance of the geometric conservation law for flow computations on moving meshes*, Comput. Methods Appl. Engrg **190** (2000), 1467–1482.

[137] Colette Guillopé, Samer Israwi, and Raafat Talhouk, *Large-time existence for one-dimensional green-naghdi equations with vorticity*, Discrete & Continuous Dynamical Systems-S **14** (2021), no. 8, 2947.

[138] P. Guyenne, *A high-order spectral method for nonlinear water waves in the presence of a linear shear current*, Computers and Fluids **154** (2017), 224–235.

[139] C. Hague and C. Swan, *A multiple flux boundary element method applied to the description of surface water waves*, J. Comput. Phys. **228** (2009), no. 14, 5111–5128.

[140] A. Haidar, F. Marche, and F. Vilar, *A posteriori finite-volume local subcell correction of high-order discontinuous Galerkin schemes for the nonlinear shallow-water equations*, J. Comput. Phys. **452** (2022), 110902.

[141] ———, *Free-boundary problems for wave-structure interactions in shallow-water: Dg-ale description and local subcell correction*, J. Sci. Comput. **98** (2024), no. 45.

[142] ———, *An ALE numerical method for the waterline evolution of a floating object with non-linear shallow-water equations*, Journal of Scientific Computing **102** (2025), no. 2, 53.

[143] R. Harris, Z.J. Wang, and Y. Liu, *Efficient Quadrature-Free High-Order Spectral Volume Method on Unstructured Grids: Theory and 2D Implementation*, J. Comp. Phys. **227** (2008), 1620–1642.

[144] R Henderson, *Design, simulation, and testing of a novel hydraulic power take-off system for the pelamis wave energy converter*, Renewable energy **31** (2006), no. 2, 271–283.

[145] Joakim Hermansson and Peter Hansbo, *A variable diffusion method for mesh smoothing*, Communications in Numerical Methods in Engineering **19** (2003), no. 11, 897–908.

[146] C. W. Hirt, A. A. Amsden, and J. L. Cook, *An arbitrary lagrangian-eulerian computing method for all flow speeds*, Journal of Computational Physics **14** (1974), no. 3, 227–253.

[147] J. Horillo, Z. Kowalik, and Y. Shigihara, *Wave dispersion study in the indian ocean tsunami of december 26*, Marine Geodesy **29** (2006), no. 1, 149–166.

[148] A. Huerta, E. Casoni, and J. Peraire, *A simple shock-capturing technique for high-order discontinuous galerkin methods*, Int. J. Numer. Meth. Fluids **69** (2012), 1614–1632.

[149] T. J. R. Hughes, W. K. Liu, and T. K. Zimmermann, *Lagrangian-eulerian finite element formulation for incompressible viscous flows*, Computer Methods in Applied Mechanics and Engineering **29** (1981), no. 3, 329–349.

[150] H. T. Huynh, Z. J. Wang, and P. E. Vincent, *High-order methods for computational fluid dynamics: A brief review of compact differential formulations on unstructured grids*, J. Comp. Phys. **98** (2014), 209–220.

[151] Hung T Huynh, *A flux reconstruction approach to high-order schemes including discontinuous galerkin methods*, 18th AIAA computational fluid dynamics conference, 2007, p. 4079.

[152] Tatsuo Iguchi and David Lannes, *The moving contact line problem for the* 2*d nonlinear shallow water equations*, 2025.

[153] M. Ioriatti and M. Dumbser, *A posteriori sub-cell finite volume limiting of staggered semi-implicit discontinuous Galerkin schemes for the shallow water equations*, Applied Numerical Mathematics **135** (2019), 443–480.

[154] M. Iskandarani, D. B. Haidvogel, and J. P. Boyd, *A staggered spectral element model with application to the oceanic shallow water equations*, Int J Numer Methods Fluids **20** (1995), no. 5, 393–414.

[155] S. Israwi, *Large time existence for 1D Green-Naghdi equations*, Nonlinear Analysis **74** (2011), no. 81-93.

[156] J. S. Park and C. Kim, *Hierarchical multi-dimensional limiting strategy for correction procedure via reconstruction*, J. Comp. Phys. **308** (2016), 57–80.

[157] J. S. Park and S.-H. Yoon and C. Kim, *Multi-dimensional limiting process for hyperbolic conservation laws on unstructured grids*, J. Comp. Phys. **229** (2010), 788–812.

[158] J. Jaffre, C. Johnson, and A. Szepessy, *Convergence of the discontinuous galerkin finite element method for hyperbolic conservation laws*, Mathematical Models and Methods in Applied Sciences **5** (1995), no. 03, 367–386.

