Numerical simulation of geophysical flows using high-order and well-balanced Lagrange-Projection methods

Alessia del Grosso

université PARIS-SACLAY

Numerical simulation of geophysical flows using high-order and well-balanced Lagrange-Projection methods

Simulation numérique d'écoulements géophysiques à l'aide de méthodes numériques de type Lagrange-Projection équilibres et d'ordre élevé

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 574 : Mathématique Hadamard (EDMH)
Spécialité de doctorat: Mathématiques appliquées
Graduate School : Mathématiques
Référent : Université de Versailles Saint-Quentin-en-Yvelines

Thèse préparée au LMV (Université Paris-Saclay, UVSQ, CNRS), sous la direction de Christophe CHALONS, Professeur à l'université Paris-Saclay

Thèse soutenue à Versailles, le 29 septembre 2022, par

Alessia DEL GROSSO

Composition du jury

Muriel BOULAKIA Professeur, Université Paris-Saclay	Présidente
Raphaël LOUBÈRE Directeur de Recherche, Université de Bordeaux	Rapporteur & Examinateur
Nicolas SEGUIN Professeur, Université de Rennes 1	Rapporteur & Examinateur
Samuel KOKH Ingénieur de Recherche, CEA Saclay	Examinateur
Tomás MORALES DE LUNA Professeur associé, Université de Málaga	Examinateur
Christophe CHALONS Professeur, Université Paris-Saclay	Directeur de thèse

THESE DE DOCTORAT
NNT : 2022UPASM017

université PARIS-SACLAY | ÉCOLE DOCTORALE de mathématiques Hadamard (EDMH)

Titre: Simulation numérique d'écoulements géophysiques à l'aide de méthodes numériques de type Lagrange-Projection équilibres et d'ordre élevé

Mots clés: Systèmes hyperboliques, méthodes des volumes finis, écoulements géophysiques, équations du flux sanguin, décomposition de type Lagrange-projection, propriété équilibre

Résumé: Dans cette thèse, nous considérons principalement la simulation numérique des écoulements géophysiques à l'aide de modèles hydrodynamiques. Nous sommes particulièrement intéressés par les applications au transport de sédiments, aux processus d'érosion-dépôt des sédiments et aux écoulements stratifiés. Des modèles avancés basés sur les équations de Saint-Venant (à une ou deux couches) ont été utilisés pour prendre en compte ces phénomènes. Cependant, comme point de départ de cette thèse, nous avons également realisé des schémas numériques pour les équations du flux sanguin. En effet, afin de pouvoir simuler de tels modèles, nous cherchons à concevoir et analyser une stratégie numérique efficace d'ordre supérieur et équilibre, en se basant sur un formalisme de type Lagrange-Projection. En quelques mots, de tels algorithmes se décomposent en deux étapes : l'étape lagrangienne prend en compte les effets de compressibilité tandis que l'étape de projection est dédiée aux phénomènes de transport. Une telle décomposition présente un intérêt particulier dans de nombreux écoulements géophysiques car elle fournit une décomposition naturelle des échelles de temps et conduit à des schémas implicites-explicites très efficaces.

Title: Numerical simulation of geophysical flows using high-order and well-balanced Lagrange-Projection methods

Keywords: Hyperbolic systems, finite volume methods, geophysical flows, blood flow equations, Lagrange-projection splitting, well-balanced property

Abstract: In this PhD thesis, we mainly consider the numerical simulation of geophysical flows by means of hydrodynamics models. We are especially interested in applications to sediment transport, sediment erosion-deposition processes and stratified flows. Advanced models based on (one or two-layer) shallow-water equations have been used to take these phenomena into account. However, as a starting point of this thesis, we also built numerical schemes for the blood flow equations.
In order to be able to simulate such flows, we aim to design, analyze and implement high-order and well-balanced efficient numerical methods that are based on a Lagrange-Projection formalism. In a few words, such algorithms are composed of two steps: the Lagrangian step takes into account the compressibility effects of the flow, while the Projection step is dedicated to the transport phenomena. Such a splitting is of particular interest in many geophysical flows as it provides a natural time scale decomposition leading to very efficient implicit-explicit schemes.


UVSQ
université PARIS-SACLAY


LMV
Laboratoire de mathématiques
de Versailles - CNRS UMR 8100
UNIVERSITÉ DE
VERSAILLES


Fondation mathématique
FMJH
Jacques Hadamard


MATH
INNOV


Région
îledeFrance

A mia madre Nadia e a mia zia Anna Rita.

Remerciements

Après trois ans, je me retrouve enfin à écrire les remerciements pour cette thèse de doctorat (avec quelques larmes dans les yeux, je l'avoue).

La première personne que je dois remercier est certainement mon directeur de thèse, Christophe Chalons. Une personne d'une incroyable gentillesse, toujours disponible et prête à m'aider pour tout problème. Christophe, merci de toujours m'encourager et de me pousser à croire en mes travaux de recherche. Je ne pense pas que j'aurais pu avoir un meilleur directeur de thèse que toi.

I would also like to thank Manuel J. Castro Díaz and Tomás Morales de Luna for welcoming me in Málaga. Not only because of your always exhaustive explanations but also because you have always been extremely kind and helpful to me throughout our collaboration. I very much hope that we will continue to work together in the future as well!

Mes remerciements vont maintenant à mes deux rapporteurs de thèse, Raphaël Loubère et Nicolas Seguin. J'espère que la lecture de ma thèse n'a pas été une tâche trop lourde cet été !
Je remercie également tous les autres membres du jury pour avoir accepté d'en faire partie: Muriel Boulakia, Samuel Kokh et Tomás Morales de Luna.

En outre, je tiens à remercier Raphaël Loubère et Pierre-Henri Maire pour m'avoir invitée à Bordeaux et m'avoir donner l'opportunité de faire un post-doc avec eux. Je suis impatiente de commencer ce nouveau chapitre !

I would also like to thank Eleuterio F. Toro, whom I was privileged to have as supervisor for my master's thesis. Thank you for having found the time to read and comment the manuscript about the blood flow equations.

Je souhaite aussi remercier tous les membres du laboratoire de mathématiques de Versailles que j'ai pu rencontrer au fil des années. Et, en particulier, tous les doctorants, post-docs et ATER avec lesquels j'ai partagé une partie de cette expérience. Louis, Melek, Monica, Taher, pour n'en citer que quelques-uns.
Bien sûr, je dois aussi remercier Guillermo Moreno-Socías, avec qui j'ai eu de très belles conversations au cours de ces trois années. Merci d'avoir accepté d'être le garant technique pour ma soutenance et d'avoir toujours été prêt à m'aider lorsque j'avais des problèmes informatiques.

È arrivato il momento di ringraziare la mia famiglia, stretta e allargata.
Grazie papà per spingermi sempre a fare nuove esperienze, anche se mi fanno paura.

Mamma, senza di te penso che non sarei proprio arrivata alla fine di questo dottorato. Sei sempre stata lì, pronta ad incoraggiarmi ed aiutarmi, anche quando mi disperavo per dei motivi un po' futili. Mamma e papà, per questo e molto altro, grazie.

Ovviamente devo ringraziare mia sorella Sabrina e il suo compagno Gabriele per tutti i bei momenti passati assieme, anche (e non solo) tra i giochi da tavolo! Sabri, lo so che ti ho fatto un po' penare da bambina e adolescente ma eri (e resti sempre) uno dei miei punti di riferimento.

Marta, sei la mia migliore amica da più di dieci anni, è scontato dire che se sono arrivata dove sono è anche merito tuo. Grazie e mille volte grazie per essermi sempre stata vicina in questi anni ed essermi venuta a trovare qui in Francia più volte. Ringrazio il fato che ci ha fatto sedere vicine quel primo giorno di liceo.

Un immenso grazie a Fabrizio. Grazie per aver accettato di iniziare con me questa avventura. Grazie per i tuoi continui incoraggiamenti e per riuscire a farmi sempre ridere, anche nei momenti di totale sconforto.

Vorrei finire col ringraziare mia zia Anna Rita e famiglia per il loro affetto, e ovviamente anche per tutte le belle serate passate insieme a giocare a carte. In particolare, devo ringraziare Eleonora. Grazie per le nostre chiacchierate al telefono, per i tuoi consigli e, soprattutto, perché so di poter sempre contare su di te, nonostante la distanza che ci separa.

Finally, I would like to thank all my friends and acquaintances who I could not mention here and who have supported me, thank you all!

Contents

Chapter 1

Introduction

1.1 Contexte et positionnement

Cette thèse de doctorat porte sur la conception, l'analyse et l'implémentation de méthodes de type Lagrange-Projection (LP) pour des systèmes d'équations aux dérivées partielles hyperboliques. Plus précisément, nous considérons plusieurs modèles mathématiques et nous adaptons la stratégie LP de manière à satisfaire certaines propriétés ou caractéristiques essentielles. Parmi ces propriétés, nous nous intéressons notamment a des propriétés équilibre ou "well-balanced", d'ordre d'approximation élevé ou encore de stabilité à grands pas de temps.

En ce qui concerne les modèles mathématiques, nous considérons principalement les écoulements géophysiques à l'aide de modèles hydrodynamiques. Nous sommes particulièrement intéressés par les applications au transport de sédiments, aux processus d'érosion-dépôt des sédiments et aux écoulements stratifiés. Des modèles avancés basés sur les équations en eaux peu profondes (à une ou deux couches) ont été utilisés pour prendre en compte ces phénomènes. En outre, comme point de départ de cette thèse, nous avons également construit des schémas numériques pour la simulation numérique du flux sanguin dans les artères humaines.

Cette thèse a été réalisée au Laboratoire de Mathématiques de Versailles de l'Université de Versailles Saint-Quentin-en-Yvelines et a été soutenu par l'attribution d'une allocation de recherche Région Île-de-France (subvention DIM MathInnov).

Les chapitres 3 et 5 ont été réalisés en collaboration avec M. J. Castro Dìaz et T. Morales de Luna, membres du groupe de recherche EDANYA de l'Université de Malaga. Ces chapitres ont été partiellement soutenus par le Programme Visibilité Scientifique Junior du FMJH (Fondation Mathématique Jacques Hadamard), le Spanish Government et FEDER à travers les projets de recherche RTI2018-096064-B-C1 et RTI2018-096064-B-C2, le projet de recherche P18-RT-3163 de la Junta de Andalucía, le projet de recherche UMA18-FEDERJA-16 de la Junta de Andalucia-FEDER-University of Málaga et l'Université de Málaga.

1.1.1 Modèles mathématiques

Dans cette thèse, nous considérons des équations aux dérivées partielles non linéaires de la forme suivante,

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \mathbf{B}(\mathbf{Q})\partial_x \mathbf{Q} = \mathbf{S}(\mathbf{Q}) \tag{1.1.1}$$

où $\mathbf{Q}(x,t) \in \Omega$ est le vecteur des inconnues dépendant du temps $t > 0$ et de l'espace $x \in \mathbb{R}$, avec $\Omega \subset \mathbb{R}^n$ un ensemble ouvert et convexe. Après, $\mathbf{F} : \Omega \to \mathbb{R}^n$ est le flux physique, $\mathbf{B}(\mathbf{Q})\partial_x \mathbf{Q}$ est le terme non conservatif du modèle avec $\mathbf{B}(\mathbf{Q}) : \Omega \to \mathcal{M}_n(\mathbb{R})$ une matrice et, enfin, $\mathbf{S} : \Omega \to \mathbb{R}^n$ est le terme source. En particulier, si $\mathbf{B}(\mathbf{Q})$ est la matrice nulle, les équations

(1.1.1) se réduisent à un système de lois de conservation avec terme source. Si, en plus, il n'y a pas de terme source $\mathbf{S}$, on obtient un système de lois de conservation [17]. Le système (1.1.1) peut aussi être reformulé comme suit,

$$\partial_t \mathbf{Q} + \mathbf{A}(\mathbf{Q})\partial_x \mathbf{Q} = \mathbf{S}(\mathbf{Q}) \tag{1.1.2}$$

où nous avons introduit la matrice non-conservative

$$\mathbf{A}(\mathbf{Q}) = \frac{\partial \mathbf{F}(\mathbf{Q})}{\partial \mathbf{Q}} + \mathbf{B}(\mathbf{Q}).$$

On dit que le système (1.1.2) est *hyperbolique* si $\mathbf{A}$ est $\mathbb{R}$-diagonalisable et *strictement hyperbolique* si, en plus, les valeurs propres sont toutes distinctes [63]. Les équations aux dérivées partielles hyperboliques font l'objet d'études et d'analyses intenses car de nombreuses applications existent en biologie, aérodynamique, dynamique des fluides et optique, pour n'en citer que quelques-unes. À titre d'exemple, un système hyperbolique très connu est donné par les équations d'Euler, qui modélisent un matériau compressible (comme un gaz ou un liquide) [63].

Comme des opérateurs différentiels sont présents dans les systèmes de forme (1.1.1), nous sommes implicitement en train de supposer que la solution est suffisamment régulière pour qu'il soit possible d'appliquer de tels opérateurs. Cependant, il est bien connu que les solutions du système (1.1.1) peuvent développer des discontinuités en un temps fini, même si des conditions initiales régulières sont envisagées [46]. Par conséquent, la notion de solutions classiques (c'est-à-dire des fonctions différentiables avec des dérivées partielles continues) doit être généralisée au cas où des discontinuités pourraient être présentes dans la solution. A cette fin, nous introduisons brièvement le concept de *solution faible* [38].

En commençant par un système de lois de conservation (1.1.1), notamment avec $\mathbf{S}(\mathbf{Q}) = \mathbf{0}$ et $\mathbf{B}(\mathbf{Q})$ matrice nulle, nous intégrons le système (1.1.1) dans le volume de contrôle $[x_1, x_2] \times [t_0, t_1]$, ce qui donne la forme intégrale des lois de conservation,

$$\int_{x_1}^{x_2} \mathbf{Q}(x, t_1)\mathrm{d}x = \int_{x_1}^{x_2} \mathbf{Q}(x, t_0)\mathrm{d}x - \left(\int_{t_0}^{t_1} \mathbf{F}(\mathbf{Q}(x_2, t))\mathrm{d}t - \int_{t_0}^{t_1} \mathbf{F}(\mathbf{Q}(x_1, t))\mathrm{d}t \right). \tag{1.1.3}$$

Cette relation exprime que la différence de la quantité $\mathbf{U}$ dans l'intervalle $[x_1, x_2]$ au temps t_1 et celle au temps t_0 est égale à une différence des intégrales temporelles du flux aux points x_2 et x_1. Dans ce cadre, nous définissons une solution faible comme une fonction de $\mathbf{Q}$ qui satisfait (1.1.3) et donc, qui n'a plus besoin d'être continue [63]. En particulier, une discontinuité d'une solution faible satisfait aux conditions de saut de Rankine-Hugoniot, qui donnent la vitesse de la discontinuité [63].

Cependant, étant donnée une condition initiale, une solution faible n'est généralement pas unique. Par conséquent, un critère devrait être envisagé pour sélectionner une solution qui soit cohérente avec la physique du problème, notamment en recherchant une condition d'entropie. Dans cette thèse, nous ne nous concentrons pas sur un tel problème et nous nous référons simplement à [46, 47] et aux références qui s'y trouvent pour plus de détails.

Même si la notion de solution faible a été définie pour les systèmes de lois de conservation, une telle définition ne peut plus être utilisée dans le cas du système non conservatif complet (1.1.1). En effet, puisque des masses de Dirac peuvent apparaître en présence de discontinuités, le produit non conservatif $\mathbf{B}(\mathbf{Q})\partial_x \mathbf{Q}$ n'aurait plus de sens dans le cadre distributionnel. En outre, ce problème peut également apparaître si le terme source est de la forme $\mathbf{S}(\mathbf{Q}) = \tilde{\mathbf{S}}(\mathbf{Q})\partial_x \sigma$ avec $\tilde{\mathbf{S}}(\mathbf{Q}) \neq 0$ et $\sigma : \mathbb{R} \to \mathbb{R}$ une fonction discontinue connue. Afin de contourner ce problème, Dal Maso, LeFloch et Murat [28] ont développé une théorie : elle permet de définir une solution faible en supposant que $\mathbf{Q}$ est une fonction à variation bornée, de sorte

que le produit $\mathbf{B}(\mathbf{Q})\partial_x\mathbf{Q}$ (ou $\tilde{\mathbf{S}}(\mathbf{Q})\partial_x\sigma$) aurait un sens en tant que mesure localement bornée. Récapitulons les lignes directrices d'une telle théorie. À cette fin, nous considérons directement les équations de forme (1.1.2) sans terme source. En effet, si un terme source de la forme $\tilde{\mathbf{S}}(\mathbf{Q})\partial_x\sigma$ est présent, il peut être facilement inclus en considérant l'équation supplémentaire $\partial_t\sigma = 0$ et en ajoutant σ au vecteur des inconnues. Ainsi, nous cherchons à définir l'intégrale

$$\int_{x_1}^{x_2} \mathbf{A}(\mathbf{Q}(x,t))\partial_x\mathbf{Q}(x,t)\mathrm{d}x\mathrm{d}t \tag{1.1.4}$$

lorsque $\mathbf{Q}$ est discontinue. Tout d'abord, nous devons définir une famille de chemins continus de Lipschitz $\Phi : [0,1]\times\Omega\times\Omega \to \Omega$ satisfaisant les propriétés suivantes :

$$\Phi(0;\mathbf{Q}_L,\mathbf{Q}_R) = \mathbf{Q}_L, \quad \Phi(1;\mathbf{Q}_L,\mathbf{Q}_R) = \mathbf{Q}_R, \quad \text{et} \quad \Phi(s;\ \mathbf{Q},\mathbf{Q}) = \mathbf{Q}.$$

À titre d'exemple, le chemin le plus facile est celui des lignes droites,

$$\Phi(s;\mathbf{Q}_L,\mathbf{Q}_R) = \mathbf{Q}_L + s(\mathbf{Q}_R - \mathbf{Q}_L).$$

Ensuite, nous définissons l'intégrale (1.1.4) par

$$\fint_{x_1}^{x_2} \mathbf{A}(\mathbf{Q}(x))\partial_x\mathbf{Q}(x)\mathrm{d}x = \int_{x_1}^{x_2} \mathbf{A}(\mathbf{Q}(x))\partial_x\mathbf{Q}(x)\mathrm{d}x + \sum_l \int_0^1 \mathbf{A}(\Phi(s;\mathbf{Q}_l^-,\mathbf{Q}_l^+))\frac{\partial\Phi}{\partial s}(s;\mathbf{Q}_l^-,\mathbf{Q}_l^+)\mathrm{d}s$$

où $\mathbf{Q}_l^-$ et $\mathbf{Q}_l^+$ sont respectivement la limite de $\mathbf{Q}$ à gauche et à droite de la $l$ième discontinuité. Ainsi, nous disons qu'une solution faible du système (1.1.2) est une fonction qui satisfait à

$$\int_{x_1}^{x_2} \mathbf{Q}(x,t_1)\mathrm{d}x = \int_{x_1}^{x_2} \mathbf{Q}(x,t_0)\mathrm{d}x - \int_{t_0}^{t_1} \fint_{x_1}^{x_2} \mathbf{A}(\mathbf{Q}(x,t))\partial_x\mathbf{Q}(x,t)\mathrm{d}x\mathrm{d}t.$$

Cependant, un problème crucial est maintenant donné par le choix du chemin. En effet, la solution pourrait changer en fonction du chemin choisi. En outre, même si le "bon" chemin est choisi, la solution numérique pourrait converger vers la mauvaise solution et non vers la solution physique, en raison de la viscosité numérique de la méthode. Selon [48, 60], la définition du chemin correct pourrait résider dans la régularisation convective-diffusive du système. En bref, et en se référant également à [17], si le système (1.1.2) est la limite de diffusion à zéro du système suivant,

$$\partial_t\mathbf{Q} + \mathbf{A}(\mathbf{Q})\partial_x\mathbf{Q} = \lambda\partial_x(\mathbf{D}(\mathbf{Q})\partial_x\mathbf{Q}) \tag{1.1.5}$$

avec $\mathbf{D}(\mathbf{Q})$ una matrice de diffusion, le choix du chemin correct est lié au profil visqueux du système (1.1.5). Encore une fois, ici, nous ne donnons pas plus de détails mais nous nous référons simplement à [17, 43, 48, 60].

Nous soulignons que, dans cette thèse, nous étudions d'une part des systèmes de lois de conservation avec termes sources, à savoir le système d'écoulement sanguin dans le chapitre 2 et les équations en eaux peu profondes dans le chapitre 6. Ensuite, du chapitre 3 au chapitre 5, on ne considère que des modèles non conservatifs, bien que toujours liés au système en eaux peu profondes. Par conséquent, comme presque tous les modèles examinés dans cette thèse sont basés sur le système en eaux peu profondes, on le présente ici brièvement. Ce modèle a fait l'objet d'études et d'analyses approfondies car, malgré sa simplicité, il est efficace pour la simulation numérique des écoulements de fluides dans le cadre de la dynamique atmosphérique et océanique [66, 65]. En outre, il est utilisé comme base pour simuler des phénomènes complexes tels que le transport de sédiments [33], les courants de turbidité [55], les écoulements

multicouches [1, 3], les tsunamis [26], les ruptures de barrage [64], les glissements de terrain [45], les inondations [31] et outre encore.

Le système en eaux peu profondes à une dimension d'espace est également appelé système de Saint-Venant car il a été décrit pour la première fois par Adhémar Jean Claude Barré de Saint-Venant en 1843. Ce système a été dérivé des équations de Navier-Stokes [62] sous les hypothèses suivantes : l'échelle verticale est beaucoup plus petite que l'échelle horizontale, le fluide est homogène et incompressible, la pression est hydrostatique, il n'y a pas de forces visqueuses et la vitesse du fluide ne doit dépendre que de l'espace x et du temps t et non de la profondeur [66, 36]. En tenant compte de toutes ces hypothèses et en intégrant les équations de Navier-Stokes sur la profondeur, nous obtenons le système en eaux peu profondes à une dimension d'espace :

$$\begin{cases} \partial_t h + \partial_x(hu) = 0 \\ \partial_t(hu) + \partial_x(hu^2 + \dfrac{gh^2}{2}) = -gh\partial_x z. \end{cases} \tag{1.1.6}$$

En particulier, la première et deuxième équation sont respectivement l'équation de conservation de la masse et de la quantité de mouvement, où $h(x,t)$ est la hauteur totale de la colonne d'eau, $u(x,t)$ est la vitesse moyenne et $z(x)$ est l'élévation du lit. Ensuite, la pression est donnée par $p = \dfrac{gh^2}{2}$ où g représente l'accélération gravitationnelle. Le système (1.1.6) est clairement un système de lois de conservation avec terme source de forme (1.1.1), où

$$\mathbf{Q} = \begin{pmatrix} h \\ hu \end{pmatrix}, \quad \mathbf{F}(\mathbf{Q}) = \begin{pmatrix} hu \\ hu^2 + \dfrac{gh^2}{2} \end{pmatrix}, \quad \mathbf{S}(\mathbf{Q}) = \begin{pmatrix} 0 \\ -gh\partial_x z \end{pmatrix}$$

et $\mathbf{B}$ matrice nulle. De plus, il est simple de prouver qu'il s'agit d'un système strictement hyperbolique avec les valeurs propres réelles $u \pm c$, où $c = \sqrt{\partial_h p} = \sqrt{gh}$ est la vitesse du son. Ainsi, la condition de Courant-Friedrichs-Lewy (CFL) sur le pas de temps pour un schéma explicite de type Godunov appliqué au système de Saint-Venant (1.1.6) est généralement donnée par

$$\Delta t \leq \frac{\Delta x}{2} \frac{1}{\max\{|u| + c\}}.$$

Pour plus de détails sur les équations en eaux peu profondes et leur approximation numérique, voir par exemple [66, 2, 8, 36, 35, 23].

Enfin, dans le but de concevoir des schémas numériques stables et comme mentionné précédemment, nous nous concentrons également sur les solutions stationnaires du modèle (1.1.1), qui satisfont la relation suivante

$$\partial_x \mathbf{F}(\mathbf{Q}) + \mathbf{B}(\mathbf{Q})\partial_x \mathbf{Q} = \mathbf{S}(\mathbf{Q}). \tag{1.1.7}$$

Par conséquent, nous recherchons des méthodes numériques capables de préserver ces états stationnaires, sinon il est bien connu que des instabilités peuvent apparaître dans les résultats numériques lorsque la solution est proche d'une état stationnaire. Un tel problème a été présenté pour la première fois dans le travail [5] de Bermudez et Vazquez pour le système en eaux peu profondes. Ils y ont nommé *C-property* la capacité d'un schéma à préserver les solutions stationnaires à vitesse nulle. Cette propriété a été renommée *équilibre* (*well-balanced* en anglais) par Greenberg et LeRoux dans leur article [41]. Plus tard, Gosse [39] a étendu la notion de cette propriété en considérant un schéma capable de préserver également les solutions stationnaires mobiles, à savoir celles dont la vitesse est non nulle. Dans ce cas, on parle de méthode numérique *complètement équilibre* (*fully well-balanced* en anglais).

Pour fixer les idées, considérons le système de Saint-Venant (1.3.6). Ses états stationnaires sont donnés par

$$hu = \text{constant} = q_0 \quad \text{et} \quad \frac{q_0^2}{2h^2} + g(h+z) = \text{constant}.$$

Si nous prenons ceux dont la vitesse est nulle, on retrouve la solution stationnaire "lac au repos" ("lake at rest" en anglais), qui se lit comme suit

$$u = 0 \quad \text{et} \quad h + z = \text{constant}.$$

Il existe de nombreux travaux sur les méthodes équilibre et complètement équilibre pour le système de Saint-Venant. Par exemple, dans [2], les auteurs ont introduit la stratégie de reconstruction hydrostatique afin d'avoir une méthode numérique capable de préserver la solution stationnaire "lac au repos" du système en eaux peu profondes. D'autres références intéressantes pour les schémas équilibres du premier ordre appliqués aux systèmes de lois de conservation avec terme source sont [6, 30]. Alors que pour les schémas équilibre de second ordre (ou plus), nous nous référons à [8, 7, 13, 18, 57, 51, 56, 59]. Des méthodes capables de préserver un état stationnaire donné (pas nécessairement avec une vélocité nulle) ont également été conçues. Voir par exemple [44], où les auteurs ont considéré les équations d'Euler avec la gravité. À cet égard, nous renvoyons également à [4], où une méthode équilibre d'ordre supérieur pour les lois de conservation avec terme source multidimensionnelles a été présentée. Les auteurs y expliquent comment préserver exactement une solution générale qui est connue a priori et qui n'est pas nécessairement un état stationnaire du modèle. Quant à la propriété équilibre pour les systèmes non-conservatifs, on peut mentionner le papier [17]. Ensuite, lorsqu'il s'agit de la propriété complètement équilibre pour le système en eaux peu profondes, nous pourrions nous référer à [52]. Dans ce papier, les auteurs ont décrit un schéma basé sur une définition particulière d'un solveur de Riemann approché. En effet, il a été construit de manière à ce que le schéma associé de type Godunov satisfasse la propriété d'équilibre. D'autres schémas complètement équilibre ont été décrits par exemple dans [53, 10, 16, 9, 37].

Dans cette thèse, nous ne nous intéressons qu'à la propriété équilibre (et non complètement équilibre) car elle se révèle généralement plus facile à préserver tout en améliorant considérablement les résultats numériques. En particulier, tous nos schémas sont basés sur la décomposition acoustique-transport (ou de type Lagrange-projection). Nous verrons que dans la plupart des cas, nous construirons un solveur de Riemann approché équilibre pour les équations acoustiques afin que la méthode numérique complète soit bien équilibrée. En effet, en général, nous n'avons pas à modifier la discrétisation numérique de la partie de transport du système.

1.1.2 Décomposition de Lagrange-projection ou acoustique-transport

Comme déjà mentionné, la décomposition de type Lagrange-projection et le splitting acoustique-transport sont des concepts clés de cette thèse de doctorat. En particulier, ils nous permettent de découpler les effets de compressibilité des phénomènes de transport du modèle mathématique considéré. Par conséquent, nous obtenons deux systèmes d'équations différents que nous devons résoudre numériquement [21, 14, 19]. En pratique, nous pouvons résumer cette stratégie numérique comme suit:

1. Résoudre (explicitement ou implicitement) le système acoustique ;
2. Utiliser la solution acoustique comme condition initiale pour résoudre le système de transport.

Par souci de clarté, on présente une telle stratégie pour le système de Saint-Venant (1.1.6). Après avoir appliqué la dérivée d'une fonction composée aux dérivées spatiales comme suit

$$\begin{cases} \partial_t h + h\partial_x u + u\partial_x h = 0 \\ \partial_t (hu) + hu\partial_x u + u\partial_x hu + \partial_x p = -gh\partial_x z, \end{cases}$$

nous découplons les différents phénomènes du modèle, obtenant ainsi les systèmes acoustiques et de transport :

$$\begin{cases} \partial_t h + h\partial_x u = 0 \\ \partial_t (hu) + hu\partial_x u + \partial_x p = -gh\partial_x z, \end{cases} \tag{1.1.8}$$

et

$$\begin{cases} \partial_t h + u\partial_x h = 0 \\ \partial_t (hu) + u\partial_x (hu) = 0. \end{cases} \tag{1.1.9}$$

En résumé, une fois qu'on a obtenu la solution du système acoustique (1.1.8), nous l'utilisons comme condition initiale pour pouvoir résoudre le système (1.1.9) et trouver la solution finale. Cependant, il est clair que les équations acoustiques (1.1.8) sont sous une forme non conservative, contrairement au système de Saint-Venant (1.1.6). Par conséquent, plutôt que de considérer un tel système, nous préférons faire une reformulation, en introduisant le volume spécifique $\tau = \dfrac{1}{h}$ et la variable de masse m donnée par $\partial_m = \tau\partial_x$. Ainsi, nous obtenons le système de lois de conservation avec terme source suivant

$$\begin{cases} \partial_t \tau - \partial_m u = 0 \\ \partial_t u + \partial_m p = -\dfrac{g}{\tau}\partial_m z. \end{cases} \tag{1.1.10}$$

Nous pouvons facilement prouver que ce système est strictement hyperbolique avec des valeurs propres données par $\pm hc$. De plus, pour le système de transport (1.1.9), la vitesse u est la seule valeur propre. Il est donc clair que le splitting des phénomènes acoustiques et de transport a conduit à la décomposition des ondes acoustiques et matérielles du système de Saint-Venant (1.1.6). Ces ondes jouent évidemment un rôle important dans la définition de la condition CFL pour le pas de temps Δt. En effet, ayant maintenant deux systèmes différents, (1.1.10) et (1.1.9), nous devons imposer deux conditions CFL différentes sur le pas de temps, respectivement

$$\Delta t \leq \frac{\Delta m}{2\max\{hc\}} \tag{1.1.11}$$

et

$$\Delta t \leq \frac{\Delta x}{2\max\{|u|\}}. \tag{1.1.12}$$

Il faut alors prendre la valeur minimale de Δt donnée par les conditions (1.1.11) et (1.1.12). L'intérêt de disposer de deux conditions CFL différentes est lié au fait qu'il existe des situations dans lesquelles les ondes acoustiques sont beaucoup plus rapides que les ondes de transport. Cela se produit dans les régimes subsoniques ou à proximité d'écoulements à faible nombre de Froude, lorsque la valeur absolue du nombre de Froude $F = \dfrac{u}{c}$ est inférieure à un. Dans de telles situations, la condition CFL habituelle pour le pas de temps des schémas de type Godunov est entrainée par les ondes acoustiques rapides et peut, donc, être très restrictive. Ainsi, en exploitant une approximation implicite du système acoustique, nous pourrions obtenir un schéma IMplicit-EXplicit (IMEX) très naturel et à grand pas de temps avec une restriction CFL basée uniquement sur les ondes de transport lentes [27].

Présentons maintenant une interprétation différente du splitting acoustique-transport, à savoir l'approche de type Lagrange-projection. L'idée est la suivante. En commençant par le modèle mathématique choisi, nous le reformulons en coordonnées lagrangiennes. Une fois que la solution lagrangienne a été trouvée, nous devons la projeter en coordonnées eulériennes [54]. Ainsi, nous avons à nouveau une stratégie numérique composée de deux étapes: celle lagrangienne et celle de projection. Ces deux étapes correspondraient respectivement aux étapes acoustique et de transport que nous avons décrites précédemment. En pratique, nous verrons que ces deux stratégies peuvent être considérées comme les deux faces d'une même pièce, même si elles ne sont pas toujours strictement équivalentes. Pour illustrer ce lien, considérons à nouveau le système de Saint-Venant (1.1.6) à une dimension d'espace. Dans un premier temps, il faut donc le reformuler en coordonnées lagrangiennes. Pour pouvoir le faire, nous suivons la particule de fluide dans la position ξ et nous définissons les courbes caractéristiques

$$\begin{cases} \dfrac{\partial x}{\partial t}(\xi,t) = u(x(\xi,t),t) \\ x(\xi,0) = \xi, \end{cases} \tag{1.1.13}$$

qui définissent la trajectoire : $t \to x(\xi,t)$ de ξ au fil du temps. Ainsi, toute fonction : $(x,t) \to \varphi(x,t)$ en coordonnées eulériennes peut être écrite en coordonnées lagrangiennes,

$$\overline{\varphi}(\xi,t) = \varphi(x(\xi,t),t).$$

Ensuite, nous introduisons le rapport de volume $L(\xi,t)$, qui est défini par

$$L(\xi,t) = \frac{\partial x}{\partial \xi}(\xi,t) \tag{1.1.14}$$

et qui satisfait

$$\begin{cases} \dfrac{\partial L}{\partial t}(\xi,t) = \partial_\xi u(x(\xi,t),t) \\ L(\xi,0) = 1. \end{cases} \tag{1.1.15}$$

Par conséquent, les dérivées temporelles et spatiales d'une fonction φ en coordonnées lagrangiennes sont données par

$$\partial_t \overline{\varphi}(\xi,t) = \partial_t \varphi(x,t) + u(x,t)\partial_x \varphi(x,t) \quad \text{et} \quad \partial_\xi \overline{\varphi}(\xi,t) = L(\xi,t)\partial_x \varphi(x,t).$$

En laissant les calculs au lecteur, nous présentons directement les équations de Saint-Venant en coordonnées lagrangiennes,

$$\begin{cases} \partial_t(L\bar{h}) = 0 \\ \partial_t(L\overline{hu}) + \partial_\xi \bar{p} = -g\bar{h}\partial_\xi \bar{z}. \end{cases} \tag{1.1.16}$$

Une fois trouvée la solution lagrangienne du système (1.1.16), il suffit donc de la projeter en coordonnées eulériennes. Pour plus de détails sur les coordonnées lagrangiennes à 2 ou 3 dimensions d'espace, nous nous référons à [29, 49, 50].

À ce stade, nous pouvons remarquer que, dans la deuxième équation du système (1.1.16), le flux physique est uniquement donné par le terme de pression. Néanmoins, la relation entre le système (1.1.16) et le système acoustique (1.1.10) pourrait ne pas être tout à fait claire. Afin de l'illustrer, nous commençons par remarquer que $L\bar{h}$ ne dépend pas du temps, à savoir

$$(L\bar{h})(\xi,t) = (L\bar{h})(\xi,0) = h(\xi,0) = h_0 \quad \text{et} \quad L(\xi,t) = \frac{h_0}{\bar{h}(\xi,t)},$$

grâce auquel on trouve

$$\begin{cases} \partial_t h_0 = 0 \\ \partial_t L - \partial_\xi \bar{u} = 0 \\ \partial_t (h_0 \bar{u}) + \partial_\xi \bar{p} = -g\bar{h}\partial_\xi \bar{z}. \end{cases}$$

Enfin, il suffit de réintroduire les variables $\bar{\tau} = \dfrac{1}{\bar{h}}$ et m telle que $\partial_m = \dfrac{1}{h_0}\partial_\xi$, pour que nous obtenions

$$\begin{cases} \partial_t \bar{\tau} - \partial_m \bar{u} = 0 \\ \partial_t \bar{u} + \partial_m \bar{p} = -\dfrac{g}{\bar{\tau}}\partial_m \bar{z}, \end{cases} \tag{1.1.17}$$

qui rappelle le système acoustique (1.1.10) manifestement. Ainsi, même s'il est clair que les deux systèmes (1.1.17) et (1.1.10) ne sont pas strictement les mêmes car ils ont été obtenus selon des procédures différentes, lorsqu'il s'agit de l'approximation numérique, nous utiliserons les ingrédients de chacune d'entre elles dans le but de construire des schémas numériques performants.

Nous avons déjà clarifié pourquoi les méthodes de type Lagrange-projection sont particulièrement utiles lorsqu'on considère les régimes subsoniques. Soulignons toutefois que cette classe de méthodes ouvre généralement la voie à de nombreuses nouvelles possibilités d'un point de vue numérique. En effet, le fait d'avoir deux ensembles d'équations différents (dans le splitting acoustique-transport) ou de travailler en coordonnées lagrangiennes (dans l'approche LP) signifie que les propriétés des méthodes numériques pourraient être préservées différemment (et plus facilement).

Donnons donc quelques exemples de méthodes de type Lagrange-projection (ou acoustique-transport). Dans [19, 20], les auteurs ont décrit un schéma de type Lagrange-projection tout-régime, respectivement en 1D et 2D. Par tout-régime, ils désignent un schéma numérique capable de calculer des solutions approchées précises avec un maillage et un pas de temps beaucoup plus grands que le nombre de Mach (analogue du nombre de Froude dans le cadre des eaux peu profondes). Par conséquent, la méthode résultante est également implicite-explicite et peut conduire à des simulations rapides. Un autre schéma semi-implicite a été développé dans [21], où la propriété équilibre du schéma a également été analysée, ce qui signifie que les solutions stationnaires à vitesse nulle du modèle mathématique sont préservées. Ce problème a été approfondi dans [14], où les auteurs ont décrit une méthode LP complètement équilibre. Ensuite, ce schéma a également été étendu à un ordre de précision arbitraire dans [54], en utilisant des reconstructions polynomiales et la stratégie Runge-Kutta. Les méthodes de type Lagrange-projection d'ordre élevé ont également été traitées dans le papier [32], où l'ordre élevé de précision a été atteint en utilisant des techniques différentes, notamment les développements de Taylor et la procédure de Cauchy-Kovalevskaya. Une autre référence intéressante est [58], où un schéma de type Lagrange-projection implicite-explicite et d'ordre élévé a été développé. En particulier, les auteurs ont obtenu un ordre de précision élevé en espace en utilisant un schéma de type Galerkin-discontinu, tandis que pour l'approximation dans le temps, ils ont considéré la méthode d'Euler "backward". Enfin, des méthodes de type Lagrange-projection préservant l'asymptotique ont également été conçues, ce qui signifie que le schéma est capable de reproduire au niveau discret le comportement asymptotique satisfait par les équations continues, voir [22, 12]. De plus, les méthodes numériques satisfont également une inégalité d'entropie entièrement discrète.

Dans cette thèse, nous sommes particulièrement attentifs à concevoir de méthodes numériques de type Lagrange-projection dotées de la propriété équilibre, donc seules les solutions stationnaires à vitesse nulle sont préservées. Une autre propriété fondamentale des méthodes numériques est qu'elles doivent être capables de préserver la positivité stricte de la solution

[34] : par exemple, de la hauteur d'eau dans les modèles en eaux peu profondes [8], de la section transversale du vaisseau dans les équations de flux sanguin [37], de la densité dans les équations de dynamique des gaz [34], et ainsi de suite.

Après, il est clair que la solution numérique doit être une bonne approximation de la solution physique et ne pas être trop diffusive. Cependant, en fonction du modèle mathématique et de la méthode numérique, cette requête n'est pas du tout triviale. En effet, les schémas du premier ordre sont généralement très diffusifs et peuvent nécessiter l'utilisation d'un maillage très fin pour obtenir une approximation acceptable de la solution, ce qui entraîne un coût de calcul excessif. Cela se produit par exemple lors de l'utilisation de schémas numériques basés sur des approches qui négligent les ondes intermédiaires d'un modèle, ne considérant que la vitesse de propagation des ondes externes, comme dans le solveur HLL (Harten-Lax-van Leer) [42]. Il est clair qu'un tel problème ne se pose pas pour les modèles avec une structure à deux ondes. Néanmoins, ici, nous essayons de concevoir des schémas numériques qui prennent en compte la structure complète de la solution. Nous verrons que cette requête est particulièrement simplifiée par l'exploitation d'une version relaxée [61, 11] du système acoustique. Enfin, avec tout cela à l'esprit et dans un deuxième temps, nous concevons également une extension d'ordre supérieur des méthodes numériques, visant à une convergence plus rapide de la solution numérique vers la solution physique [63].

1.2 Présentation des travaux de thèse

Cette thèse est structurée comme suit.

Dans le **chapitre 2**, nous appliquons la stratégie de type Lagrange-projection aux équations du flux sanguin avec rigidité artérielle et surface de la section transversale à l'équilibre non constantes en espace, obtenant un système de lois de conservation avec terme source. En raison de ces derniers paramètres, il n'est pas simple de préserver la propriété équilibre, et nous nous concentrons sur la préservation de la solution stationnaire "l'homme au repos éternel". Nous montrons deux façons de la préserver, l'une basée sur une modification du solveur de Riemann approché pour le système acoustique et la seconde sur une stratégie de reconstruction hydrostatique [2]. Des efforts pour étendre la stratégie numérique au second ordre de précision sont également faits en utilisant des techniques classiques comme la reconstruction polynomiale [63] et le schéma de Runge-Kutta [40]. Une fois de plus, d'autres modifications sont appliquées aux schémas numériques afin de préserver la propriété équilibre, à cette fin, nous utiliserons la notion de fluctuations [54].

Puis, à partir du chapitre 3, seulement les équations de Saint-Venant et les modèles associés sont étudiés. En particulier, le **chapitre 3** est consacré à l'approximation numérique du modèle non conservatif de Saint-Venant à deux couches [1]. Autrement dit, on considère que le fluide est composé de deux couches superposées de liquides non miscibles. La difficulté liée à ce modèle est de deux natures. D'une part, la présence de deux vitesses différentes rend moins évidente l'utilisation d'une stratégie de type Lagrange-projection. D'autre part, non seulement le modèle n'est pas conservatif mais il n'est aussi que conditionnellement hyperbolique, ce qui rend la simulation numérique de ce système généralement plus difficile. Néanmoins, nous avons développé une méthode de type Lagrange-projection implicite-explicite équilibre pour ce système et obtenu des résultats encourageants. En particulier, nous proposons un solveur de Riemann approché pour le système acoustique et l'utilisons pour le schéma associé de type Godunov. Une telle méthode peut également être comprise comme une approximation du système lagrangien. Enfin, la version implicite-explicite du schéma est facilement obtenue

en résolvant un système linéaire dans l'étape acoustique.

Dans le **chapitre 4**, nous étudions le couplage du système de Saint-Venant avec l'équation d'Exner, où cette dernière est utilisée pour simuler le transport de sédiments du lit [33]. Le système non-conservatif qui en résulte se révèle particulièrement difficile à approcher. En effet, si une approche découplée est utilisée, des instabilités pourraient être facilement trouvées dans les simulations numériques en raison de la structure propre du modèle complet. Nous étudions un tel problème et nous concevons trois différents schémas de type Lagrange-projection équilibres pour ce système. En particulier, l'équation d'Exner pourrait être complètement prise en compte soit dans une seule des deux étapes soit dans les deux. Dans tous les cas, nous proposons une solution de Riemann approchée pour les équations acoustiques. De plus, ces trois stratégies sont également étendues au second ordre de précision. Enfin, à la fin de ce chapitre, nous présentons également la version bidimensionnelle du modèle et de l'une des stratégies numériques.

Le **chapitre 5** est une extension naturelle du travail présenté dans le chapitre 4. En effet, nous considérons maintenant le transport de sédiments en général, ce qui signifie que les particules de sédiments pourraient non seulement se déplacer le long du fond mais aussi être en suspension dans l'eau si une fraction suffisamment fine est présente. De plus, puisque nous savons que le temps caractéristique associé aux sédiments est beaucoup plus grand que celui correspondant au fluide, nous construisons une méthode implicite-explicite afin d'obtenir des simulations rapides.

Enfin, dans cette thèse et à ce stade, nous avons considéré soit des schémas implicites-explicites du premier ordre, soit des méthodes explicites du second ordre. Ce n'est que dans le **chapitre 6** que nous développons une approche très générale permettant de proposer un schéma implicite-explicite du second ordre pour le système de Saint-Venant. En particulier, le second ordre de précision est atteint en utilisant des techniques différentes par rapport aux schémas numériques du second ordre présentés dans les chapitres précédents, à savoir en utilisant des developpements de Taylor et la procédure de Cauchy-Kovalevskaya. Nous soulignons que nous n'avons besoin que de résoudre des systèmes linéaires afin de trouver la solution, de sorte que le coût de calcul n'est pas élevé. Cependant, il est bien connu que des oscillations parasites peuvent apparaître en présence de discontinuités lors de l'utilisation de schémas d'ordre élevé. Afin de supprimer ces oscillations, nous utilisons une approche de limitation *a posteriori*, ressemblant à la stratégie MOOD (Multi-dimensional Optimal Order Detection) [24, 25]. Nous soulignons que l'utilisation d'une approche a posteriori nous permet de ne pas modifier la stratégie pour le second ordre de précision et donc les systèmes linéaires. De cette façon, le coût de calcul ne devrait pas augmenter excessivement.

Publications

Les travaux présentés dans ce manuscrit ont été soit publiés, soit soumis.

- A. Del Grosso and C. Chalons. *Second-order well balanced Lagrange-Projection schemes for Blood Flow Equations.* Calcolo 58, 43, 2021. 10.1007/s10092-021-00434-5

- C. Chalons and A. Del Grosso, *A second-order well-balanced Lagrange-projection numerical scheme for Shallow Water Exner equations in 1D and 2D.* 2022. Communications in Mathematical Sciences. 20(7): 1839-1873, 2022. 10.4310/CMS.2022.v20.n7.a3

- C. Chalons and A. Del Grosso, *Exploring different possibilities for second-order well-balanced*

Lagrange-projection numerical schemes applied to shallow water Exner equations. International Journal for Numerical Methods in Fluids. 1- 31, 2022. 10.1002/fld.5064

- A. Del Grosso, M. J. Castro Dìaz, C. Chalons and T. Morales de Luna. *On well-balanced implicit-explicit Lagrange-projection schemes for two-layer shallow water equations.* Soumis dans le Journal "Applied Mathematics and Computation" en Mai 2022.
- A. Del Grosso, M. J. Castro Dìaz, C. Chalons and T. Morales de Luna. *Lagrange-projection methods for shallow water equations with movable bottom and erosion-deposition processes.* Soumis dans le Journal Journal "Communications in Mathematics and Applications" en Mars 2022.
- C. Chalons and A. Del Grosso. *Second-order Well-Balanced Implicit-Explicit Scheme for Systems of Balance Laws.* Soumis dans le Journal "Numerische Mathematik" en Janvier 2022.

Conférences et séminaires

- HYP 2022 - XVIII International Conference on Hyperbolic Problems - *Juin 20-24, 2022.*
- CANUM - 45eme Congrès National d'Analyse Numérique - *Juin 13-17, 2022.*
- SHARK-FV 2022 (Sharing Higher-order Advanced Know-how on Finite Volume) workshop - *Mai 23-27, 2022.*
- Séminaire à l'Institut de Mathématiques de Bordeaux - *Mars 25, 2022.*
- Séminaire doctorant au Laboratoire Amiénois de Mathématique Fondamentale et Appliquée - *Décembre 1, 2021.*
- NumHyp 2021 - Numerical Methods for Hyperbolic Problems - *Juillet 26-30, 2021.*
- SMAI 2021 - 10 ième Biennale Française des Mathématiques Appliquées et Industrielles - *Juin 21-25, 2021.*
- CEDYA 2020 - XXVI Congreso de Ecuaciones Diferenciales y Aplicaciones XVI Congreso de Matemática Aplicada - *Juin 14-18, 2021.*
- CAN-J 2020 - congrès d'analyse numérique pour les jeunes 2020 - *Décembre 3-4 2020.*

SEME

En février 2022, j'ai participé à la Semaine Maths-Entreprises à l'Institut Polytechnique de Paris. En particulier, j'ai travaillé dans le projet "Piloter le traitement de gros volumes de signaux sismiques pour améliorer l'imagerie des sous-sols".

Activités d'enseignement

Au cours de la deuxième (2020-2021) et de la troisième (2021-2022) année du doctorat, j'ai enseigné 64 heures de travaux dirigés en L1 pour le module Mathématiques Générales 2.

Introduction

1.3 Context and positioning

This PhD thesis deals with the design, analysis and implementation of Lagrange-Projection (LP) methods for hyperbolic partial differential equations. More precisely, we consider several mathematical models and we adapt the LP strategy to them in such a way to satisfy essential properties. Among them, we are especially interested in the well-balanced one, in the high order of accuracy and also in the stability of the schemes at large time steps.

Concerning the mathematical models, we mainly consider geophysical flows by means of hydrodynamics models. We are especially interested in applications to sediment transport, sediment erosion-deposition processes and stratified flows. Advanced models based on (one or two-layer) shallow-water equations have been used to take into account these phenomena. Furthermore, as a starting point of this thesis, we have also built numerical schemes for the numerical simulation of blood flow in human arteries.

This PhD thesis has been realized at the Université de Versailles Saint-Quentin-en-Yvelines' Laboratoire de Mathématiques de Versailles and it has been supported by a grant from Région Île-de-France (DIM MathInnov's grant).

Chapters 3 and 5 have been done in collaboration with M. J. Castro Dìaz and T. Morales de Luna, members of the EDANYA research group of the University of Malaga. These chapters have been partially supported by the Junior Scientific Visibility Program offered by the FMJH (Fondation Mathématique Jacques Hadamard), the Spanish Government and FEDER through the coordinated Research projects RTI2018-096064-B-C1 and RTI2018-096064-B-C2, the Junta de Andalucía research project P18-RT-3163, the Junta de Andalucia-FEDER-University of Málaga research project UMA18-FEDERJA-16 and the University of Málaga.

1.3.1 Mathematical models

In this PhD thesis, we consider hyperbolic non-linear partial differential equations of the following form,

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \mathbf{B}(\mathbf{Q})\partial_x \mathbf{Q} = \mathbf{S}(\mathbf{Q}) \tag{1.3.1}$$

where $\mathbf{Q}(x,t) \in \Omega$ is the vector of unknowns depending on the time $t > 0$ and space $x \in \mathbb{R}$, with $\Omega \subset \mathbb{R}^n$ open convex set. Then, $\mathbf{F} : \Omega \to \mathbb{R}^n$ is the physical flux, $\mathbf{B}(\mathbf{Q})\partial_x \mathbf{Q}$ is the non-conservative term of the model with matrix $\mathbf{B} : \Omega \to \mathcal{M}_n(\mathbb{R})$ and, finally, $\mathbf{S} : \Omega \to \mathbb{R}^n$ is the source term. In particular, if $\mathbf{B}$ is the null matrix, equations (1.3.1) reduce themselves to a system of balance laws. In addition, if there is no source term $\mathbf{S}$, we obtain a system of conservation laws [17]. System (1.3.1) can also be reformulated as

$$\partial_t \mathbf{Q} + \mathbf{A}(\mathbf{Q})\partial_x \mathbf{Q} = \mathbf{S}(\mathbf{Q}) \tag{1.3.2}$$

where we have introduced the non-conservative matrix

$$\mathbf{A}(\mathbf{Q}) = \frac{\partial \mathbf{F}(\mathbf{Q})}{\partial \mathbf{Q}} + \mathbf{B}(\mathbf{Q}).$$

System (1.3.2) is *hyperbolic* if $\mathbf{A}$ is $\mathbb{R}$-diagonalizable and strictly hyperbolic if, in addition, the eigenvalues are all distinct [63]. Hyperbolic partial differential equations are the subject of intense study and analysis, as plenty of applications exist in biology, aerodynamics, fluid dynamics and optics, to name but a few. As an example, a very well-known hyperbolic system is given by the Euler equations, which models a compressible material (as a gas or a liquid) [63].

With differential operators being present in systems of form (1.3.1), we are implicitly assuming that the solution is regular enough that it is possible to apply such operators. However, it is well known that solutions of system (1.3.1) may develop discontinuities in a finite time, even if regular initial conditions are envisaged [46]. Hence, the notion of classical solutions (i.e. differentiable functions with continuous partial derivatives) should be generalized to the case in which discontinuities could be present in the solution. For this purpose, we need to briefly introduce the concept of *weak solution* [38].

Starting by a system of conservation laws (1.3.1), namely with $\mathbf{S}(\mathbf{Q}) = \mathbf{0}$ and $\mathbf{B}(\mathbf{Q})$ the null matrix, we integrate system (1.3.1) in the control volume $[x_1, x_2] \times [t_0, t_1]$, obtaining the integral form of conservation laws,

$$\int_{x_1}^{x_2} \mathbf{Q}(x, t_1)\mathrm{d}x = \int_{x_1}^{x_2} \mathbf{Q}(x, t_0)\mathrm{d}x - \left(\int_{t_0}^{t_1} \mathbf{F}(\mathbf{Q}(x_2, t))\mathrm{d}t - \int_{t_0}^{t_1} \mathbf{F}(\mathbf{Q}(x_1, t))\mathrm{d}t \right). \tag{1.3.3}$$

This relation expresses that the difference of the amount of $\mathbf{Q}$ in the interval $[x_1, x_2]$ at time t_1 and at time t_0 is equal to a difference of time integrals of the flux at the points x_2 and x_1. In this framework, we define a weak solution as a function of $\mathbf{Q}$ which satisfies (1.3.3) and thus, which does not have to be continuous anymore [63]. In particular, a discontinuity of a weak solution satisfies the Rankine-Hugoniot jump conditions, which provide the speed of the discontinuity [63].

However, given an initial data, a weak solution is generally not unique. Hence, a criterion should be exploited to select a solution which is consistent with the physics of the problem, namely we could look for an entropy condition. In this PhD thesis, we do not focus on such a problem and we simply refer to [46, 47] and the references therein for more details.

Even if the notion of weak solution has been defined for systems of conservation laws, such a definition cannot longer be used when it comes to the complete non-conservative system (1.3.1). Indeed, since Dirac's delta functions could appear in presence of discontinuities, the non-conservative product $\mathbf{B}(\mathbf{Q})\partial_x \mathbf{Q}$ would no longer make sense in the distributional framework. Furthermore, this problem may also appear if the source term is of the form $\mathbf{S}(\mathbf{Q}) = \tilde{\mathbf{S}}(\mathbf{Q})\partial_x \sigma$ with $\tilde{\mathbf{S}}(\mathbf{Q}) \neq 0$ and $\sigma : \mathbb{R} \to \mathbb{R}$ is a known discontinuous function. In order to overcome this problem, Dal Maso, LeFloch and Murat [28] have developed a theory: it allows to define a weak solution by assuming $\mathbf{Q}$ to be a function with bounded variation, so that the product $\mathbf{B}(\mathbf{Q})\partial_x \mathbf{Q}$ (or $\tilde{\mathbf{S}}(\mathbf{Q})\partial_x \sigma$) has sense as a locally bounded measure. Let us look at the guidelines of such a theory in a nutshell. For this purpose, we directly consider equations of form (1.3.2) with no source term. Indeed, if a source term of the form $\tilde{\mathbf{S}}(\mathbf{Q})\partial_x \sigma$ is present, it can be easily included by considering the additional equation $\partial_t \sigma = 0$ and adding σ to the vector of unknowns. Therefore, we look for a way to define the integral

$$\int_{x_1}^{x_2} \mathbf{A}(\mathbf{Q}(x, t))\partial_x \mathbf{Q}(x, t)\mathrm{d}x\mathrm{d}t \tag{1.3.4}$$

when $\mathbf{Q}$ is discontinuous. Firstly, we need to define a family of Lipschitz continuous paths $\Phi : [0,1] \times \Omega \times \Omega \to \Omega$ satisfying the following properties:

$$\Phi(0; \mathbf{Q}_L, \mathbf{Q}_R) = \mathbf{Q}_L, \quad \Phi(1; \mathbf{Q}_L, \mathbf{Q}_R) = \mathbf{Q}_R, \quad \text{et} \quad \Phi(s;\ \mathbf{Q}, \mathbf{Q}) = \mathbf{Q}.$$

As an example, the easiest path is the one of straight lines,

$$\Phi(s; \mathbf{Q}_L, \mathbf{Q}_R) = \mathbf{Q}_L + s(\mathbf{Q}_R - \mathbf{Q}_L).$$

Then, we define integral (1.3.4) by

$$\text{⨍}_{x_1}^{x_2} \mathbf{A}(\mathbf{Q}(x))\partial_x \mathbf{Q}(x)\mathrm{d}x = \int_{x_1}^{x_2} \mathbf{A}(\mathbf{Q}(x))\partial_x \mathbf{Q}(x)\mathrm{d}x + \sum_l \int_0^1 \mathbf{A}(\Phi(s; \mathbf{Q}_l^-, \mathbf{Q}_l^+))\frac{\partial \Phi}{\partial s}(s; \mathbf{Q}_l^-, \mathbf{Q}_l^+)\mathrm{d}s$$

where $\mathbf{Q}_l^-$ and $\mathbf{Q}_l^+$ are, respectively, the limit of $\mathbf{Q}$ to the left and to the right of the lth discontinuity. As such, we say that a weak solution of system (1.3.2) is a function that satisfies

$$\int_{x_1}^{x_2} \mathbf{Q}(x, t_1)\mathrm{d}x = \int_{x_1}^{x_2} \mathbf{Q}(x, t_0)\mathrm{d}x - \int_{t_0}^{t_1} \text{⨍}_{x_1}^{x_2} \mathbf{A}(\mathbf{Q}(x,t))\partial_x \mathbf{Q}(x,t)\mathrm{d}x\mathrm{d}t.$$

However, a crucial problem is now given by the choice of the path. Namely, the solution could change depending on the chosen path. On top of that, even if the "correct" path is chosen, the numerical output could converge to the wrong solution and not the physical one, due to the numerical viscosity of the method. According to [48, 60], the definition of the correct path could lie in the convective-diffusion regularization of the system. In a nutshell and also referring to [17], if system (1.3.2) is the vanishing diffusion limit of the following system,

$$\partial_t \mathbf{Q} + \mathbf{A}(\mathbf{Q})\partial_x \mathbf{Q} = \lambda \partial_x (\mathbf{D}(\mathbf{Q})\partial_x \mathbf{Q}) \tag{1.3.5}$$

with $\mathbf{D}(\mathbf{Q})$ a diffusion matrix, the choice of the correct path is related to the viscous profile of system (1.3.5). Once again, here we do not provide further details but we simply refer to [17, 43, 48, 60].

We highlight that, in this PhD thesis, on one hand we study systems of balance laws, namely the blood flow system in chapter 2 and the shallow water equations in chapter 6. On the other hand, from chapter 3 to 5, only non-conservative models related to the shallow water system are analyzed. Hence, with almost all the models considered in this PhD thesis being based on the shallow water system, we briefly introduce it here. Such a model has been extensively studied and analyzed as, despite being simple, it is yet effective when it comes to the numerical simulation of fluid flows in atmospheric and oceanic dynamics [66, 65]. Furthermore, it is used as a basis to simulate complex phenomena as sediment transport [33], turbidity currents [55], multilayer flows [1, 3], tsunamis [26], dam breaks [64], landslides [45], floods [31] and more.

The one-dimensional shallow water system is also called the Saint-Venant system as it has been described for the first time by Adhémar Jean Claude Barré de Saint-Venant in 1843. Such a system has been derived from the Navier-Stokes equations [62] under the following hypothesis: the vertical scale is much smaller than the horizontal one, the fluid is homogeneous and incompressible, the pressure is hydrostatic, there are no viscous forces and the velocity of the fluid should only depend on the space x and time t and not on the depth [66, 36]. Taking into account all these hypothesis and integrating the Navier-Stokes equations over the depth, we shall obtain the one-dimensional shallow water system:

$$\begin{cases} \partial_t h + \partial_x (hu) = 0 \\ \partial_t (hu) + \partial_x (hu^2 + \dfrac{gh^2}{2}) = -gh\partial_x z. \end{cases} \tag{1.3.6}$$

In particular, the first and second equations are the continuity and momentum equations respectively, where $h(x,t)$ is the total depth of water column, $u(x,t)$ is the averaged velocity and $z(x)$ is the bed elevation. Then, the pressure is given by $p = \dfrac{gh^2}{2}$ where g stays for the gravitational acceleration. System (1.3.6) is clearly a system of balance laws with form (1.3.1), where

$$\mathbf{Q} = \begin{pmatrix} h \\ hu \end{pmatrix}, \quad \mathbf{F}(\mathbf{Q}) = \begin{pmatrix} hu \\ hu^2 + \dfrac{gh^2}{2} \end{pmatrix}, \quad \mathbf{S}(\mathbf{Q}) = \begin{pmatrix} 0 \\ -gh\partial_x z \end{pmatrix}$$

and zero $\mathbf{B}$ matrix. Moreover, it is simple to prove that it is a strictly hyperbolic system with real eigenvalues $u \pm c$, where $c = \sqrt{\partial_h p} = \sqrt{gh}$ is the sound speed. Therefore, the Courant–Friedrichs–Lewy (CFL) condition on the time step for an explicit Godunov-type scheme applied to the Saint-Venant system (1.3.6) is generally given by

$$\Delta t \leq \frac{\Delta x}{2} \frac{1}{\max\{|u| + c\}}.$$

For more details about shallow water equations and their numerical approximation, see for instance [66, 36, 2, 8, 35, 23].

Lastly, for the purpose of designing stable numerical schemes and as previously mentioned, we also focus on the stationary solutions of the model (1.3.1), which satisfy the following relation

$$\partial_x \mathbf{F}(\mathbf{Q}) + \mathbf{B}(\mathbf{Q})\partial_x \mathbf{Q} = \mathbf{S}(\mathbf{Q}). \tag{1.3.7}$$

Hence, we look for numerical methods able to preserve these steady states, otherwise it is well-known that instabilities could appear in the numerical results when the solution is close to a stationary one. Such an issue has been presented for the first time in the work [5] of Bermudez and Vazquez for the shallow water system. There, they named *C-property* the capability of a scheme of preserving stationary solutions with zero-velocity. This property has been renamed *well-balanced* by Greenberg and LeRoux in their work [41]. Later, Gosse [39] has extended the notion of this property by considering a scheme able to also preserve the moving stationary solutions, namely the ones with non-zero velocity. In this case, we talk about *fully well-balanced* numerical method.

To fix ideas, let us consider the shallow water system (1.3.6). Its stationary solutions are given by

$$hu = \text{constant} = q_0 \quad \text{and} \quad \frac{q_0^2}{2h^2} + g(h + z) = \text{constant}.$$

If we take the ones with zero velocity, we find the so-called "lake at rest" stationary solution, which reads

$$u = 0 \quad \text{and} \quad h + z = \text{constant}.$$

There exist plenty of works about well-balanced and fully well-balanced methods for the Saint-Venant system. For instance, in [2], the authors have introduced the well-known hydrostatic reconstruction in order to have a numerical method able to preserve the "lake at rest" stationary solution of the shallow water system. Other interesting references for first-order well-balanced schemes applied to systems of balance laws are [6, 30]. While for second (or higher) order well-balanced schemes, we refer to [8, 7, 13, 18, 57, 51, 56, 59]. Methods that are able to preserve a given steady state (not necessarily with zero velocity) have also been designed. See for instance [44], where the authors considered the Euler equations with gravity. In this regard, we also refer to [4], where a high-order well-balanced method for multi-dimensional balance laws has been presented. There, the authors explain how to exactly preserve a general

solution which is known a priori and it is not necessarily a steady state of the model. As for the well-balanced property for non-conservative systems, we can mention paper [17]. Then, when it comes to the fully well-balanced property for the shallow water system, we could refer to [52]. In this work, the authors described a scheme based on a particular definition of an approximate Riemann solver. Indeed, it has been built so that the associated Godunov-type scheme satisfies the well-balanced property. Other fully well-balanced schemes are described for instance in [53, 10, 16, 9, 37].

In this PhD thesis, we are only interested in the well-balanced property (and not fully well-balanced) as it generally reveals itself easier to preserve but still improves significantly the numerical results. In particular, all our schemes are based on the acoustic-transport (or Lagrange-projection) decomposition. We will see that, in most of the cases, we shall build a well-balanced approximate Riemann solver for the acoustic equations so that the complete numerical method is well-balanced. Indeed, in general we do not have to modify the numerical discretization of the transport part of the system.

1.3.2 Acoustic-transport or Lagrange-projection splitting

As already mentioned, key concepts of this PhD thesis are the Lagrange-projection decomposition and the acoustic-transport splitting. They allow the decoupling of the compressibility effects of the flow from the transport phenomena of the considered mathematical model. In this way, we obtain two different systems of equations that we need to solve numerically [21, 14, 19]. In practice, we sum up the numerical strategy as follows:

1. Solve the acoustic system, either explicitly or implicitly;

2. Exploit the acoustic solution as initial condition to solve the transport system.

For the sake of clarity, let us explain such a strategy for the shallow water system (1.3.6). After having applied the chain rule to the space derivatives as follows

$$\begin{cases} \partial_t h + h\partial_x u + u\partial_x h = 0 \\ \partial_t (hu) + hu\partial_x u + u\partial_x hu + \partial_x p = -gh\partial_x z, \end{cases}$$

we decouple the acoustic phenomena from the transport ones, obtaining the acoustic and transport systems:

$$\begin{cases} \partial_t h + h\partial_x u = 0 \\ \partial_t (hu) + hu\partial_x u + \partial_x p = -gh\partial_x z, \end{cases} \tag{1.3.8}$$

and

$$\begin{cases} \partial_t h + u\partial_x h = 0 \\ \partial_t (hu) + u\partial_x (hu) = 0 \end{cases} \tag{1.3.9}$$

respectively. To sum up, once the solution of the acoustic system (1.3.8) has been found, we use it as initial condition to solve system (1.3.9) and find the transport (and final) solution. However, it is clear that the acoustic equations (1.3.8) are in non-conservative form, contrarily to the shallow water system (1.3.6). Hence, rather than considering such a system, we prefer doing a reformulation introducing the specific volume $\tau = \dfrac{1}{h}$ and the mass variable m given by $\partial_m = \tau\partial_x$. Consequently, we obtain the following system of balance laws

$$\begin{cases} \partial_t \tau - \partial_m u = 0 \\ \partial_t u + \partial_m p = -\dfrac{g}{\tau}\partial_m z. \end{cases} \tag{1.3.10}$$

We can easily prove that this system is strictly hyperbolic with eigenvalues given by $\pm hc$. On the other hand, for the transport system (1.3.9), the velocity u is the only eigenvalue. Therefore, it is clear that the splitting of the acoustic and transport phenomena led to the decomposition of the acoustic and material waves of the shallow water system (1.3.6). Such waves obviously play an important role in determining the CFL condition on the time step Δt. Indeed, having now two different systems, (1.3.10) and (1.3.9), we need to impose two different CFL conditions on the time step,

$$\Delta t \leq \frac{\Delta m}{2\max\{hc\}} \tag{1.3.11}$$

and

$$\Delta t \leq \frac{\Delta x}{2\max\{|u|\}} \tag{1.3.12}$$

respectively. Then, the Δt minimum value between conditions (1.3.11) and (1.3.12) should be taken. The interest in having two different CFL conditions is related to the fact that there exist situations in which the acoustic waves are much faster than the transport ones. This happens in subsonic regimes or near low-Froude number flows, when the absolute value of the Froude number $F = \dfrac{u}{c}$ is smaller than one. Under this condition, the usual CFL time step limitation of Godunov-type schemes is driven by the fast acoustic waves and can be very restrictive. Hence, exploiting an implicit approximation of the acoustic system, we could obtain a very natural IMplicit-EXplicit (IMEX) and large time step scheme with a CFL restriction only based on the slow transport waves [27].

Let us now present a different interpretation of the acoustic-transport splitting, namely the Lagrange-projection approach. The idea is as follows. Starting from the chosen mathematical model, we formulate it in Lagrangian coordinates. Once the Lagrangian solution is found, we should project it into Eulerian coordinates [54]. Therefore, we have once again a numerical strategy composed of two steps: the Lagrangian and projection ones. These two stages would correspond to the acoustic and transport steps respectively. In practice, we will see that these two strategies can be seen as two sides of the same coin, even though they are not always strictly equivalent. To illustrate this link, let us consider once again the one-dimensional shallow water system (1.3.6). Hence, as a first step, we need to reformulate it in Lagrangian coordinates. In order to be able to do this, we follow the fluid particle in the position ξ and we define the characteristic curves

$$\begin{cases} \dfrac{\partial x}{\partial t}(\xi, t) = u(x(\xi, t), t) \\ x(\xi, 0) = \xi, \end{cases} \tag{1.3.13}$$

which define the trajectory : $t \to x(\xi, t)$ of ξ as time goes on. Therefore, any function : $(x, t) \to \varphi(x, t)$ in Eulerian coordinates can be written in Lagrangian coordinates,

$$\overline{\varphi}(\xi, t) = \varphi(x(\xi, t), t).$$

Then, we introduce the volume ratio $L(\xi, t)$, which is defined by

$$L(\xi, t) = \frac{\partial x}{\partial \xi}(\xi, t) \tag{1.3.14}$$

and which satisfies

$$\begin{cases} \dfrac{\partial L}{\partial t}(\xi, t) = \partial_\xi u(x(\xi, t), t) \\ L(\xi, 0) = 1. \end{cases} \tag{1.3.15}$$

Hence, the time and space derivatives of a function φ in Lagrangian coordinates are given by

$$\partial_t \overline{\varphi}(\xi,t) = \partial_t \varphi(x,t) + u(x,t)\partial_x \varphi(x,t) \quad \text{and} \quad \partial_\xi \overline{\varphi}(\xi,t) = L(\xi,t)\partial_x \varphi(x,t).$$

Leaving the few computations to the reader, we directly present the shallow water equations in Lagrangian coordinates, which read

$$\begin{cases} \partial_t(L\bar{h}) = 0 \\ \partial_t(L\overline{hu}) + \partial_\xi \bar{p} = -g\bar{h}\partial_\xi \bar{z}. \end{cases} \tag{1.3.16}$$

Therefore, once the Lagrangian solution of system (1.3.16) has been found, we only need to project it into Eulerian coordinates. For details about Lagrangian coordinates in 2 or 3 dimensions, we refer to [29, 49, 50].

At this stage, we could remark that, in the second equation of system (1.3.16), the physical flux is only given by the pressure term. Nonetheless, it may not yet be entirely clear the relation between system (1.3.16) and the acoustic one (1.3.10). In order to illustrate it, we start by remarking that $L\bar{h}$ does not depend on time, namely

$$(L\bar{h})(\xi,t) = (L\bar{h})(\xi,0) = h(\xi,0) = h_0 \quad \text{and} \quad L(\xi,t) = \frac{h_0}{\bar{h}(\xi,t)},$$

thanks to which we obtain

$$\begin{cases} \partial_t h_0 = 0 \\ \partial_t L - \partial_\xi \bar{u} = 0 \\ \partial_t(h_0\bar{u}) + \partial_\xi \bar{p} = -g\bar{h}\partial_\xi \bar{z}. \end{cases}$$

Then, we only need to re-introduce the variables $\bar{\tau} = \dfrac{1}{\bar{h}}$ and the variable m such that $\partial_m = \dfrac{1}{h_0}\partial_\xi$, so that we finally get

$$\begin{cases} \partial_t \bar{\tau} - \partial_m \bar{u} = 0 \\ \partial_t \bar{u} + \partial_m \bar{p} = -\dfrac{g}{\bar{\tau}}\partial_m \bar{z}, \end{cases} \tag{1.3.17}$$

which clearly evokes the acoustic system (1.3.10). It is clear that the two systems (1.3.17) and (1.3.10) are not strictly the same as they have been obtained following different procedures. However, when it comes to the numerical approximation, we will exploit ingredients of both interpretations for the purpose of designing efficient numerical schemes.

We have already clarified why Lagrange-projection methods are particularly useful when considering subsonic regimes. However, let us point out that this kind of method generally opens up a lot of new possibilities from a numerical point of view. Indeed, either having two different sets of equations (in the acoustic-transport splitting) or working in Lagrangian coordinates (in the LP approach) means that properties of the numerical methods could be differently (and more easily) preserved.

Let us give a few examples of Lagrange-projection (or acoustic-transport) schemes. In [19, 20], the authors described an all-regime Lagrange-projection scheme in 1D and 2D respectively. By all-regime they denote a numerical scheme able to compute accurate approximate solutions with a mesh size and time step much bigger than the Mach number (analogous of the Froude number in the shallow water framework). Hence, the resulting method is also implicit-explicit and able to lead to fast simulations. Another semi-implicit scheme was developed in [21], where the well-balanced property of the scheme was also analyzed, meaning that

the zero-velocity stationary solutions of the mathematical model are preserved. Such a property has been further developed in [14], where the authors described a fully well-balanced LP method. Then, this scheme has also been extended to an arbitrary order of accuracy in [54], by using polynomial reconstruction and Runge-Kutta strategy. The issue of high-order Lagrange-projection methods has been treated in [32] as well, where the high order of accuracy is reached by means of different techniques, namely by using Taylor expansions and the Cauchy-Kovalevskaya procedure. Another interesting reference is [58], where a high-order implicit-explicit Lagrange-projection scheme was developed. In particular, the authors obtained the high order of accuracy in space using the discontinuous Galerkin scheme while, for the approximation in time, they considered the backward Euler method. Finally, asymptotic-preserving Lagrange-projection methods have also been designed, meaning that the scheme is able to reproduce the asymptotic behavior satisfied by the continuous equations at the discrete level, see [22, 12]. There, the numerical methods also satisfy a fully discrete entropy inequality.

In this PhD thesis, we are particularly careful about building Lagrange-projection numerical methods endowed with the well-balanced property. Hence, only the zero-velocity stationary solutions are preserved. Another essential property of the numerical methods is the positivity-preserving one, meaning that the schemes should be able to preserve the strict positivity of the solution [34]: for instance, of the water height in shallow water models [8], of the cross-sectional area of the vessel in blood flow equations [37], of the density in the gas dynamic equations [34] and so on.

Subsequently, it is clear that the numerical solution should be a good approximation of the physical solution and not overly diffusive. However, depending on the mathematical model and the numerical method, this is not a trivial request at all. Indeed, many first-order schemes are generally very diffusive and may need the use of a very fine mesh to obtain an acceptable approximation of the solution, leading to an excessive computational cost. This happens for instance when using numerical schemes based on approaches which neglect the intermediate waves of a model, considering only the propagation speeds of the external ones, as in the HLL (Harten-Lax-van Leer) solver [42]. Clearly, such a problem is not present for models with a two-waves structure. Nevertheless, here we generally try to design numerical schemes which take into account the complete structure of the solution. We will see that this request is made particularly easier by exploiting a relaxation version [61, 11] of the acoustic system. Finally, with all this in mind and as a second step, we also design a higher order extension of the numerical methods, aiming for a faster convergence of the numerical solution to the physical one [63].

1.4 Thesis synopsis

This thesis is structured as follows.

In **chapter 2**, we apply the Lagrange-projection strategy to the blood flow equations with non-constant in space arterial stiffness and cross-sectional area at equilibrium, obtaining a system of balance laws. Due to these parameters, it is not straightforward to preserve the well-balanced property, hence we focus on the preservation of the so-called "man at eternal rest solution". Two ways of preserving it are shown, one based on a modification of the approximate Riemann solver for the acoustic system and the second one on the well-known hydrostatic reconstruction's strategy [2]. Efforts to extend the numerical strategy to second-order of accuracy are also made by using classical techniques as polynomial reconstruction [63] and Runge-Kutta scheme [40]. Once again, further modifications are applied to the numerical schemes in order to preserve the well-balanced property, for this purpose we will use

the notion of fluctuations [54].

Then from chapter 3, only shallow water equations and related models are studied. In particular, **chapter 3** is devoted to the numerical approximation of the non-conservative two-layer shallow water model [1]. That is to say, we consider the fluid to be composed of two superposed layers of immiscible liquids. The difficulty related to this model is twofold. On one hand, the presence of two different velocities makes less clear how to apply the Lagrange-projection strategy to this system. Secondly, not only the model is non-conservative, but it is also only conditionally hyperbolic, due to which the numerical simulation of this system is generally more difficult. Nevertheless, we propose a well-balanced implicit-explicit Lagrange-projection method for this system, obtaining promising results. In particular, we define an approximate Riemann solver for the acoustic system and use the associate Godunov-type scheme. Such a method can also be understood as an approximation of the Lagrangian system. Finally, the implicit-explicit version of the Lagrange-projection scheme is easily obtained, by solving a linear system in the acoustic step.

In **chapter 4**, we study the coupling of the shallow water system with the Exner equation, where the latter is used to simulate the bedload sediment transport [33]. The resulting non-conservative system appears to be particularly difficult to approximate. Indeed, if a decoupled approach is used, instabilities could be easily found in the numerical simulations due to the eigenstructure of the complete model. We investigate such a problem and we design three different well-balanced Lagrange-projection schemes for this system. In particular the Exner equation could be completely taken into account in only one of the two steps or in both of them. In any case, we provide an approximate Riemann solution for the acoustic equations. These three strategies are extended to second-order of accuracy as well. Finally, at the end of this chapter, we also present the two-dimensional version of the model and of one numerical strategy.

Chapter 5 is the natural extension of the work presented in chapter 4. Indeed, we now consider sediment transport in general, meaning that the sediment particles could not only move along the bottom but also go into suspension in the water if a fine enough fraction is present. Furthermore, since we know that the characteristic time associated with sediments is much larger than the one corresponding to fluid, we build an implicit-explicit method in order to obtain fast simulations.

Finally, in this thesis and at this stage, we considered either first-order implicit-explicit schemes or second-order explicit methods. Only in **chapter 6** we develop a general approach which allows us to design a second-order implicit explicit scheme for the shallow water system. Here, the second-order of accuracy is attained using different techniques with respect to the second-order numerical schemes presented in the previous chapters, namely by means of Taylor expansions and the Cauchy-Kovalevskaya procedure. We highlight that we only need to solve linear systems in order to find the solution, so that the computational cost is not elevated. However, it is well-known that there could appear spurious oscillations in presence of discontinuities when using high-order schemes. In order to remove such oscillations, we employ *a posteriori* limiting approach, resembling the MOOD (Multi-dimensional Optimal Order Detection) strategy [24, 25]. We emphasize that, using a posteriori approach, allows us not to modify the strategy for the second-order of accuracy and thus the linear systems. In this way, the computational cost should not increase excessively.