[159] G. Jiang and C.-W. Shu, *Efficient implementation of weighted eno schemes*, J. Comput. Phys. **126** (1996), no. 1, 202–228.

[160] Guang-Shan Jiang and Chi-Wang Shu, *Efficient implementation of weighted ENO schemes*, Journal of Computational Physics **126** (1996), no. 1, 202–228.

[161] T. Jiang, *Ship waves in shallow water*, Verkehrstechnik, Fahrzeugtechnik, Fortschritt-Berichte VDIReihe, 2001.

[162] T. Jiang, R. Henn, and S.D. Sharma, *Wash waves generated by ships moving on fairways of varying topography*, 24th Symposium on Naval Hydrodynamics Fukuoka, JAPAN, 2002.

[163] F. John, *On the motion of floating bodies i*, Communications on Pure and Applied Mathematics **2** (1949), 13–57.

[164] A. A. Johnson and T. E. Tezduyar, *Mesh update strategies in parallel finite element computations of flow problems with moving boundaries and interfaces*, Computer Methods in Applied Mechanics and Engineering **119** (1994), no. 1–2, 73–94.

[165] C. Johnson and J. Pitkäranta, *An analysis of the discontinuous galerkin method for a scalar hyperbolic equation*, Mathematics of computation **46** (1986), no. 173, 1–26.

[166] N. Kaewbanjak, S. Grilli, J. Kirby, and P. Watts, *Modeling the 26 december 2004 indian ocean tsunami: Case study of impact in thailand*, Journal of Geophysical Research (2007).

[167] T. Kato, *The Cauchy problem for quasi-linear symmetric hyperbolic systems*, Arch. Ration. Mech. Anal. **58** (1975), no. 3, 181–205.

[168] D. Kay, V. Styles, and E. Sulli, *Discontinuous Galerkin finite element approximation of the Cahn-Hilliard equation with convection*, SIAM J. Numer. Anal. **47** (2009), no. 4, 2660–2685.

[169] G. Kesserwani and Q. Liang, *A discontinuous Galerkin algorithm for the two-dimensional shallow water equations*, Comput. Methods Appl. Mech. Engrg. **199** (2010), no. 49-52, 3356–3368.

[170] ———, *Well-balanced RKDG2 solutions to the shallow water equations over irregular domains with wetting and drying*, Comput. Fluids **39** (2010), no. 10, 2040–2050.

[171] G. Kesserwani, Q. Liang, J. Vazquez, and R. Mose, *Well-balancing issues related to the RKDG2 scheme for the shallow water equations*, Int. J. Numer. Meth. Fluids **62** (2010), 428–448.

[172] R.M. Kirby and S.J. Sherwin, *Stabilisation of spectral / hp element methods through spectral vanishing viscosity: Application to fluid mechanics*, Comput. Methods Appl. Mech. Engrg. **195** (2006), 3128–3144.

[173] E.V. Koutandos, T.V. Karambas, and C.G. Koutitas, *Floating breakwater response to waves action using a boussinesq model coupled with a 2dv elliptic solver*, J. Waterw. Port Coastal Ocean Eng. **130** (2004), 243–255.

[174] E.E. Kriezi, T. Karambas, P. Prinos, and C. Koutitas, *Interaction of floating breakwaters with waves in shallow waters*, Proc., Int. Conf. IAHR, Beijing, 2001.

[175] L. Krivodonova, *Limiters for high-order discontinuous Galerkin methods*, J. Comp. Phys. **226** (2007), 879–896.

[176] M. Kucharik, J. Breil, S. Galera, P.-H. Maire, M. Berndt, and M. Shashkov, *Hybrid remap for multi-material ALE*, Computers & Fluids **46** (2011), 293–297.

[177] A. Kurganov and G. Petrova, *A second-order well-balanced positivity preserving central-upwind scheme for the saint-venant system*, Commun. Math. Sci. **5(1)** (2007), 133–160.

[178] D. Kuzmin, *A vertex-based hierarchical slope limiter for p-adaptive discontinuous galerkin methods*, J. Comput. Applied. Math. **233** (2010), no. 12.

[179] D. Kuzmin and M.Q. de Luna, *Subcell flux limiting for high-order bernstein finite element discretizations of scalar hyperbolic conservation laws*, J. Comp. Phys. **411** (2020), 109411.

[180] D. Kuzmin, R. Lohner, and S. Turek, *Flux-corrected transport - Principles, Algorithms and Applications*, Springer, 2012.

[181] D. Lannes, *The water waves problem: mathematical analysis and asymptotics*, Mathematical Surveys and Monographs, no. 188, American Mathematical Society, 2013.

[182] ———, *On the dynamics of floating structures*, Annals of PDE **3** (2017), no. 1, 11.

[183] D. Lannes and F. Marche, *A new class of fully nonlinear and weakly dispersive green–naghdi models for efficient 2d simulations*, Journal of Computational Physics **282** (2015), 238–268.

[184] D. Lannes and G. Métivier, *The shoreline problem for the one-dimensional shallow water and green-naghdi equations*, Journal de l'École polytechnique—Mathématiques **5** (2018), 455–518.

[185] David Lannes and Tatsuo Iguchi, *The 2d nonlinear shallow water equations with a partially immersed obstacle*, 2023.

[186] David Lannes and Mei Ming, *Well posedness of f. john's floating body problem for a fixed object*, 2025.

[187] David Lannes and Martin Oen Paulsen, *The f. john model and cummins' equations for freely floating objects*, 2025.

[188] O. Le Métayer, S. Gavrilyuk, and S. Hank, *A numerical scheme for the Green-Naghdi model*, J. Comput. Phys. **34** (2010), no. 229, 2034–2045.

[189] Alain Yves Le Roux, *Riemann solvers for some hyperbolic problems with a source term*, ESAIM: Proceedings **6** (1999), 75–90.

[190] C. Lee and J.N. Newman, *Computation of wave effects using the panel method*, Numerical Models in Fluid-Structure Interaction, 2005.

[191] B. Van Leer, *Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method*, J. Comput. Phys. **135** (1997), no. 2, 227–248.

[192] A.-Y. Leroux, *A new general Riemann solver for the shallow water equations with friction and topography*, 2009.

[193] P. Lesaint, *Finite element methods for symmetric hyperbolic equations*, Numerische Mathematik **21** (1973), no. 3, 244–255.

[194] P. Lesaint and P.-A. Raviart, *On a finite element method for solving the neutron transport equation*, Publications mathématiques et informatique de Rennes (1974), no. S4, 1–40.

[195] F. Levholt, S. Glimsdal, P. Lynett, and G. Pedersen, *Simulating tsunami propagation in fjords with long wave models*, Nat. Hazards Earth Syst. Sci. (2014).

[196] H. Li and R.-X. Liu, *The discontinuous Galerkin finite element method for the 2d shallow water equations*, Mathematics and Computers in Simulation **56** (2001), 223–233.

[197] L. Li and Q. Zhang, *A new vertex-based imiting approach for nodal discontinuous galerkin methods on arbitrary unstructured meshes*, Comput. Fluids **159** (2017), 316–326.

[198] Y.L. Li and Y.-H. Yu, *A synthesis of numerical methods for modeling wave energy converter-point absorbers*, Renewable and Sustainable Energy Reviews **16** (2012), 4352–4364.

[199] C. Liang, F. Ham, and E. Johnsen, *Discontinuous galerkin method with weno limiter for flows with discontinuity*, Center for Turbulence Research 335 Annual Research Briefs 2009 (2009).

[200] Q. Liang and F. Marche, *Numerical resolution of well-balanced shallow water equations with complex source terms*, Advances in Water Resources **32** (2009), no. 6, 873 – 884.

[201] Y. Lin and J. Chan, *High order entropy stable discontinuous galerkin spectral element methods through subcell limiting*, J. Comp. Phys. **498** (2024), 112677.

[202] R. Liska, M. Kuchařík, J. Limpouch, O. Renner, P. Váchal, L. Bednárik, and J. Velechovský, *Ale method for simulations of laser-produced plasmas*, Finite Volumes for Complex Applications VI Problems & Perspectives (Berlin, Heidelberg) (Jaroslav Fořt, Jiří Fürst, Jan Halama, Raphaèle Herbin, and Florence Hubert, eds.), Springer Berlin Heidelberg, 2011, pp. 857–873.

[203] Xin Liu, Abdolmajid Mohammadian, Alexander Kurganov, and Julio Angel Infante Sedano, *Well-balanced central-upwind scheme for a fully coupled shallow water system modeling flows over erodible bed*, Journal of Computational Physics **300** (2015), 202–218.

[204] Xu-Dong Liu, Stanley Osher, and Tony Chan, *Weighted essentially non-oscillatory schemes*, Journal of Computational Physics **115** (1994), no. 1, 200–212.

[205] I. Lomtev, R.M. Kirby, and G.E. Karniadakis, *A discontinuous Galerkin ALE method for compressible viscous flows in moving domains*, J. Comput. Phys. **155** (1999), no. 1, 128–159.