Publications

The work presented in this manuscript has been either published or submitted.

- A. Del Grosso and C. Chalons. *Second-order well balanced Lagrange-Projection schemes for Blood Flow Equations.* Calcolo 58, 43, 2021. 10.1007/s10092-021-00434-5
- C. Chalons and A. Del Grosso, *A second-order well-balanced Lagrange-projection numerical scheme for Shallow Water Exner equations in 1D and 2D.* 2022. Communications in Mathematical Sciences. 20(7): 1839-1873, 2022. 10.4310/CMS.2022.v20.n7.a3
- C. Chalons and A. Del Grosso, *Exploring different possibilities for second-order well-balanced Lagrange-projection numerical schemes applied to shallow water Exner equations.* International Journal for Numerical Methods in Fluids. 1- 31, 2022. 10.1002/fld.5064
- A. Del Grosso, M. J. Castro Dìaz, C. Chalons and T. Morales de Luna. *On well-balanced implicit-explicit Lagrange-projection schemes for two-layer shallow water equations.* Submitted to the Journal "Applied Mathematics and Computation" in May 2022.
- A. Del Grosso, M. J. Castro Dìaz, C. Chalons and T. Morales de Luna. *Lagrange-projection methods for shallow water equations with movable bottom and erosion-deposition processes.* Submitted to the Journal "Communications in Mathematics and Applications" in March 2022.
- C. Chalons and A. Del Grosso. *Second-order Well-Balanced Implicit-Explicit Scheme for Systems of Balance Laws.* Submitted to the Journal "Numerische Mathematik" in January 2022.

Conferences and seminars

- HYP 2022 - XVIII International Conference on Hyperbolic Problems - *June 20-24, 2022.*
- CANUM - 45eme Congrès National d'Analyse Numérique - *June 13-17, 2022.*
- SHARK-FV 2022 (Sharing Higher-order Advanced Know-how on Finite Volume) workshop - *May 23-27, 2022.*
- Seminar at the Institut de Mathématiques de Bordeaux - *March 25, 2022.*
- PhD seminar at the Laboratoire Amiénois de Mathématique Fondamentale et Appliquée - *December 1, 2021.*
- NumHyp 2021 - Numerical Methods for Hyperbolic Problems - *July 26-30, 2021.*
- SMAI 2021 - 10 ième Biennale Française des Mathématiques Appliquées et Industrielles - *June 21-25, 2021.*
- CEDYA 2020 - XXVI Congreso de Ecuaciones Diferenciales y Aplicaciones XVI Congreso de Matemática Aplicada - *June 14-18, 2021.*
- CAN-J 2020 - congrès d'analyse numérique pour les jeunes 2020 - *December 3-4 2020.*

SEME

In February 2022, I have participated to the Semaine Maths-Entreprises (Week Maths-Enterprises) at the Polytechnic Institute of Paris. In particular, in the group project "Piloting the treatment of large volumes of seismic signals to improve subsurface imaging".

Teaching activities

In both the second (2020-2021) and the third (2021-2022) year of PhD, I have taught 64 hours of practical work for the course General Mathematics 2.

1.5 Bibliography

[1] E. Audusse. *A multilayer Saint-Venant model: derivation and numerical validation.* Discrete Contin. Dyn. Syst., Ser. B. 5: 189–214, 2005.

[2] 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: 2050-2065, 2004. 10.1137/S1064827503431090.

[3] E. Audusse, M.O. Bristeau, B. Perthame and J. Sainte-Marie. *A multilayer Saint-Venant system with mass exchanges for Shallow Water flows. Derivation and numerical validation.* ESAIM Mathematical Modelling and Numerical Analysis. 45, 2009. 10.1051/m2an/2010036.

[4] J. P. Berberich, P. Chandrashekar, C. Klingenberg. *High order well-balanced finite volume methods for multi-dimensional systems of hyperbolic balance laws.* Computers & Fluids, 219: 104858 2021. 10.1016/j.compfluid.2021.104858.

[5] A. Bermudez, M. E. Vazquez, *Upwind methods for hyperbolic conservation laws with source terms.* Computers & Fluids, 23(8): 1049-1071, 1994. 10.1016/0045-7930(94)90004-3.

[6] C. Berthon, A. Duran, F. Foucher, K. Saleh, Khaled and J. Zabsonré. *Improvement of the Hydrostatic Reconstruction Scheme to Get Fully Discrete Entropy Inequalities.* Journal of Scientific Computing. 80, 2019. 10.1007/s10915-019-00961-y.

[7] C. Berthon, F. Foucher. *Hydrostatic Upwind Schemes for Shallow-Water Equations.* Springer Proceedings in Mathematics, 4: 97-105, 2011. 10.1007/978-3-642-20671-9_11

[8] C. Berthon, F. Foucher. *Efficient well–balanced hydrostatic upwind schemes for shallow–water equations.* Journal of Computational Physics. 231(15): 4993–5015, 2012. 10.1016/j.jcp.2012.02.031

[9] C. Berthon, R. Loubère and V. Michel-Dansac. *Second-order well-balanced scheme for the shallow-water equations with topography.* XVI International Conference on Hyperbolic Problems Theory, Numerics, Applications, Aachen (Germany), Springer Proceed, 2016. 10.1007/978-3-319-91545-6_13

[10] C. Berthon, V. Michel-Dansac. *A simple fully well-balanced and entropy preserving scheme for the shallow-water equations.* Applied Mathematics Letters. 86: 284-290, 2018. 10.1016/j.aml.2018.07.013.

[11] F. Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources.* Frontiers in mathematics, 2004. 10.1007/b93802.

[12] F. Bouchut, C. Chalons, S. Guisset. *An entropy satisfying two-speed relaxation system for the barotropic Euler equations. Application to the numerical approximation of low Mach number flows.* Numerische Mathematik, Springer Verlag, 145: 35-76, 2020. 10.1007/s00211-020-01111-5.

[13] A. Canestrelli, A. Siviglia, M. Dumbser, 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(6): 834-844, 2009. 10.1016/j.advwatres.2009.02.006.

[14] M. J. Castro Díaz, C. Chalons and T. Morales De Luna. *A fully well-balanced Lagrange-Projection type scheme for the Shallow-water equations.* SIAM J. Numer. Anal., 56(5): 3071–3098, 2018. 10.1137/17M1156101.

[15] M.J. Castro, P.G. LeFloch, M.L. Munoz, C. Parés. *Why many theories of shock waves are necessary: convergence error in formally path-consistent schemes.* J. Comput. Phys 227: 8107–8129, 2008.

[16] M.J. Castro Díaz, J.A. López-García, Carlos Parés, *High order exactly well-balanced numerical methods for shallow water systems.* Journal of Computational Physics, 246: 242-264, 2013.

[17] M. J. Castro, T. Morales de Luna, C. Parés Madroñal. *Well-Balanced Schemes and Path-Conservative Numerical Methods.* Handbook of Numerical Analysis, Elsevier, 18: 131-175, 2017. 10.1016/bs.hna.2016.10.002.

[18] M. J. Castro and C. Parés. *Well-Balanced High-Order Finite Volume Methods for Systems of Balance Laws.* Journal of Scientific Computing. 82: 48, 2020. 10.1007/s10915-020-01149-5. 10.1016/j.jcp.2013.03.033.

[19] C. Chalons, M. Girardin, S. Kokh. *An All-Regime Lagrange-Projection like scheme for the gas dynamics equations on unstructured meshes.* Communications in Computational Physics. 20(1): 188-233 2016. 10.4208/cicp.260614.061115a.

[20] C. Chalons, M. Girardin, S. Kokh. *An all-regime Lagrange-Projection like scheme for 2D homogeneous models for two-phase flows on unstructured meshes.* Journal of Computational Physics, Elsevier, 335: 885-904, 2017. 10.1016/j.jcp.2017.01.017.

[21] C. Chalons, P. Kestener, S. Kokh, and M. Stauffert. *A large time-step and well-balanced Lagrange-Projection type scheme for the Shallow-water equations.* Communications in Mathematical Sciences. 15(3): 765–788, 2017. 10.4310/CMS.2017.v15.n3.a9.

[22] C. Chalons, S. Kokh, M. Girardin. *Large Time Step and Asymptotic Preserving Numerical Schemes for the Gas Dynamics Equations with Source Terms.* SIAM Journal on Scientific Computing. 35(6): A2874–A2902, 2013. 10.1137/130908671.

[23] A. Chinnayya, A.-Y. LeRoux, N. Seguin. *A well-balanced numerical scheme for the approximation of the shallow-water equations with topography: the resonance phenomenon.* International Journal on Finite Volume (electronic), 1(1), 2004. hal-00017378

[24] S. Clain, S. Diot, R. Loubère. *A high-order finite volume method for hyperbolic systems: Multi-dimensional Optimal Order Detection (MOOD).* Journal of Computational Physics, Elsevier, 230, 2011. ffhal-00518478v2f

[25] S. Clain, S. Diot and R. Loubère. *Improved Detection Criteria for the Multi-dimensional Optimal Order Detection (MOOD) on unstructured meshes with very high-order polynomials.* Computer and Fluids, 64 : 43-63, 2012. 10.1016/j.compfluid.2012.05.004

[26] S. Clain, C. Reis, R. Costa, J. Figueiredo, M. Baptista, J. Miranda. *Second-order finite volume with hydrostatic reconstruction for tsunami simulation. The Lisbon 1969 tsunami.* Journal of Advances in Modeling Earth Systems. 8, 2016. 10.1002/2015MS000603.

[27] F. Coquel, Q. Nguyen, M. Postel, and Q. Tran. *Entropy-satisfying relaxation method with large time-steps for Euler IBVPs.* Mathematics of Computation, 79(271):1493–1533, 2010. 10.1090/S0025-5718-10-02339-2.

[28] G. Dal Maso, P. G. Lefloch, and F. Murat. *Definition and weak stability of nonconservative products.* Journal de Mathématiques Pures et Appliquées, 74(6):483–548, 1995.

[29] B. Després. *Numerical Methods for Eulerian and Lagrangian Conservation Laws.* 2017. 10.1007/978-3-319-50355-4.

[30] V. Desveaux, M. Zenk, C. Berthon and C. Klingenberg. *A well-balanced scheme to capture non-explicit steady states in the Euler equations with gravity.* International Journal for Numerical Methods in Fluids. 81(2) : 104-127, 2015. 10.1002/fld.4177.

[31] G. A. M. de Almeida, P. Bates, J. E. Freer, and M. Souvignet. *Improving the stability of a simple formulation of the shallow water equations for 2-D flood modeling.* Water Resour. Res., 48, W05528, 2012. doi:10.1029/2011WR011570.

[32] F. Duboc, C. Enaux, S. Jaouen, H. Jourdren, M. Wolff. *High-order dimensionally split Lagrange-remap schemes for compressible hydrodynamics.* Comptes Rendus Mathematique, 348(1–2):105-110, 2010. 10.1016/j.crma.2009.12.008.

[33] F. M. Exner. *Über die wechselwirkung zwischen wasser und geschiebe in flüssen.* Akad. Wiss. Wien Math. Naturwiss. Klasse, 134(2a): 165-204, 1925.

[34] G. Gallice. *Positive and entropy stable Godunov-type schemes for gas dynamics and MHD equations in Lagrangian or Eulerian coordinates.* Numer. Math. 94(4): 673-713, 2003. 10.1007/s00211-002-0430-0

[35] T. Gallouët, J.-M. Hérard, N. Seguin. *Some approximate Godunov schemes to compute shallow-water equations with topography.* Computers and Fluids, 32(4) : 479-513, 2003. 10.1016/S0045-7930(02)00011-7

[36] J. Gerbeau and B. Perthame. *Derivation of Viscous Saint-Venant System for Laminar Shallow Water; Numerical Validation.* Discrete Contin. Dyn. Syst., Ser. B. 1, 89-102, 2000.

[37] B. Ghitti, C. Berthon, M. H. Le, E. F. Toro. *A fully well-balanced scheme for the 1D blood flow equations with friction source term.* Journal of Computational Physics, 421 : 109750, 2020. 10.1016/j.jcp.2020.109750.

[38] E. Godlewski, P.A. Raviart. *Numerical Approximation of Hyperbolic Systems of Conservation Laws.* Springer, 1996.

[39] L. Gosse. *A well-balanced flux-vector splitting scheme designed for hyperbolic systems of conservation laws with source terms.* Computers & Mathematics with Applications, 39(9–10): 135-159, 2000. 10.1016/S0898-1221(00)00093-6.

[40] S. Gottlieb and C.-W. Shu. *Total variation diminishing RUNGE-KUTTA schemes.* Mathematics of Computation. 67, 1996. 10.1090/S0025-5718-98-00913-2.

[41] J. M. Greenberg and A. Y. LeRoux. *A Well-Balanced Scheme for the Numerical Processing of Source Terms in Hyperbolic Equations.* SIAM Journal on Numerical Analysis, 33(1): 1–16, 1996. http://www.jstor.org/stable/2158421

[42] A. Harten, P. Lax and B. van Leer. *On Upstream Differencing and Godunov-Type Schemes for Hyperbolic Conservation Laws.* SIAM Rev. 25: 35-61, 1983. 10.1137/1025002.

[43] T. Y. Hou, P. G. Le Floch. *Why Nonconservative Schemes Converge to Wrong Solutions: Error Analysis.* Mathematics of Computation, 62(206): 497–530, 1994. 10.2307/2153520

[44] C. Klingenberg, G. Puppo, Gabriella, M. Semplice. *Arbitrary Order Finite Volume Well-Balanced Schemes for the Euler Equations with Gravity.* SIAM Journal on Scientific Computing. 41: A695-A721, 2019. 10.1137/18M1196704.

[45] C.Y. Kuo, Y.C. Tai, F. Bouchut, A. Mangeney, M. Pelanti, R.F. Chen, K.J. Chang, *Simulation of Tsaoling landslide, Taiwan, based on Saint Venant equations over general topography*. Engineering Geology, 104(3–4): 181-189, , 2009, ISSN 0013-7952. 10.1016/j.enggeo.2008.10.003.

[46] P. Lax. *Hyperbolic Systems of Conservation Laws and Mathematical Theory Shock Waves*. Society for Industrial and Applied Mathematics, 1973. 10.1137/1.9781611970562.

[47] P. LeFloch. *Entropy weak solutions to nonlinear hyperbolic systems under nonconservative form*. Communications in Partial Differential Equations. 13: 669-727, 1988. 10.1080/03605308808820557.

[48] P. LeFloch. *Shock Waves for Nonlinear Hyperbolic Systems in Nonconservative Form*. Institute for Math. and its Appl., Minneapolis, Preprint #593, 1989.

[49] R. Loubère. *Une Méthode Particulaire Lagrangienne de type Galerkin Discontinu Application à la Mécanique des Fluides et l'Interaction Laser/Plasma*, These doctorale, 2002.

[50] P. H. Maire, R. Abgrall, J. Breil, J. Ovadia. *A Cell-Centered Lagrangian Scheme for Two-Dimensional Compressible Flow Problems*. SIAM Journal on Scientific Computing, 2007. 10.1137/050633019

[51] F. Marche, P. Bonneton, P. Fabrie, N. Seguin. *Evaluation of well-balanced bore-capturing schemes for 2D wetting and drying processes*. International Journal for Numerical Methods in Fluids, 53(5), 2007. 10.1002/fld.1311

[52] V. Michel-Dansac, C. Berthon, S. Clain, F. Foucher. *A well-balanced scheme for the shallow-water equations with topography*. Computers and Mathematics with Applications, Elsevier, 72: 568 - 593, 2016. 10.1016/j.camwa.2016.05.015.

[53] V. Michel-Dansac, C. Berthon, S. Clain, F. Foucher. *A well-balanced scheme for the shallow-water equations with topography or Manning friction*. Journal of Computational Physics, Elsevier, 335: 115-154, 2017. 10.1016/j.jcp.2017.01.009.

[54] T. Morales De Luna, M. J. Castro Díaz and C. Chalons. *High order fully well-balanced Lagrange-Projection scheme for Shallow-water*. Commun. Math. Sci., 18(3): 781–807, 2020. 10.4310/CMS.2020.v18.n3.a9

[55] T. Morales de Luna, M. J. Castro Díaz, C. Parés Madroñal, and E. D. Fernández Nieto. *On a shallow water model for the simulation of turbidity currents*. Communications in Computational Physics, 6(4): 848-882, 2009. 10.4208/cicp.2009.v6.p848

[56] L. O. Müller, C. Parés and E. F. Toro. *Well-balanced high-order numerical schemes for one-dimensional blood flow in vessels with varying mechanical properties*. Journal of Computational Physics 242: 53-85, 2013. 10.1016/j.jcp.2013.01.050.

[57] S. Noelle, N. Pankratz, G. Puppo, and J. Natvig. *Well-balanced finite volume schemes of arbitrary order of accuracy for shallow water flows*. Journal of Computational Physics. 213: 474-499, 2006. 10.1016/j.jcp.2005.08.019.

[58] F. Renac. *A robust high-order Lagrange-projection like scheme with large time steps for the isentropic Euler equations*. Numerische Mathematik. 135, 2017. 10.1007/s00211-016-0807-0.

[59] G. Russo and A. Khe. *High order well balanced schemes for systems of balance laws*. Proceedings of Symposia in Applied Mathematics. 67: 919-928, 2009. 10.1090/psapm/067.2/2605287.

[60] L. Sainsaulieu. *Traveling Waves Solution of Convection–Diffusion System whose Convection Terms are Weakly Nonconservative: Application to the Modeling of Two-Phase Fluid Flows*. SIAM Journal on Applied Mathematics, 55(6), 1552–1576, 1995. http://www.jstor.org/stable/2102503

[61] I. Suliciu. *On the thermodynamics of fluids with relaxation and phase transitions. Fluids with relaxation.* Int. J. Engag. Sci. 36: 921-947, 1998.

[62] R. Temam. *The Navier-Stokes Equations: Theory and Numerical Methods.* American Mathematical Society. 343, 2001.

[63] E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics.* Third Edition. Springer–Verlag, 2009. 10.1007/b79761_5.

[64] A. Valiani, V. Caleffi, A Zanni. *Case Study: Malpasset Dam-Break Simulation using a Two-Dimensional Finite Volume Method.* Journal of Hydraulic Engineering. 128: 460-472, 2002. 10.1061/(ASCE)0733-9429(2002)128:5(460).

[65] G. K. Vallis. *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*, (2nd ed.). Cambridge: Cambridge University Press, 2017. 10.1017/9781107588417

[66] C. B. Vreugdenhil. *Numerical Methods for Shallow Water Flow.* Dordrecht; Boston: Kluwer Academic Publishers, 1994. Part of the Water Science and Technology Library book series (WSTL, volume 13).

Chapter 2

Schémas de type Lagrange-projection équilibres et de second ordre pour les équations du flux sanguin

Ce chapitre a fait l'objet d'une publication dont la référence est: A. Del Grosso and C. Chalons. *Second-order well balanced Lagrange-Projection schemes for Blood Flow Equations.* Calcolo 58, 43, 2021. https://doi.org/10.1007/s10092-021-00434-5

Dans ce chapitre, nous nous concentrons sur le développement de schémas de type Lagrange-projection équilibres pour les équations du flux sanguin. Ici, nous négligeons les forces de friction alors que le terme source est dû à la présence de paramètres variables en espace tels que la rigidité artérielle et la surface de la section transversale à l'équilibre. Par schéma équilibre, nous entendons une méthode capable de préserver la solution "homme au repos éternel". À cette fin, nous présentons deux stratégies différentes : la première nécessite une définition cohérente du terme source basée sur un solveur de Riemann approché, tandis que la seconde utilise une stratégie de reconstruction hydrostatique. Nous expliquons ensuite comment atteindre le second ordre de précision pour les deux procédures. Des simulations numériques sont effectuées afin de montrer le bon ordre de précision et le bon comportement des schémas.

Enfin, nous soulignons que, si dans les chapitres suivants tous les modèles considerés sont basés sur le système de Sant-Venant, ce n'est pas le cas dans ce premier chapitre. Le choix de commencer cette thèse par l'étude du modèle du flux sanguin doit en fait être recherché dans le parcours de formation de la thésarde. En effet, pour le mémoire du master, un stage a été effectué sous la direction du professeur C. Chalons et du professeur E. F. Toro, portant sur l'étude de l'analyse de sensibilité pour les équations du flux sanguin. Il a donc été estimé qu'une manière progressive de rentrer dans le sujet de la thèse était de commencer à appliquer une méthode de type Lagrange-projection à un modèle mathématique déjà traité. Cette idée a été renforcée par le fait qu'on sait que le nombre de Shapiro (ou de manière équivalente le nombre de Froude pour les équations de Saint-Venant) dans les artères humaines est en général de l'ordre de 10^{-2}. Pour cette raison, il pouvait être intéressant d'envisager une méthode de type Lagrange-projection appliquée à ce modèle.

Nous citons ici l'article qui a été rédigé pendant le stage susmentionné : C. Chalons, A. Del Grosso and E. F. Toro. *Numerical approximation and uncertainty quantification for arterial blood flow models with viscoelasticity.* Journal of Computational Physics, Volume 457, 2022, 111071, ISSN 0021-9991, https://doi.org/10.1016/j.jcp.2022.111071

Second-order well-balanced Lagrange-projection schemes for blood flow equations

Abstract

We focus on the development of well-balanced Lagrange-projection schemes applied to the one-dimensional blood flow system of balance laws. Here we neglect the friction forces and the source term comes from the presence of variable parameters in space, like the cross-sectional area at the equilibrium and the arterial stiffness. By well-balanced we mean that the method preserves the "man at eternal rest" solution. For this purpose we present two different strategies: the first requires a consistent definition of the source term based on an approximate Riemann solver, while the second one exploits the well-established hydrostatic reconstruction. We then explain how to reach the second-order of accuracy for both procedures. Numerical simulations are carried out in order to show the right order of accuracy and the good behaviour of the schemes.

2.1 Introduction

This work focuses on the construction of second-order well-balanced Lagrange-projection schemes applied to the 1D Blood Flow Equations (BFE). This model turns out to be extremely useful when dealing with the study of the cardiovascular system and related diseases. Indeed, it proved to be effective in the computation of averaged quantities as the cross-sectional area A of the vessel, the blood flow q and internal pressure p. There is therefore a huge amount of works about this system, for which we refer the reader to [17, 32] and the references therein. Here we study the model as applied to arteries, in the particular case in which the cross-sectional area at equilibrium and the wall stiffness could be non-constant in space, see for instance [15, 20, 36, 34, 35]. Indeed, there exist physiological and pathological situations in which geometrical and mechanical parameters can vary locally, as in presence of stenoses or aneurysms and tapering of blood vessels. However, considering non-constant parameters leads to the presence of a non-zero source term and, therefore, we aim to develop numerical schemes for hyperbolic systems of balance laws.

We are also interested in preserving the so-called "man at eternal rest" stationary solution, namely in the well-balancedness of the numerical method. As a matter of fact, this property is of critical relevance as a non well-balanced scheme could produce non-physical spurious oscillations in certain cases, namely when the solution is close to a steady state. In particular,

the "man at eternal rest" condition is characterized by zero-velocity; if the numerical scheme also preserves the moving stationary solutions, it is called fully well-balanced. Many studies have been done about well-balanced methods in general, as in [4, 6, 7, 8]. For application to shallow water models, see [1, 3, 23, 24]. As far as the blood flow equations are concerned, we refer for instance to the work of Delestre and collaborators [15], in which they considered the cross-sectional area at equilibrium to be non-constant in space. In particular, they developed a first-order well-balanced scheme basing themselves on the well-known hydrostatic reconstruction procedure, introduced for the first time by Audusse et al. in [1] in the context of the shallow water equations. In [20], Delestre et al. expanded their work considering variable values in space for the arterial wall rigidity as well. In [26], Müller et al. followed the generalized hydrostatic reconstruction to build a high-order well-balanced path-conservative numerical method for blood flow equations with mechanical properties which could vary in space. Then, in [27], Müller and Toro presented another high-order well-balanced path-conservative scheme for BFE. There, discontinuous values for the cross-sectional area at rest and external pressure were also taken into account, in addition to the wall stiffness.

Last but not least, we design numerical schemes based on the Lagrange-projection formalism, which allows us to split up the system into two different ones, and in particular to separately take into account the acoustic (Lagrangian step) and the transport (projection step) parts of the system. This also implies the decoupling of fast and slow waves, which plays an important role in determining the CFL time step. Indeed, a fast wave leads to a very restrictive time step, contrarily to a slow wave. However, as far as human arteries are concerned, the ratio between the velocity and the wave speed, i.e. the Shapiro number (or equivalently the Froude number for the shallow water equations), is in general of order 10^{-2}. As such, it could be interesting to develop a numerical method in which the Lagrangian step is solved implicitly. While in this work we start by describing explicit schemes, we also aim to develop implicit-explicit schemes in future works. For Lagrange-projection schemes, we refer the reader to [12, 10, 9] and the references therein, while for well-balanced Lagrange-projection methods see [5, 11, 25].

In conclusion, we present two second-order well-balanced Lagrange-projection schemes for the 1D BFE. We start by introducing the numerical method in the case of constant parameters, and thus for a system of conservation laws. Then, when variable properties are considered, two different ways of preserving the "man at eternal rest" stationary solution are described. On one hand, referring to the work of Suliciu [29], we relax the Lagrangian system introducing a new variable, which stands for a linearization of the pressure term. Then, following the theory of Gallice [18, 19], we easily solve the associated Riemann problem. Alternatively and as a second strategy, we exploit the hydrostatic reconstruction.

Chapter outline. In the next section we present the 1D mathematical model for the blood flow equations. We also introduce the Lagrange-projection decomposition which leads to two different systems, the acoustic and transport ones. We describe an approximate Riemann solver for the acoustic system as well. In sections 2.3 and 2.4, we present the first and second-order well-balanced schemes respectively. In both sections, we explain two different strategies for preserving the "man at eternal rest" solution. In section 2.5, numerical simulations are carried out. Finally conclusions and perspectives are drawn in section 2.6.

2.2 The mathematical model

Given the axial coordinate x along the longitudinal axis of the vessel and the time $t > 0$, the general one-dimensional blood flow model consists of two equations: the mass conservation

and momentum balance equation. Hence, the system reads

$$\begin{cases} \partial_t A + \partial_x q = 0 \\ \partial_t q + \partial_x \Big(\hat{\alpha} \dfrac{q^2}{A} \Big) + \dfrac{A}{\rho} \partial_x p = f, \end{cases} \tag{2.2.1}$$

where $A(x,t) > 0$ is the cross-sectional area of the vessel, $q = Au$ is the blood flow, with $u(x,t)$ being the averaged velocity of blood at cross section, and finally $p(x,t)$ is the averaged internal pressure at cross section. Furthermore, ρ represents the constant blood density, while $\hat{\alpha}$ is determined by the velocity profile, which is assumed to be flat in this work, thus we take $\hat{\alpha} = 1$. Lastly, f accounts for the friction forces, although we will neglect them in the rest of the work. Note that we assume the initial area $A(x, t = 0)$ and initial velocity $u(x, t = 0)$ to be given at initial time $t = 0$. For more details about the derivation of system (2.2.1), we refer to [17] and [32].

Since in system (2.2.1) there are three unknowns but only two equations, we need a closure condition, namely a tube law or more specifically a relation between the internal pressure and the cross-sectional area. In this work we refer to [20] and we consider the blood vessels to be purely elastic arteries. As such, the tube law reads

$$p(x,t) = p_{ext} + K(x)(\sqrt{A(x,t)} - \sqrt{A_0(x)}), \tag{2.2.2}$$

where p_{ext} is the constant external pressure, $A_0(x)$ is the cross-sectional area at equilibrium and $K(x)$ is a parameter related to arterial stiffness. In particular, K is a positive function depending on the vessel thickness $h_0(x)$ and the Young modulus $E(x)$ (see again [20]). Note also that condition (2.2.2) is only valid for blood flow in arteries and not in veins. For a more general tube law, we refer for instance to [17, 32]. Equipped with this closure condition (2.2.2), we are now able to show the system of balance laws that we will investigate in the rest of this work, namely

$$\begin{cases} \partial_t A + \partial_x q = 0 \\ \partial_t q + \partial_x \Big(\dfrac{q^2}{A} + \gamma A^{\frac{3}{2}} \Big) = s, \end{cases} \tag{2.2.3}$$

with $\gamma = \dfrac{K}{3\rho}$ and

$$s = s(A; A_0, K) = \frac{A}{\rho} \partial_x (K\sqrt{A_0}) - \frac{2A}{3\rho} \sqrt{A} \partial_x K. \tag{2.2.4}$$

In compact form this system reads

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) = \mathbf{S}(\mathbf{Q}; A_0, K) \tag{2.2.5}$$

where

$$\mathbf{Q} = \begin{pmatrix} A \\ q \end{pmatrix}, \quad \mathbf{F}(\mathbf{Q}) = \begin{pmatrix} q \\ \dfrac{q^2}{A} + \gamma A^{\frac{3}{2}} \end{pmatrix} \quad \text{and} \quad \mathbf{S}(\mathbf{Q}; A_0, K) = \begin{pmatrix} 0 \\ s \end{pmatrix}.$$

Note that, if both $K(x)$ and $A_0(x)$ are constant, then $s = 0$ and equations (2.2.5) reduce themselves to a system of conservation laws. We can easily see that the two eigenvalues of system (2.2.5) are $\lambda^{\pm} = u \pm c$, where c is the wave speed defined by

$$c = \sqrt{\frac{3}{2} \gamma \sqrt{A}}.$$

Consequently, the convective part of (2.2.5) is strictly hyperbolic as long as $\lambda^{\pm}$ are real and distinct, namely if the vector of unknowns $\mathbf{Q}$ belongs to the phase space $\Omega = \{(A, Au)^t \in$

$\mathbb{R}^2|A > 0\}$. Finally, both characteristic fields are genuinely non-linear and the Riemann invariants associated with $\lambda^{\pm}$ are given by $I^- = u + 4c$ and $I^+ = u - 4c$ respectively. For more details refer to [31].

In this work we are specially interested in developing second-order well-balanced Lagrange-projection methods. As such, we hereafter introduce both the well-balanced property and the Lagrange-projection decomposition.

The well-balanced property. A numerical scheme is (fully) well-balanced if it is able to preserve the smooth stationary solutions of the system, that is to say the steady states which satisfy the ordinary differential equations

$$\partial_x \mathbf{F}(\mathbf{Q}) = \mathbf{S}(\mathbf{Q}; A_0, K),$$

and hence

$$q = q_0 = \text{constant}, \quad \frac{q_0^2}{2A^2} + \frac{K}{\rho}(\sqrt{A} - \sqrt{A_0}) = \text{constant}, \tag{2.2.6}$$

where the quantity $E = \dfrac{q^2}{2A^2} + \dfrac{K}{\rho}(\sqrt{A} - \sqrt{A_0})$ can be referred as the energy discharge. As we already said, a scheme able to preserve the steady states (2.2.6) is called fully well-balanced, while a method which only conserves the stationary solutions with zero velocity ($u = 0$) is defined as well-balanced. We are interested in a scheme endowed with the latter property, and thus in the "man at eternal rest" solution,

$$q = 0, \quad K(\sqrt{A} - \sqrt{A_0}) = \text{constant}. \tag{2.2.7}$$

For more details about well-balanced schemes for blood flow equations, see again [15, 20, 27, 36].

The Lagrangian coordinates. Observing that system (2.2.3) is given in Eulerian coordinates, we now want to express it using the Lagrangian coordinates, which describe the flow following the fluid motion. While with Eulerian coordinates the viewer has a fixed position and watches the flow from the exterior, with Lagrangian coordinates he focuses on a single "fluid particle" in the position ξ. Hence, we introduce the characteristic curves

$$\begin{cases} \dfrac{\partial x}{\partial t}(\xi, t) = u(x(\xi, t), t) \\ x(\xi, 0) = \xi. \end{cases} \tag{2.2.8}$$

Then, given the trajectory : $t \to x(\xi, t)$, any function : $(x, t) \to \mathbf{Q}(x, t)$ in Eulerian coordinates can be written in Lagrangian coordinates,

$$\overline{\mathbf{Q}}(\xi, t) = \mathbf{Q}(x(\xi, t), t).$$

Moreover, defining the volume ratio

$$L(\xi, t) = \frac{\partial x}{\partial \xi}(\xi, t) \tag{2.2.9}$$

which satisfies

$$\begin{cases} \dfrac{\partial L}{\partial t}(\xi, t) = \partial_\xi u(x(\xi, t), t) \\ L(\xi, 0) = 1 \end{cases} \tag{2.2.10}$$

and

$$\partial_t L(\xi,t) = \partial_\xi u(x(\xi,t),t) = \partial_\xi \bar{u}(\xi,t),$$

we easily find time and space derivatives in Lagrangian coordinates,

$$\partial_\xi \overline{\mathbf{Q}}(\xi,t) = L(\xi,t)\partial_x \mathbf{Q}(x,t) \quad \text{and} \quad \partial_t \overline{\mathbf{Q}}(\xi,t) = \partial_t \mathbf{Q}(x,t) + u(x,t)\partial_x \mathbf{Q}(x,t).$$

Hence, using the chain rule and defining $\tilde{p} = \tilde{p}(A;K) = \gamma A^{\frac{3}{2}}$, from system (2.2.3) we write

$$\begin{cases} \partial_t A + A\partial_x u + u\partial_x A = 0 \\ \partial_t (Au) + u\partial_x (Au) + Au\partial_x u + \partial_x \tilde{p} = s, \end{cases} \tag{2.2.11}$$

and multiplying by $L(\xi,t)$, we obtain

$$\begin{cases} L\partial_t \overline{A} + \overline{A}\partial_t L = 0 \\ L\partial_t (\overline{Au}) + \overline{Au}\partial_t L + \partial_\xi \bar{\tilde{p}} = \bar{s}, \end{cases}$$

where $\bar{s} = \dfrac{A}{\rho}\partial_\xi(\overline{K\sqrt{A_0}}) - \dfrac{2A}{3\rho}\sqrt{A}\partial_\xi \overline{K}$. We finally find that, in Lagrangian coordinates, system (2.2.3) reads,

$$\begin{cases} \partial_t (L\overline{A}) = 0 \\ \partial_t (L\overline{Au}) + \partial_\xi \bar{\tilde{p}} = \bar{s}. \end{cases} \tag{2.2.12}$$

Hence, the Lagrange-projection algorithm consists of two steps:

1. Solve system (2.2.12) written in Lagrangian coordinates;
2. Project the solution of system (2.2.12) in Eulerian coordinates.

For more details about the Lagrange-projection decomposition, once again we refer the reader to [5, 10, 9, 11, 16, 25] and the references therein.

2.2.1 Lagrange-projection splitting and relaxation formulation

At this stage, we present the Lagrange-projection decomposition in a different way, which will prove to be extremely useful for the numerical method we are going to describe in sections 2.3.2.1 and 2.4.2.1. In particular, we split system (2.2.3) into two different ones, the acoustic (Lagrangian) system and the transport system (projection step). We take into account the acoustic effects and parameters variations in the former and transport phenomena in the latter, see also [5, 11]. In particular, starting from formulation (2.2.11), we find that the acoustic equations read

$$\begin{cases} \partial_t A + A\partial_x u = 0 \\ \partial_t (Au) + Au\partial_x u + \partial_x \tilde{p} = s, \end{cases} \tag{2.2.13}$$

while the transport system is given by

$$\begin{cases} \partial_t A + u\partial_x A = 0 \\ \partial_t q + u\partial_x q = 0. \end{cases} \tag{2.2.14}$$

The latter can also simply be seen as a system of two equations of the form $\partial_t X + u\partial_x X = 0$, with either $X = A$ or $X = q$.

Then, let us show a useful reformulation of the acoustic system (2.2.13). We rewrite it as

$$\begin{cases} -\dfrac{1}{A^2}(\partial_t A + A\partial_x u) = 0 \\ A\partial_t u + u\partial_t A + Au\partial_x u + \partial_x \tilde{p} = s \end{cases}$$

and defining the specific volume $\tau = \dfrac{1}{A}$ and the mass variable m such that $\dfrac{1}{A}\partial_x = \partial_m$, the acoustic system also reads

$$\begin{cases} \partial_t \tau - \partial_m u = 0 \\ \partial_t u + \partial_m \tilde{p} = \tilde{s}, \end{cases} \tag{2.2.15}$$

with $\tilde{s} = \dfrac{A}{\rho}\partial_m(K\sqrt{A_0}) - \dfrac{2A}{3\rho}\sqrt{A}\partial_m K$. Observe that, from a numerical point of view, (2.2.15) will be equivalent to (2.2.12). System (2.2.15) has two eigenvalues $\lambda_\pm = \pm Ac$, and it is strictly hyperbolic in the same phase space of system (2.2.3), namely when $A > 0$, with the two characteristic fields being genuinely non-linear.

Relaxation formulation. At this stage, we are interested in finding an approximate solution of a Riemann Problem (RP) for system (2.2.15). For this purpose we exploit the Suliciu relaxation approach, which allows us to enlarge equations (2.2.15) to a strictly hyperbolic system with only linearly degenerate characteristic fields, which is well-known to be easier to solve. For the Suliciu relaxation approach and related applications refer to [29, 4, 2, 12, 13, 14] and the references therein.

Thus, we introduce the relaxation parameter λ and the new variable Π such that at least formally

$$\lim_{\lambda\to\infty} \Pi = \tilde{p},$$

where Π can be interpreted as a linearization of the pressure $\tilde{p}$. We observe that $\partial_\tau \tilde{p}(\tau) = -A^2\partial_A \tilde{p}(A)$ and $\partial_t \tilde{p} = \partial_\tau \tilde{p}(\tau)\partial_t \tau$ so that, multiplying the first equation of system (2.2.15) by $\partial_\tau \tilde{p}(\tau)$, we find that $\partial_t \tilde{p} + A^2c^2\partial_m u = 0$. The latter motivates the relaxation system

$$\begin{cases} \partial_t \tau - \partial_m u = 0 \\ \partial_t u + \partial_m \Pi = \tilde{s} \\ \partial_t \Pi + a^2\partial_m u = \lambda(\tilde{p}(\tau) - \Pi) \end{cases} \tag{2.2.16}$$

where a^2 is a constant which linearizes A^2c^2 and it should be taken as $a^2 \geq A^2c^2$ according to the sub-characteristic condition. Indeed, this condition entails that the information in the relaxation model (2.2.16) propagates faster than in the original one (2.2.15), see also [29, 11, 12].