[206] M. Lukacova, S. Noelle, and M. Kraft, *Well-balanced finite volume evolution galerkin methods for the shallow water equations*, J. Comput. Phys. **221** (2007), no. 1, 122–147.

[207] G. Ma, J.T. Kirby, and F. Shi, *Numerical simulation of tsunami waves generated by deformable submarine landslides*, Ocean Modelling **69** (2013), 146–165.

[208] Hong Ma, *A spectral element basin model for the shallow water equations*, J. Comput. Phys. **109** (1993), no. 1, 133 – 149.

[209] Per A Madsen and Ole R Sørensen, *A new form of the boussinesq equations with improved linear dispersion characteristics. part 2. a slowly-varying bathymetry*, Coastal engineering **18** (1992), no. 3-4, 183–204.

[210] F. Marche, *Derivation of a new two-dimensional viscous shallow water model with varying topography, bottom friction and capillary effects*, Eur. J. Mech. B Fluids **26** (2007), no. 1, 49–63.

[211] ———, *Combined Hybridizable Discontinuous Galerkin (HDG) and Runge-Kutta Discontinuous Galerkin (RK-DG) formulations for Green-Naghdi equations on unstructured meshes*, J. Comput. Phys. **418** (2020), 109637.

[212] S. Marras, M.A. Kopera, E.M. Constantinescu, J. Suckale, and F.X. Giraldo, *A continuous/discontinuous Galerkin solution of the shallow water equations with dynamic viscosity, high-order wetting and drying, and implicit time integration*, preprint (2016).

[213] A. Meister and S. Ortleb, *A positivity preserving and well-balanced DG scheme using finite volume subcells in almost dry regions*, Appl. Math. Comp. **272** (2016), 259–273.

[214] E. Mendoza, R. Silva, B. Zanuttigh, E. Angelelli, T.L. Andersen, L. Martinelli, J.Q.H. Norgaard, and P. Ruol, *Beach response to wave energy converter farms acting as coastal defence*, Coastal Engineering **87** (2014), 97–111.

[215] V. Michel-Dansac, *Development of high-order well-balanced schemes for geophysical flows*, Ph.D. thesis, Université de Nantes, 2016.

[216] ———, *A well-balanced scheme for the shallow-water equations with topography*, Computers and Mathematics with Applications **72** (2016), no. 3, 568–593.

[217] V. Michel-Dansac, C. Berthon, S. Clain, and F. Foucher, *A well-balanced scheme for the shallow-water equations with topography*, Comput. Math. Appl. **72** (2016), Comput. Math. Appl.

[218] Victor Michel-Dansac, Christophe Berthon, Stéphane Clain, and Françoise Foucher, *A two-dimensional high-order well-balanced scheme for the shallow water equations with topography and manning friction*, Computers and Fluids **230** (2021), 105152.

[219] Paul Moujaes and Dmitri Kuzmin, *Monolithic convex limiting and implicit pseudo-time stepping for calculating steady-state solutions of the euler equations*, Journal of Computational Physics **523** (2025), 113687.

[220] R. D. Nair, S. J. Thomas, and R. D. Loft, *A discontinuous galerkin global shallow water model*, Monthly Weather Review **133** (2004), 876–888.

[221] I.M. Navon, *Finite-element simulation of the shallow-water equations model on a limited-area domain*, Appl. Math. Modelling **3** (1979).

[222] Joachim Nitsche, *Über ein variationsprinzip zur lösung von dirichlet-problemen bei verwendung von teilräumen, die keinen randbedingungen unterworfen sind*, Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg **36** (1971), 9–15.

[223] S. Noelle, N. Pankratz, G. Puppo, and J.R. Natvig, *Well-balanced finite volume schemes of arbitrary order of accuracy for shallow water flows*, J. Comput. Phys. **213** (2006), no. 2, 474–499.

[224] 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), no. 1, 29–58.

[225] Okey Nwogu, *Nonlinear evolution of directional wave spectra in shallow water*, Coastal Engineering 1994, 1995, pp. 467–481.

[226] T.F. Ogilvie, *Second-order hydrodynamic effects on ocean platforms*, Proc. Intl. Workshop on Ship and Platform Motions, ed. R. W. Yeung, University of California, Berkeley, 1983, pp. 205–265.

[227] M.D. Orzech, F. Shi, J. Veeramony, Bateman S., J. Calantoni, and J.T. Kirby, *Incorporating floating surface objects into a fully dispersive surface wave model*, Ocean Modelling **102** (2016), 14–26.

[228] H. T. Ozkan-Haller and J.T.Kirby, *A fourier-chebyshev collocation method for the shallow water equations including shoreline*, Applied Ocean Research **19** (1997), 21–34.

[229] P.-O.Persson and J.Peraire, *Sub-cell shock capturing for discontinuous galerkin methods*, AIAA Aerospace Sciences Meeting and Exhibit **112** (2006).

[230] H. park, D.T. Cox, P.J. Lynett, D.M. Wiebe, and S. Shin, *Tsunami inundation modeling in constructed environments: A physical and numerical comparison of free-surface elevation, velocity, and momentum flux*, Coastal Engineering **79** (2013), 9–21.

[231] J. Patera and V. Nassehi, *A new two-dimensional finite element model for the shallow water equations using a lagrangian framework constructed along fluid particle trajectories.*, Int. J. Numer. Meth. Fluids **39** (1996), 4159–4182.

[232] D Howell Peregrine, *Long waves on a beach*, Journal of fluid mechanics **27** (1967), no. 4, 815–827.

[233] P.-O. Persson, J. Peraire, and J. Bonet, *Discontinuous Galerkin solution of the Navier–Stokes equations on deformable domains*, Comp. Meth. App. Mech. Eng. **198** (2009), 1585–1595.

[234] Benoît Perthame and Youchun Qiu, *A variant of Van Leer's method for multidimensional systems of conservation laws*, J. Comput. Phys. **112** (1994), no. 2, 370–381.

[235] C.S. Peskin, *The immersed boundary method the immersed boundary method*, Acta Numerica (2002), 479–517.

[236] J. Qiu and C.-W. Shu, *A comparison of troubled-cell indicators for Runge-Kutta discontinuous Galerkin methods using weighted essentially nonoscillatory limiters*, SIAM J. Sci. Comput. **27** (2005), 995–1013.

[237] J. Qiu and C.-W. Shu, *Runge Kutta discontinuous Galerkin method using WENO limiters*, SIAM J. Sci. Comput. **26** (2005), 907–929.

[238] W.H. Reed and T.R. Hill, *Triangular mesh methods for the neutron transport equation*, Tech. report, Los Alamos Scientific Lab., N. Mex.(USA), 1973.

[239] M. Ricchiuto, *An explicit residual based approach for shallow water flows*, J. Comput. Phys. **280** (2015), 306–344.

[240] M. Ricchiuto and A. Bollermann, *Stabilized residual distribution for shallow water simulations*, J. Comput. Phys. **228** (2009), 1071–1115.

[241] B. Rogers, M. Fujihara, and A. Borthwick, *Adaptive Q-tree Godunov-type scheme for shallow water equations*, Int. J. Numer. Methods Fluids **35** (2001), 247–280.

[242] A. Rueda-Ramirez, B. Bolm, D. Kuzmin, and G. Gassner, *Monolithic convex limiting for legendre-gauss-lobatto discontinuous galerkin spectral-element methods*, Commun. Appl. Math. Comput. (2024).

[243] Andres M. Rueda-Ramirez, Will Pazner, and Gregor J. Gassner, *Subcell limiting strategies for discontinuous galerkin spectral element methods*, Computers and Fluids **247** (2022), 105627.

[244] D. Schwanenberg and M. Harms, *Discontinuous galerkin finite-element method for transcritical two-dimensional shallow water flows*, J.Hydraul. Eng. **130** (2004), no. 5, 412–421.

[245] François Serre, *Contribution à l'étude des écoulements permanents et variables dans les canaux*, La Houille Blanche (1953), no. 6, 830–872.

[246] Alexander Shamanskiy, *Mesh moving techniques in fluid-structure interaction*, Computational Mechanics **67** (2021), 583–603.

[247] F. Shi, J.T. Kirby, J.C. Harris, J.D. Geiman, and S.T. Grilli, *A high-order adaptive time-stepping TVD solver for Boussinesq modeling of breaking waves and coastal inundation.*, Ocean Modelling **43-44** (2012), 36–51.

[248] T. Song, A. Main, G. Scovazzi, and M. Ricchiuto, *The shifted boundary method for hyperbolic systems: Embedded domain computations of linear waves and shallow water flows*, Tech. report, Inria, 2017.

[249] M. Sonntag and C.D. Munz, *Shock capturing for discontinuous Galerkin methods using finite volume subcells*, Finite Volumes for Complex Applications VII-Elliptic, Parabolic and Hyperbolic Problems, 2014, pp. 945–953.

[250] K. Stein, T. Tezduyar, and R. Benney, *Mesh moving techniques for fluid-structure interactions with large displacements*, Journal of Applied Mechanics **70** (2003), no. 1, 58–63.