Considering that the initial data for Π is well-prepared in the sense that $\Pi = \tilde{p}$, it is natural to introduce a more compact notation for system (2.2.16), which reads

$$\partial_t \mathbf{U} + \partial_m \mathbf{G}(\mathbf{U}) = \tilde{\mathbf{S}}$$

with

$$\mathbf{U} = \begin{pmatrix} \tau \\ u \\ \Pi \end{pmatrix}, \quad \mathbf{G}(\mathbf{U}) = \begin{pmatrix} -u \\ \Pi \\ a^2 u \end{pmatrix} \quad \text{and} \quad \tilde{\mathbf{S}} = \begin{pmatrix} 0 \\ \tilde{s} \\ 0 \end{pmatrix},$$

where we also note that the relaxation source term in the evolution equation for Π is not present anymore. Rewriting this system in quasi-linear form, it gives

$$\partial_t \mathbf{U} + \mathbf{A}(\mathbf{U})\partial_m \mathbf{U} = \tilde{\mathbf{S}}$$

where $\mathbf{A}(\mathbf{U})$ is the Jacobian matrix of the flux vector $\mathbf{G}(\mathbf{U})$, that is

$$\mathbf{A}(\mathbf{U}) = \frac{\partial \mathbf{G}}{\partial \mathbf{U}} = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & a^2 & 0 \end{pmatrix}.$$

Then, the eigenvalues of the matrix $\mathbf{A}(\mathbf{U})$ are given by

$$\lambda_- = -a, \quad \lambda_0 = 0 \quad \text{and} \quad \lambda_+ = a,$$

where $\pm a$ can be seen as constant approximations of the eigenvalues of system (2.2.15). Therefore system (2.2.16) is strictly hyperbolic as long as a is real and $a \neq 0$. The corresponding right eigenvectors read

$$\mathbf{R}_- = \begin{pmatrix} 1 \\ a \\ -a^2 \end{pmatrix}, \quad \mathbf{R}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{R}_+ = \begin{pmatrix} 1 \\ -a \\ -a^2 \end{pmatrix}.$$

Since all the three associated characteristic fields are linearly degenerate, the three waves are contact discontinuities and, as such, the Riemann problem can be solved using the Riemann invariants. In order to find them, we observe that the equations in phase space associated with $\lambda_\pm$ are given by

$$\frac{d\tau}{1} = \frac{du}{\mp a} = \frac{d\Pi}{-a^2}$$

and they lead to the Riemann invariants $RI_{\pm,1} = a\tau \pm u$ and $RI_{\pm,2} = \Pi \mp au$. Finally, the corresponding equations to λ_0 in phase space are

$$du = 0 \quad \text{and} \quad d\Pi = 0$$

and the associated Riemann invariants are $RI_{0,1} = u$ and $RI_{0,2} = \Pi$.

At this stage, we can look for an approximate solution of system (2.2.16) for a Riemann problem. We will use the theory of Gallice [18, 19], which consists in an extension of the well-known Harten, Lax and van Leer formalism [22] for systems of conservation laws.

2.2.2 Approximate Riemann solver for the acoustic system

Let us start by giving some general notions about the approximate Riemann solver and consistency with the integral form as described by Gallice [18, 19]. We will then focus on our specific case, namely system (2.2.16).

We briefly consider a general system of form

$$\partial_t \mathbf{U} + \partial_m \mathbf{G}(\mathbf{U}) = \tilde{\mathbf{S}} \tag{2.2.17}$$

where $\mathbf{U}$ is the vector of h unknowns, $\mathbf{G}(\mathbf{U})$ is the physical flux and $\tilde{\mathbf{S}}$ is the source term. We want to solve the Riemann Problem (RP) with the following initial condition,

$$\mathbf{U}(m, t=0) = \begin{cases} \mathbf{U}_L & \text{if} \quad m < 0 \\ \mathbf{U}_R & \text{if} \quad m > 0, \end{cases} \tag{2.2.18}$$

for any given $\mathbf{U}_L, \mathbf{U}_R$ in the phase space. Assuming we have h discontinuities with velocities λ_k, $k = 1, \ldots, h$, the solution of the Riemann problem consists of $h + 1$ states separated by

the discontinuities, hence

$$\mathbf{U}(\frac{m}{t},\mathbf{U}_L,\mathbf{U}_R)=\begin{cases}\mathbf{U}_1=\mathbf{U}_L & \text{if} \quad \frac{m}{t}<\lambda_1\\ \vdots & \vdots\\ \mathbf{U}_k & \text{if} \quad \lambda_k<\frac{m}{t}<\lambda_{k+1}\\ \vdots & \vdots\\ \mathbf{U}_{h+1}=\mathbf{U}_R & \text{if} \quad \frac{m}{t}>\lambda_h\end{cases} \tag{2.2.19}$$

Given the space and time steps Δm and Δt, the approximate solution (2.2.19) of the RP (2.2.17)-(2.2.18) is defined as being consistent with the integral form of (2.2.17) if in the interval $[0,\Delta m]$ we have

$$\mathbf{G}(\mathbf{U}_R)-\mathbf{G}(\mathbf{U}_L)-\Delta m\tilde{\mathbf{S}}(\Delta m,\Delta t;\mathbf{U}_L,\mathbf{U}_R)=\sum_{k=1}^{h}\lambda_k(\mathbf{U}_{k+1}-\mathbf{U}_k), \tag{2.2.20}$$

where the source term $\tilde{\mathbf{S}}(\Delta m,\Delta t;\mathbf{U}_L,\mathbf{U}_R)$ also has to satisfy a consistency property, namely

$$\lim_{\mathbf{U}_L,\mathbf{U}_R\to\mathbf{U};\Delta m,\Delta t\to 0}\tilde{\mathbf{S}}(\Delta m,\Delta t;\mathbf{U}_L,\mathbf{U}_R)=\tilde{\mathbf{S}}(\mathbf{U}). \tag{2.2.21}$$

2.2.2.1 Riemann solver for the system of conservation laws

Next, we focus on system (2.2.16). We start by neglecting the source term, hence we want to solve the following Riemann problem,

$$\begin{cases}\partial_t\mathbf{U}+\partial_m\mathbf{G}(\mathbf{U})=0\\ \mathbf{U}(m,t=0)=\begin{cases}\mathbf{U}_L & \text{if} \quad m<0\\ \mathbf{U}_R & \text{if} \quad m>0\end{cases}\end{cases} \tag{2.2.22}$$

where, in particular,

$$\mathbf{U}_L=\begin{pmatrix}\tau_L\\ u_L\\ \Pi_L\end{pmatrix} \quad \text{and} \quad \mathbf{U}_R=\begin{pmatrix}\tau_R\\ u_R\\ \Pi_R\end{pmatrix}.$$

In this specific case, the solution can be exactly computed and takes the following form

$$\mathbf{U}(\frac{m}{t},\mathbf{U}_L,\mathbf{U}_R)=\begin{cases}\mathbf{U}_L & \text{if} \quad \frac{m}{t}<\lambda_-=-a\\ \mathbf{U}_{*,L} & \text{if} \quad \lambda_-<\frac{m}{t}<0\\ \mathbf{U}_{*,R} & \text{if} \quad 0<\frac{m}{t}<\lambda_+=a\\ \mathbf{U}_R & \text{if} \quad \frac{m}{t}>\lambda_+\end{cases} \tag{2.2.23}$$

with

$$\mathbf{U}_{*,L}=\begin{pmatrix}\tau_{*,L}\\ u_{*,L}\\ \Pi_{*,L}\end{pmatrix} \quad \text{and} \quad \mathbf{U}_{*,R}=\begin{pmatrix}\tau_{*,R}\\ u_{*,R}\\ \Pi_{*,R}\end{pmatrix}. \tag{2.2.24}$$

In order to find the star values $\mathbf{U}_{*,L}$ and $\mathbf{U}_{*,R}$, we exploit the consistency conditions (2.2.20) together with the Rankine-Hugoniot conditions associated with the velocities $\pm a$, namely

$$a(\mathbf{U}_{*,L}-\mathbf{U}_L)+\mathbf{G}(\mathbf{U}_{*,L})-\mathbf{G}(\mathbf{U}_L)=0 \quad \text{and} \quad -a(\mathbf{U}_R-\mathbf{U}_{*,R})+\mathbf{G}(\mathbf{U}_{*,R})-\mathbf{G}(\mathbf{U}_R)=0, \tag{2.2.25}$$

and with the discontinuity of null velocity, i.e. $u_{*,L} = u_{*,R} = u_*$, $\Pi_{*,L} = \Pi_{*,R} = \Pi_*$. As such, we can easily find the following algebraic system of 6 relations,

$$\begin{cases} a\tau_{*,L} - u_{*,L} = a\tau_L - u_L \\ au_{*,L} + \Pi_{*,L} = au_L + \Pi_L \\ u_{*,L} = u_{*,R} = u_* \\ \Pi_{*,L} = \Pi_{*,R} = \Pi_* \\ a\tau_{*,R} + u_{*,R} = a\tau_R + u_R \\ au_{*,R} - \Pi_{*,R} = au_R - \Pi_R, \end{cases}$$

whose resolution leads us to the following values in the star regions,

$$\begin{cases} \tau_{*,L} = \tau_L + \dfrac{1}{a}(u_* - u_L) = \tau_L + \dfrac{1}{2a}(u_R - u_L) - \dfrac{1}{2a^2}(\Pi_R - \Pi_L) \\ \tau_{*,R} = \tau_R - \dfrac{1}{a}(u_* - u_R) = \tau_R + \dfrac{1}{2a}(u_R - u_L) + \dfrac{1}{2a^2}(\Pi_R - \Pi_L) \\ u_* = \dfrac{1}{2}(u_L + u_R) - \dfrac{1}{2a}(\Pi_R - \Pi_L) \\ \Pi_* = \dfrac{1}{2}(\Pi_L + \Pi_R) - \dfrac{a}{2}(u_R - u_L). \end{cases} \tag{2.2.26}$$

2.2.2.2 Riemann solver for the system of balance laws

Let us now consider the source term due to the spatial variations of the cross-sectional area at equilibrium A_0 and arterial stiffness K as well. Therefore, we look for an approximate solution of the following Riemann Problem,

$$\begin{cases} \partial_t \mathbf{U} + \mathbf{A}(\mathbf{U})\partial_m \mathbf{U} = \tilde{\mathbf{S}}(\mathbf{Q}; A_0, K) \\ \mathbf{U}(m, t=0) = \begin{cases} \mathbf{U}_L & \text{if} \quad m < 0 \\ \mathbf{U}_R & \text{if} \quad m > 0. \end{cases} \end{cases} \tag{2.2.27}$$

Once again we assume that the $m-t$ plane is divided in four different zones by the three waves and the solution has the form given by (2.2.23) and (2.2.24). In order to find such a solution made of 6 unknowns, we need to impose different conditions. Starting from the consistency relations (2.2.20), we have

$$\begin{cases} -u_R + u_L = -a(\tau_{*,L} - \tau_L) + a(\tau_R - \tau_{*,R}) \\ \Pi_R - \Pi_L - \Delta m \tilde{s} = -a(u_{*,L} - u_L) + a(u_R - u_{*,R}) \\ a^2(u_R - u_L) = -a(\Pi_{*,L} - \Pi_L) + a(\Pi_R - \Pi_{*,R}). \end{cases} \tag{2.2.28}$$

Then, the Rankine-Hugoniot conditions associated with the mass equation read

$$\begin{cases} -(u_R - u_{*,R}) = a(\tau_R - \tau_{*,R}) \\ u_L - u_{*,L} = -a(\tau_{*,L} - \tau_L) \\ u_{*,L} = u_{*,R} = u_* \end{cases} \tag{2.2.29}$$

which only give us two additional conditions, as the first relation of (2.2.28) is a linear combination of (2.2.29). Consequently, we have only found five conditions for six unknowns and as such they are not sufficient to define the approximate Riemann solution. Also note that $\tilde{s} = \tilde{s}(\Delta m, \Delta t; \mathbf{U}_L, \mathbf{U}_R)$ in (2.2.28) has to be specified such that (2.2.21) holds true. In particular, $\tilde{s}$ should be determined in such a way that it is equal to zero if A_0 and K are both

constant. With this requirement, the solution of (2.2.27) would degenerate to the solution of the homogeneous Riemann problem (2.2.22). Finally, two relations are still missing to define our solution.

Assume we have two other equations, namely $\partial_t A_0 = 0$ and $\partial_t K = 0$. For these two equations, the solutions are $A_{0,L}$ if $m < 0$, $A_{0,R}$ if $m > 0$ and K_L if $m < 0$, K_R if $m > 0$. Hence, we ask for the jump condition across the middle discontinuity associated with the momentum equation, that is to say

$$\Pi_{*R} - \Pi_{*L} + \mathcal{M} = 0, \tag{2.2.30}$$

where the function

$$\mathcal{M} = -\frac{\Delta m_L + \Delta m_R}{2}\tilde{s}(\Delta m, \Delta t; \mathbf{U}_L, \mathbf{U}_R)$$

has to be defined such that it satisfies $\mathcal{M} = 0$ if $A_{0,L} = A_{0,R}$ and $K_L = K_R$. Thus, after few algebraic computations, we obtain the following solution for the Riemann problem (2.2.27),

$$\begin{cases} \tau_{*,L} = \tau_L + \dfrac{1}{a}(u_* - u_L) \\ \tau_{*,R} = \tau_R - \dfrac{1}{a}(u_* - u_R) \\ u_* = \dfrac{1}{2}(u_L + u_R) - \dfrac{1}{2a}(\Pi_R - \Pi_L) - \dfrac{\mathcal{M}}{2a} \\ \Pi_{*L} = \dfrac{1}{2}(\Pi_L + \Pi_R) - \dfrac{a}{2}(u_R - u_L) + \dfrac{\mathcal{M}}{2} \\ \Pi_{*R} = \dfrac{1}{2}(\Pi_L + \Pi_R) - \dfrac{a}{2}(u_R - u_L) - \dfrac{\mathcal{M}}{2} \\ \Pi_* = \dfrac{1}{2}(\Pi_{*L} + \Pi_{*R}) = \dfrac{1}{2}(\Pi_L + \Pi_R) - \dfrac{a}{2}(u_R - u_L). \end{cases} \tag{2.2.31}$$

We have yet to define $\mathcal{M}$ in a consistent way, and it is clear that

$$\mathcal{M} = -\left\{\frac{A}{\rho}\right\}(K_R\sqrt{A_{0,R}} - K_L\sqrt{A_{0,L}}) + \left\{\frac{2A}{3\rho}\sqrt{A}\right\}(K_R - K_L)$$

is relevant, provided that $\left\{\dfrac{A}{\rho}\right\}$ and $\left\{\dfrac{2A}{3\rho}\sqrt{A}\right\}$ are consistent approximations of $\dfrac{A}{\rho}$ and $\dfrac{2A}{3\rho}\sqrt{A}$ respectively. Let us discuss this issue. Firstly, we point out that since $\mathbf{U}_L$ and $\mathbf{U}_R$ are taken to be well prepared, we have $\Pi_L = \dfrac{K_L}{3\rho}A_L^{\frac{3}{2}}$ and $\Pi_R = \dfrac{K_R}{3\rho}A_R^{\frac{3}{2}}$. Since we are interested in preserving the "man at eternal rest" solution, we also ask for the well-balanced condition: $\mathbf{U}_{*,L} = \mathbf{U}_L$ and $\mathbf{U}_{*,R} = \mathbf{U}_R$ if $\mathbf{U}_L$ and $\mathbf{U}_R$ verify the "man at eternal rest" solution, namely

$$u_L = u_R = 0, \quad K_L(\sqrt{A_L} - \sqrt{A_{0,L}}) = K_R(\sqrt{A_R} - \sqrt{A_{0,R}}). \tag{2.2.32}$$

If we have (2.2.32), then

$$\begin{aligned} \mathcal{M} &= -\left\{\frac{A}{\rho}\right\}(K_R\sqrt{A_{0,R}} - K_L\sqrt{A_{0,L}}) + \left\{\frac{2A}{3\rho}\sqrt{A}\right\}(K_R - K_L) \\ &= -\left\{\frac{A}{\rho}\right\}(K_R\sqrt{A_R} - K_L\sqrt{A_L}) + \left\{\frac{2A}{3\rho}\sqrt{A}\right\}(K_R - K_L) \\ &= -\left\{\frac{A}{\rho}\right\}\frac{K_R + K_L}{2}(\sqrt{A_R} - \sqrt{A_L}) - \left\{\frac{A}{\rho}\right\}\frac{\sqrt{A_R} + \sqrt{A_L}}{2}(K_R - K_L) + \left\{\frac{2A}{3\rho}\sqrt{A}\right\}(K_R - K_L) \end{aligned}$$

and thanks to (2.2.30), using formula $[XY]_L^R = \overrightarrow{X}[Y]_L^R + \overleftarrow{Y}[X]_L^R$, where $[X]_L^R = X_R - X_L$,

$\overrightarrow{X} = \alpha X_L + (1-\alpha)X_R$, $\overleftarrow{X} = \alpha X_R + (1-\alpha)X_L$ and $\alpha = \frac{1}{2}$, we get

$$\begin{aligned}\mathcal{M} &= -(\Pi_R - \Pi_L) = -\frac{1}{3\rho}(K_R A_R\sqrt{A_R} - K_L A_L\sqrt{A_L}) \\ &= -\frac{1}{6\rho}(K_R + K_L)(A_R\sqrt{A_R} - A_L\sqrt{A_L}) - \frac{1}{6\rho}(A_L\sqrt{A_L} + A_R\sqrt{A_R})(K_R - K_L) \\ &= -\frac{1}{3\rho}\frac{K_R + K_L}{2}(A_L + \sqrt{A_R}\sqrt{A_L} + A_R)(\sqrt{A_R} - \sqrt{A_L}) + \\ &\qquad - \frac{1}{6\rho}(A_L\sqrt{A_L} + A_R\sqrt{A_R})(K_R - K_L).\end{aligned}$$

Therefore, it is clearly sufficient to set

$$\left\{\frac{A}{\rho}\right\} = \frac{A_L + A_R + \sqrt{A_L}\sqrt{A_R}}{3\rho}$$

and

$$\begin{aligned}\left\{\frac{2A}{3\rho}\sqrt{A}\right\} &= \frac{\sqrt{A_L} + \sqrt{A_R}}{2}\left\{\frac{A}{\rho}\right\} - \frac{1}{6\rho}(A_L\sqrt{A_L} + A_R\sqrt{A_R}) \\ &= \frac{\sqrt{A_L}\sqrt{A_R}(\sqrt{A_L} + \sqrt{A_R})}{3\rho}.\end{aligned}$$

Finally, we get

$$\begin{aligned}&\mathcal{M} = \mathcal{M}((A_L, A_{0,L}, K_L); (A_R, A_{0,R}, K_R)) \\ &= -\frac{A_L + A_R + \sqrt{A_L}\sqrt{A_R}}{3\rho}\Big(K_R\sqrt{A_{0,R}} - K_L\sqrt{A_{0,L}} - \frac{\sqrt{A_L}\sqrt{A_R}(\sqrt{A_L} + \sqrt{A_R})}{A_L + A_R + \sqrt{A_L}\sqrt{A_R}}(K_R - K_L)\Big).\end{aligned} \tag{2.2.33}$$

Lastly, let us note that the definition of $\mathcal{M}$ is consistent:

$$\lim_{A_L, A_R \to A} \mathcal{M} = -\frac{A}{\rho}\Big(K_R\sqrt{A_{0,R}} - K_L\sqrt{A_{0,L}}\Big) + \frac{2A\sqrt{A}}{3\rho}(K_R - K_L).$$

2.3 First-order well-balanced scheme

The next step is to present the first-order well-balanced scheme. We start by giving a first-order approximation of the homogeneous version of system (2.2.3) and then we modify it in order to satisfy the well-balanced property and include the source term at the same time. In particular, for the latter step, we show two different ways of preserving the stationary state (2.2.7), one of them exploiting the approximate solution of the Riemann problem (2.2.27) and the other exploiting the well-known hydrostatic reconstruction procedure, for which we refer to [11] and [25] respectively.

Let us now introduce some notations. First of all, we define the constant space step Δx and constant time step Δt. The mesh interfaces are given by $x_{j+1/2} = j\Delta x$ for $j \in \mathbb{Z}$ and the intermediate times by $t^n = n\Delta t$ for $n \in \mathbb{N}$. As usual in the finite volume framework, we seek for an approximation $\mathbf{Q}_j^n$ of the solution in the interval $[x_{j-1/2}, x_{j+1/2})$, $j \in \mathbb{Z}$, at each time t^n. Therefore, a piecewise constant approximate solution $x \to \mathbf{Q}_{\Delta t, \Delta x}(x, t^n)$ of the solution $\mathbf{Q}$ is given by

$$\mathbf{Q}_{\Delta t, \Delta x}(x, t^n) = \mathbf{Q}_j^n \text{ for all } x \in C_j = [x_{j-1/2}; x_{j+1/2}), \quad j \in \mathbb{Z}, \quad n \in \mathbb{N}.$$

As far as the variable ξ is concerned, we use the same space discretization of x, hence $\Delta x = \Delta\xi$, $x_{j+\frac{1}{2}} = \xi_{j+\frac{1}{2}}$ and $x_j = \xi_j \ \forall j$.

2.3.1 First-order approximation with constant parameters

We start by describing the first-order scheme by assuming A_0 and K constant parameters, therefore the source term s disappears. As we already pointed out, the Lagrange-projection scheme is composed of two steps:

1. Solve system (2.2.12) written in Lagrangian coordinates, or equivalently update $\mathbf{Q}^n$ to $\mathbf{Q}^{n+1-}$ approximating the solution of system (2.2.13);
2. Project the solution of system (2.2.12) in Eulerian coordinates, namely update $\mathbf{Q}^{n+1-}$ to $\mathbf{Q}^{n+1}$ by solving system (2.2.14).

For more details about this procedure in the shallow water context, refer to [5, 11, 25].

Lagrangian step. Given a system of form $\partial_t \mathbf{U} + \partial_m \mathbf{G}(\mathbf{U}) = 0$ as in (2.2.22), the first-order Godunov-type scheme associated with the Riemann solver of section 2.2.2.1 reads

$$\mathbf{U}_j^{n+1-} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta m_j}(\mathbf{G}_{j+\frac{1}{2}}^n - \mathbf{G}_{j-\frac{1}{2}}^n), \tag{2.3.1}$$

with

$$\mathbf{G}_{j+\frac{1}{2}}^n = \mathbf{G}(\mathbf{U}_j^n, \mathbf{U}_{j+1}^n), \tag{2.3.2}$$

and

$$\mathbf{G}(\mathbf{U}_L, \mathbf{U}_R) = \frac{1}{2}\Big(\mathbf{G}(\mathbf{U}_L) + \mathbf{G}(\mathbf{U}_R) - \sum_{k=1}^{h} |\lambda_k|(\mathbf{U}_{k+1} - \mathbf{U}_k)\Big), \tag{2.3.3}$$

where λ_k the speeds of the discontinuities and $\mathbf{U}_k$ the intermediate states. In more details, we get that the natural discretization of the homogeneous version of the first two equations of system (2.2.16) is given by

$$\begin{cases} \tau_j^{n+1-} = \tau_j^n + \dfrac{\Delta t}{\Delta m_j}(u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}) \\ u_j^{n+1-} = u_j^n - \dfrac{\Delta t}{\Delta m_j}(\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}), \end{cases} \tag{2.3.4}$$

with $\Delta m_j = \dfrac{\Delta x}{\tau_j^n}$ and $a^n_{j+\frac{1}{2}} = \max((Ac)_j^n, (Ac)_{j+1}^n)$. The numerical fluxes are given by $u^*_{j+\frac{1}{2}}$ and $\Pi^*_{j+\frac{1}{2}}$ at time t^n and, in particular, exploiting formulae (2.2.26),

$$\begin{aligned} u^*_{j+\frac{1}{2}} &= u^*_{j+\frac{1}{2}}(\mathbf{U}_j^n, \mathbf{U}_{j+1}^n) = \frac{1}{2}(u_{j+1}^n + u_j^n) - \frac{1}{2a^n_{j+\frac{1}{2}}}(\Pi_{j+1}^n - \Pi_j^n), \\ \Pi^*_{j+\frac{1}{2}} &= \Pi^*_{j+\frac{1}{2}}(\mathbf{U}_j^n, \mathbf{U}_{j+1}^n) = \frac{1}{2}(\Pi_{j+1}^n + \Pi_j^n) - \frac{a^n_{j+\frac{1}{2}}}{2}(u_{j+1}^n - u_j^n). \end{aligned} \tag{2.3.5}$$

Note that the first equation of system (2.3.4) is equivalent to

$$A_j^n = A_j^{n+1-}\big(1 + \frac{\Delta t}{\Delta x}(u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}})\big). \tag{2.3.6}$$

For more details we refer again to [5, 10, 9, 11, 18, 19, 25].

It is useful to show a numerical discretization for the Lagrangian step starting from system (2.2.12) with no source term, namely

$$\begin{cases} L_j^{n+1-} A_j^{n+1-} = L_j^n A_j^n \\ L_j^{n+1-}(Au)_j^{n+1-} = L_j^n (Au)_j^n - \dfrac{\Delta t}{\Delta x}(\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}) \end{cases} \tag{2.3.7}$$

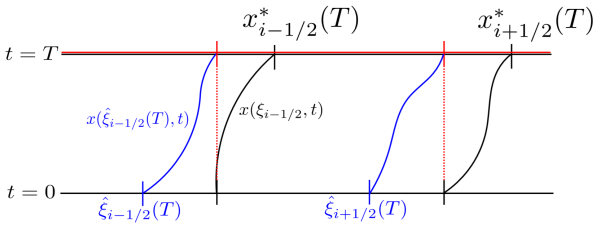


Figure 2.1: Connection between Lagrangian and Eulerian coordinates, see [25].

with $\Pi^*_{j\pm\frac{1}{2}}$ given by (2.3.5); let us remark that this approximation (2.3.7) is equivalent to (2.3.4). Moreover, we note that we have neglected the bars over the unknowns for the sake of simplicity. Lastly, observe from system (2.2.10) that a natural discretization for $L(\xi, t)$ is the following,

$$L_j^{n+1-} = L_j^n + \frac{\Delta t}{\Delta x}(u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}) \quad \text{with} \quad L_j^n = 1, \tag{2.3.8}$$

$u^*_{j\pm\frac{1}{2}}$ given by (2.3.5) and, from its definition (2.2.9),

$$L_j(t) = \frac{1}{\Delta x}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \frac{\partial x}{\partial \xi}(\xi, t)d\xi = \frac{x^*_{j+\frac{1}{2}} - x^*_{j-\frac{1}{2}}}{\Delta x}$$

with $x^*_{j\pm\frac{1}{2}} = x(\xi_{j\pm\frac{1}{2}}, t) \simeq \xi_{j\pm\frac{1}{2}} + t u^*_{j\pm\frac{1}{2}}$ according to (2.2.8). See also figure 2.1 for a better understanding.

Projection step. As already explained, the subsequent step is to project the solution of system (2.2.12) (here with no source term) in Eulerian coordinates. With this in mind, let us first note that, by definition of L,

$$\int_{x(\xi_l,t)}^{x(\xi_r,t)} \mathbf{Q}(x, t)dx = \int_{\xi_l}^{\xi_r} L(\xi, t)\overline{\mathbf{Q}}(\xi, t)d\xi.$$

This relation allows to make a simple link between $\mathbf{Q}$ in Eulerian and Lagrangian coordinates. In order to use it at the discrete level, it just remains to define $\hat{\xi}_{j+\frac{1}{2}}(t)$ such that for all j (see again figure 2.1)

$$x(\hat{\xi}_{j+\frac{1}{2}}(T), T) = x_{j+\frac{1}{2}}, \quad \text{with} \quad T \geq 0,$$

and the corresponding trajectories

$$\begin{cases} \dfrac{\partial x}{\partial t}(\hat{\xi}_{j+\frac{1}{2}}(T), t) = u(x(\hat{\xi}_{j+\frac{1}{2}}(T), t), t) \\ x(\hat{\xi}_{j+\frac{1}{2}}(T), 0) = \hat{\xi}_{j+\frac{1}{2}}(T), \end{cases}$$

leading to

$$\mathbf{Q}_j(t) = \frac{1}{\Delta x}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{Q}(x, t)dx = \frac{1}{\Delta x}\int_{x(\hat{\xi}_{j-\frac{1}{2}},t)}^{x(\hat{\xi}_{j+\frac{1}{2}},t)} \mathbf{Q}(x, t)dx = \frac{1}{\Delta x}\int_{\hat{\xi}_{j-\frac{1}{2}}}^{\hat{\xi}_{j+\frac{1}{2}}} L(\xi, t)\overline{\mathbf{Q}}(\xi, t)d\xi. \tag{2.3.9}$$

Note that we can approximate $x_{j+\frac{1}{2}}$ at first-order:

$$x_{j+\frac{1}{2}} = x(\hat{\xi}_{j+\frac{1}{2}}(T), T) \simeq x(\hat{\xi}_{j+\frac{1}{2}}(T), 0) + T\partial_t x(\hat{\xi}_{j+\frac{1}{2}}(T), 0) \simeq \hat{\xi}_{j+\frac{1}{2}} + T u^*_{j+\frac{1}{2}}, \tag{2.3.10}$$

for a fixed time $T \geq 0$ and observe that $x_{j+\frac{1}{2}} - x^*_{j+\frac{1}{2}} \simeq \hat{\xi}_{j+\frac{1}{2}} - \xi_{j+\frac{1}{2}}$. In order to define $\mathbf{Q}_j^{n+1}$ using $(L\overline{\mathbf{Q}})_j^{n+1-}$, we suggest to split the integral (2.3.9) to obtain

$$\begin{aligned}\mathbf{Q}_j^{n+1} =& \frac{1}{\Delta x}\int_{\hat{\xi}_{j-\frac{1}{2}}}^{\xi_{j-\frac{1}{2}}} L(\xi, t^{n+1-})\overline{\mathbf{Q}}(\xi, t^{n+1-})dx + \\ &+ \frac{1}{\Delta x}\int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} L(\xi, t^{n+1-})\overline{\mathbf{Q}}(\xi, t^{n+1-})dx + \frac{1}{\Delta x}\int_{\xi_{j+\frac{1}{2}}}^{\hat{\xi}_{j+\frac{1}{2}}} L(\xi, t^{n+1-})\overline{\mathbf{Q}}(\xi, t^{n+1-})dx.\end{aligned} \tag{2.3.11}$$

Then, the middle integral can be clearly replaced by $\frac{1}{\Delta x}\int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} L(\xi,t)\overline{\mathbf{Q}}(\xi,t)dx = (L\overline{\mathbf{Q}})_j^{n+1-}$, while the others two integrals can be approximated as in the following. Let us state $\frac{1}{\Delta x}\int_{\hat{\xi}_{j-\frac{1}{2}}}^{\xi_{j-\frac{1}{2}}} L(\xi,t)\overline{\mathbf{Q}}(\xi,t)dx = \frac{\xi_{j-\frac{1}{2}} - \hat{\xi}_{j-\frac{1}{2}}}{\Delta x}(L\overline{\mathbf{Q}})_{j-\frac{1}{2}}^{n+1-}$, where we set

$$(L\overline{\mathbf{Q}})_{j-\frac{1}{2}}^{n+1-} = \begin{cases}(L\overline{\mathbf{Q}})_{j-1}^{n+1-} & \text{if} \quad \xi_{j-\frac{1}{2}} > \hat{\xi}_{j-\frac{1}{2}} \\ (L\overline{\mathbf{Q}})_{j}^{n+1-} & \text{if} \quad \xi_{j-\frac{1}{2}} \leq \hat{\xi}_{j-\frac{1}{2}}\end{cases}$$

or equivalently

$$(L\overline{\mathbf{Q}})_{j-\frac{1}{2}}^{n+1-} = \begin{cases}(L\overline{\mathbf{Q}})_{j-1}^{n+1-} & \text{if} \quad x^{*,n+1-}_{j-\frac{1}{2}} > x_{j-\frac{1}{2}} \\ (L\overline{\mathbf{Q}})_{j}^{n+1-} & \text{if} \quad x^{*,n+1-}_{j-\frac{1}{2}} \leq x_{j-\frac{1}{2}}\end{cases} \tag{2.3.12}$$

with

$$x^{*,n+1-}_{j+\frac{1}{2}} = x_{j+\frac{1}{2}} + \Delta t u^*_{j+\frac{1}{2}} \quad \text{and} \quad x_{j+\frac{1}{2}} = \hat{\xi}_{j+\frac{1}{2}} + \Delta t u^*_{j+\frac{1}{2}},$$

and analogously for the third integral appearing in (2.3.11). Consequently, the projection step reads

$$\begin{aligned}\mathbf{Q}_j^{n+1} &= \frac{\xi_{j-\frac{1}{2}} - \hat{\xi}_{j-\frac{1}{2}}}{\Delta x}(L\mathbf{Q})_{j-\frac{1}{2}}^{n+1-} + (L\mathbf{Q})_j^{n+1-} + \frac{\hat{\xi}_{j+\frac{1}{2}} - \xi_{j+\frac{1}{2}}}{\Delta x}(L\mathbf{Q})_{j+\frac{1}{2}}^{n+1-} \\ &= (L\mathbf{Q})_j^{n+1-} - \frac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}(L\mathbf{Q})_{j+\frac{1}{2}}^{n+1-} - u^*_{j-\frac{1}{2}}(L\mathbf{Q})_{j-\frac{1}{2}}^{n+1-}\big),\end{aligned} \tag{2.3.13}$$

with $u^*_{j+\frac{1}{2}}$ defined as in the Lagrangian step and where we have neglected the bars over $\mathbf{Q}$. Let us also remark that (2.3.13) can be seen as a discretization of system (2.2.14).

Overall discretization. It can be easily proved that the whole scheme is conservative using formulae (2.3.7) and (2.3.13). Indeed, one can recover the following final form:

$$\begin{cases}A_j^{n+1} = A_j^n - \dfrac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}A_{j+\frac{1}{2}}^{n+1-} - u^*_{j-\frac{1}{2}}A_{j-\frac{1}{2}}^{n+1-}\big) \\ q_j^{n+1} = q_j^n - \dfrac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}q_{j+\frac{1}{2}}^{n+1-} + \Pi^*_{j+\frac{1}{2}} - (u^*_{j-\frac{1}{2}}q_{j-\frac{1}{2}}^{n+1-} + \Pi^*_{j-\frac{1}{2}})\big),\end{cases} \tag{2.3.14}$$

with

$$X_{j+\frac{1}{2}}^{n+1-} = \begin{cases}(LX)_j^{n+1-} & \text{if} \quad x^{*,n+1-}_{j+\frac{1}{2}} > x_{j+\frac{1}{2}} \\ (LX)_{j+1}^{n+1-} & \text{if} \quad x^{*,n+1-}_{j+\frac{1}{2}} \leq x_{j+\frac{1}{2}},\end{cases}$$

and X either A or q.

2.3.2 First-order approximation with variable parameters

As far as the source term is concerned, it is different from zero only when considering non-constant parameters, due to the presence of the spatial derivatives of A_0 and K. Since the source term is taken into consideration at the level of the Lagrangian step, its inclusion in the numerical scheme only requires to modify the first step and not the remap one. Note, however, that the projection step generally has to be modified in order to obtain a fully well-balanced numerical method, see [5, 25]. We show that this is not required in our particular case as we want to only preserve the "man at eternal rest" solution. In the rest of this section, we present two different ways to obtain a well-balanced Lagrangian step in which the source term is included.

2.3.2.1 Based on the approximate Riemann solver

The first approach we present requires the use of the approximate Riemann solver described in section 2.2.2.2. In particular the well-balanced property is achieved approximating the source term in a special way, namely exploiting formula (2.2.33).

The first-order Godunov-type method associated with the Riemann solver proposed in section 2.2.2.2 now reads

$$\mathbf{U}_j^{n+1-} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta m_j}(\mathbf{G}^n_{j+\frac{1}{2}} - \mathbf{G}^n_{j-\frac{1}{2}}) + \Delta t \tilde{\mathbf{S}}_j^n \tag{2.3.15}$$

with the numerical flux $\mathbf{G}^n_{j+\frac{1}{2}}$ as in (2.3.2)-(2.3.3) and the source term defined as the average of the source at the interfaces,

$$\tilde{\mathbf{S}}_j^n = \frac{1}{2}\Big(\frac{\Delta m_{j+\frac{1}{2}}}{\Delta m_j}\tilde{\mathbf{S}}^n_{j+\frac{1}{2}} + \frac{\Delta m_{j-\frac{1}{2}}}{\Delta m_j}\tilde{\mathbf{S}}^n_{j-\frac{1}{2}}\Big), \quad \tilde{\mathbf{S}}^n_{j+\frac{1}{2}} = \tilde{\mathbf{S}}(\Delta m_{j+\frac{1}{2}}, \Delta t; \mathbf{U}_j^n, \mathbf{U}_{j+1}^n) \tag{2.3.16}$$

with $\Delta m_{j+1/2} = \frac{\Delta m_j + \Delta m_{j+1}}{2}$. For more details we refer again to Gallice [18, 19]. Hence, we approximate system (2.2.15) by

$$\begin{cases} \tau_j^{n+1-} = \tau_j^n + \dfrac{\Delta t}{\Delta m_j}(u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}) \\ u_j^{n+1-} = u_j^n - \dfrac{\Delta t}{\Delta m_j}(\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}) + \Delta t \tilde{s}_j^n \end{cases} \tag{2.3.17}$$

or equivalently the Lagrangian system (2.2.12) by

$$\begin{cases} L_j^{n+1-} A_j^{n+1-} = L_j^n A_j^n \\ L_j^{n+1-}(Au)_j^{n+1-} = L_j^n (Au)_j^n - \dfrac{\Delta t}{\Delta x}(\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}) + \Delta t s_j^n, \end{cases} \tag{2.3.18}$$

with L_j^{n+1-} defined as in (2.3.8). Similarly, we observe that the numerical fluxes are now given by

$$\begin{cases} u^*_{j+\frac{1}{2}} = \dfrac{1}{2}(u^n_{j+1} + u^n_j) - \dfrac{1}{2a^n_{j+\frac{1}{2}}}(\Pi^n_{j+1} - \Pi^n_j) + \dfrac{\Delta m_{j+1/2}}{2a_{j+1/2}}\tilde{s}^n_{j+1/2} \\ \Pi^*_{j+\frac{1}{2}} = \dfrac{1}{2}(\Pi^n_{j+1} + \Pi^n_j) - \dfrac{a^n_{j+\frac{1}{2}}}{2}(u^n_{j+1} - u^n_j). \end{cases} \tag{2.3.19}$$

As far as the source term (2.2.4) is concerned, we state

$$\tilde{s}_j^n = \frac{1}{2}\Big(\frac{\Delta m_{j+\frac{1}{2}}}{\Delta m_j}\tilde{s}^n_{j+1/2} + \frac{\Delta m_{j-\frac{1}{2}}}{\Delta m_j}\tilde{s}^n_{j-1/2}\Big) \quad \text{with} \quad \tilde{s}^n_{j+1/2} = -\frac{\mathcal{M}^n_{j+1/2}}{\Delta m_{j+1/2}}, \tag{2.3.20}$$

and $\mathcal{M}^n_{j+1/2} = \mathcal{M}((A^n_j; A_{0,j}, K_j); (A^n_{j+1}; A_{0,j+1}, K_{j+1}))$ given by (2.2.33). Note that, in (2.3.18), $s^n_j = L^n_j A^n_j \tilde{s}^n_j = A^n_j \tilde{s}^n_j = \dfrac{\Delta m^n_j}{\Delta x}\tilde{s}^n_j$ or equivalently

$$s^n_j = \frac{1}{2}\Big(s^n_{j+1/2} + s^n_{j-1/2}\Big) \quad \text{with} \quad s^n_{j+1/2} = -\frac{\mathcal{M}^n_{j+1/2}}{\Delta x} \quad \forall j. \tag{2.3.21}$$

Theorem 1. *The Lagrangian step (2.3.18)-(2.3.21) (or equivalently (2.3.17)-(2.3.20)) here described is well-balanced under the "man at eternal rest" condition (2.2.7).*

Proof. Let us assume that $\mathbf{Q}^n_j$ satisfies the "man at eternal rest" condition (2.2.7). Thus, $u^n_j = 0\ \forall j$ and, thanks to condition (2.2.30), we also have that $\Pi^n_{j+1} - \Pi^n_j + \mathcal{M}_{j+1/2} = 0$ and, consequently, $-(\Pi^n_{j+1} - \Pi^n_j) + \Delta m_{j+1/2}\tilde{s}^n_{j+1/2} = 0$. Hence, $u^*_{j+\frac{1}{2}} = 0\,\forall j$ and $\Pi^*_{j+\frac{1}{2}} = \frac{1}{2}(\Pi^n_{j+1} + \Pi^n_j)$. Similarly, s^n_j compensates $-\frac{1}{\Delta x}(\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}})$ and we find that $A^{n+1-}_j = A^n_j$ and $q^{n+1-}_j = q^n_j$.