[251] Pei Su and Marius Tucsnak, *Shallow water waves generated by a floating object: a control theoretical perspective*, Mathematical Control and Related Fields **13** (2023), no. 4, 1722–1744.

[252] C.E. Synolakis, E.N. Bernard, V.V. Titov, U Kanoglu, and F.I. Gonzalez, *Standards, criteria, and procedures for noaa evaluation of tsunami numerical models*, NOAA Tech. Memo. **OAR PMEL-135** (2007).

[253] P.A. Tassi, O. Bokhove, and C.A. Vionnet, *Space discontinuous galerkin method for shallow water flows—kinetic and hllc flux, and potential vorticity generation*, Advances in water resources **30** (2007), no. 4, 998–1015.

[254] P.D. Thomas and C.K. Lombard, *Geometric conservation law and its applications to flow computations on moving grids*, AIAA Journal **17** (1979), 1030–1037.

[255] V. V. Titov and C.E. Synolakis, *Numerical modeling of tidal wave runup*, J. Wtrwy. Port Coast. and Oc. Engrg. **124** (1998), no. 4, 157–171.

[256] P. Tonon, S. R. Mathur, M.-C. Hsu, and T. E. Tezduyar, *A linear-elasticity-based mesh moving method with no cycle-to-cycle accumulated distortion*, Computational Mechanics **67** (2021), 413–434.

[257] E.F. Toro, *Shock-capturing methods for free-surface shallow flows*, Chichester: John Wiley and Sons, 2001.

[258] T. Utnes, *A finite element solution of the shallow-water wave equations*, Appl. Math. Modelling **14** (1990), 20–29.

[259] J.J.W. van de Vegt and Y. Xu, *Space-time discontinuous galerkin method for nonlinear water waves*, J. Comput. Phys. **224** (2007), 17–39.

[260] J.J.W. Van der Vegt and H. Van der Ven, *Space-time discontinuous Galerkin finite element method with dynamic grid motion for inviscid compressible flows*, J. Comput. Phys. **182** (2002), 546–585.

[261] C.S. Ventakasubban, *A new finite element formulation for ALE (arbitrary Lagrangian Eulerian) compressible fluid mechanics*, Int. J. Engrg. Sci. **33** (1995), 1743–1762.

[262] F. Vilar, *A high-order discontinuous Galerkin discretization for solving two-dimensional Lagrangian hydrodynamics*, Ph.D. thesis, Université Bordeaux I, 2012.

[263] F. Vilar, *Cell-centered discontinuous galerkin discretization for two-dimensional lagrangian hydrodynamics*, Computers and Fluids **64** (2012), 64–73.

[264] ———, *A posteriori correction of high-order discontinuous galerkin scheme through subcell finite volume formulation and flux reconstruction*, J. Comput. Phys. **387** (2019), 245–279.

[265] ———, *Local subcell monolithic DG/FV convex property preserving scheme on unstructured grids and entropy consideration*, J. Comput. Phys. **521** (2025), no. 1.

[266] F. Vilar, P.H. Maire, and R. Abgrall, *Cell-centered discontinuous Galerkin discretizations for two-dimensional scalar conservation laws on unstructured grids and for one-dimensional Lagrangian hydrodynamics*, Comput. Fluids **46** (2011), 498–504.

[267] François Vilar and Rémi Abgrall, *A posteriori local subcell correction of high-order discontinuous galerkin scheme for conservation laws on two-dimensional unstructured grids*, SIAM Journal on Scientific Computing **46** (2024), no. 2, A851–A883.

[268] François Vilar, Chi-Wang Shu, and Pierre-Henri Maire, *Positivity-preserving cell-centered Lagrangian schemes for multi-material compressible flows: From first-order to high-orders. Part I: The one-dimensional case*, Journal of Computational Physics **312** (2016), 385–415.

[269] ———, *Positivity-preserving cell-centered Lagrangian schemes for multi-material compressible flows: From first-order to high-orders. Part II: The two-dimensional case*, Journal of Computational Physics **312** (2016), 416–442.

[270] P. E. Vincent, P. Castonguay, and A. Jameson, *A New Class of High-Order Energy Stable Flux Reconstruction Schemes*, J. Sci. Comput. **47** (2011), 50–72.

[271] C. Vinchon, A.-L. Agenais, N. Baron-Yelles, E. Berthelier, C. Hérivaux, L. Maton, C. Meur-Ferec, H. Rey-Vaalette, Y. Balouin, J.C. Charles, E. Delaavalee, H. Flanquart, M. Garcin, A.-P. Hellequin, Y. Krien, L. Kuhfuss, S. Lecacheux, F. Le Cozannet, G. annd Marche, V. Mazeiraud, R. Pedreros, B. Poisson, E. Romieu, and B. Rulleau, *MISEEVA: set up of a transdisciplinary approach to assess vulnerability of the coastal zone to marine inundation at regional and local scale, within a global change context.*, Coastal Futures 2010: review and future trends, London, United Kingdom, 2010.