□

2.3.2.2 Based on the hydrostatic reconstruction

Let us consider once again a discretization of the form (2.3.18) for the Lagrangian system (2.2.12). We now want to exploit the well-known hydrostatic reconstruction approach [1, 5]. Thus, we are going to modify the numerical flux and source term to be employed in either system (2.3.17) or (2.3.18). Hence, we perform a reconstruction of the variables values $\mathbf{Q}$ at the cells interfaces. Considering relations (2.2.7) locally, we first write (for all j)

$$\begin{cases} q(x) = 0, \\ (K(\sqrt{A} - \sqrt{A_0}))(x) = K_j(\sqrt{A^n_j} - \sqrt{A_{0,j}}), \end{cases}$$

which can be understood as a reconstruction procedure of a space-dependent stationary solution inside the $j-$th cell. Then, following [20], we define cross-sectional areas at the cell interfaces ($x = x_{j\pm\frac{1}{2}}$) by

$$\begin{cases} (K\sqrt{A})^n_{j+\frac{1}{2},L} = \max\big(K_j(\sqrt{A^n_j} - \sqrt{A_{0,j}}) + (K\sqrt{A_0})_{j+\frac{1}{2}}, 0\big) \\ (K\sqrt{A})^n_{j+\frac{1}{2},R} = \max\big(K_{j+1}(\sqrt{A^n_{j+1}} - \sqrt{A_{0,j+1}}) + (K\sqrt{A_0})_{j+\frac{1}{2}}, 0\big), \end{cases} \tag{2.3.22}$$

where the maximum has been added in order to preserve the non-negativity, and

$$\begin{cases} \sqrt{A^n_{j+\frac{1}{2},L}} = \max\Big(\dfrac{1}{K^*_{j+\frac{1}{2}}}\big(K_j(\sqrt{A^n_j} - \sqrt{A_{0,j}}) + (K\sqrt{A_0})_{j+\frac{1}{2}}\big), 0\Big) \\ \sqrt{A^n_{j+\frac{1}{2},R}} = \max\Big(\dfrac{1}{K^*_{j+\frac{1}{2}}}\big(K_{j+1}(\sqrt{A^n_{j+1}} - \sqrt{A_{0,j+1}}) + (K\sqrt{A_0})_{j+\frac{1}{2}}\big), 0\Big) \end{cases}$$

where we have imposed

$$(K\sqrt{A_0})_{j+\frac{1}{2}} = \min\big(K_j\sqrt{A_{0,j}}, K_{j+1}\sqrt{A_{0,j+1}}\big)$$

and

$$K^*_{j+\frac{1}{2}} = \max\big(K_j, K_{j+1}\big). \tag{2.3.23}$$

Hence, the reconstructed values at cell interfaces are defined by

$$\mathbf{Q}^n_{j+\frac{1}{2},L} = \begin{pmatrix} A^n_{j+\frac{1}{2},L} \\ A^n_{j+\frac{1}{2},L}u^n_i \end{pmatrix} \quad \text{and} \quad \mathbf{Q}^n_{j+\frac{1}{2},R} = \begin{pmatrix} A^n_{j+\frac{1}{2},R} \\ A^n_{j+\frac{1}{2},R}u^n_{i+1} \end{pmatrix}. \tag{2.3.24}$$

At this stage, we can define the star values for the velocity u and linearized pressure Π as in (2.3.5), exploiting the values at interfaces (2.3.24):

$$u^*_{j+\frac{1}{2}} = u^*(\mathbf{Q}^n_{j+\frac{1}{2},L}, \mathbf{Q}^n_{j+\frac{1}{2},R}) \quad \text{and} \quad \Pi^*_{j+\frac{1}{2}} = \Pi^*(\mathbf{Q}^n_{j+\frac{1}{2},L}, \mathbf{Q}^n_{j+\frac{1}{2},R})$$

with

$$\begin{cases} u^*_{j+\frac{1}{2}} = \dfrac{1}{2}(u^n_{j+1} + u^n_j) - \dfrac{1}{2a^n_{j+\frac{1}{2}}}(\Pi^n_{j+\frac{1}{2},R} - \Pi^n_{j+\frac{1}{2},L}) \\ \Pi^*_{j+\frac{1}{2}} = \dfrac{1}{2}(\Pi^n_{j+\frac{1}{2},R} + \Pi^n_{j+\frac{1}{2},L}) - \dfrac{a^n_{j+\frac{1}{2}}}{2}(u^n_{j+1} - u^n_j) \end{cases} \tag{2.3.25}$$

and

$$a_{j+\frac{1}{2}} = \max\left(A^n_{j+\frac{1}{2},L} c^n_{j+\frac{1}{2},L}, A^n_{j+\frac{1}{2},R} c^n_{j+\frac{1}{2},R}\right).$$

In particular, (2.3.25) has to be used in (2.3.17)-(2.3.18) instead of (2.3.19). Finally, since we want to preserve the steady states with zero velocity, namely the ones satisfying $u = 0$ and $\partial_x \tilde{p} = s$, for the source term s^n_j we suggest the following,

$$\begin{aligned} s^n_j &= \frac{1}{\Delta x}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} s(\mathbf{Q}; A_0, K)(x,t)dx = \frac{1}{\Delta x}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \partial_x \tilde{p}(x,t)dx = \frac{1}{\Delta x}\left(\tilde{p}(x_{j+\frac{1}{2}}) - \tilde{p}(x_{j-\frac{1}{2}})\right) \\ &= \frac{\tilde{p}_{j+\frac{1}{2},L} - \tilde{p}_{j-\frac{1}{2},R}}{\Delta x}. \end{aligned} \tag{2.3.26}$$

Note that the first equality comes from space-dependent reconstruction of a solution inside the cell, while the second equality holds as we are exploiting the reconstructed values (2.3.24) to define the pressures $\tilde{p}$ at the interfaces. More precisely, let us remark that to define $\tilde{p}^n_{j+\frac{1}{2},L} = \tilde{p}(A^n_{j+\frac{1}{2},L}, K)$, $\tilde{p}^n_{j+\frac{1}{2},R} = \tilde{p}(A^n_{j+\frac{1}{2},R}, K)$ in (2.3.26), and $\Pi^n_{j+\frac{1}{2},L}$, $\Pi^n_{j+\frac{1}{2},R}$ at equilibrium in (2.3.25), namely $\Pi^n_{j+\frac{1}{2},L} = \tilde{p}(A^n_{j+\frac{1}{2},L}, K)$, $\Pi^n_{j+\frac{1}{2},R} = \tilde{p}(A^n_{j+\frac{1}{2},R}, K)$, we can use either the values $K = K(x_{j+\frac{1}{2}})$ or $K = K^*_{j+\frac{1}{2}}$ as in (2.3.23) with the only requirement to define the source term (2.3.26) accordingly.

Theorem 2. *The Lagrangian step (2.3.18), (2.3.25)-(2.3.26) is well-balanced under the "man at eternal rest" condition (2.2.7).*

Proof. Let us assume that $\mathbf{Q}^n_j$ satisfies the "man at eternal rest" condition (2.2.7). Thus, $u^n_j = 0\ \forall j$, $A^n_{j+\frac{1}{2},L} = A^n_{j+\frac{1}{2},R}$, and consequently, $\mathbf{Q}^n_{j+\frac{1}{2},L} = \mathbf{Q}^n_{j+\frac{1}{2},R}$. Hence, $u^*_{j+\frac{1}{2}} = 0\ \forall j$ and $-\frac{1}{\Delta x}(\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}) + \{s\}^n_j = 0$. Therefore, we obtain $A^{n+1-}_j = A^n_j$ and $q^{n+1-}_j = q^n_j$.

□

Remark 1. *As it has been shown in the proofs of theorems 1 and 2, $u^*_{j+\frac{1}{2}} = 0$ under the "man at eternal rest" condition. Hence, the projection step (2.3.13) preserves the stationary solution (2.2.7) and the whole numerical scheme (Lagrangian plus projection step) does not have to be further modified.*

2.4 Second-order well-balanced scheme

At this stage, we are interested in second (or higher) order extension of the Lagrange-projection schemes. We proceed as for the first-order scheme: we first explain how to reach the second-order of accuracy in the case of constant parameters, which is already non-trivial due to the presence of two steps in the Lagrange-Projection procedure. Then, we extend the

strategy to the case of variable parameters, A_0 and K. Once again, in the latter case we pay attention to the well-balanced property.

Here we focus on a second-order simplified version of the scheme applied to (2.2.3) which can also be easily extended to higher order of accuracy following [25, 6]. In particular, we make use of polynomial reconstruction and Runge-Kutta TVD scheme [21] in order to reach second-order of accuracy in space and time respectively.

2.4.1 Second-order approximation with constant parameters

First of all, we explain how to reach the second-order of accuracy in space in both the Lagrangian and projection steps. Then, we comment on the Runge-Kutta TVD scheme for the second-order approximation in time. The time is hereafter assumed to be left continuous for the sake of clarity.

Space discretization of the Lagrangian step. Given a time t, a j-th cell and the cell value $\mathbf{Q}_j(t)$, this step aims at defining evolved values $\mathbf{Q}_{j+\frac{1}{2}L,R}(t)$ at the cell interface $x_{j+\frac{1}{2}}$ by means of polynomial data reconstructions. More precisely, using a reconstructed polynomial vector $\mathbf{P}_j^t(x)$ for each cell I_j, such as

$$\mathbf{P}_j^t(x) = \mathbf{Q}_j(t) + \Delta_j^t(x - x_j), \tag{2.4.1}$$

where $\Delta_j^t = \Delta_j^t(\mathbf{Q}_{j-1}(t), \mathbf{Q}_j(t), \mathbf{Q}_{j+1}(t))$ is the slope, either ENO [30] or MINMOD [33] one, we define

$$\mathbf{Q}_{j+\frac{1}{2}L}(t) = \mathbf{P}_j^t(x_{j+\frac{1}{2}}) \quad \text{and} \quad \mathbf{Q}_{j+\frac{1}{2}R}(t) = \mathbf{P}_{j+1}^t(x_{j+\frac{1}{2}}). \tag{2.4.2}$$

Therefore, once again we use formulae (2.3.5), computed in $(\mathbf{Q}_{j+\frac{1}{2}L}(t), \mathbf{Q}_{j+\frac{1}{2}R}(t))$, namely

$$u^*_{j+\frac{1}{2}}(t) = u^*_{j+\frac{1}{2}}(\mathbf{Q}_{j+\frac{1}{2}L}(t), \mathbf{Q}_{j+\frac{1}{2}R}(t)) \quad \text{and} \quad \Pi^*_{j+\frac{1}{2}}(t) = \Pi^*_{j+\frac{1}{2}}(\mathbf{Q}_{j+\frac{1}{2}L}(t), \mathbf{Q}_{j+\frac{1}{2}R}(t)),$$

with

$$a_{j+\frac{1}{2}} = \max\left(A^n_{j+\frac{1}{2},L}c^n_{j+\frac{1}{2},L}, A^n_{j+\frac{1}{2},R}c^n_{j+\frac{1}{2},R}\right).$$

Note that the polynomial $\mathbf{P}_j^t(x)$ should satisfy the conservation property, which reads

$$\frac{1}{\Delta x}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{P}_j^t(x) = \mathbf{Q}_j(t).$$

Space discretization of the remap step. In order to obtain the second-order of accuracy in space, we exploit relations (2.3.11)-(2.3.13) seen in section 2.3.1. Once again the middle integral in (2.3.11) can be substituted with $(L\mathbf{Q})_j(t)$ thanks to the conservation property. Then, for the other two integrals, instead of considering the values $(L\mathbf{Q})_j(t)$, we reconstruct them using the polynomial $\mathbf{P}_j^t(x)$. Thus, we introduce

$$(L\mathbf{P})_j^t(\xi) = (L\mathbf{Q})_j(t) + \Delta_j^t(\xi - \xi_j), \tag{2.4.3}$$

with $\Delta_j^t = \Delta_j^t((L\mathbf{Q})_{j-1}(t), (L\mathbf{Q})_j(t), (L\mathbf{Q})_{j+1}(t))$ and

$$(L\mathbf{P})^t_{j-\frac{1}{2}}(\xi) = \begin{cases} (L\mathbf{P})^t_{j-1}(\xi) & \text{if} \quad \xi_{j-\frac{1}{2}} > \hat{\xi}_{j-\frac{1}{2}} \\ (L\mathbf{P})^t_{j}(\xi) & \text{if} \quad \xi_{j-\frac{1}{2}} \leq \hat{\xi}_{j-\frac{1}{2}} \end{cases}$$

in place of $(L\mathbf{Q})_{j-\frac{1}{2}}(t)$, where we remark that $x_{j+\frac{1}{2}} = \hat{\xi}_{j+\frac{1}{2}} + \Delta t u^*_{j+\frac{1}{2}}$. Finally, since $(L\mathbf{P})^t(\xi)$ is not constant, in order to be able to evaluate its integral $\frac{1}{\Delta x}\int_{\hat{\xi}_{j-\frac{1}{2}}}^{\xi_{j-\frac{1}{2}}} (L\mathbf{P})^t_{j-\frac{1}{2}}(\xi)dx$, one can

exploit either the mid-point rule (only for second-order of accuracy) or a Gauss quadrature formula with nodes $\xi_{j-\frac{1}{2},k}$ and weights ω_k for $k = 1, \dots, m$. In the former case we find

$$\begin{aligned}\mathbf{Q}_j(t) &= (L\mathbf{Q})_j(t) + \frac{\xi_{j-\frac{1}{2}} - \hat{\xi}_{j-\frac{1}{2}}}{\Delta x}(L\mathbf{P})^t_{j-\frac{1}{2}}\Big(\frac{\xi_{j-\frac{1}{2}} + \hat{\xi}_{j-\frac{1}{2}}}{2}\Big) + \frac{\hat{\xi}_{j+\frac{1}{2}} - \xi_{j+\frac{1}{2}}}{\Delta x}(L\mathbf{P})^t_{j+\frac{1}{2}}\Big(\frac{\xi_{j+\frac{1}{2}} + \hat{\xi}_{j+\frac{1}{2}}}{2}\Big) \\ &= (L\mathbf{Q})_j(t) - \frac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}(L\mathbf{P})^t_{j+\frac{1}{2}}\Big(\frac{\xi_{j+\frac{1}{2}} + \hat{\xi}_{j+\frac{1}{2}}}{2}\Big) - u^*_{j-\frac{1}{2}}(L\mathbf{P})^t_{j-\frac{1}{2}}\Big(\frac{\xi_{j-\frac{1}{2}} + \hat{\xi}_{j-\frac{1}{2}}}{2}\Big)\big),\end{aligned} \tag{2.4.4}$$

while in the latter we find

$$\begin{aligned}\mathbf{Q}_j(t) &= (L\mathbf{Q})_j(t) + \frac{\xi_{j-\frac{1}{2}} - \hat{\xi}_{j-\frac{1}{2}}}{\Delta x}\sum_{k=1}^m \omega_k (L\mathbf{P})^t_{j-\frac{1}{2}}(\xi_{j-\frac{1}{2},k}) + \frac{\hat{\xi}_{j+\frac{1}{2}} - \xi_{j+\frac{1}{2}}}{\Delta x}\sum_{k=1}^m \omega_k (L\mathbf{P})^t_{j+\frac{1}{2}}(\xi_{j+\frac{1}{2},k}) \\ &= (L\mathbf{Q})_j(t) - \frac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}\sum_{k=1}^m \omega_k (L\mathbf{P})^t_{j+\frac{1}{2}}(\xi_{j+\frac{1}{2},k}) - u^*_{j-\frac{1}{2}}\sum_{k=1}^m \omega_k (L\mathbf{P})^t_{j-\frac{1}{2}}(\xi_{j-\frac{1}{2},k})\big),\end{aligned}$$

where the other integral $\frac{1}{\Delta x}\int_{\hat{\xi}_{j+\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} (L\mathbf{P})^t_{j+\frac{1}{2}}(\xi)dx$ has been evaluated in a similar way.

Second-order approximation in time. With the last step of the second-order method, we aim to obtain the right accuracy in time. In order to do this, we simply use the Runge-Kutta TVD scheme at second order [21]. However, we have to specify that it has to be applied to the overall scheme (Lagrangian and remap steps together) in order to avoid diffusion due to the splitting.

2.4.2 Second-order approximation with variable parameters

As we have done before for the well-balanced first-order scheme, here we present two different well-balanced second-order methods, the first exploiting the approximate Riemann solver of section 2.2.2 and the second one exploiting the hydrostatic reconstruction approach.

Again, it is sufficient to focus on the Lagrangian step and no changes are needed for the second-order projection step, as described above, and the Runge-Kutta TVD procedure. Indeed, it will easily be seen that they preserve the "man at eternal rest" solution. However, we specify that the projection step would need to be modified if we were interested in preserving a different stationary solution. For more details refer to [11].

2.4.2.1 Based on the approximate Riemann solver

Here we describe the second-order extension of the Lagrangian step explained in section 2.3.2.1, which makes use of the approximate Riemann solver of section 2.2.2.2 in order to maintain the well-balancedness of the method.

Thus, in order to obtain a second-order approximation in space, the idea is to exploit the reconstructed values at cell interfaces and then apply the usual updating formulae. However, we cannot simply use the reconstructed polynomial (2.4.1), as it would prevent the scheme from preserving the stationary solutions. Thus, the idea is to compute the slopes in such a way that they become equal to zero when the "man at eternal rest" condition (2.2.7) is satisfied. With this in mind, we suggest to make use of the so-called fluctuations $\mathbf{D}$, refer to [25, 6]. Let us get into the details. Given a time t and the j-th cell, we need the values $\mathbf{Q}_{j-1}(t), \mathbf{Q}_j(t), \mathbf{Q}_{j+1}(t)$ to determine the slopes at second-order of accuracy; thus for $k = j-1, j, j+1$, we define the so-called fluctuations

$$\mathbf{D}_{k,j}(t) = \mathbf{Q}_k(t) - \frac{1}{\Delta x}\int_{x_{k-\frac{1}{2}}}^{x_{k+\frac{1}{2}}} \mathbf{Q}^{t,e}_j(x)dx, \tag{2.4.5}$$

where $\mathbf{Q}_j^{t,e}(x)$ denotes a reconstructed stationary solution we want to preserve and which satisfies

$$\frac{1}{\Delta x}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{Q}_j^{t,e}(x)dx = \mathbf{Q}_j(t). \tag{2.4.6}$$

Note that, since we are interested in a second-order accurate scheme, we solve the integral in (2.4.5) using the mid-point rule in space. Usually, it is not straightforward to compute $\mathbf{Q}_j^{t,e}(x)$ with the constraint (2.4.6), however, since we only want to preserve the "man at eternal rest" solution, we can automatically define it such that

$$\big(K\sqrt{A_j^{t,e}}\big)(x) = K_j(\sqrt{A_j} - \sqrt{A_{0,j}}) + (K\sqrt{A_0})(x) \quad \text{and} \quad u_j^{t,e}(x) = u_j^t. \tag{2.4.7}$$

Consequently, we denote the reconstruction operator as

$$\mathbf{P}_j^t(x) = \mathbf{P}_j^t(x; \mathbf{Q}_j(t), \mathbf{D}_{j-1,j}(t), \mathbf{D}_{j,j}(t), \mathbf{D}_{j+1,j}(t)) = \mathbf{Q}_j(t) + \Delta_j^t(x - x_j) \tag{2.4.8}$$

where, for the sake of clarity, we specify that $\Delta_j^t = \Delta_j^t(\mathbf{D}_{j-1,j}(t), \mathbf{D}_{j,j}(t), \mathbf{D}_{j+1,j}(t))$, for the definition of which we will use either the ENO or the MINMOD slopes. Let us observe that in our specific case we always have $\mathbf{D}_{j,j}(t) = 0$, while $\mathbf{D}_{j\pm1,j}(t) = 0$ when $\mathbf{Q}_j(t)$ satisfies a stationary solution. Therefore it is clear that the slopes equal zero when the "man at eternal rest" condition is satisfied.

Equipped with the definition for the slopes and the same definition (2.4.2) for the left and right traces, we can now compute $u^*_{j+\frac{1}{2}}$ and $\Pi^*_{j+\frac{1}{2}}$ as in (2.3.19), namely

$$u^*_{j+\frac{1}{2}}(t) = u^*(\mathbf{Q}_{j+\frac{1}{2},L}(t), \mathbf{Q}_{j+\frac{1}{2},R}(t)) \quad \text{and} \quad \Pi^*_{j+\frac{1}{2}}(t) = \Pi^*(\mathbf{Q}_{j+\frac{1}{2},L}(t), \mathbf{Q}_{j+\frac{1}{2},R}(t)) \tag{2.4.9}$$

with

$$\mathbf{Q}_{j+\frac{1}{2}L}(t) = \mathbf{P}_j^t(x_{j+\frac{1}{2}}) \quad \text{and} \quad \mathbf{Q}_{j+\frac{1}{2}R}(t) = \mathbf{P}_{j+1}^t(x_{j+\frac{1}{2}}).$$

More specifically, we state

$$\begin{cases} u^*_{j+\frac{1}{2}}(t) = \dfrac{1}{2}(u^t_{j+\frac{1}{2}R} + u^t_{j+\frac{1}{2}L}) - \dfrac{1}{2a^t_{j+\frac{1}{2}}}(\Pi^t_{j+\frac{1}{2}R} - \Pi^t_{j+\frac{1}{2}L}) + \dfrac{\Delta m_{j+1/2}}{2a_{j+1/2}} s^t_{j+1/2} \\ \Pi^*_{j+\frac{1}{2}}(t) = \dfrac{1}{2}(\Pi^t_{j+\frac{1}{2}R} + \Pi^t_{j+\frac{1}{2}L}) - \dfrac{a^t_{j+\frac{1}{2}}}{2}(u^t_{j+\frac{1}{2}R} - u^t_{j+\frac{1}{2}L}). \end{cases}$$

Lastly, one needs to specify the value of K used in the definition $\Pi^t_{j+\frac{1}{2}R,L}$. It turns out that using the natural value of K in $x_{j+\frac{1}{2}}$ leads to the loss of the well-balanced property. Therefore we propose to reconstruct K exploiting the usual reconstruction polynomial and corresponding slope Δ_j^t. However, since the equilibrium part of K reads $K_j^{t,e}(x) = K(x)$ as K is known and does not depend on time, its fluctuations are null and, thus, we can simply state that

$$K_{j+\frac{1}{2}L} = K_j \quad \text{and} \quad K_{j+\frac{1}{2}R} = K_{j+1},$$

which is very convenient as with this choice of K we do not have to do further modifications to preserve the well-balancedness of the scheme.

Finally, for the source term at second-order of accuracy, we simply consider (2.3.20)-(2.3.21) with $\mathcal{M}^t_{j+1/2} = \mathcal{M}\big((A_{j+\frac{1}{2},L}(t); A_{0,j}, K_j); (A_{j+\frac{1}{2},R}(t); A_{0,j+1}, K_{j+1})\big)$ defined as in (2.2.33), and the updating formulae (2.3.18).

Theorem 3. *The Lagrangian step (2.4.5)-(2.4.9) with updating formulae (2.3.18) described here is well-balanced under the "man at eternal rest" condition (2.2.7).*

Proof. Once again, let us assume that $\mathbf{Q}_j^n$ satisfies the "man at eternal rest" condition (2.2.7) and, as such, $u_j^n = 0 \, \forall j$. Due to definitions (2.4.5)-(2.4.7) for the fluctuations, we have that the slope satisfies $\Delta_j^t = 0$ and thus, as for the first-order scheme, we have $u^*_{j+\frac{1}{2}} = 0$ and $A_j^{n+1-} = A_j^n$. As for the proof of the corresponding first-order scheme, simple algebraic computations show that $q_j^{n+1-} = q_j^n$.

□

Lastly, a summary of this second order scheme is provided below for the reader's sake.

Step	Second order scheme based on the approximate Riemann solver.
1	Lagrangian step: look for the reconstructed stationary solution $\mathbf{Q}_j^{t,e}(x)$ that satisfies (2.4.6).
2	Exploit it to compute the fluctuations $\mathbf{D}_{k,j}(t)$ as in (2.4.5).
3	Define the reconstruction operator $\mathbf{P}_j^t(x)$ as in (2.4.8).
4	Find the reconstructed values $\mathbf{Q}_{j+\frac{1}{2}L}(t) = \mathbf{P}_j^t(x_{j+\frac{1}{2}})$, $\mathbf{Q}_{j+\frac{1}{2}R}(t) = \mathbf{P}_{j+1}^t(x_{j+\frac{1}{2}})$.
5	Compute $u^*_{j+\frac{1}{2}}$, $\Pi^*_{j+\frac{1}{2}}$ as in (2.3.19)-(2.4.9) and the source term as in (2.3.20)-(2.3.21).
6	Solve system (2.3.18) written in Lagrangian coordinates obtaining $(L\mathbf{Q})(t)$.
7	Continue with the remap step; define the polynomial $\mathbf{P}_j^t(x)$ as in (2.4.3).
8	Update $\mathbf{Q}(t)$ using formula (2.4.4).
9	Apply the Runge-Kutta scheme in order to reach the second order of accuracy in time.

2.4.2.2 Based on the hydrostatic reconstruction

Let us now see how to modify the second-order accurate scheme exploiting the hydrostatic reconstruction, already introduced in section 2.3.2.2. Once again, in order to have a well-balanced scheme, we only have to modify the Lagrangian step.

Hence, in this case, we would like to combine two different kinds of reconstruction, one for the well-balancedness and one for the high-order of accuracy. Thus, at the end we will have a unique reconstruction function for $\mathbf{Q}_j^t(x)$ which will consist of two parts: the fluctuation $\mathbf{P}_j^t(x)$ and the equilibrium $\mathbf{Q}_j^{t,e}(x)$ ones,

$$\mathbf{Q}_j^t(x) = \mathbf{Q}_j^{t,e}(x) + \mathbf{P}_j^t(x), \tag{2.4.10}$$

as suggested in [25]. Subsequently, we will use the values $\mathbf{Q}^t_{j+\frac{1}{2}L} = \mathbf{Q}_j^t(x_{j+\frac{1}{2}})$ and $\mathbf{Q}^t_{j+\frac{1}{2}R} = \mathbf{Q}^t_{j+1}(x_{j+\frac{1}{2}})$ to compute $u^*_{j+\frac{1}{2}}$ and $\Pi^*_{j+\frac{1}{2}}$ in (2.3.5) and to find $(L\mathbf{Q})_j(t)$ according to (2.3.18).

At this stage, we have to define $\mathbf{P}_j^t(x)$ and $\mathbf{Q}_j^{t,e}(x)$. For the latter, since we are only interested in preserving the "man at eternal rest solution", we simply use the well-balanced reconstructed values (2.3.24), in particular

$$\mathbf{Q}_j^{t,e}(x_{j+\frac{1}{2}}) = \mathbf{Q}^{t,e}_{j+\frac{1}{2},L} = \begin{pmatrix} A^{t,e}_{j+\frac{1}{2},L} \\ A^{t,e}_{j+\frac{1}{2},L} u_i^t \end{pmatrix} \quad \text{and} \quad \mathbf{Q}^{t,e}_{j+1}(x_{j+\frac{1}{2}}) = \mathbf{Q}^{t,e}_{j+\frac{1}{2},R} = \begin{pmatrix} A^{t,e}_{j+\frac{1}{2},R} \\ A^{t,e}_{j+\frac{1}{2},R} u_{i+1}^t \end{pmatrix},$$

with $A^{t,e}_{j+\frac{1}{2},L}$, $A^{t,e}_{j+\frac{1}{2},R}$ computed as in (2.3.22)-(2.3.23). See either section 2.3.2.2 or [20] for more details. As far as the reconstruction polynomial $\mathbf{P}_j^t(x)$ is concerned, we use a similar but not equal strategy to the one explained in the previous section. Indeed, here we write $\mathbf{P}_j^t(x)$ only depending on the fluctuations $\mathbf{D}_{j,j}(t)$, namely

$$\mathbf{P}_j^t(x) = \mathbf{P}_j^t(x; \mathbf{D}_{j-1,j}(t), \mathbf{D}_{j,j}(t), \mathbf{D}_{j+1,j}(t)) = \mathbf{D}_{j,j}(t) + \Delta_j^t(x - x_j), \tag{2.4.11}$$

with $\Delta_j^t = \Delta_j^t(\mathbf{D}_{j-1,j}(t), \mathbf{D}_{j,j}(t), \mathbf{D}_{j+1,j}(t))$ and $\mathbf{D}_{k,j}(t)$, with $k = j-1, j, j+1$, defined as in (2.4.5). Let us remark that we always have $\mathbf{D}_{j,j}(t) = 0$ and thus, $\mathbf{Q}_j^t(x)$ reads

$$\mathbf{Q}_j^t(x) = \mathbf{Q}_j^{t,e}(x) + \Delta_j^t(x - x_j) = \mathbf{Q}_j^{t,e}(x) + \mathbf{Q}_j^{t,f}(x), \tag{2.4.12}$$

where we have renamed $\mathbf{Q}_j^{t,f}(x) = \Delta_j^t(x - x_j)$ for the fluctuations part. Then, let us compare (2.4.8) and (2.4.12). Indeed the term $\mathbf{Q}_j(t)$ that appears in the right hand side of (2.4.8) can be understood as $\mathbf{Q}_j^{t,e}(x)$ in (2.4.12) but replaced by $\mathbf{Q}_j(t)$ using (2.4.6) and the mid-point rule.

Finally, we only need to specify the definition of the source term, which in general is defined by

$$s_j^t = \frac{1}{\Delta x} \int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} s(\xi, t) d\xi$$

with $s(A; A_0, K)$ given by (2.2.4). Since we aim to reach the second order of accuracy, in this case the mid-point rule in space suffices. Thus, using the equilibrium and fluctuation decomposition (2.4.12) for the cross-sectional area A, we can write

$$\begin{aligned}
s_j^t &= \frac{1}{\Delta x} \int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} s(A; A_0, K)(\xi, t) d\xi = \frac{1}{\Delta x} \int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} s(A^e + A^f; A_0, K)(\xi, t) d\xi \\
&= \frac{1}{\Delta x} \int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} (s(A^e + A^f; A_0, K)(\xi, t) - s(A^e; A_0, K)(\xi, t)) d\xi + \frac{1}{\Delta x} \int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} s(A^e; A_0, K)(\xi, t) d\xi \\
&= \frac{1}{\Delta x} \int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} s(A^e; A_0, K)(\xi, t) d\xi,
\end{aligned}$$

where the last quality holds as, when applying the mid-point rule, the fluctuations part A^f disappears, leaving only $\frac{\xi_{j+\frac{1}{2}} - \xi_{j-\frac{1}{2}}}{\Delta x}(s(A^e + A^f; A_0, K)(x_j, t) - s(A^e; A_0, K)(x_j, t)) = \frac{\xi_{j+\frac{1}{2}} - \xi_{j-\frac{1}{2}}}{\Delta x}(s(A^e + 0; A_0, K)(x_j, t) - s(A^e; A_0, K)(x_j, t)) = 0$. Hence, similarly to the first-order scheme, the source term finally reads

$$s_j^t = \frac{1}{\Delta x} \int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} s(A^e; A_0, K)(\xi, t) d\xi = \frac{\tilde{p}^e_{j+\frac{1}{2},L}(t) - \tilde{p}^e_{j-\frac{1}{2},R}(t)}{\Delta x}. \tag{2.4.13}$$

Theorem 4. *The Lagrangian step (2.4.10)-(2.4.13) is well-balanced under the "man at eternal rest" condition (2.2.7).*

Proof. It is straightforward to see that, under the "man at eternal rest" condition (2.2.7), the fluctuations part in (2.4.12) satisfies $\mathbf{Q}_j^{t,f}(x) = 0$. Consequently, the Lagrangian step is reduced to the first-order one, which we already proved to be well-balanced.

□

As in the previous section, an algorithm is provided to to summarize this second order method. Let us note that the projection and Runge-Kutta steps are the same of the previous scheme.

2.5 Numerical simulations

In this section, we carry out different numerical simulations that aim to show the good behaviour of the proposed numerical schemes. First of all, we numerically prove that, in the case

Step	Second order scheme based on the hydrostatic reconstruction.
1	Start with the Lagrangian step; compute the well-balanced reconstructed values $\mathbf{Q}_j^{t,e}(x)$ as in (2.3.24).
2	Find the fluctuations $\mathbf{D}_{k,j}(t)$ by (2.4.5).
3	Define the reconstruction polynomial $\mathbf{P}_j^t(x)$ only depending on the fluctuations as in (2.4.11).
4	Use $\mathbf{Q}_j^{t,e}(x)$ and $\mathbf{P}_j^t(x)$ to determine the reconstruction function $\mathbf{Q}_j^t(x)$ by (2.4.10).
5	Find the reconstructed values $\mathbf{Q}_{j+\frac{1}{2}L}^t = \mathbf{Q}_j^t(x_{j+\frac{1}{2}})$ and $\mathbf{Q}_{j+\frac{1}{2}R}^t = \mathbf{Q}_{j+1}^t(x_{j+\frac{1}{2}})$.
6	Compute $u_{j+\frac{1}{2}}^*$ and $\Pi_{j+\frac{1}{2}}^*$ as in (2.3.5) exploiting the reconstructed values.
7	Define the source term s_j^t according to (2.4.13).
8	Solve system (2.3.18) written in Lagrangian coordinates obtaining $(L\mathbf{Q})(t)$.
9	Continue with the remap step; define the polynomial $\mathbf{P}_j^t(x)$ as in (2.4.3).
10	Update $\mathbf{Q}(t)$ using formula (2.4.4).
11	Apply the Runge-Kutta scheme in order to reach the second order of accuracy in time.

of both systems of conservation and balance laws, the numerical schemes reach the required order of convergence. A Riemann problem is also presented in the case of constant parameters. Then, different test cases are introduced in order to assess the well-balancedness and the wave-capturing properties of the numerical methods when applied to the non-conservative system.

Time step and CFL condition. Since the Lagrange-projection approach leads to a splitting of the original system (2.2.3) into the acoustic (2.2.13) and the transport (2.2.14) ones, the time step is computed as the minimum between the two time steps obtained from (2.2.13) and (2.2.14). As far as the Lagrangian system is concerned, the Courant-Friedrichs-Lewy (CFL) condition reads

$$\Delta t \leq \mathrm{CFL}_l \frac{\Delta x}{\max\limits_j \{\max(\tau_j^n, \tau_{j+1}^n) a_{j+\frac{1}{2}}\}}, \tag{2.5.1}$$

while, for the transport system

$$\Delta t \leq \mathrm{CFL}_t \frac{\Delta x}{\max\limits_j \{u_{j-\frac{1}{2}}^+ - u_{j+\frac{1}{2}}^-\}}, \tag{2.5.2}$$

where CFL_l and CFL_t are, respectively, the CFL number for the Lagrangian and the transport systems, and finally

$$u_{j-\frac{1}{2}}^+ = \max(u_{j-\frac{1}{2}}^*, 0) \quad \text{and} \quad u_{j+\frac{1}{2}}^- = \min(u_{j+\frac{1}{2}}^*, 0).$$

When considering a first-order scheme, we can take $\mathrm{CFL}_l \leq 0.5$ and $\mathrm{CFL}_t < 1$. For more details refer to [9, 11, 12].

Remark 2. *It is not difficult to prove that the first-order approximation (2.3.14) we presented preserves the strict positivity of the cross-sectional area A under the CFL condition (2.5.2) with $\mathrm{CFL}_t < \frac{1}{2}$, see also [5]. This statement remains true even if the parameters K and A_0 are not constant in space.*

2.5.1 System of conservation laws

Convergence study. In this section we start by assessing that the second-order scheme of section 2.4.1 reaches the right order of accuracy. For this purpose, we need to compare the numerical solution with the exact one, specifying that, in order to obtain the correct order of accuracy, the smoothness of the exact solution is required. Since in general such solution is

not known for system (2.2.5), we have to exploit the method of the manufactured solution, for which we refer to [28]. In a nutshell, given an acceptable smooth function $\hat{\mathbf{Q}}$, we have to modify the sought system in such a way that $\hat{\mathbf{Q}}$ is actually one of its solutions. This is achieved by adding a source term to the starting system, namely passing from the homogeneous version of system (2.2.5), $\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) = 0$, to

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) = \hat{\mathbf{S}}(\mathbf{Q}).$$

$\hat{\mathbf{S}}(\mathbf{Q})$ is usually found through an algebraic manipulator, and thus we will not report the modified source term here. Referring to [27], we have considered the following solution,

$$\hat{\mathbf{Q}} = \begin{pmatrix} \hat{A} \\ \hat{q} \end{pmatrix} = \begin{pmatrix} \tilde{A} + \tilde{a}\sin(\frac{2\pi}{L}x)\cos(\frac{2\pi}{T_0}t) \\ \tilde{q} - \tilde{a}\frac{L}{T_0}\cos(\frac{2\pi}{L}x)\sin(\frac{2\pi}{T_0}t) \end{pmatrix} \tag{2.5.3}$$

with the cross-sectional area at equilibrium and the wall stiffness given by $\hat{a_0} = \tilde{A}$ and $\hat{K} = \tilde{K}$ respectively, where $\tilde{A} = 4.0 \times 10^{-4}\text{m}^2$, $\tilde{a} = 4.0 \times 10^{-5}\text{m}^2$, $\tilde{q} = 0\text{m}^3\text{s}^{-1}$, $\tilde{K} = 2500\text{kPa}$, $T_0 = 1.0\text{s}$, the length of the vessel being $L = 1.0\text{m}$ and $\rho = 1050.0\text{kg m}^{-3}$. We also take CFL $= 0.25$ and exploit the MINMOD slope. With (2.5.3) being a periodic solution, the boundary conditions we use are periodic. As initial condition, we take $\hat{\mathbf{Q}}$ at initial time $t = 0$.

At this stage, let us give the definition of the p-norm of the global error E^n,

$$||E(t^n, \Delta x)||_p = \big(\Delta x \sum_{i=-\infty}^{\infty} |v_i^n - v(x_i, t^n)|^p\big)^{\frac{1}{p}},$$

where v_i^n is the numerical solution and $v(x_i, t^n)$ is the exact solution computed in (x_i, t^n). Note that we will use $p = 1$, $p = 2$ and $p = +\infty$. In our case we take either $v = A$ or $v = Au$. Given an increasing sequence of mesh M_k, with $k = 1, \dots, N$, and respective dimension Δx_k, we can now define the empirical order of accuracy p_{k+1} as:

$$p_{k+1} = \frac{\ln\big(\frac{E_{k+1}(T_{out}, \Delta x_{k+1})}{E_k(T_{out}, \Delta x_k)}\big)}{\ln\big(\frac{\Delta x_{k+1}}{\Delta x_k}\big)},$$

with ending time $t = T_{\text{out}}$. p_k should tend to the theoretical order of accuracy p, for sufficiently large k.

Lastly, in table 2.1, we show the errors and order of convergence of the variables A and $q = Au$ at ending time $T_{\text{out}} = 0.08\text{s}$. Indeed, we observe that the non-well-balanced numerical scheme described in section 2.4.1 reaches the second order of accuracy.