[272] S. Vukovic, *Eno and weno schemes with the exact conservation property for one-dimensional shallow water equations*, J. Comput. Phys. **179** (2002), no. 2, 593–621.

[273] Z.J. Wang, *High-Order Spectral Volume Method for Benchmark Aeroacoustic Problems*, AIAA Paper **2003-0880** (2003).

[274] W.C. Webster, W. Duan, and B. Zhao, *Green-naghdi theory, part A: Green-Naghdi (GN) equations for shallow water waves*, J. Mar. Sci. Appl. **10** (2011), 253:258.

[275] Mary Fanett Wheeler, *An elliptic collocation-finite element method with interior penalties*, SIAM Journal on Numerical Analysis **15** (1978), no. 1, 152–161.

[276] G.B. Whitham, *Linear and nonlinear waves*, John Wiley & Sons, 2011.

[277] Thomas Wick, *Fluid-structure interactions using different mesh motion techniques*, Computers & Structures **89** (2011), no. 13–14, 1456–1467.

[278] A. N. Williams and W. G. McDougal, *Flexible floating break-water*, J. Waterw. Port Coastal Ocean Eng. **117** (1991), no. 5, 429–450.

[279] A. M. Winslow, *Numerical solution of the quasilinear poisson equation in a nonuniform triangle mesh*, Journal of Computational Physics **1** (1966), no. 2, 149–172.

[280] D. Wirasaet, E.J. Kubatko, C.E. Michoski, S. Tanaka, and J.J. Westerink, *Discontinuous Galerkin methods with nodal and hybrid modal/nodal triangular, quadrilateral, and polygonal elements for nonlinear shallow water flow*, Comput. Methods Appl. Mech. Engrg. **270** (2014), 113–149.

[281] Xinhui Wu, Nathaniel Trask, and Jesse Chan, *Entropy stable discontinuous galerkin methods for the shallow water equations with subcell positivity preservation*, Numerical Methods for Partial Differential Equations **40** (2024), no. 6, e23129.

[282] 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), 567–598.

[283] ———, *A new approach of high order well-balanced finite volume WENO schemes and discontinuous Galerkin methods for a class of hyperbolic systems with source terms*, Commun. Comput. Phys. **1** (2006), 100–134.

[284] ———, *A survey of high order schemes for the shallow water equations*, J. Math. Study (2014).

[285] Y. Xing and X. Zhang, *Positivity-preserving well-balanced discontinuous Galerkin methods for the shallow water equations on unstructured triangular meshes*, J. Sci. Comput. **57** (2013), 19–41.

[286] Y. Xing, X. Zhang, and C.-W. Shu, *Positivity-preserving high order well-balanced discontinuous galerkin methods for the shallow water equations*, Advances in Water Resources **33** (2010), no. 12, 1476 – 1493.

[287] G.Q. Yang, O.M. Faltinsen, and R. Zhao, *Wash of ships in finite water depth*, Proceedings of the FAST 2001, Southhampton, UK, 2001.

[288] M. Yang and Z.J. Wang, *A parameter-free generalized moment limiter for high-order methods on unstrucured grids*, Adv. Appl. Math. Mech. **4** (2009), 451–480.

[289] Y.H. Yu and L. Ye, *Reynolds-averaged navier stokes simulation of the heave performance of a two-body floating-point absorber wave energy system*, Computers and Fluids **73** (2013), 104–114.

[290] Steven T. Zalesak, *Fully multidimensional Flux-Corrected Transport algorithms for fluids*, Journal of Computational Physics **31** (1979), no. 3, 335–362.

[291] Olindo Zanotti, Francesco Fambri, Michael Dumbser, and Arturo Hidalgo, *Space-time adaptive ADER discontinuous galerkin finite element schemes with a posteriori sub-cell finite volume limiting*, Computers & Fluids **118** (2015), 204–224.

[292] M. Zefzouf and F. Marche, *A new symmetric interior penalty discontinuous galerkin formulation for the serre-green-naghdi equations*, Numer. Methods Partial Differential Eq. **39** (2023), no. 2, 1478–1503.

[293] Weijie Zhang, Yinhua Xia, and Yan Xu, *Positivity-preserving well-balanced arbitrary lagrangian–eulerian discontinuous galerkin methods for the shallow water equations*, Journal of Scientific Computing **88** (2021), no. 2, 38.