Riemann problem: the ideal tourniquet. Since Riemann problems are simple and idealized test cases but still useful for a better understanding of the numerical schemes, here we present the following problem, the ideal tourniquet, for which we refer to [15, 36]. In it, a tourniquet is placed and then immediately removed. As such, we consider the following initial data

$$\mathbf{Q}(x, t = 0) = \begin{cases} \mathbf{Q}_L & \text{if} \quad x < L/2 \\ \mathbf{Q}_R & \text{if} \quad x > L/2 \end{cases}$$

with initial velocity $u_L = u_R = 0\text{m/s}$, initial radius $R_L = 5 \times 10^{-3}\text{m}$, $R_R = 4 \times 10^{-3}\text{m}$ and initial area computed as $A = \pi R^2$. As for the other parameters we take $K = \frac{1}{\sqrt{\pi}} \times 10^7\text{Pa/m}$, the length of the vessel being $L = 0.08\text{m}$ and $\rho = 1060.0\text{kg m}^{-3}$. For the first and second-order schemes we use $\text{CFL}_l = 0.45$ and $\text{CFL}_l = 0.25$ respectively. Finally, the ending time is given by $T_{\text{out}} = 0.005\text{s}$ and once again we exploit the MINMODE slope. In figure 2.2, we

Variable	Mesh M	err $\mathbf{L}^1$	err $\mathbf{L}^2$	err $\mathbf{L}^\infty$	$O(\mathbf{L}^1)$	$O(\mathbf{L}^2)$	$O(\mathbf{L}^\infty)$
Area A	16	0.1435×10^{-5}	0.1592×10^{-5}	0.2075×10^{-5}	–	–	–
	32	0.0362×10^{-5}	0.0428×10^{-5}	0.0703×10^{-5}	1.9873	1.8944	1.5626
	64	0.0087×10^{-5}	0.0114×10^{-5}	0.0206×10^{-5}	2.0626	1.9078	1.7697
	128	0.0022×10^{-5}	0.0030×10^{-5}	0.0053×10^{-5}	1.9732	1.9275	1.9572
	256	0.0006×10^{-5}	0.0008×10^{-5}	0.0013×10^{-5}	1.9598	1.9724	2.0123
Flow q	16	0.6610×10^{-5}	0.6622×10^{-5}	0.7288×10^{-5}	–	–	–
	32	0.1487×10^{-5}	0.1514×10^{-5}	0.1832×10^{-5}	2.1519	2.1292	1.9924
	64	0.0313×10^{-5}	0.0320×10^{-5}	0.0394×10^{-5}	2.2501	2.2412	2.2152
	128	0.0071×10^{-5}	0.0075×10^{-5}	0.0105×10^{-5}	2.1372	2.1012	1.9032
	256	0.0017×10^{-5}	0.0018×10^{-5}	0.0028×10^{-5}	2.1058	2.0761	1.9092

Table 2.1: Errors and empirical convergence rates for norms $\mathbf{L}^1$, $\mathbf{L}^2$ and $\mathbf{L}^\infty$. Mesh of size $M = (16, 32, 64, 128, 256)$. Second-order scheme of section 2.4.1.

compare the result for the first and second-order schemes against the exact solution. We used $M = 100$ and $M = 500$ cells on the left and the right respectively, where $\Delta x = \frac{L}{M}$. Both schemes approximate the exact solution well; obviously the second-order scheme results to be less diffusive than the first-order one. On the right, we can see that the numerical solution converges to the exact one.

2.5.2 System of balance laws

For the numerical simulations in this section we refer to [20, 36]. We distinguish the well-balanced schemes based on the approximate Riemann solver and the hydrostatic reconstruction as WB-ARS and WB-HR respectively. If second-order accurate, we call them WB-ARS2 and WB-HR2. Given below are numerical details we will use for the subsequent test problems, unless it is specified otherwise. In general, as initial condition we take

$$\begin{cases} A(x, t = 0) = A_0(x) \\ q(x, t = 0) = 0. \end{cases}$$

As far as the cross-sectional radius at equilibrium R_0 and wall rigidity K are concerned, we use the following relations

$$\begin{aligned} R_0(x) &= \begin{cases} R_{in} & \text{if} \quad x < x_s \quad \text{or} \quad x > x_f \\ R_{in}(1 - \frac{\Delta \mathbf{G}}{2}(1 + cos(\pi + 2\pi \frac{x - x_s}{x_f - x_s}))) & \text{if} \quad x_s \leq x \leq x_f, \end{cases} \\ K(x) &= \begin{cases} K_{in} & \text{if} \quad x < x_s \quad \text{or} \quad x > x_f \\ K_{in}(1 + \frac{\Delta \mathbf{G}}{2}(1 + cos(\pi + 2\pi \frac{x - x_s}{x_f - x_s}))) & \text{if} \quad x_s \leq x \leq x_f, \end{cases} \end{aligned} \tag{2.5.4}$$

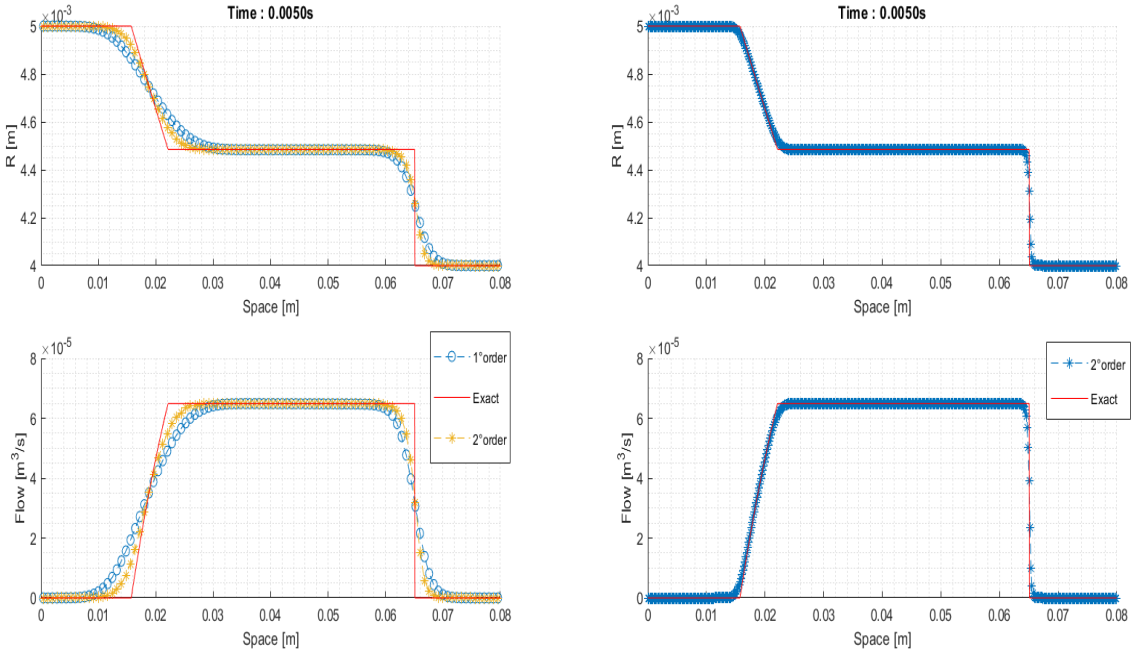
where $A_0 = \pi R_0^2$, $x_s = \frac{3L}{10}$, $x_f = \frac{7L}{10}$ and $\Delta \mathbf{G} \in \{1\%, 10\%, 30\%, 60\%\}$. The other parameters' values can be found in table 2.2. Regarding the boundary conditions, we impose the following flow at the inlet of the domain,

$$q_{in} = Shap A_{in} c_{in}$$

where a value for A_{in} consistent with the Shapiro number ($Shap$) has been estimated in [20] to be

$$A_{in} = A_0(x = 0)(1 + Shap)^2.$$

The Shapiro number is the analogous to the Mach number for the compressible Euler equations and it is defined as $Shap = \dfrac{u}{c}$. In particular, we take $Shap = 0, 10^{-2}, 10^{-3}$. Let us note that



(a) $M = 100$ cells

(b) $M = 500$ cells

Figure 2.2: Ideal tourniquet problem, radius (left) and flow (right). First-order (-o symbol), second-order (-* symbol) and exact (red line) solution.

Parameter	Value
L	10 [cm]
R_{in}	0.5 [cm]
K_{in}	1×10^5 [gcm^{-2}s^{-2}]
ρ	1 [gcm^{-3}]

Table 2.2: Parameters values for test cases in section 2.5.2.

Numerical method	Variable	Mesh M	err $\mathbf{L}^1$	err $\mathbf{L}^2$	err $\mathbf{L}^\infty$	$O(\mathbf{L}^1)$	$O(\mathbf{L}^2)$	$O(\mathbf{L}^\infty)$
WB-ARS2	Energy	16	45.9648	20.2167	15.2702	–	–	–
		32	12.1161	5.3498	4.5128	1.9236	1.9180	1.7586
		64	2.9137	1.3085	1.1747	2.0560	2.0315	1.9418
		128	0.7133	0.3244	0.2970	2.0302	2.0121	1.9835
	Flow	16	0.0978	0.0525	0.0403	–	–	–
		32	0.0280	0.0147	0.0112	1.8034	1.8334	1.8475
		64	0.0070	0.0037	0.0029	2.0016	1.9787	1.9542
		128	0.0018	0.0010	0.0007	1.9621	1.9439	1.9875
WB-HR2	Energy	16	44.4265	18.8607	13.4593	–	–	–
		32	11.4229	5.0448	4.3839	1.9595	1.9025	1.6183
		64	2.7538	1.2443	1.1764	2.05254	2.0194	1.8978
		128	0.7072	0.3185	0.3001	1.9613	1.9660	1.9711
	Flow	16	0.0973	0.0523	0.0377	–	–	–
		32	0.0281	0.0148	0.0111	1.7910	1.8237	1.7686
		64	0.0070	0.0037	0.0029	2.0057	1.9883	1.9403
		128	0.0018	0.0010	0.0007	1.9634	1.9448	1.9854

Table 2.3: Errors and empirical convergence rates for norms $\mathbf{L}^1$, $\mathbf{L}^2$ and $\mathbf{L}^\infty$ for the energy discharge E and flow q. Mesh of size $M = (16, 32, 64, 128)$. WB-ARS2 (top) and WB-HR2 (bottom) methods.

with $Shap = 10^{-2}$ we are already in the subsonic regime, thus it could be interesting to use an implicit scheme in this case. In arteries, the average value for the Shapiro number is indeed usually of order $Shap = 10^{-2}$. We refer again to [20] for more details. Then, we can find the boundary values for the A_{in} cross-sectional area exploiting the right Riemann invariant, i.e. $I^+ = u - 4c$, and imposing $I^+(\mathbf{Q}_1) = I^+(\mathbf{Q}_{in})$, where the subscript 1 indicates the values in the first cell of the computational domain. Whereas, regarding the right boundary condition, we impose

$$A_{out} = A_0(x = L)(1 + Shap)^2,$$

and then the flow value q_{out} is found exploiting the left Riemann invariant, namely $I^-(\mathbf{Q}_{end}) = I^-(\mathbf{Q}_{out})$, where with $\mathbf{Q}_{end}$ we mean the value in the last cell. For more details about this test cases we refer again to [20].

Stationary solution. First of all, to assess the well-balanced property, we take $Shap = 0$, and check that the numerical schemes preserve the stationary solution $A = A_0$, $q = 0$. Indeed, we observe that they maintain it up to an error of order 10^{-12}.

Convergence study. Then, in order to check that the well-balanced schemes also reach the right order of accuracy, we take $Shap = 10^{-2}$, $\Delta\mathbf{G} = 10\%$ and final time $tEnd = 100.0$s. Note that here the exact solution is a steady state with non-zero velocity, namely it is given by

$$\begin{cases} q_{ex} = q_{in} \\ E_{ex} = \frac{q_{in}^2}{2A_{out}^2} + \frac{K(end)}{\rho}(\sqrt{A_{out}} - \sqrt{A_0(end)}) \end{cases} \tag{2.5.5}$$

with E being the energy and $q_{ex} \neq 0$. Indeed, our schemes are able to only preserve stationary solutions with zero-velocity, thus the numerical solutions should converge to (2.5.5) when refining the mesh.

In table 2.3 the numerical errors and orders of convergence are shown in norms $\mathbf{L}^1$, $\mathbf{L}^2$ and $\mathbf{L}^\infty$ for both WB-ARS2 and WB-HR2; the results seem to be satisfying.

Wave propagation test case. In this test we assess the wave-capturing properties of the well-balanced schemes. We assume that a single wave propagates in the vessel, with parameters defined by (2.5.4) and table 2.2. Namely, we impose the following unsteady inlet flow

$$q_{\text{in}}(t) = \begin{cases} q_{\text{pulse}} \sin\left(2\pi \frac{t}{t_{\text{pulse}}}\right) & \text{if} \quad t \leq \frac{t_{\text{pulse}}}{2} \\ 0 & \text{otherwise} \end{cases}$$

where once again we define q_{pulse} as $q_{\text{pulse}} = ShapA_{in}c_{in}$. For the right boundary condition we simply use the transmissive one. Finally $t_{\text{pulse}} = 0.04$s and we take $t_{Out} = 0.045$s as ending time. We compute a reference solution with the WB-HR2 and $M = 2048$ cells. In figure 2.3, we insert the results only for WB-ARS and WB-ARS2 as the ones obtained with WB-HR/2 are similar. Of course solutions obtained with first-order schemes are more diffusive than the ones found exploiting higher order methods, but both outputs tend to the reference one. We also observe that there are no spurious oscillations in the results.

Propagation of a pulse to/from an expansion. Here we want to consider two different cases, a pulse propagating to and from an expansion. In the former case the initial radius is

$$R(x, t=0) = \begin{cases} R_0(x)(1 + \varepsilon \sin(\frac{100}{20L}\pi(x - \frac{65L}{100}))) & \text{if} \quad \frac{65L}{100} \leq x \leq \frac{85L}{100} \\ R_0 & \text{if} \quad \text{else}, \end{cases} \tag{2.5.6}$$

while in the second one

$$R(x, t=0) = \begin{cases} R_0(x)(1 + \varepsilon \sin(\frac{100}{20L}\pi(x - \frac{15L}{100}))) & \text{if} \quad \frac{15L}{100} \leq x \leq \frac{35L}{100} \\ R_0 & \text{if} \quad \text{else} \end{cases} \tag{2.5.7}$$

where $\varepsilon = 5.0 \times 10^{-3}$. In this last numerical problem we assume the K wall rigidity to be constant, while the radius at equilibrium is given by

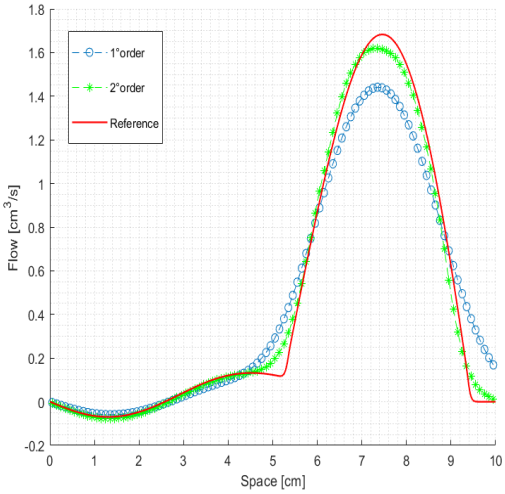
$$R_0(x) = \begin{cases} R_{in} + \Delta R & \text{if} \quad x < x_s \\ R_{in} + \frac{\Delta R}{2}(1 + cos(\pi \frac{x - x_s}{x_f - x_s})) & \text{if} \quad x_s \leq x \leq x_f \\ R_{in} & \text{if} \quad \text{else} \end{cases}$$

where $A_0 = \pi R_0^2$, $x_s = \frac{19L}{40}$ and $x_f = \frac{L}{2}$. Note that $A_1 > A_{end}$. The other parameters' values can be found in table 2.4. We use transmissive boundary conditions.

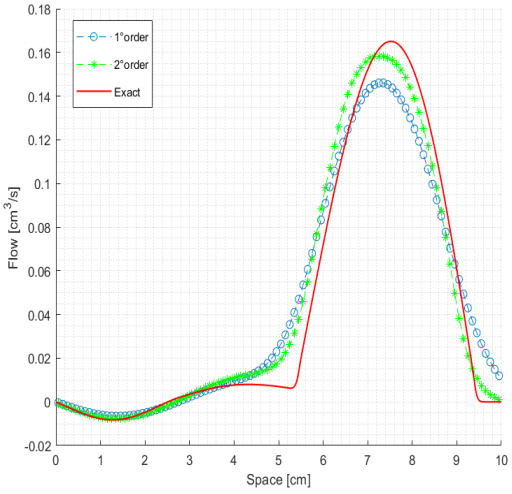
In figure 2.4, we present the outputs for this two problems: for the results on the left and right we use initial conditions (2.5.6) and (2.5.7) respectively. We compare the outputs of WB-HR2 for $M = 200$ cells against a reference solution attained with WB-HR2 and $M = 2048$ cells. Once again, we show the results of only one of the numerical schemes as their outputs are very similar. Indeed, the numerical solutions are satisfying and comparable with the ones of [15, 36].

2.6 Concluding remarks

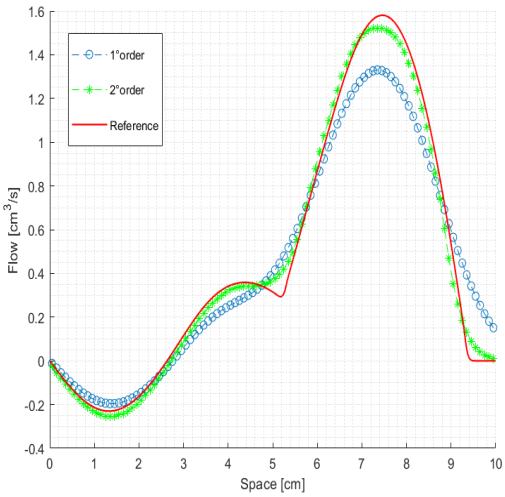
In this work, we have presented two different second-order well-balanced numerical schemes for the 1D blood flow equations, where the source term comes from variable mechanical and geometrical properties. By well-balanced we mean that the numerical method is able to preserve the zero-velocity "man at eternal rest" stationary solution. The first scheme is based on



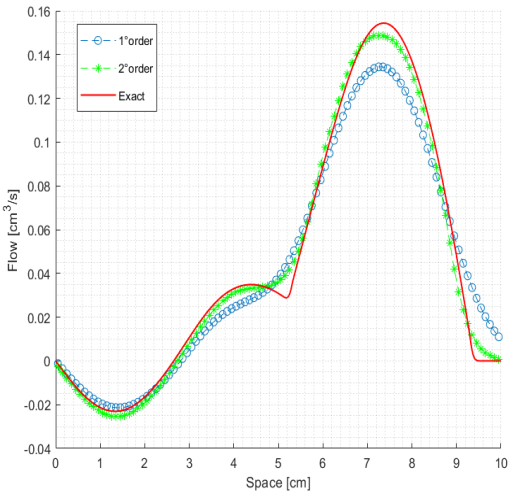
(a) $\Delta G = 10\%$, $Shap = 10^{-2}$



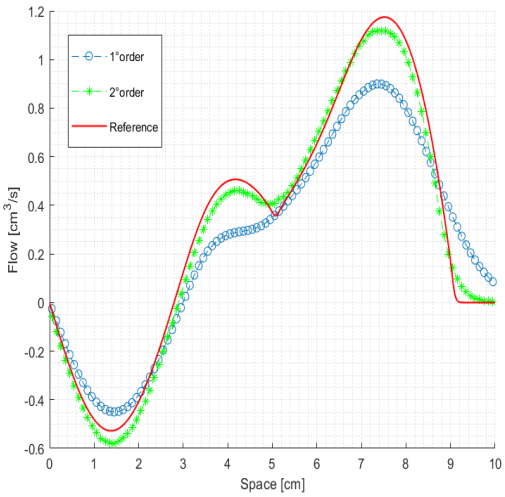
(b) $\Delta G = 10\%$, $Shap = 10^{-3}$



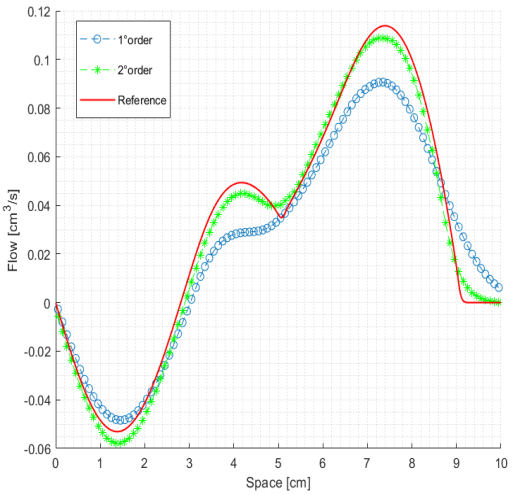
(c) $\Delta G = 30\%$, $Shap = 10^{-2}$



(d) $\Delta G = 30\%$, $Shap = 10^{-3}$



(e) $\Delta G = 60\%$, $Shap = 10^{-2}$

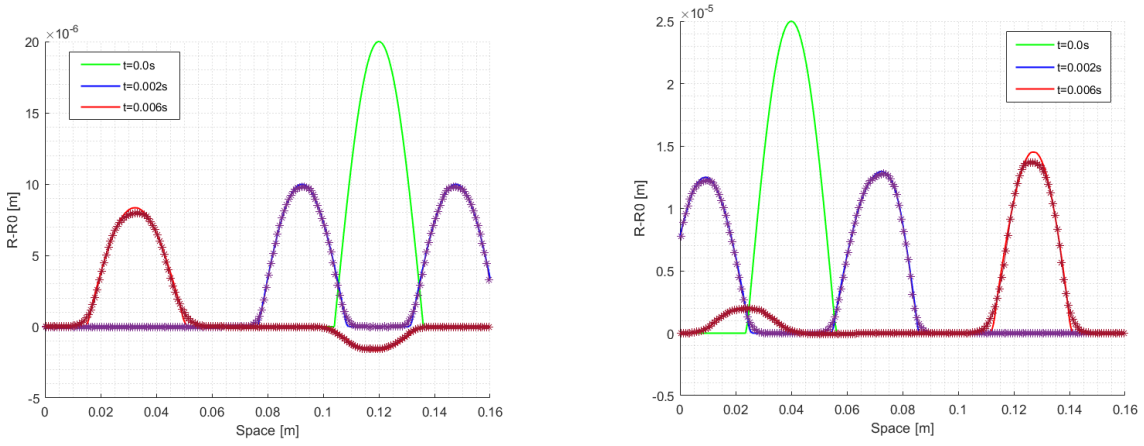


(f) $\Delta G = 60\%$, $Shap = 10^{-3}$

Figure 2.3: Wave propagation problem. We used $\Delta G = 10\%$ (top), $\Delta G = 30\%$ (middle), $\Delta G = 60\%$ (bottom) and $Shap = 10^{-2}$ (left), $Shap = 10^{-3}$(right). Comparison among WB-ARS (-o blue symbol), WB-ARS2 (-* green symbol) and reference (red line) solution. $M = 100$ cells.

Parameter	Value
L	0.16 [m]
R_{in}	4.0×10^{-3} [m]
ΔR	1.0×10^{-3} [m]
K	$\frac{1}{\sqrt{\pi}} \times 10^{8}$ [Pam]
ρ	1060 [Kgm^{-3}]

Table 2.4: Parameters values for test cases in section 2.5.2.



(a) IC (2.5.6)

(b) IC (2.5.7)

Figure 2.4: Propagation of a pulse to (left) and from (right) an expansion. Radius minus radius at equilibrium at time $t = 0.0s$ (green), $t = 0.002s$ (blue), $t = 0.006s$(red). Comparison between WB-HR2 (-* symbol) and reference (line) solution. $M = 200$ cells.

an approximate Riemann solver and the numerical source is defined in such a way that it is consistent in the integral sense with the source term, whereas the second method exploits the hydrostatic reconstruction approach. Only the Lagrangian step has to be modified in order to satisfy the well-balanced property. Both numerical schemes proved to be satisfying and their results are almost identical. On one hand, future works are expected to deal with an implicit formulation of the Lagrangian step as it would speed up the numerical methods. Indeed, the arteries are known to have an average Shapiro number of order 10^{-2} and this could imply a restriction on the time step value. On the other hand, efforts could be made to obtain fully well-balanced schemes able to also preserve stationary solutions with non-zero velocity. In this regard, we refer to [5, 25] for fully well-balanced Lagrange-projection schemes applied to the shallow water system at first and high order of accuracy respectively. Another interesting reference in this framework is [20], in which the authors considered the low-Shapiro number steady states, which could be more easily preserved than the classical steady states with non-zero velocity.

2.7 Bibliography

[1] 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: 2050-2065, 2004. 10.1137/S1064827503431090

[2] M. Baudin, C. Berthon, F. Coquel, R. Masson, Q. HuyTran. *A relaxation method for two-phase flow models with hydrodynamic closure law.* Numerische Mathematik. 99: 411-440, 2005. 10.1007/s00211-004-0558-1.

[3] C. Berthon, F. Foucher. *Efficient well–balanced hydrostatic upwind schemes for shallow–water equations.* Journal of Computational Physics. 231: 4993–5015, 2012. 10.1016/j.jcp.2012.02.031.

[4] F. Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources.* Frontiers in mathematics, 2004. 10.1007/b93802.

[5] M. J. Castro Díaz, C. Chalons and T. Morales De Luna. *A fully well-balanced Lagrange-Projection type scheme for the Shallow-water equations.* SIAM J. Numer. Anal., 56(5): 3071–3098, 2018. 10.1137/17M1156101.

[6] M. J. Castro, T. Morales de Luna, C. Parés Madroñal. *Well-Balanced Schemes and Path-Conservative Numerical Methods.* Handbook of Numerical Analysis, Elsevier, 18: 131-175, 2017. 10.1016/bs.hna.2016.10.002.

[7] M. J. Castro Díaz, A. Pardo Milanés and C. Parés. *Well-balanced numerical schemes based on a generalized hydrostatic reconstruction technique.* Mathematical Models and Methods in Applied Sciences. 5: 2055-2113, 2007. 10.1142/S021820250700256X.

[8] M. J. Castro and C. Parés. *Well-Balanced High-Order Finite Volume Methods for Systems of Balance Laws.* Journal of Scientific Computing. 82: 48, 2020. 10.1007/s10915-020-01149-5.

[9] C. Chalons, M. Girardin, S. Kokh. *An All-Regime Lagrange-Projection like scheme for the gas dynamics equations on unstructured meshes.* Communications in Computational Physics. 20(1): 188-233 2016. 10.4208/cicp.260614.061115a.

[10] C. Chalons, M. Girardin, S. Kokh. *An all-regime Lagrange-Projection like scheme for 2D homogeneous models for two-phase flows on unstructured meshes.* Journal of Computational Physics, Elsevier, 335: 885-904, 2017. 10.1016/j.jcp.2017.01.017.

[11] C. Chalons, P. Kestener, S. Kokh, and M. Stauffert. *A large time-step and well-balanced Lagrange-Projection type scheme for the Shallow-water equations.* Communications in Mathematical Sciences. 15(3): 765–788, 2017. 10.4310/CMS.2017.v15.n3.a9.

[12] C. Chalons, S. Kokh, M. Girardin. *Large Time Step and Asymptotic Preserving Numerical Schemes for the Gas Dynamics Equations with Source Terms.* SIAM Journal on Scientific Computing. 35(6): A2874–A2902, 2013. 10.1137/130908671.

[13] F. Coquel, E. Godlewski, B. Perthame, A. In, and P. Rascle. *Some new Godunov and relaxation methods for two-phase flow problems.* In: Toro E.F. (eds) Godunov Methods. Springer, New York, NY, pages 179–188, 2001. 10.1007/978-1-4615-0663-8_18

[14] F. Coquel and B. Perthame. *Relaxation of energy and approximate Riemann solvers for general pressure laws in fluid dynamics.* SIAM Journal on Numerical Analysis, 35(6): 2223–2249, 1998. 10.1137/S0036142997318528.

[15] O. Delestre, P.-Y. Lagrée. *A "well-balanced" finite volume scheme for blood flow simulation.* International Journal for Numerical Methods in Fluids, Wiley, 72(2): 177-205, 2013. 10.1002/fld.3736.

[16] F. Duboc, C. Enaux, S. Jaouen, H. Jourdren, M. Wolff. *High-order dimensionally split Lagrange-remap schemes for compressible hydrodynamics.* Comptes Rendus Mathematique, 348(1–2): 105-110, 2010. 10.1016/j.crma.2009.12.008.

[17] L. Formaggia, A. Quarteroni, A. Veneziani. *Cardiovascular Mathematics: Modeling and Simulation of the Circulatory System.* Springer-Verlag: Italia, Milano, 2009. 10.1007/978-88-470-1152-6.

[18] G. Gallice. *Solveurs simples positifs et entropiques pour les systèmes hyperboliques avec terme source.* C. R. Math. Acad. Sci. Paris 334(8): 713-716, 2002. 10.1016/S1631-073X(02)02307-5.

[19] G. Gallice. *Positive and entropy stable Godunov-type schemes for gas dynamics and MHD equations in Lagrangian or Eulerian coordinates.* Numer. Math. 94(4): 673-713, 2003. 10.1007/s00211-002-0430-0

[20] A. R. Ghigo, O. Delestre, J.-M. Fullana, P.-Y. Lagrée. *Low-Shapiro hydrostatic reconstruction technique for blood flow simulation in large arteries with varying geometrical and mechanical properties.* Journal of Computational Physics, Elsevier, 331: 108 - 136, 2017. 10.1016/j.jcp.2016.11.032.

[21] S. Gottlieb and C.-W. Shu. *Total variation diminishing RUNGE-KUTTA schemes.* Mathematics of Computation. 67, 1996. 10.1090/S0025-5718-98-00913-2.

[22] A. Harten, P. Lax and B. van Leer. *On Upstream Differencing and Godunov-Type Schemes for Hyperbolic Conservation Laws.* SIAM Rev. 25: 35-61, 1983. 10.1137/1025002.

[23] V. Michel-Dansac, C. Berthon, S. Clain, F. Foucher. *A well-balanced scheme for the shallow-water equations with topography.* Computers and Mathematics with Applications, Elsevier, 72: 568 - 593, 2016. 10.1016/j.camwa.2016.05.015.

[24] V. Michel-Dansac, C. Berthon, S. Clain, F. Foucher. *A well-balanced scheme for the shallow-water equations with topography or Manning friction.* Journal of Computational Physics, Elsevier, 335: 115-154, 2017. 10.1016/j.jcp.2017.01.009.

[25] T. Morales De Luna, M. J. Castro Díaz and C. Chalons. *High order fully well-balanced Lagrange-Projection scheme for Shallow-water.* Commun. Math. Sci., 18(3): 781–807, 2020. 10.4310/CMS.2020.v18.n3.a9

[26] L. O. Müller, C. Parés and E. F. Toro. *Well-balanced high-order numerical schemes for one-dimensional blood flow in vessels with varying mechanical properties.* Journal of Computational Physics 242: 53-85, 2013. 10.1016/j.jcp.2013.01.050.

[27] L. O. Müller and E. F. Toro. *Well-balanced high-order solver for blood flow in networks of vessels with variable properties.* Int. J. Numer. Meth. Biomed. Engng. 29: 1388–1411, 2013. 10.1002/cnm.2580.

[28] K. Salari, P. Knupp. *Code Verification by the Method of Manufactured Solutions.* Report, 2000. https://digital.library.unt.edu/ark:/67531/metadc702130/

[29] I. Suliciu. *On the thermodynamics of fluids with relaxation and phase transitions. Fluids with relaxation.* Int. J. Engag. Sci. 36: 921-947, 1998.

[30] E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics*, Third Edition. Springer–Verlag, 2009. 10.1007/b79761_5.

[31] E. F. Toro. *Brain venous haemodynamics, neurological diseases and mathematical modelling. A review.* Applied Mathematics and Computation, 272: 542–579, 2016. 10.1016/j.amc.2015.06.066.

[32] E. F. Toro. *Lecture notes on computational haemodynamics.* Mathematics Department, University of Trento, Italy, 2017.

[33] E. F. Toro and A. Siviglia. *PRICE: Primitive Centred Schemes for Hyperbolic Systems.* Int. J. Numer. Meth. in Fluids, 42: 1263–1291, 2003. 10.1002/fld.491.

[34] E. F. Toro and A. Siviglia. *Simplified blood flow model with discontinuous vessel properties: analysis and exact solutions.* Modeling, Simulation and Applications, 5, 2011. 10.1007/978-88-470-1935-5_2.

[35] E. F. Toro and A. Siviglia. *Flow in collapsible tubes with discontinuous mechanical properties: mathematical model and exact solutions.* Communications in Computational Physics, 13(2): 361–385, 2013. 10.4208/cicp.210611.240212a.

[36] Z. Wang, G. Li and O. Delestre. *Well-balanced finite difference weighted essentially non-oscillatory schemes for the blood flow model.* Int. J. Numer. Meth. Fluids, 82, 2016. 10.1002/fld.4232.

Chapter 3

Sur les schémas implicites-explicites équilibres de Lagrange-projection pour les équations de Saint-Venant à deux couches

Ce chapitre a été soumis pour publication dans le Journal "Applied Mathematics and Computation" sous la référence: A. Del Grosso, M. J. Castro Dìaz, C. Chalons and T. Morales de Luna. *On well-balanced implicit-explicit Lagrange-projection schemes for two-layer shallow water equations.*

Le manuscrit est disponible sous la forme d'un preprint sur l'archive ouverte HAL avec ID: hal-03655011 .

Ce chapitre concerne l'étude des schémas de type Lagrange-projection équilibres appliqués au système de Saint-Venant à deux couches. Ce modèle est connu pour être difficile à approcher en raison de sa forme non-conservative et du fait qu'il n'est que conditionnellement hyperbolique. En particulier, on propose une formulation du modèle mathématique en coordonnées lagrangiennes. La méthode HLL est ensuite appliquée à une version simplifiée du système lagrangien résultant. De plus, nous interprétons la décomposition de type Lagrange-projection comme un splitting acoustique-transport. En utilisant cette nouvelle interprétation, un autre solveur de Riemann approché pour l'étape acoustique-lagrangienne est décrit. Une méthode explicite et une méthode implicite-explicite sont proposées, cette dernière pouvant permettre des simulations très rapides dans des régimes sous-critiques. Enfin, nous présentons des simulations numériques dans lesquelles les résultats des méthodes de type Lagrange-projection sont comparés à ceux de la méthode IFCP.

Deux annexes sont ensuite insérées afin de clarifier certains aspects de ce chapitre. Dans l'annexe 3.A nous montrons comment écrire le système linéaire associé à l'approximation implicite du système acoustique. Ensuite, puisque l'approche HLL a également été appliquée au système lagrangien, des détails à ce sujet sont donnés dans l'annexe 3.B. En particulier, un aperçu rapide des schémas numeriques préservants les chemins ("path-preserving") est donné.

On well-balanced implicit-explicit Lagrange-projection schemes for two-layer shallow water equations

Abstract

This works concerns the study of well-balanced Lagrange-projection schemes applied to the two-layer shallow water system. In particular, a formulation of the mathematical model in Lagrangian coordinates is proposed. The HLL method is then applied to a simplified version of the resulting Lagrangian system. Furthermore, based on the acoustic-transport splitting interpretation, another approximate Riemann solver for the acoustic-Lagrangian step is described. Both an explicit and an implicit-explicit method are proposed, where the latter can allow very fast simulations in sub-critical regimes. Finally, we show some numerical simulations in which the outputs are compared with the IFCP method's results.

3.1 Introduction and mathematical model

In this work we are interested in the numerical approximation of the 1D two-layer shallow water system, which models a fluid composed of two superposed layers of immiscible liquids where the upper one has a smaller density ρ_1. Thus, using the subscript $j = 1, 2$ to indicate the jth layer, we state $\rho_1 < \rho_2$. This kind of situations can occur when there are two liquids of different densities or even with a single fluid present at two different temperatures, as in oceanic flows. Referring for instance to [1, 3, 7, 9, 10, 11, 23, 24] and also to figure 3.1 for the notations, the two-layer shallow water system is given by

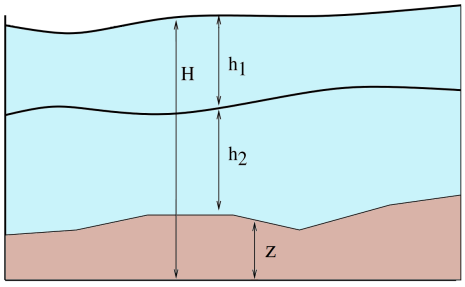


Figure 3.1: Sketch of the two-layer shallow water: h_1, h_2 water heights, z topography and H free surface.

$$\begin{cases} \partial_t h_1 + \partial_x(h_1 u_1) = 0 \\ \partial_t(h_1 u_1) + \partial_x\left(h_1 u_1^2 + \dfrac{g h_1^2}{2}\right) + g h_1 \partial_x h_2 = -g h_1 \partial_x z \\ \partial_t h_2 + \partial_x(h_2 u_2) = 0 \\ \partial_t(h_2 u_2) + \partial_x\left(h_2 u_2^2 + \dfrac{g h_2^2}{2}\right) + g\dfrac{\rho_1}{\rho_2} h_2 \partial_x h_1 = -g h_2 \partial_x z \end{cases} \tag{3.1.1}$$

where $t > 0$ represents the time and x the axial coordinate. Then, $h_j(x,t) > 0$ is the water depth of the corresponding layer, $u_j(x,t)$ the averaged horizontal velocity and finally $z(x)$ the bed elevation. Regarding the parameters, we define the gravitational acceleration g and we denote $r = \dfrac{\rho_1}{\rho_2}$ the ratio of densities. These equations (3.1.1) can also be reformulated in a more compact way, namely

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \mathbf{B}(\mathbf{Q})\partial_x \mathbf{Q} = \mathbf{S}(\mathbf{Q})$$

where $\mathbf{Q} = (h_1, h_1u_1, h_2, h_2u_2)^T$ is the vector of unknowns,

$$\mathbf{F}(\mathbf{Q}) = \begin{pmatrix} h_1 u_1 \\ h_1 u_1^2 + \dfrac{g h_1^2}{2} \\ h_2 u_2 \\ h_2 u_2^2 + \dfrac{g h_2^2}{2} \end{pmatrix}, \quad \mathbf{B}(\mathbf{Q}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & g h_1 & 0 \\ 0 & 0 & 0 & 0 \\ r g h_2 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{S}(\mathbf{Q}) = \begin{pmatrix} 0 \\ -g h_1 \partial_x z \\ 0 \\ -g h_2 \partial_x z \end{pmatrix}.$$

Then, few computations show that the characteristic equation of the non-conservative matrix $\mathbf{A}(\mathbf{Q}) = \dfrac{\partial \mathbf{F}(\mathbf{Q})}{\partial \mathbf{Q}} + \mathbf{B}(\mathbf{Q})$ is given by

$$(\lambda^2 + u_1^2 - c_1^2 - 2\lambda u_1)(\lambda^2 + u_2^2 - c_2^2 - 2\lambda u_2) = r g^2 h_1 h_2$$

where $c_j = \sqrt{g h_j}$ would correspond to the sound speed of the jth layer in absence of the other layer. In particular, it is easy to see that we have null eigenvalue when

$$G^2 = F_1^2 + F_2^2 - (1-r) F_2^2 F_1^2 = 1$$

where F_j such that $F_j^2 = \dfrac{u_j^2}{(1-r)c_j^2}$ are the internal Froude numbers and G is the composite Froude number, see also [23]. It is worth to specify that if $G^2 < 1$, we say to be in a subcritical regime while $G^2 > 1$ indicates a supercritical flow. Depending on the value of r we may be able to explicitly define the eigenvalues of system (3.1.1) or not. Indeed, if $r = 0$, it is clear that the eigenvalues correspond to the shallow water system's eigenvalues for each layer separately. Thus, if $r \approx 0$, the two layers of fluids behave almost independently. However, we are interested in situations in which $r \approx 1$, which often happen in geophysical flows. In this case, thanks to [27], the following first-order approximation of the eigenvalues are available

$$\lambda_{\text{Ext}}^{\pm} = \frac{h_1 u_1 + h_2 u_2}{h_1 + h_2} \pm \sqrt{g(h_1 + h_2)}$$

$$\lambda_{\text{Int}}^{\pm} = \frac{h_1 u_2 + h_2 u_1}{h_1 + h_2} \pm \sqrt{g_1 \frac{h_1 h_2}{h_1 + h_2}\Big(1 - \frac{(u_1 - u_2)^2}{g_1(h_1 + h_2)}\Big)}$$

where $g_1 = g(1-r)$ is the reduced gravity. It is a well-known fact that the two-layer system may lose its hyperbolic character and complex eigenvalues may arise. Indeed, in view of the

first-order approximation of the eigenvalues, it is clear that the fulfillment of the following condition

$$\frac{(u_1 - u_2)^2}{g_1(h_1 + h_2)} > 1$$

would lead to complex internal eigenvalues and thus we would lose the hyperbolicity of system (3.1.1). Physically speaking, this corresponds to situations where the mixing of the two layers would occur, leading to the appearance of shear instabilities. In practice, this mixture would partially dissipates the energy. To simulate such an effect, we could include friction in the mathematical model, otherwise, in the numerical simulations, the interface disturbances may grow and lead to a wrong solution [9]. Hence, it is clear that model (3.1.1) is not well-adapted to those situations and a more complex one would be needed. When we consider numerical tests, the loss of hyperbolicity could be accepted only in occasional situations to make sure we do not depart from the correct solution. For further informations about the two-layer shallow water model and its loss of hyperbolicity, see for instance [9, 7, 11, 10].