[294] J. Zhu, J. Qiu, C.-W. Shu, and M. Dumbser, *Runge–Kutta discontinuous Galerkin method using WENO limiters II: Unstructured meshes*, J. Comput. Phys. **227** (2008), 4330–4353.

[295] J. Zhu, X. Zhong, C.-W. Shu, and J. Qiu, *Runge Kutta discontinuous Galerkin method using a new type of WENO type limiters on unstructured meshes*, J. Comp. Phys. **248** (2013), 200–220.

[296] M. Zijlema and G.S. Stelling, *Efficient computation of surf zone waves using the nonlinear shallow water equations with non-hydrostatic pressure*, Coastal Engineering **55** (2008), 780–790.

I was lying in my bed
A creature void of form
Been so afraid of everything
I need a chance to be reborn

I never wanted anything
That someone had to give
I don't live here anymore
I went along the wind

When I think about the old days, babe
You're always on my mind
I know it ain't like I remember
I guess my memories run wild

Like when we went to see Bob Dylan
We danced to "Desolation Row"
But I don't live here anymore
But I got no place to go

Beating like a heart
I'm gonna walk through every doorway, I can't stop
I need some time, I need control, I need your love
I wanna find out everything I need to know

I'm gonna say everything that there is to say
Although you've taken everything I need away
I'm gonna make it to the place I need to go
We're all just walkin' through this darkness on our own

Time surrounds me like an ocean
My memories like waves
Is life just dying in slow motion
Or getting stronger everyday?

I never took our love for granted
You never left me wanting more
But you'd never recognise me, babe
I don't live here anymore

Beating like a heart
I'm gonna walk through every doorway, I can't stop
I need some time, I need control, I need your love
I wanna find out everything I need to know

I'm gonna say everything that I need to say
Although you've taken everything I need away
I'm gonna take you to the place I need to go
We're all just walkin' through this darkness on our own

Beating like a heart
I'm gonna walk through every doorway, I can't stop
I need some time, I need control, I need your love
I wanna find out everything I need to know

I'm gonna say everything that I need to say
Although you've taken everything I need away
I'm gonna take you to the place I need to go
We're all just walkin' through this darkness on our own

"I Don't Live Here Anymore", The War on Drugs (feat. Lucius) — *I Don't Live Here Anymore (2021)*

High-order numerical methods for shallow-water flows on unstructured meshes: moving boundaries and wave-structure interactions

Predicting how ocean waves interact with floating structures like ships, offshore wind turbines, or wave energy converters, is a major challenge in coastal engineering. However, the mathematical equations governing these fluid-structure dynamics are highly complex and notoriously difficult to solve, especially during extreme wave events.

This thesis develops new mathematical models and numerical methods to approximate these complex moving-boundary flows with both accuracy and robustness. The work combines and improves several state-of-the-art techniques designed to respect the main physical constraints of the water-wave equations. These methods are then extended to wave-structure interactions problems, where the motion of the water and the constraints imposed by floating or immersed structures must be coupled in a consistent way. Such a framework may also contribute to the design, testing, and optimization of coastal engineering structures.

Keywords. Wave-structure interactions – Shallow-water asymptotics – High-order methods – Arbitrary Lagrangian–Eulerian – Subcell schemes – Unstructured meshes

---

Méthodes numériques d'ordre élevé pour les écoulements en eaux peu profondes sur maillages non-structurés : frontières mobiles et interactions vagues-structure

Prédire comment les vagues interagissent avec des structures flottantes, telles que navires, éoliennes en mer ou convertisseurs d'énergie houlomotrice, est un défi majeur en ingénierie côtière. Cependant, les équations régissant ces dynamiques sont d'une grande complexité et particulièrement difficiles à résoudre, surtout lors d'événements maritimes violents.

Cette thèse développe de nouveaux modèles mathématiques et méthodes numériques pour approcher ces écoulements complexes avec précision et robustesse. Le travail combine et améliore plusieurs techniques avancées conçues pour respecter les principales contraintes physiques des équations des vagues. Ces méthodes sont ensuite étendues aux problèmes d'interactions vague-structure, où le mouvement de l'eau et les contraintes imposées par des structures flottantes ou immergées doivent être couplés de manière cohérente. Un tel cadre peut également contribuer à la conception, au test et à l'optimisation de futures installations d'ingénierie marine.

Mots-clés. Interactions vagues-structure – Modèles asymptotiques en eaux peu profondes – Méthodes d'ordre élevé – Arbitraire Lagrangien-Eulérien – Schémas de sous-mailles – Maillages non-structurés