Concerning the numerical strategy, here we aim to design and implement well-balanced implicit-explicit Lagrange-projection schemes. So far Lagrange-Projection (LP) methods have been studied for different mathematical models as the shallow water system [18] and related models [14, 15], the gas dynamic equations [16, 17] and the blood flow system [22]. However, up to our knowledge, they have never been employed to numerically approximate the two-layer shallow water equations. Indeed, due the presence of two velocities u_1, u_2, it is not straightforward to understand how to apply the Lagrange-projection strategy to this system. Indeed, a first idea could be to implement the LP approach for each layer and then to couple them. However, it is known that a method that applies an arbitrary scheme to each layer usually leads to the presence of oscillations in the numerical results [11]. In [13], the authors described a first attempt to apply the Lagrange-projection strategy to a two-phase system, in particular to the two-fluid two-pressure (or seven-equation) model. There, the coupling terms of the system have not been considered directly inside the Lagrange-projection decomposition but in a third step. In this work, we propose a different approach from the ones mentioned above.

Moreover, we also consider a different interpretation of the Lagrange-projection approach, namely the acoustic-transport splitting, refer again to the previous references. Indeed, by decomposing the different phenomena of the mathematical model, we obtain two different systems, the acoustic and transport ones. For the former, we design an approximate Riemann solver based on a relaxation approach and then the associated Godunov-type scheme is used. We also explain how the resulting approximation can be exploited for the Lagrangian system. Furthermore, let us recall that the acoustic-transport splitting (or equivalently the Lagrange-projection decomposition) can be particularly interesting in subsonic regimes, where the acoustic waves are much faster than the transport ones. This means that an implicit approximation applied to the acoustic system could lead to the construction of very fast numerical schemes as we would neglect the acoustic time step condition. For this reason, we propose both an explicit and an implicit strategy for the acoustic equations, while keeping an explicit approximation for the transport step. For implicit-explicit Lagrange-projection methods refer for instance to [19, 18].

Last but not least, we are interested in the well-balanced property of the numerical schemes, meaning that the numerical methods are able to preserve the stationary solutions of the mathematical model, at least for the so-called lake-at-rest solutions. Indeed, it is well-known that we could otherwise observe spurious oscillations in the numerical results when close to a steady state, refer for instance to [8, 14, 15, 18, 22, 26] and to [2, 25, 20, 5, 6, 4] for well-balanced schemes with and without the Lagrange-projection decomposition respectively. Therefore,

let us see which are the stationary solutions of the model. They are generally given by the following relations

$$\begin{cases} h_j u_j = q_j^0 = \text{constant}, \quad \text{with} \quad j = 1, 2 \\ \dfrac{q_1^0}{2h_1^2} + g(h_1 + h_2 + z) = \text{constant} \\ \dfrac{q_2^0}{2h_2^2} + g(rh_1 + h_2 + z) = \text{constant}. \end{cases} \tag{3.1.2}$$

Here we are particularly interested in preserving only the steady states with zero velocity, usually known as lake-at-rest stationary solutions, namely

$$\begin{cases} u_j = 0, \quad \text{with} \quad j = 1, 2 \\ h_1 = \text{constant} \\ h_2 + z = \text{constant}. \end{cases} \tag{3.1.3}$$

For fully well-balanced Lagrange-projection methods, refer to [8, 26].

Chapter outline. To conclude this section, we give a brief outline of the manuscript. In section 3.2, we formulate the mathematical model (3.1.1) in Lagrangian coordinates and we analyze a simplified version of it. In section 3.3, the acoustic-transport decomposition is presented, leading to the description of an approximate Riemann solver for the acoustic system. Section 3.4 is devoted to the presentation of both the explicit and implicit-explicit numerical strategies. In particular, the well-balanced property is proved in both cases. Finally, we show numerical simulations and draw conclusions in sections 3.5 and 3.6 respectively.

3.2 Lagrangian coordinates

This section is devoted to the description of the mathematical model (3.1.1) in Lagrangian coordinates. After the introduction of an arbitrary fluid particle located at ξ, the usual procedure is to describe the corresponding characteristic curves. Hence, due to the presence of two different velocities, one for each layer, it is convenient to define two different trajectories x_j, $j = 1, 2$, such that

$$\begin{cases} \dfrac{\partial x_j}{\partial t}(\xi, t) = u_j(x_j(\xi, t), t) \\ x_j(\xi, 0) = \xi. \end{cases} \tag{3.2.1}$$

As a consequence, we define the volume ratio $L_j(\xi, t)$ for each layer as follows

$$L_j(\xi, t) = \frac{\partial x_j}{\partial \xi}(\xi, t) \quad \text{such that} \quad \begin{cases} \dfrac{\partial L_j}{\partial t}(\xi, t) = \partial_\xi u_j(x_j(\xi, t), t) \\ L_j(\xi, 0) = 1. \end{cases} \tag{3.2.2}$$

Then, any function : $(x_j, t) \to \varphi(x_j, t)$ (associated to the jth trajectory) in Eulerian coordinates can be expressed in Lagrangian coordinates,

$$\overline{\varphi}^{(j)}(\xi, t) = \varphi(x_j(\xi, t), t).$$

This means that we have to introduce new additional variables. Indeed, we generally need to distinguish between $\overline{\varphi}^{(1)}(\xi, t) = \varphi(x_1(\xi, t), t)$ and $\overline{\varphi}^{(2)}(\xi, t) = \varphi(x_2(\xi, t), t)$. Then, defining the space and time derivatives

$$\partial_\xi \overline{\varphi}^{(j)}(\xi, t) = L_j(\xi, t)\partial_x \varphi(x_j, t) \quad \text{and} \quad \partial_t \overline{\varphi}^{(j)}(\xi, t) = \partial_t \varphi(x_j, t) + u_j(x_j, t)\partial_x \varphi(x_j, t),$$

we are able to reformulate system (3.1.1) in Lagrangian coordinates as follows,

$$\begin{cases} \partial_t(L_1\overline{h}_1^{(1)}) = 0 \\ \partial_t(L_1\overline{h}_1^{(1)}\overline{u}_1^{(1)}) + \partial_\xi\overline{p}_1^{(1)} + g\overline{h}_1^{(1)}L_1\partial_x\overline{h}_2^{(1)} = -g\overline{h}_1^{(1)}\partial_\xi\overline{z}^{(1)} \\ \partial_t(L_2\overline{h}_2^{(2)}) = 0 \\ \partial_t(L_2\overline{h}_2^{(2)}\overline{u}_2^{(2)}) + \partial_\xi\overline{p}_2^{(2)} + gr\overline{h}_2^{(2)}L_2\partial_x\overline{h}_1^{(2)} = -g\overline{h}_2^{(2)}\partial_\xi\overline{z}^{(2)}, \end{cases} \tag{3.2.3}$$

with pressures $p_j^{(j)} = g\dfrac{(h_j^{(j)})^2}{2}$. Observe that here we wrote the evolution equations for the variables $\overline{h}_1^{(1)}$, $\overline{h}_1^{(1)}\overline{u}_1^{(1)}$, $\overline{h}_2^{(2)}$ and $\overline{h}_2^{(2)}\overline{u}_2^{(2)}$ but four additional equations would be needed for the unknowns $\overline{h}_1^{(2)}$, $\overline{h}_1^{(2)}\overline{u}_1^{(2)}$, $\overline{h}_2^{(1)}$ and $\overline{h}_2^{(1)}\overline{u}_2^{(1)}$. However, in the following we do an approximation in order to be able to neglect such variables. Indeed, after few manipulations and neglecting the superscript when it is equal to the subscript, we obtain

$$\begin{cases} \partial_t(L_1\overline{h}_1) = 0 \\ \partial_t(L_1\overline{h}_1\overline{u}_1) + \partial_\xi\overline{p}_1 + g\overline{h}_1\Big(\dfrac{L_1}{L_2}\partial_\xi\overline{h}_2 + L_1\partial_x(\overline{h}_2^{(1)} - \overline{h}_2)\Big) = -g\overline{h}_1\partial_\xi\overline{z}^{(1)} \\ \partial_t(L_2\overline{h}_2) = 0 \\ \partial_t(L_2\overline{h}_2\overline{u}_2) + \partial_\xi\overline{p}_2 + gr\overline{h}_2\Big(\dfrac{L_2}{L_1}\partial_\xi\overline{h}_1 + L_2\partial_x(\overline{h}_1^{(2)} - \overline{h}_1)\Big) = -g\overline{h}_2\partial_\xi\overline{z}^{(2)}. \end{cases}$$

At this stage, we do an approximation and assume the terms $\partial_x(\overline{h}_2^{(1)} - \overline{h}_2)$ and $\partial_x(\overline{h}_1^{(2)} - \overline{h}_1)$ to be null as at initial time we have $\overline{h}_1\big|_{t=0} = \overline{h}_1^{(2)}\big|_{t=0} = h_1(x, t=0)$ and $\overline{h}_2\big|_{t=0} = \overline{h}_2^{(1)}\big|_{t=0} = h_2(x, t=0)$. From a numerical point of view, this implies that we are approximating these terms explicitly and at first-order of accuracy. Furthermore, we generally expect these two terms to be rather small when the velocities u_1 and u_2 are close. Hence, from now on, we consider the following Lagrangian system

$$\begin{cases} \partial_t(L_1\overline{h}_1) = 0 \\ \partial_t(L_1\overline{h}_1\overline{u}_1) + \partial_\xi\overline{p}_1 + g\overline{h}_1\dfrac{L_1}{L_2}\partial_\xi\overline{h}_2 = -g\overline{h}_1\partial_\xi\overline{z}^{(1)} \\ \partial_t(L_2\overline{h}_2) = 0 \\ \partial_t(L_2\overline{h}_2\overline{u}_2) + \partial_\xi\overline{p}_2 + gr\overline{h}_2\dfrac{L_2}{L_1}\partial_\xi\overline{h}_1 = -g\overline{h}_2\partial_\xi\overline{z}^{(2)}. \end{cases} \tag{3.2.4}$$

It is clear that further studies are required in order to include the unknowns $\overline{h}_1^{(2)}$, $\overline{h}_1^{(2)}\overline{u}_1^{(2)}$, $\overline{h}_2^{(1)}$ and $\overline{h}_2^{(1)}\overline{u}_2^{(1)}$ in system (3.2.3). More details about the Lagrange-projection decomposition applied to the shallow water system can be found for instance in [26, 8].

Neglecting the bars over the unknowns, we reformulate equations (3.2.4) together with system (3.2.2) in a more compact way,

$$\partial_t\mathbf{LQ} + \mathbf{A}(\mathbf{LQ})\partial_\xi\mathbf{LQ} = \mathbf{S}(\mathbf{LQ}, z)$$

where

$$\mathbf{LQ} = \begin{pmatrix} L_1 \\ L_1h_1 \\ L_1h_1u_1 \\ L_2 \\ L_2h_2 \\ L_2h_2u_2 \end{pmatrix}, \quad \mathbf{A}(\mathbf{LQ}) = \begin{pmatrix} 0 & \frac{u_1}{L_1h_1} & -\frac{1}{L_1h_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -g\frac{h_1^2}{L_1} & g\frac{h_1}{L_1} & 0 & -g\frac{L_1h_1h_2}{L_2^2} & g\frac{L_1h_1}{L_2^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{u_2}{L_2h_2} & -\frac{1}{L_2h_2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -gr\frac{L_2h_2h_1}{L_1^2} & gr\frac{L_2h_2}{L_1^2} & 0 & -g\frac{h_2^2}{L_2} & g\frac{h_2}{L_2} & 0 \end{pmatrix}$$

and

$$\mathbf{S}(\mathbf{LQ}, z) = \begin{pmatrix} 0 \\ 0 \\ -gh_1\partial_\xi \overline{z}^{(1)} \\ 0 \\ 0 \\ -gh_2\partial_\xi \overline{z}^{(2)} \end{pmatrix}.$$

Moreover, observing that $L_j\overline{h}_j$ does not depend on time, we find

$$L_j\overline{h}_j(\xi,t) = L_j\overline{h}_j(\xi,0) = h_j(\xi,0) = h_j^0 \quad \text{and consequently} \quad L_j = \frac{h_j^0}{\overline{h}_j},$$

and as such

$$\frac{L_1}{L_2} = \frac{h_1^0}{h_2^0}\frac{\overline{h}_2}{\overline{h}_1}.$$

Thus, defining the variables $\overline{\tau}_j = \dfrac{1}{\overline{h}_j}$, we propose another formulation for system (3.2.4),

$$\begin{cases} \partial_t h_1^0 = 0 \\ \partial_t(h_1^0\overline{\tau}_1) - \partial_\xi \overline{u}_1 = 0 \\ \partial_t(h_1^0\overline{u}_1) + \partial_\xi \overline{p}_1 + \dfrac{h_1^0}{h_2^0}\partial_\xi \overline{p}_2 = -\dfrac{g}{\overline{\tau}_1}\partial_\xi \overline{z}^{(1)} \\ \partial_t h_2^0 = 0 \\ \partial_t(h_2^0\overline{\tau}_2) - \partial_\xi \overline{u}_2 = 0 \\ \partial_t(h_2^0\overline{u}_2) + \partial_\xi \overline{p}_2 + r\dfrac{h_2^0}{h_1^0}\partial_\xi \overline{p}_1 = -\dfrac{g}{\overline{\tau}_2}\partial_\xi \overline{z}^{(2)} \end{cases}$$

or equivalently,

$$\begin{cases} \partial_t \overline{\tau}_1 - \dfrac{1}{h_1^0}\partial_\xi \overline{u}_1 = 0 \\ \partial_t \overline{u}_1 + \dfrac{1}{h_1^0}\partial_\xi \overline{p}_1 + \dfrac{1}{h_2^0}\partial_\xi \overline{p}_2 = -\dfrac{g}{h_1^0\overline{\tau}_1}\partial_\xi \overline{z}^{(1)} \\ \partial_t \overline{\tau}_2 - \dfrac{1}{h_2^0}\partial_\xi \overline{u}_2 = 0 \\ \partial_t \overline{u}_2 + \dfrac{1}{h_2^0}\partial_\xi \overline{p}_2 + \dfrac{r}{h_1^0}\partial_\xi \overline{p}_1 = -\dfrac{g}{h_2^0\overline{\tau}_2}\partial_\xi \overline{z}^{(2)}. \end{cases} \tag{3.2.5}$$

The matrix of the latter system reads

$$\begin{pmatrix} 0 & -\frac{1}{h_1^0} & 0 & 0 \\ \frac{1}{h_1^0}\partial_{\tau_1} p_1 & 0 & \frac{1}{h_2^0}\partial_{\tau_2} p_2 & 0 \\ 0 & 0 & 0 & -\frac{1}{h_2^0} \\ \frac{r}{h_1^0}\partial_{\tau_1} p_1 & 0 & \frac{1}{h_2^0}\partial_{\tau_2} p_2 & 0 \end{pmatrix},$$

whose characteristic polynomial is given by

$$\begin{aligned} &(\lambda^2 + \frac{\partial_{\tau_1} p_1}{h_{0,1}^2})(\lambda^2 + \frac{\partial_{\tau_2} p_2}{h_{0,2}^2}) - r\frac{\partial_{\tau_1} p_1}{h_{0,1}^2}\frac{\partial_{\tau_2} p_2}{h_{0,2}^2} = \\ &= \lambda^4 + (\frac{\partial_{\tau_1} p_1}{h_{0,1}^2} + \frac{\partial_{\tau_2} p_2}{h_{0,2}^2})\lambda^2 + (1-r)\frac{\partial_{\tau_1} p_1 \partial_{\tau_2} p_2}{h_{0,1}^2 h_{0,2}^2} = \lambda^4 + (\alpha_1+\alpha_2)\lambda^2 + (1-r)\alpha_1\alpha_2 \end{aligned}$$

with $\alpha_1 = \dfrac{\partial_{\tau_1} p_1}{h_{0,1}^2} = -g\dfrac{h_1^3}{h_{0,1}^2}$, $\alpha_2 = \dfrac{\partial_{\tau_2} p_2}{h_{0,2}^2} = -g\dfrac{h_2^3}{h_{0,2}^2}$ and thus $\alpha_1, \alpha_2 \leq 0$. It is easy to see that the eigenvalues are then given by

$$\lambda_{ext}^{\pm} = \pm\sqrt{\frac{-(\alpha_1+\alpha_2) + \sqrt{(\alpha_1+\alpha_2)^2 - 4(1-r)\alpha_1\alpha_2}}{2}},$$

$$\lambda_{int}^{\pm} = \pm\sqrt{\frac{-(\alpha_1+\alpha_2) - \sqrt{(\alpha_1+\alpha_2)^2 - 4(1-r)\alpha_1\alpha_2}}{2}}$$

which can be proved to be always real. Indeed, we first observe that

$$(\alpha_1+\alpha_2)^2 - 4(1-r)\alpha_1\alpha_2 = (\alpha_1-\alpha_2)^2 + 4r\alpha_1\alpha_2 \geq 0,$$

from which it follows that $\lambda_{ext}^{\pm}$ are real. Regarding $\lambda_{int}^{\pm}$, it is enough to note that

$$-(\alpha_1+\alpha_2) - \sqrt{(\alpha_1+\alpha_2)^2 - 4(1-r)\alpha_1\alpha_2} \geq 0$$

which holds true as $\alpha_1\alpha_2 \geq 0$. Thus it is interesting to remark that this simplified version of the Lagrangian system has always real eigenvalues.

Finally, let us remark that, in systems (3.2.4) and (3.2.5), we still have $\overline{z}^{(1)}$ and $\overline{z}^{(2)}$, which are generally different. However, since in section 3.4 we treat these topography source terms explicitly, once again we do an approximation and simply assume $\partial_\xi \overline{z}^{(1)} = \partial_\xi \overline{z}^{(2)} = \partial_\xi z$.

3.3 Acoustic transport splitting

As it is already known, the Lagrange-projection splitting and the acoustic-transport one can be interpreted as two different ways of describing the same kind of decomposition. Indeed, it has already been illustrated the relation between the acoustic and Lagrangian numerical approximation, see for instance the following papers applied to the shallow water system [18], the gas dynamics equations [16] and the blood flow system [22]. Therefore, following the lines of these works, we present the acoustic-transport splitting for the two-layer shallow water model, as it will be useful for the development of the numerical strategy. Hence, we decouple the different terms of the model, obtaining the acoustic and transport systems given by

$$\begin{cases} \partial_t h_1 + h_1\partial_x u_1 = 0 \\ \partial_t(h_1u_1) + h_1u_1\partial_x u_1 + \partial_x\dfrac{gh_1^2}{2} + gh_1\partial_x h_2 = -gh_1\partial_x z \\ \partial_t h_2 + h_2\partial_x u_2 = 0 \\ \partial_t(h_2u_2) + h_2u_2\partial_x u_2 + \partial_x\dfrac{gh_2^2}{2} + grh_2\partial_x h_1 = -gh_2\partial_x z \end{cases} \tag{3.3.1}$$

and

$$\begin{cases} \partial_t h_1 + u_1\partial_x h_1 = 0 \\ \partial_t(h_1u_1) + u_1\partial_x h_1u_1 = 0 \\ \partial_t h_2 + u_2\partial_x h_2 = 0 \\ \partial_t(h_2u_2) + u_2\partial_x h_2u_2 = 0 \end{cases} \tag{3.3.2}$$

respectively. We also observe that the latter simply reads as $\partial_t X_j + u_j\partial_x X_j = 0$ with $X = h, hu$ and $j = 1, 2$. On the other hand, after few computations, system (3.3.1) can be reformu-

lated as

$$\begin{cases} \partial_t h_1 + h_1 \partial_x u_1 = 0 \\ \partial_t u_1 + \dfrac{1}{h_1}\partial_x \dfrac{g h_1^2}{2} + g\partial_x h_2 = -g\partial_x z \\ \partial_t h_2 + h_2 \partial_x u_2 = 0 \\ \partial_t u_2 + \dfrac{1}{h_2}\partial_x \dfrac{g h_2^2}{2} + gr\partial_x h_1 = -g\partial_x z \end{cases}$$

and again

$$\begin{cases} \partial_t \tau_1 - \tau_1 \partial_x u_1 = 0 \\ \partial_t u_1 + \tau_1 \partial_x \dfrac{g}{2\tau_1^2} + \tau_2 \partial_x \dfrac{g}{2\tau_2^2} = -g\partial_x z \\ \partial_t \tau_2 - \tau_2 \partial_x u_2 = 0 \\ \partial_t u_2 + r\tau_1 \partial_x \dfrac{g}{2\tau_1^2} + \tau_2 \partial_x \dfrac{g}{2\tau_2^2} = -g\partial_x z. \end{cases} \tag{3.3.3}$$

where we have introduced the variables $\tau_j = \dfrac{1}{h_j}$, $j = 1, 2$. It is then evident the similarity between the Lagrangian system (3.2.5) and the acoustic equations (3.3.3).

Next, a first difficulty is related to the fact that usually the mass variable m is introduced for the acoustic system at this stage, see for instance the previous references [18, 16, 22]. For example, in the shallow water system, this new variable m is defined such that $\partial_m = \dfrac{1}{h}\partial_x$ with h the water height. Subsequently, it allows to obtain a conservation form of the equations (at least when there is no source term), making easier the definition of the numerical strategy. However, having in this case two different water heights h_1 and h_2, it is not clear how to include such a device and thus, we do not consider it here.

Since we do not know the general exact solution of a Riemann Problem (RP) associated with the acoustic system (3.3.3), we look for an approximate Riemann solver. In order to be able to define it, it is convenient to start by applying a linearization of the non-linear terms present in the equations. Therefore, considering the following form for the acoustic system

$$\begin{cases} \partial_t \tau_1 - \tau_1 \partial_x u_1 = 0 \\ \partial_t u_1 + \partial_x \dfrac{g}{\tau_1} + \partial_x \dfrac{g}{\tau_2} = -g\partial_x z \\ \partial_t \tau_2 - \tau_2 \partial_x u_2 = 0 \\ \partial_t u_2 + r\partial_x \dfrac{g}{\tau_1} + \partial_x \dfrac{g}{\tau_2} = -g\partial_x z, \end{cases} \tag{3.3.4}$$

we introduce the relaxation parameter λ and two new variables $\mathcal{C}_1, \mathcal{C}_2$ such that they satisfy

$$\lim_{\varepsilon \to 0} \mathcal{C}_1 = g\frac{1}{\tau_1} = g h_1 \quad \text{and} \quad \lim_{\varepsilon \to 0} \mathcal{C}_2 = g\frac{1}{\tau_2} = g h_2 \tag{3.3.5}$$

at least formally. Hence, we are able to define the following relaxation system

$$\begin{cases} \partial_t \tau_1 - \tau_1 \partial_x u_1 = 0 \\ \partial_t u_1 + \partial_x \mathcal{C}_1 + \partial_x \mathcal{C}_2 = -g\partial_x z \\ \partial_t \tau_2 - \tau_2 \partial_x u_2 = 0 \\ \partial_t u_2 + r\partial_x \mathcal{C}_1 + \partial_x \mathcal{C}_2 = -g\partial_x z. \\ \partial_t \mathcal{C}_1 + a_1^2 \partial_x u_1 = \dfrac{1}{\varepsilon}\left(g\dfrac{1}{\tau_1} - \mathcal{C}_1 \right) \\ \partial_t \mathcal{C}_2 + a_2^2 \partial_x u_2 = \dfrac{1}{\varepsilon}\left(g\dfrac{1}{\tau_2} - \mathcal{C}_2 \right). \end{cases} \tag{3.3.6}$$

where a_1, a_2 are constant parameters.

Proposition 1. *The relaxation system (3.3.6) is stable under the sub-characteristic condition* $a_j^2 \geq {c_j}^2 = gh_j$, $j = 1, 2$.

Proof. To prove it, we start by writing the following first-order correction for $\mathcal{C}_j$,

$$\mathcal{C}_j = g\frac{1}{\tau_j} + \varepsilon \mathcal{C}_j^{(1)} + \mathcal{O}(\varepsilon^2), \tag{3.3.7}$$

which we insert into the evolution equations for $\mathcal{C}_j$, obtaining

$$-\mathcal{C}_j^{(1)} + \mathcal{O}(\varepsilon) = g\partial_t \frac{1}{\tau_j} + a_j^2 \partial_x u_j = -g\frac{1}{\tau_j^2}\partial_t \tau_j + a_j^2 \partial_x u_j = (-g\frac{1}{\tau_j} + a_j^2)\partial_x u_j. \tag{3.3.8}$$

Afterwards, neglecting the source term, we consider relations (3.3.8) together with equations (3.3.6), which lead to the following system

$$\partial_t \mathbf{W} + \mathbf{E}(\mathbf{W})\partial_x \mathbf{W} = \varepsilon \partial_x \big(\mathbf{D}(\mathbf{W})\partial_x \mathbf{W}\big) + \mathcal{O}(\varepsilon^2)$$

with $\mathbf{W} = (\tau_1, u_1, \tau_2, u_2)^T$ and

$$\mathbf{D}(\mathbf{W}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -g\frac{1}{\tau_1} + a_1^2 & 0 & -g\frac{1}{\tau_2} + a_2^2 \\ 0 & 0 & 0 & 0 \\ 0 & r(-g\frac{1}{\tau_1} + a_1^2) & 0 & -g\frac{1}{\tau_2} + a_2^2 \end{pmatrix}.$$

Finally, we need to impose that $\mathbf{D}(\mathbf{W})$ is a diffusion matrix, namely that its eigenvalues are non-negative. From this condition, we obtain the following sub-characteristic conditions $a_j^2 \geq gh_j$. □

Moreover, since we assume $\mathcal{C}_j$ to be well-prepared in the sense that $\mathcal{C}_j(x, t = 0) = gh_j(x, t = 0)$, we neglect the source terms $\frac{1}{\lambda}(g\frac{1}{\tau_j} - \mathcal{C}_j)$ in the evolution equations for $\mathcal{C}_j$. Then, we rewrite system (3.3.6) in a more compact form as

$$\partial_t \mathbf{U} + \mathbf{A}(\mathbf{U})\partial_x \mathbf{U} = \tilde{\mathbf{S}}(\mathbf{U}, z)$$

where $\mathbf{U} = (\tau_1, u_1, \tau_2, u_2, \mathcal{C}_1, \mathcal{C}_2)^T$,

$$\mathbf{A}(\mathbf{U}) = \begin{pmatrix} 0 & -\tau_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -\tau_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & r & 1 \\ 0 & a_1^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_2^2 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \tilde{\mathbf{S}}(\mathbf{U}, z) = \begin{pmatrix} 0 \\ -g\partial_x z \\ 0 \\ -g\partial_x z \\ 0 \\ 0 \end{pmatrix}.$$

and then we find its characteristic polynomial

$$\lambda^2(\lambda^4 - (a_1^2 + a_2^2)\lambda^2 + (1 - r)a_1^2 a_2^2).$$

Its roots are given by $\lambda_0 = 0$,

$$\lambda_{ext}^{\pm} = \pm\sqrt{\frac{a_1^2 + a_2^2 + \sqrt{(a_1^2 + a_2^2)^2 - 4(1 - r)a_1^2 a_2^2}}{2}} \tag{3.3.9}$$

and

$$\lambda_{int}^{\pm} = \pm\sqrt{\frac{a_1^2 + a_2^2 - \sqrt{(a_1^2 + a_2^2)^2 - 4(1-r)a_1^2 a_2^2}}{2}},$$

where once again it can be easily proved that the eigenvalues are always real since we assume the water heights to be positive. Moreover, it is also clear that the eigenvalues are ordered a priori, namely

$$\lambda_{ext}^{-} < \lambda_{int}^{-} < \lambda_0 = 0 < \lambda_{int}^{+} < \lambda_{ext}^{+}.$$

Subsequently, we look for the right eigenvectors of form

$$R = (r_1, r_2, r_3, r_4, r_5, r_6)^t.$$

Then, associated with the zero eigenvalue λ_0, we easily obtain

$$\mathbf{R}_1^{\lambda_0} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{R}_2^{\lambda_0} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

otherwise we have

$$\mathbf{R}^{\lambda} = \begin{pmatrix} 1 \\ -\dfrac{\lambda}{\tau_1} \\ \beta \\ -\dfrac{\lambda}{\tau_2}\beta \\ -\dfrac{a_1^2}{\tau_1} \\ -\dfrac{a_2^2}{\tau_2}\beta \end{pmatrix}$$

for $\lambda = \lambda_{ext}^{\pm}, \lambda_{int}^{\pm}$ and where $\beta = \dfrac{\tau_2}{\tau_1}\dfrac{a_1^2}{a_2^2}\left(\dfrac{\lambda^2}{a_1^2} - 1\right)$. Finally, few computations give us the Riemann invariants. Across the zero-discontinuity we get

$$\mathrm{RI}_{u_j}^{\lambda_0} = u_j, \quad \mathrm{RI}_{\mathcal{C}_j}^{\lambda_0} = \mathcal{C}_j,$$

while for the other waves with speed $\lambda \neq 0$,

$$\mathrm{RI}_{1,j}^{\lambda} = \mathcal{C}_j - \frac{a_j^2}{\lambda}u_j, \quad \mathrm{RI}_{2,j}^{\lambda} = \mathcal{C}_j + a_j^2 \ln \tau_j, \quad \mathrm{RI}_3^{\lambda} = \mathcal{C}_2 + \left(1 - \frac{\lambda^2}{a_1^2}\right)\mathcal{C}_1.$$

3.3.1 Approximate Riemann solver for the acoustic system

Here, we look for the approximate solution of the Riemann problem associated with system (3.3.6) and the following initial conditions

$$\mathbf{U}(x, t=0) = \begin{cases} \mathbf{U}_L & \text{if} \quad x < 0 \\ \mathbf{U}_R & \text{if} \quad x > 0 \end{cases} \tag{3.3.10}$$

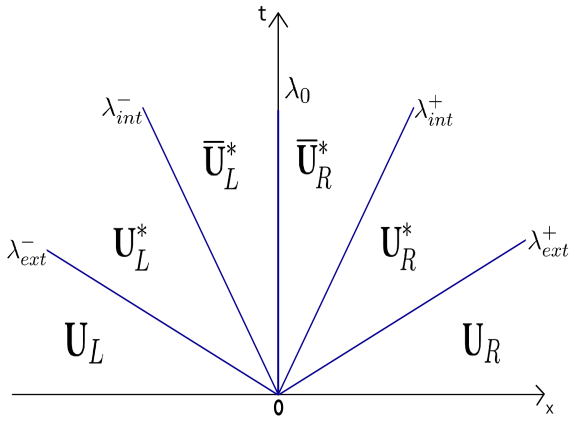


Figure 3.2: Sketch of the approximate solution for the Riemann problem.

where

$$\mathbf{U}_L = \begin{pmatrix} \tau_{1,L} \\ u_{1,L} \\ \tau_{2,L} \\ u_{2,L} \\ \mathcal{C}_{1,L} \\ \mathcal{C}_{2,L} \end{pmatrix} \quad \text{and} \quad \mathbf{U}_R = \begin{pmatrix} \tau_{1,R} \\ u_{1,R} \\ \tau_{2,R} \\ u_{2,R} \\ \mathcal{C}_{1,R} \\ \mathcal{C}_{2,R} \end{pmatrix}.$$

Since we have five discontinuities, the solution of the Riemann problem consists of six separated states, namely

$$\mathbf{U}(\frac{x}{t};\mathbf{U}_L,\mathbf{U}_R) = \begin{cases} \mathbf{U}_L & \text{if} \quad \frac{x}{t} < \lambda^-_{ext} \\ \mathbf{U}^*_L & \text{if} \quad \lambda^-_{ext} < \frac{x}{t} < \lambda^-_{int} \\ \overline{\mathbf{U}}^*_L & \text{if} \quad \lambda^-_{int} < \frac{x}{t} < \lambda_0 \\ \overline{\mathbf{U}}^*_R & \text{if} \quad \lambda_0 < \frac{x}{t} < \lambda^+_{int} \\ \mathbf{U}^*_R & \text{if} \quad \lambda^+_{int} < \frac{x}{t} < \lambda^+_{ext} \\ \mathbf{U}_R & \text{if} \quad \frac{x}{t} > \lambda^+_{ext} \end{cases} \tag{3.3.11}$$

where, in particular,

$$\mathbf{U}^*_L = \begin{pmatrix} \tau^*_{1,L} \\ u^*_{1,L} \\ \tau^*_{2,L} \\ u^*_{2,L} \\ \mathcal{C}^*_{1,L} \\ \mathcal{C}^*_{2,L} \end{pmatrix} \quad \overline{\mathbf{U}}^*_L = \begin{pmatrix} \overline{\tau}^*_{1,L} \\ \overline{u}^*_{1,L} \\ \overline{\tau}^*_{2,L} \\ \overline{u}^*_{2,L} \\ \overline{\mathcal{C}}^*_{1,L} \\ \overline{\mathcal{C}}^*_{2,L} \end{pmatrix} \quad \mathbf{U}^*_R = \begin{pmatrix} \tau^*_{1,R} \\ u^*_{1,R} \\ \tau^*_{2,R} \\ u^*_{2,R} \\ \mathcal{C}^*_{1,R} \\ \mathcal{C}^*_{2,R} \end{pmatrix} \quad \overline{\mathbf{U}}^*_R = \begin{pmatrix} \overline{\tau}^*_{1,R} \\ \overline{u}^*_{1,R} \\ \overline{\tau}^*_{2,R} \\ \overline{u}^*_{2,R} \\ \overline{\mathcal{C}}^*_{1,R} \\ \overline{\mathcal{C}}^*_{2,R} \end{pmatrix}, \tag{3.3.12}$$

refer also to figure 3.2. Due to the fact that a_j are constant, it is clear that characteristic fields associated with the eigenvalues are linearly degenerate and, thus, that all the waves are contact discontinuities. Therefore, in order to find the Riemann solution, we can exploit the Riemann invariants or equivalently the Rankine-Hugoniot jump conditions.

We highlight that the star values for u_j and $\mathcal{C}_j$ can be found without exploiting the ones for τ_j. Therefore, for the sake of conciseness and since in the numerical strategies we only need the star states for u_j and $\mathcal{C}_j$, here we do not include the definitions for the star values for τ_j but they can be found in an analogous way.

Starting with the jump conditions across the zero-discontinuity and including the topography in the solver, we impose

$$\begin{cases} \overline{u}^*_{j,L} = \overline{u}^*_{j,R} = \overline{u}^*_j \quad \text{with} \quad j = 1,2 \\ \overline{\mathcal{C}}^*_{1,R} - \overline{\mathcal{C}}^*_{1,L} + \overline{\mathcal{C}}^*_{2,R} - \overline{\mathcal{C}}^*_{2,L} + \mathcal{M} = 0 \\ r(\overline{\mathcal{C}}^*_{1,R} - \overline{\mathcal{C}}^*_{1,L}) + \overline{\mathcal{C}}^*_{2,R} - \overline{\mathcal{C}}^*_{2,L} + \mathcal{M} = 0 \end{cases}$$

where $\mathcal{M} = \Delta x\{g\partial_x z\}$ is a function that has to be defined but it should be zero when $z_L = z_R$. Thus, we simply ask for

$$\mathcal{M} = g(z_R - z_L).$$

As far as the other waves are concerned, from the Riemann invariants presented in the previous section, we get the following conditions associated with λ^-_{ext}

$$\begin{cases} \mathcal{C}_{j,L} - \dfrac{a_j^2}{\lambda^-_{ext}} u_{j,L} = \mathcal{C}^*_{j,L} - \dfrac{a_j^2}{\lambda^-_{ext}} u^*_{j,L} \quad \text{with} \quad j = 1,2 \\ \mathcal{C}^*_{2,L} + (1 - \dfrac{(\lambda^-_{ext})^2}{a_1^2})\mathcal{C}^*_{1,L} = \mathcal{C}_{2,L} + (1 - \dfrac{(\lambda^-_{ext})^2}{a_1^2})\mathcal{C}_{1,L}, \end{cases}$$

associated with λ^-_{int}

$$\begin{cases} \mathcal{C}^*_{j,L} - \dfrac{a_j^2}{\lambda^-_{int}} u^*_{j,L} = \overline{\mathcal{C}}^*_{j,L} - \dfrac{a_j^2}{\lambda^-_{int}} \overline{u}^*_j \quad \text{with} \quad j = 1,2 \\ \mathcal{C}^*_{2,L} + (1 - \dfrac{(\lambda^-_{int})^2}{a_1^2})\mathcal{C}^*_{1,L} = \overline{\mathcal{C}}^*_{2,L} + (1 - \dfrac{(\lambda^-_{int})^2}{a_1^2})\overline{\mathcal{C}}^*_1, \end{cases}$$

associated with λ^+_{int}

$$\begin{cases} \mathcal{C}^*_{j,R} - \dfrac{a_j^2}{\lambda^+_{int}} u^*_{j,R} = \overline{\mathcal{C}}^*_{j,R} - \dfrac{a_j^2}{\lambda^+_{int}} \overline{u}^*_j \quad \text{with} \quad j = 1,2 \\ \mathcal{C}^*_{2,R} + (1 - \dfrac{(\lambda^+_{int})^2}{a_1^2})\mathcal{C}^*_{1,R} = \overline{\mathcal{C}}^*_{2,R} + (1 - \dfrac{(\lambda^+_{int})^2}{a_1^2})\overline{\mathcal{C}}^*_1, \end{cases}$$

and finally associated with λ^+_{ext}

$$\begin{cases} \mathcal{C}^*_{j,R} - \dfrac{a_j^2}{\lambda^+_{ext}} u^*_{j,R} = \mathcal{C}_{j,R} - \dfrac{a_j^2}{\lambda^+_{ext}} u_{j,R} \quad \text{with} \quad j = 1,2 \\ \mathcal{C}^*_{2,R} + (1 - \dfrac{(\lambda^+_{ext})^2}{a_1^2})\mathcal{C}^*_{1,R} = \mathcal{C}_{2,R} + (1 - \dfrac{(\lambda^+_{ext})^2}{a_1^2})\mathcal{C}_{1,R}. \end{cases}$$

Since we had 16 unknowns and we found 16 relations, we are able to solve the resulting system.

Thus, after some computations, we can explicitly write the star states as follows,

$$\begin{cases}
\mathcal{C}^*_{1,R} = \mathcal{C}_{1,R} - \dfrac{a_1^2-(\lambda^+_{int})^2}{2((\lambda^+_{ext})^2-(\lambda^+_{int})^2)}\Big(\mathcal{C}_{1,R}-\mathcal{C}_{1,L}\Big) - \dfrac{a_1^2}{2((\lambda^+_{ext})^2-(\lambda^+_{int})^2)}\Big(\mathcal{C}_{2,R}-\mathcal{C}_{2,L}+\mathcal{M}+\mathcal{U}\Big)\\
\mathcal{C}^*_{1,L} = \mathcal{C}_{1,L} + \dfrac{a_1^2-(\lambda^+_{int})^2}{2((\lambda^+_{ext})^2-(\lambda^+_{int})^2)}\Big(\mathcal{C}_{1,R}-\mathcal{C}_{1,L}\Big) + \dfrac{a_1^2}{2((\lambda^+_{ext})^2-(\lambda^+_{int})^2)}\Big(\mathcal{C}_{2,R}-\mathcal{C}_{2,L}+\mathcal{M}-\mathcal{U}\Big)\\
\overline{\mathcal{C}}^*_1 = \dfrac{\mathcal{C}_{1,R}+\mathcal{C}_{1,L}}{2} - \dfrac{a_1^2}{2\lambda^+_{int}}(u_{1,R}-u_{1,L}) + \dfrac{a_1^2}{2\lambda^+_{int}}\dfrac{\mathcal{U}}{\lambda^+_{ext}+\lambda^+_{int}}\\
\mathcal{C}^*_{2,R} = \mathcal{C}_{2,R} + \dfrac{a_1^2-(\lambda^+_{ext})^2}{2a_1^2((\lambda^+_{ext})^2-(\lambda^+_{int})^2)}\Big((a_1^2-(\lambda^+_{int})^2)(\mathcal{C}_{1,R}-\mathcal{C}_{1,L}) + a_1^2(\mathcal{C}_{2,R}-\mathcal{C}_{2,L}+\mathcal{M}+\mathcal{U})\Big)\\
\mathcal{C}^*_{2,L} = \mathcal{C}_{2,L} - \dfrac{a_1^2-(\lambda^+_{ext})^2}{2a_1^2((\lambda^+_{ext})^2-(\lambda^+_{int})^2)}\Big((a_1^2-(\lambda^+_{int})^2)(\mathcal{C}_{1,R}-\mathcal{C}_{1,L}) + a_1^2(\mathcal{C}_{2,R}-\mathcal{C}_{2,L}+\mathcal{M}-\mathcal{U})\Big)\\
\overline{\mathcal{C}}^*_2 = \dfrac{\mathcal{C}_{2,R}+\mathcal{C}_{2,L}}{2} - \dfrac{a_2^2}{2\lambda^+_{int}}(u_{2,R}-u_{2,L}) - \dfrac{1}{2\lambda^+_{int}}\dfrac{(a_1^2-(\lambda^+_{ext})^2)\mathcal{U}}{\lambda^+_{ext}+\lambda^+_{int}}\\
\overline{\mathcal{C}}^*_{2,R} = \overline{\mathcal{C}}^*_2 - \dfrac{\mathcal{M}}{2}\\
\overline{\mathcal{C}}^*_{2,L} = \overline{\mathcal{C}}^*_2 + \dfrac{\mathcal{M}}{2}\\
u^*_{1,R} = u_{1,R} - \dfrac{\lambda^+_{ext}}{2((\lambda^+_{ext})^2-(\lambda^+_{int})^2)}(\mathcal{K}^+_2 + \mathcal{M} + (1-\dfrac{(\lambda^+_{int})^2}{a_1^2})\mathcal{K}^+_1)\\
u^*_{1,L} = u_{1,L} - \dfrac{\lambda^+_{ext}}{2((\lambda^+_{ext})^2-(\lambda^+_{int})^2)}(\mathcal{K}^-_2 + \mathcal{M} + (1-\dfrac{(\lambda^+_{int})^2}{a_1^2})\mathcal{K}^-_1)\\
\overline{u}^*_1 = \dfrac{u_{1,R}+u_{1,L}}{2} - \dfrac{\lambda^+_{int}}{2a_1^2}(\mathcal{C}_{1,R}-\mathcal{C}_{1,L}) - \dfrac{\mathcal{P}+\mathcal{M}}{2(\lambda^+_{ext}+\lambda^+_{int})}\\
u^*_{2,R} = u_{2,R} + \dfrac{\lambda^+_{ext}}{2((\lambda^+_{ext})^2-(\lambda^+_{int})^2)}\dfrac{a_1^2-(\lambda^+_{ext})^2}{a_2^2}(\mathcal{K}^+_2 + \mathcal{M} + (1-\dfrac{(\lambda^+_{int})^2}{a_1^2})\mathcal{K}^+_1)\\
u^*_{2,L} = u_{2,L} + \dfrac{\lambda^+_{ext}}{2((\lambda^+_{ext})^2-(\lambda^+_{int})^2)}\dfrac{a_1^2-(\lambda^+_{ext})^2}{a_2^2}(\mathcal{K}^-_2 + \mathcal{M} + (1-\dfrac{(\lambda^+_{int})^2}{a_1^2})\mathcal{K}^-_1)\\
\overline{u}^*_2 = \dfrac{u_{2,R}+u_{2,L}}{2} - \dfrac{\lambda^+_{int}}{2a_2^2}(\mathcal{C}_{2,R}-\mathcal{C}_{2,L}+\mathcal{M}) + \dfrac{a_1^2}{a_2^2}(1-\dfrac{(\lambda^+_{ext})^2}{a_1^2})\dfrac{\mathcal{P}+\mathcal{M}}{2(\lambda^+_{ext}+\lambda^+_{int})}
\end{cases} \tag{3.3.13}$$

where the following quantities were introduced to lighten the formulas

$$\mathcal{U} = \frac{a_2^2}{\lambda^+_{ext}}(u_{2,R}-u_{2,L}) + \frac{a_1^2-(\lambda^+_{int})^2}{\lambda^+_{ext}}(u_{1,R}-u_{1,L}), \quad \mathcal{P} = \mathcal{C}_{2,R}-\mathcal{C}_{2,L}+(1-\frac{(\lambda^+_{int})^2}{a_1^2})(\mathcal{C}_{1,R}-\mathcal{C}_{1,L}),$$

and

$$\mathcal{K}^\pm_j = \mathcal{C}_{j,R} - \mathcal{C}_{j,L} \pm \frac{a_j^2}{\lambda^+_{ext}}(u_{j,R}-u_{j,L}).$$

3.4 Numerical approximation

Before describing the numerical scheme, it is necessary to give some details about the space and time discretizations. We start by defining the constant space step Δx and the time step Δt. Then, the mesh interfaces are given by $x_{i+1/2} = i\Delta x$ for $i \in \mathbb{Z}$ and the intermediate times by $t^n = n\Delta t$ for $n \in \mathbb{N}$. As far as the variable ξ is concerned, we use the same space discretization of x, hence $\Delta\xi = \Delta x$, $\xi_{i+1/2} = x_{i+1/2}$ and $\xi_i = x_i\ \forall i$.

Concerning the numerical strategy, since we have two different steps (Lagrangian and projection) or systems (acoustic and transport), the numerical method is composed of two stages as well:

1. Update $\mathbf{Q}^n$ to $\mathbf{Q}^{n+1-}$ by solving the acoustic or Lagrangian system;

2. Exploit $\mathbf{Q}^{n+1-}$ to solve either the transport system or the projection step and find $\mathbf{Q}^{n+1}$.

Finally, we highlight that we will present both an explicit and implicit approximation of the first step. Clearly, depending on which formulation we use, also the Courant–Friedrichs–Lewy (CFL) condition [28] on the time step changes. Indeed, in general for the acoustic step we ask for

$$\Delta t \leq \frac{\Delta x}{2} \frac{1}{\max\limits_i \{\lambda^+_{ext,i+1/2}\}} \tag{3.4.1}$$

with λ^+_{ext} given by (3.3.9), while in the case of the transport stage we impose

$$\Delta t \leq \frac{\Delta x}{2} \frac{1}{\max\limits_i \{(\overline{u}^*_{j,i-1/2})^+ - (\overline{u}^*_{j,i+1/2})^-\}} \quad \text{for} \quad j = 1, 2 \tag{3.4.2}$$

with $(\overline{u}^*_{j,i-1/2})^+ = \max(\overline{u}^*_{j,i-1/2}, 0)$ and $(\overline{u}^*_{j,i+1/2})^- = \min(\overline{u}^*_{j,i+1/2}, 0)$. Then, the final time step should be taken as the minimum between the two. However, if we use an implicit approximation for the Lagrangian step, we could neglect condition (3.4.1) and exploit only the transport one (3.4.2).

3.4.1 Explicit approximation of the acoustic-Lagrangian system

Considering the Godunov method associated with the approximate Riemann solver of the previous section 3.3.1, the updating formula is given by

$$\begin{aligned} \mathbf{U}_i^{n+1-} =& \mathbf{U}_i^n - \frac{\Delta t}{\Delta x}\Big(\lambda^{+,n}_{ext,i-1/2}(\mathbf{U}_i^n - \mathbf{U}^*_{R,i-1/2}) + \lambda^{+,n}_{int,i-1/2}(\mathbf{U}^*_{R,i-1/2} - \overline{\mathbf{U}}^*_{R,i-1/2}) + \\ &+ \lambda^{+,n}_{ext,i+1/2}(\mathbf{U}_i^n - \mathbf{U}^*_{L,i+1/2}) + \lambda^{+,n}_{int,i+1/2}(\mathbf{U}^*_{L,i+1/2} - \overline{\mathbf{U}}^*_{L,i+1/2})\Big) \end{aligned} \tag{3.4.3}$$

which is simply given by a juxtaposition of the approximate solutions of the Riemann problems locally defined at each interface, refer for instance to [28].

As far as the acoustic system is concerned, to update the variables h_j, hu_j, we could use the Godunov method, namely (3.4.3). In practice, since we would like to use the Lagrangian variables, we do the following. Indeed, we find u_j^{n+1-} using formula (3.4.3), then for the water heights we simply exploit

$$Lh^{n+1-}_{j,i} = Lh^n_{j,i} = h^n_{j,i}$$

and finally we state

$$Lh^{n+1-}_{j,i} u^{n+1-}_{j,i} = Lh^n_{j,i} u^{n+1-}_{j,i} = h^n_{j,i} u^{n+1-}_{j,i}.$$

Thus, we do not actually use the Godunov updating formula for the evolution equations for τ_j^{n+1-}, namely $\partial_t \tau_j - \tau_j \partial_x u_j = 0$, but only for the equations for the velocities, which are written in conservative form. Indeed, we could also observe that their updating formula can be reformulated using the numerical fluxes. Hence, together with the evolution equations for $\mathcal{C}_j$, we state

$$\begin{cases} u^{n+1-}_{1,i} = u^n_{1,i} - \dfrac{\Delta t}{\Delta x}(\overline{\mathcal{C}}^{*,n}_{1,i+1/2} - \overline{\mathcal{C}}^{*,n}_{1,i-1/2} + \overline{\mathcal{C}}^{*,n}_{2,i+1/2} - \overline{\mathcal{C}}^{*,n}_{2,i-1/2}) + \Delta t \dfrac{S_{i+1/2} + S_{i-1/2}}{2} \\ u^{n+1-}_{2,i} = u^n_{2,i} - \dfrac{\Delta t}{\Delta x}(r(\overline{\mathcal{C}}^{*,n}_{1,i+1/2} - \overline{\mathcal{C}}^{*,n}_{1,i-1/2}) + \overline{\mathcal{C}}^{*,n}_{2,i+1/2} - \overline{\mathcal{C}}^{*,n}_{2,i-1/2}) + \Delta t \dfrac{S_{i+1/2} + S_{i-1/2}}{2} \\ \mathcal{C}^{n+1-}_{1,i} = \mathcal{C}^n_{1,i} - \dfrac{\Delta t}{\Delta x}(a^2_{1,i+1/2}\overline{u}^{*,n}_{1,i+1/2} - a^2_{1,i-1/2}\overline{u}^{*,n}_{1,i-1/2}) + u^n_{1,i}\dfrac{\Delta t}{\Delta x}(a^2_{1,i+1/2} - a^2_{1,i-1/2}) \\ \mathcal{C}^{n+1-}_{2,i} = \mathcal{C}^n_{2,i} - \dfrac{\Delta t}{\Delta x}(a^2_{2,i+1/2}\overline{u}^{*,n}_{2,i+1/2} - a^2_{2,i-1/2}\overline{u}^{*,n}_{2,i-1/2}) + u^n_{2,i}\dfrac{\Delta t}{\Delta x}(a^2_{2,i+1/2} - a^2_{2,i-1/2}) \end{cases} \tag{3.4.4}$$

where $S_{i+1/2} = -g\dfrac{z_{i+1} - z_i}{\Delta x}$. Then, it is straightforward to write the Lagrangian system approximation,

$$\begin{cases} (Lh)_{1,i}^{n+1-} = h_{1,i}^n \\ (Lhu)_{1,i}^{n+1-} = (hu)_{1,i}^n - h_{1,i}^n \dfrac{\Delta t}{\Delta x}(\overline{\mathcal{C}}_{1,i+1/2}^{*,n} - \overline{\mathcal{C}}_{1,i-1/2}^{*,n} + \overline{\mathcal{C}}_{2,i+1/2}^{*,n} - \overline{\mathcal{C}}_{2,i-1/2}^{*,n}) + \\ \qquad\qquad + h_{1,i}^n \Delta t \dfrac{S_{i+1/2} + S_{i-1/2}}{2} \\ (Lh)_{2,i}^{n+1-} = h_{2,i}^n \\ (Lhu)_{2,i}^{n+1-} = (hu)_{2,i}^n - h_{2,i}^n \dfrac{\Delta t}{\Delta x}(r(\overline{\mathcal{C}}_{1,i+1/2}^{*,n} - \overline{\mathcal{C}}_{1,i-1/2}^{*,n}) + \overline{\mathcal{C}}_{2,i+1/2}^{*,n} - \overline{\mathcal{C}}_{2,i-1/2}^{*,n}) + \\ \qquad\qquad + h_{2,i}^n \Delta t \dfrac{S_{i+1/2} + S_{i-1/2}}{2} \end{cases} \tag{3.4.5}$$

Let us observe that this discretization (3.4.5) is indeed consistent with system (3.2.4) as $\dfrac{\overline{\mathcal{C}}_{j,i+1/2}^{*,n} - \overline{\mathcal{C}}_{j,i-1/2}^{*,n}}{\Delta \xi}$ approximates $\partial_\xi \mathcal{C}_j$ and therefore $\partial_\xi(gh_j)$. From system (3.2.2), we remark that a natural explicit discretization for $L_j(\xi, t)$ is as follows,

$$L_{j,i}^{n+1-} = L_{j,i}^n + \frac{\Delta t}{\Delta x}(\overline{u}_{j,i+\frac{1}{2}}^{*,n} - \overline{u}_{j,i-\frac{1}{2}}^{*,n}) \quad \text{with} \quad L_{j,i}^n = 1. \tag{3.4.6}$$

Thus, it is clear that we can do the following approximation: $\dfrac{L_{1,i}^n}{L_{2,i}^n} = 1$.

Lastly, we can conclude observing that this numerical approximation of the acoustic (Lagrangian) system can be interpreted as a path-conservative numerical method. Indeed, the choice of the path is naturally driven by the presence of the linearly degenerate characteristic fields associated with the eigenvalues of the relaxation system (3.3.6). While, for the topography z, the segment path is considered in order to ensure the well-balanced property.

3.4.2 Implicit approximation of the acoustic-Lagrangian system

In order to have an implicit discretization for the Lagrangian step, the idea is to simply exploit the star values of the approximate Riemann solver computed at time t^{n+1-} instead of t^n. As a consequence, we would get a linear system which could be numerically solved, allowing us to obtain $u_j^{n+1-}, \mathcal{C}_j^{n+1-}$. We highlight that at this stage we do not need the values τ_j^{n+1-}, whose approximation could increase the complexity of the numerical method, giving a non-linear system and increasing the computational cost. Thus, the implicit approximation of the acoustic system reads

$$\begin{cases} u_{1,i}^{n+1-} = u_{1,i}^n - \dfrac{\Delta t}{\Delta x}(\overline{\mathcal{C}}_{1,i+1/2}^{*,n+1-} - \overline{\mathcal{C}}_{1,i-1/2}^{*,n+1-} + \overline{\mathcal{C}}_{2,i+1/2}^{*,n+1-} - \overline{\mathcal{C}}_{2,i-1/2}^{*,n+1-}) + \Delta t \dfrac{S_{i+1/2} + S_{i-1/2}}{2} \\ u_{2,i}^{n+1-} = u_{2,i}^n - \dfrac{\Delta t}{\Delta x}(r(\overline{\mathcal{C}}_{1,i+1/2}^{*,n+1-} - \overline{\mathcal{C}}_{1,i-1/2}^{*,n+1-}) + \overline{\mathcal{C}}_{2,i+1/2}^{*,n+1-} - \overline{\mathcal{C}}_{2,i-1/2}^{*,n+1-}) + \Delta t \dfrac{S_{i+1/2} + S_{i-1/2}}{2} \\ \mathcal{C}_{1,i}^{n+1-} = \mathcal{C}_{1,i}^n - \dfrac{\Delta t}{\Delta x}(a_{1,i+1/2}^2 \overline{u}_{1,i+1/2}^{*,n+1-} - a_{1,i-1/2}^2 \overline{u}_{1,i-1/2}^{*,n+1-}) + u_{1,i}^{n+1-} \dfrac{\Delta t}{\Delta x}(a_{1,i+1/2}^2 - a_{1,i-1/2}^2) \\ \mathcal{C}_{2,i}^{n+1-} = \mathcal{C}_{2,i}^n - \dfrac{\Delta t}{\Delta x}(a_{2,i+1/2}^2 \overline{u}_{2,i+1/2}^{*,n+1-} - a_{2,i-1/2}^2 \overline{u}_{2,i-1/2}^{*,n+1-}) + u_{2,i}^{n+1-} \dfrac{\Delta t}{\Delta x}(a_{2,i+1/2}^2 - a_{2,i-1/2}^2) \end{cases} \tag{3.4.7}$$

System (3.4.7) can be reformulated as a linear system of form

$$A\overrightarrow{Y} = \overrightarrow{b} + \overrightarrow{f} \tag{3.4.8}$$

where, in our case,

$$\overrightarrow{Y} = \left(u_{1,1}^{n+1-}, u_{2,1}^{n+1-}, \mathcal{C}_{1,1}^{n+1-}, \mathcal{C}_{2,1}^{n+1-}, \dots, u_{1,i}^{n+1-}, u_{2,i}^{n+1-}, \mathcal{C}_{1,i}^{n+1-}, \mathcal{C}_{2,i}^{n+1-}, \dots, u_{1,M}^{n+1-}, u_{2,M}^{n+1-}, \mathcal{C}_{1,M}^{n+1-}, \mathcal{C}_{2,M}^{n+1-}\right)^T,$$

$$\overrightarrow{b} = \left(u_{1,1}^{n}, u_{2,1}^{n}, \mathcal{C}_{1,1}^{n}, \mathcal{C}_{2,1}^{n}, \dots, u_{1,i}^{n}, u_{2,i}^{n}, \mathcal{C}_{1,i}^{n}, \mathcal{C}_{2,i}^{n}, \dots, u_{1,M}^{n}, u_{2,M}^{n}, \mathcal{C}_{1,M}^{n}, \mathcal{C}_{2,M}^{n}\right)^T,$$

$$\overrightarrow{f} = \begin{pmatrix} \vdots \\ \vdots \\ \Delta t \dfrac{S_{i+1/2} + S_{i-1/2}}{2} \\ \Delta t \dfrac{S_{i+1/2} + S_{i-1/2}}{2} \\ -\dfrac{\Delta t}{2}\left(\dfrac{a_{1,i+1/2}^2 S_{i+1/2}}{\lambda_{ext,i+1/2}^{+,n} + \lambda_{int,i+1/2}^{+,n}} - \dfrac{a_{1,i-1/2}^2 S_{i-1/2}}{\lambda_{ext,i-1/2}^{+,n} + \lambda_{int,i-1/2}^{+,n}}\right) \\ \dfrac{\Delta t}{2}\left(\left(\dfrac{a_1^2 - (\lambda_{ext}^{+,n})^2}{\lambda_{ext}^{+,n} + \lambda_{int}^{+,n}} S\right)_{i+1/2} - \left(\dfrac{a_1^2 - (\lambda_{ext}^{+,n})^2}{\lambda_{ext}^{+,n} + \lambda_{int}^{+,n}} S\right)_{i-1/2} - \left((\lambda_{int}^{+,n} S)_{i+1/2} - (\lambda_{int}^{+,n} S)_{i-1/2}\right)\right) \\ \vdots \\ \vdots \end{pmatrix}$$

and the matrix A defined consequently, see also appendix 3.A for more details. Thus, in each line of our matrix A, we only have 12 entries which could be different from zero. Therefore, even if the matrix dimension is large ($4\mathrm{M} \times 4\mathrm{M}$, M number of cells), since the matrix is sparse, the computational cost is not too high to solve the linear system.

Then, the variables Lh_j and Lhu_j would be updated as explained in the previous section 3.4.1, getting the following implicit approximation

$$\begin{cases} (Lh)_{1,i}^{n+1-} = h_{1,i}^n \\ (Lhu)_{1,i}^{n+1-} = (hu)_{1,i}^n - h_{1,i}^n \dfrac{\Delta t}{\Delta x}(\overline{\mathcal{C}}_{1,i+1/2}^{*,n+1-} - \overline{\mathcal{C}}_{1,i-1/2}^{*,n+1-}) - h_{1,i}^n \dfrac{L_{1,i}^n}{L_{2,i}^n} \dfrac{\Delta t}{\Delta x}(\overline{\mathcal{C}}_{2,i+1/2}^{*,n+1-} - \overline{\mathcal{C}}_{2,i-1/2}^{*,n+1-}) + \\ \qquad + h_{1,i}^n \Delta t \dfrac{S_{i+1/2} + S_{i-1/2}}{2} \\ (Lh)_{2,i}^{n+1-} = h_{2,i}^n \\ (Lhu)_{2,i}^{n+1-} = (hu)_{2,i}^n - h_{2,i}^n \dfrac{\Delta t}{\Delta x}(\overline{\mathcal{C}}_{2,i+1/2}^{*,n+1-} - \overline{\mathcal{C}}_{2,i-1/2}^{*,n+1-}) - r h_{2,i}^n \dfrac{L_{2,i}^n}{L_{1,i}^n} \dfrac{\Delta t}{\Delta x}(\overline{\mathcal{C}}_{1,i+1/2}^{*,n+1-} - \overline{\mathcal{C}}_{1,i-1/2}^{*,n+1-}) \\ \qquad + h_{2,i}^n \Delta t \dfrac{S_{i+1/2} + S_{i-1/2}}{2} \end{cases} \tag{3.4.9}$$

Let us observe that the approximation of the ratio $\dfrac{L_1}{L_2}$ in the evolution equations for $(Lhu)_1, (Lhu)_2$ is kept explicit in order to have a simpler discretization, namely $\dfrac{L_{1,i}^n}{L_{2,i}^n} = 1$.

3.4.3 Transport-projection step

Let us finally see how to discretize the transport system or projection step. The two approximations will be very similar, still different.

Regarding the former, since we have two sets of two equations of form $\partial_t X_j + u_j \partial_x X_j = 0$ with $X = h_j, h_j u_j$ and $j = 1, 2$, we observe that

$$\partial_t X_j + u_j \partial_x X_j = \partial_t X_j - X_j \partial_x u_j + \partial_x (Xu)_j = 0,$$

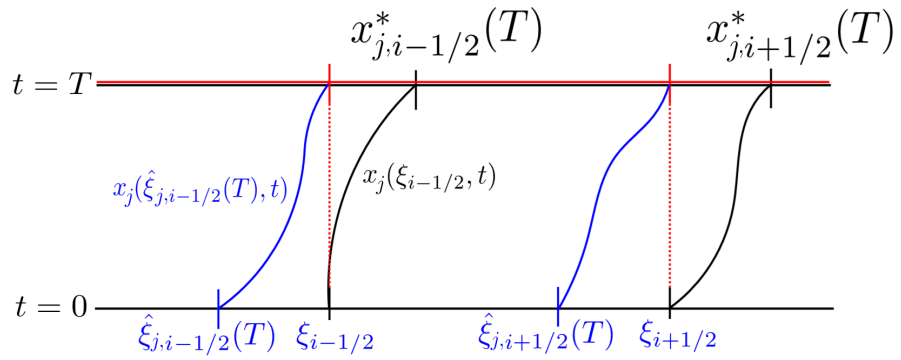


Figure 3.3: Sketch of the connection between Lagrangian and Eulerian coordinates.

thus we simply exploit the following explicit approximation

$$X^{n+1}_{j,i} = X^{n+1-}_{j,i}(1 + \frac{\Delta t}{\Delta x}(\overline{u}^*_{j,i+1/2} - \overline{u}^*_{j,i-1/2})) - \frac{\Delta t}{\Delta x}\big(\overline{u}^*_{j,i+1/2}X^{n+1-}_{j,i+\frac{1}{2}} - \overline{u}^*_{j,i-1/2}X^{n+1-}_{j,i-\frac{1}{2}}\big).$$

In particular, the latter discretization is equivalent to

$$X^{n+1}_{j,i} = (LX)^{n+1-}_{j,i} - \frac{\Delta t}{\Delta x}\big(\overline{u}^*_{j,i+1/2}X^{n+1-}_{j,i+\frac{1}{2}} - \overline{u}^*_{j,i-1/2}X^{n+1-}_{j,i-\frac{1}{2}}\big) \tag{3.4.10}$$

where we used an explicit upwind discretization for $X^{n+1-}_{j,i+1/2}$, namely

$$X^{n+1-}_{j,i+1/2} = \begin{cases} X^{n+1-}_{j,i} & \text{if} \quad \overline{u}^*_{j,i+1/2} \geq 0 \\ X^{n+1-}_{j,i+1} & \text{if} \quad \overline{u}^*_{j,i+1/2} < 0. \end{cases}$$

Next, let us move on to the projection step discretization. In order to be able to explain it, it is convenient to give few details about the link between Eulerian and Lagrangian coordinates, for which we also refer to figure 3.3. Pointing out that we use the index j for the layer and i for the cell of the mesh, we define $\hat{\xi}_{j,i+\frac{1}{2}}(t)$ such that $\forall i$

$$x_j(\hat{\xi}_{j,i+\frac{1}{2}}(T), T) = x_{i+\frac{1}{2}}, \quad \text{with} \quad T \geq 0,$$

where the corresponding trajectories are given by

$$\begin{cases} \dfrac{\partial x_j}{\partial t}(\hat{\xi}_{j,i+\frac{1}{2}}(T), t) = u_j(x_j(\hat{\xi}_{j,i+\frac{1}{2}}(T), t), t) \\ x_j(\hat{\xi}_{j,i+\frac{1}{2}}(T), 0) = \hat{\xi}_{j,i+\frac{1}{2}}(T). \end{cases}$$

Therefore, it is easy to obtain the following approximation $x_{i+\frac{1}{2}} = \hat{\xi}_{j,i+\frac{1}{2}} + \Delta t u^*_{j,i+\frac{1}{2}}$. Similarly, we also find $x^{*,n+1-}_{j,i+\frac{1}{2}} = x_{i+\frac{1}{2}} + \Delta t u^*_{j,i+\frac{1}{2}}$ and thus $x_{i+\frac{1}{2}} - x^*_{j,i+\frac{1}{2}} = \hat{\xi}_{j,i+\frac{1}{2}} - \xi_{i+\frac{1}{2}}$. Moving to the integrals of the variables, we can change coordinates as follows,

$$\int_{x_j(\xi_{j,l},t)}^{x_j(\xi_{j,r},t)} X_j(x,t)dx = \int_{\xi_{j,l}}^{\xi_{j,r}} L_j(\xi,t)\overline{X}_j(\xi,t)d\xi.$$

leading to

$$X_{j,i}(t) = \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} X_j(x,t)dx = \frac{1}{\Delta x}\int_{x_j(\hat{\xi}_{j,i-\frac{1}{2}},t)}^{x_j(\hat{\xi}_{j,i+\frac{1}{2}},t)} X_j(x,t)dx = \frac{1}{\Delta x}\int_{\hat{\xi}_{j,i-\frac{1}{2}}}^{\hat{\xi}_{j,i+\frac{1}{2}}} L_j(\xi,t)\overline{X}_j(\xi,t)d\xi. \tag{3.4.11}$$

Therefore, splitting the integrals in three parts

$$X_{j,i}^{n+1} = \frac{1}{\Delta x}\int_{\hat{\xi}_{j,i-\frac{1}{2}}}^{\xi_{i-\frac{1}{2}}} L_j(\xi, t^{n+1-})\overline{X}_j(\xi, t^{n+1-})d\xi + \\ + \frac{1}{\Delta x}\int_{\xi_{i-\frac{1}{2}}}^{\xi_{i+\frac{1}{2}}} L_j(\xi, t^{n+1-})\overline{X}_j(\xi, t^{n+1-})d\xi + \frac{1}{\Delta x}\int_{\xi_{i+\frac{1}{2}}}^{\hat{\xi}_{j,i+\frac{1}{2}}} L_j(\xi, t^{n+1-})\overline{X}_j(\xi, t^{n+1-})d\xi. \tag{3.4.12}$$

and approximating them, we obtain

$$X_{j,i}^{n+1} = (LX)_{j,i}^{n+1-} - \frac{\Delta t}{\Delta x}\big(\overline{u}_{j,i+1/2}^{*}(LX)_{j,i+\frac{1}{2}}^{n+1-} - \overline{u}_{j,i-1/2}^{*}(LX)_{j,i-\frac{1}{2}}^{n+1-}\big) \tag{3.4.13}$$

where

$$(LX)_{j,i+1/2}^{n+1-} = \begin{cases} (LX)_{j,i}^{n+1-} & \text{if} \quad \overline{u}_{j,i+1/2}^{*} \geq 0 \\ (LX)_{j,i+1}^{n+1-} & \text{if} \quad \overline{u}_{j,i+1/2}^{*} < 0 \end{cases}$$

$\forall i$. For a similar procedure applied to the shallow water system or the blood flow equations, see [26] and [22] respectively. Hence, the only difference between formulation (3.4.10) and (3.4.13) is related to the use of the variable X or LX in the definition of the numerical fluxes. Observe that, in the numerical simulations, we always use formulation (3.4.13).

3.4.4 Properties of the numerical scheme

Considering the explicit and implicit Lagrangian approximations, (3.4.5) and (3.4.9), and the projection formulation (3.4.13), we can find an overall approximation for the two-layer shallow water system (3.1.1),

$$\begin{cases} h_{1,i}^{n+1} = h_{1,i}^{n} - \dfrac{\Delta t}{\Delta x}\big(\overline{u}_{1,i+1/2}^{*,\#}(Lh)_{1,i+\frac{1}{2}}^{n+1-} - \overline{u}_{1,i-1/2}^{*,\#}(Lh)_{1,i-\frac{1}{2}}^{n+1-}\big) \\ (hu)_{1,i}^{n+1} = (hu)_{1,i}^{n} - h_{1,i}^{n}\dfrac{\Delta t}{\Delta x}(\overline{\mathcal{C}}_{1,i+1/2}^{*,\#} - \overline{\mathcal{C}}_{1,i-1/2}^{*,\#} + \overline{\mathcal{C}}_{2,i+1/2}^{*,\#} - \overline{\mathcal{C}}_{2,i-1/2}^{*,\#}) + \\ \qquad\qquad - \dfrac{\Delta t}{\Delta x}\big(\overline{u}_{1,i+1/2}^{*,\#}(Lh)_{1,i+\frac{1}{2}}^{n+1-} - \overline{u}_{1,i-1/2}^{*,\#}(Lh)_{1,i-\frac{1}{2}}^{n+1-}\big) + h_{1,i}^{n}\Delta t\dfrac{S_{i+1/2}+S_{i-1/2}}{2} \\ h_{2,i}^{n+1} = h_{2,i}^{n} - \dfrac{\Delta t}{\Delta x}\big(\overline{u}_{2,i+1/2}^{*,\#}(Lh)_{2,i+\frac{1}{2}}^{n+1-} - \overline{u}_{2,i-1/2}^{*,\#}(Lh)_{2,i-\frac{1}{2}}^{n+1-}\big) \\ (hu)_{2,i}^{n+1} = (hu)_{2,i}^{n} - h_{2,i}^{n}\dfrac{\Delta t}{\Delta x}(r(\overline{\mathcal{C}}_{1,i+1/2}^{*,\#} - \overline{\mathcal{C}}_{1,i-1/2}^{*,\#}) + \overline{\mathcal{C}}_{2,i+1/2}^{*,\#} - \overline{\mathcal{C}}_{2,i-1/2}^{*,\#}). \\ \qquad\qquad - \dfrac{\Delta t}{\Delta x}\big(\overline{u}_{2,i+1/2}^{*,\#}(Lh)_{2,i+\frac{1}{2}}^{n+1-} - \overline{u}_{2,i-1/2}^{*,\#}(Lh)_{2,i-\frac{1}{2}}^{n+1-}\big) + h_{2,i}^{n}\Delta t\dfrac{S_{i+1/2}+S_{i-1/2}}{2}. \end{cases} \tag{3.4.14}$$

with either $\# = n$ or $\# = n+1-$ depending on the explicit and implicit approximation.

Remark 3. *The numerical approximation (3.4.14) preserves the positivity of the water heights h_1, h_2. Indeed, it is enough to exploit the CFL condition (3.4.2) to prove it.*

Theorem 1. *The numerical approximation (3.4.14) with star values (3.3.13) is well-balanced in the sense that it preserves the stationary solution (3.1.3).*

Proof. We assume that the stationary solution (3.1.3) is satisfied at time t^n, namely $u_{j,i}^n = 0$, $h_{1,i}^n = h_{1,i+1}^n$ and $h_{2,i}^n + z_i = h_{2,i+1}^n + z_{i+1}$ $\forall i$. We want to prove that $h_{j,i}^{n+1} = h_{j,i}^n$ and $(hu)_{j,i}^{n+1} = (hu)_{j,i}^n$ for $j = 1, 2$ and $\forall i$.

Then, first of all we consider the explicit approximation. Since $\mathcal{C}_{2,i+1} - \mathcal{C}_{2,i} + \mathcal{M} = g(h_{2,i+1} - h_{2,i} + z_{i+1} - z_i) = 0$ and $u_{j,i} = 0$ by hypothesis, it is straightforward to see

that $\overline{u}^{*,n}_{1,i-1/2} = \overline{u}^{*,n}_{2,i-1/2} = 0$. Similarly we get $\mathcal{C}^{*,n}_{1,i+1/2} = \mathcal{C}_{1,i}$ and $\mathcal{C}^{*,n}_{2,i+1/2} = \dfrac{\mathcal{C}_{2,i} + \mathcal{C}_{2,i+1}}{2}$. Observing that $\mathcal{C}^{*,n}_{2,i+1/2} - \mathcal{C}^{*,n}_{2,i-1/2} = \frac{S_{i+1/2}+S_{i-1/2}}{2}$, we get $h^{n+1}_{j,i} = h^{n+1-}_{j,i} = h^n_{j,i}$ and $(hu)^{n+1}_{j,i} = (hu)^{n+1-}_{j,i} = (hu)^n_{j,i}$ for $j = 1, 2$ and $\forall i$. Thus, we proved that the explicit approximation is well-balanced. Let us now move to the implicit one. The heart of the proof is to show that, when we are under the lake at rest condition, system (3.4.7) can be reformulated as $A\mathbf{U}^{n+1-} = A\mathbf{U}^n$ (with a matrix A) which would lead to the conclusion of the proof. Indeed, it is easy to see such a thing as $u^n_{j,i} = 0\ \forall i$, for $j = 1, 2$ while for the water heights we have $h^n_{1,i} = h^n_{1,i+1} \quad \text{and} \quad h^n_{2,i+1} - h^n_{2,i} = -z_{i+1} + z_i \quad \forall i$. □

3.5 Numerical simulations

In the following, we are interested in comparing the numerical results of the four following numerical schemes:

- "IFCP" (Intermediate Field Capturing Parabola) method applied to the two-layer shallow water system (3.1.1), for which we refer to [23];
- "LP-ARS" scheme, for which we use the explicit acoustic approximation described in section 3.4.1 and the transport discretization (3.4.13);
- "LP-ARS-IMP" method, for which we use the implicit acoustic approximation described in section 3.4.2 and the transport discretization (3.4.13);
- "LP-HLL" scheme, where once again we use the transport discretization (3.4.13) but we approximate the Lagrangian step by applying the HLL strategy to the Lagrangian system (3.2.4), see for instance [28] or appendix 3.B.

Observe that the IFCP solutions are taken as a reference to establish if the proposed numerical strategies can be considered satisfying.

Furthermore, we use $r = 0.98$ and transmissive boundary conditions. As far as the LP-ARS-IMP method is concerned, we consider both the acoustic (3.4.1) and transport (3.4.2) CFL conditions with CFL $= 0.5$, unless otherwise stated. Finally, we point out that for all the numerical simulations, we exploited MATLAB language with a single Intel Core i7 CPU.

3.5.1 Riemann problems

Using $M = 200$ cells, we start by considering academic test cases, namely two Riemann problems with flat topography.

RP 1. Taking into account a channel of length $L = 10$ m, we consider a dam-break problem for the interface. More explicitly, we take the following initial data, $u_1(x, t = 0) = 0, u_2(x, t = 0) = 0$ and

$$h_{1,L} = 0.2, \quad h_{1,R} = 0.8, \quad h_{2,L} = 0.8, \quad h_{2,R} = 0.2,$$

refer also to [23] for more details about this and the following Riemann problem.

In figure 3.4, we show the results for the water heights h_1, h_2 and fluxes q_1, q_2 using the four above-mentioned methods. In general, we observe that all the schemes give analogous solutions, where the IFCP and LP-HLL methods are the less and the most diffusive respectively. It is not surprising the difference in accuracy between the IFCP and the LP-HLL schemes, as the HLL strategy neglects the internal waves. Nevertheless, refining the mesh, we observed that the LP-HLL solution seems to converge towards the reference output. On the other hand,

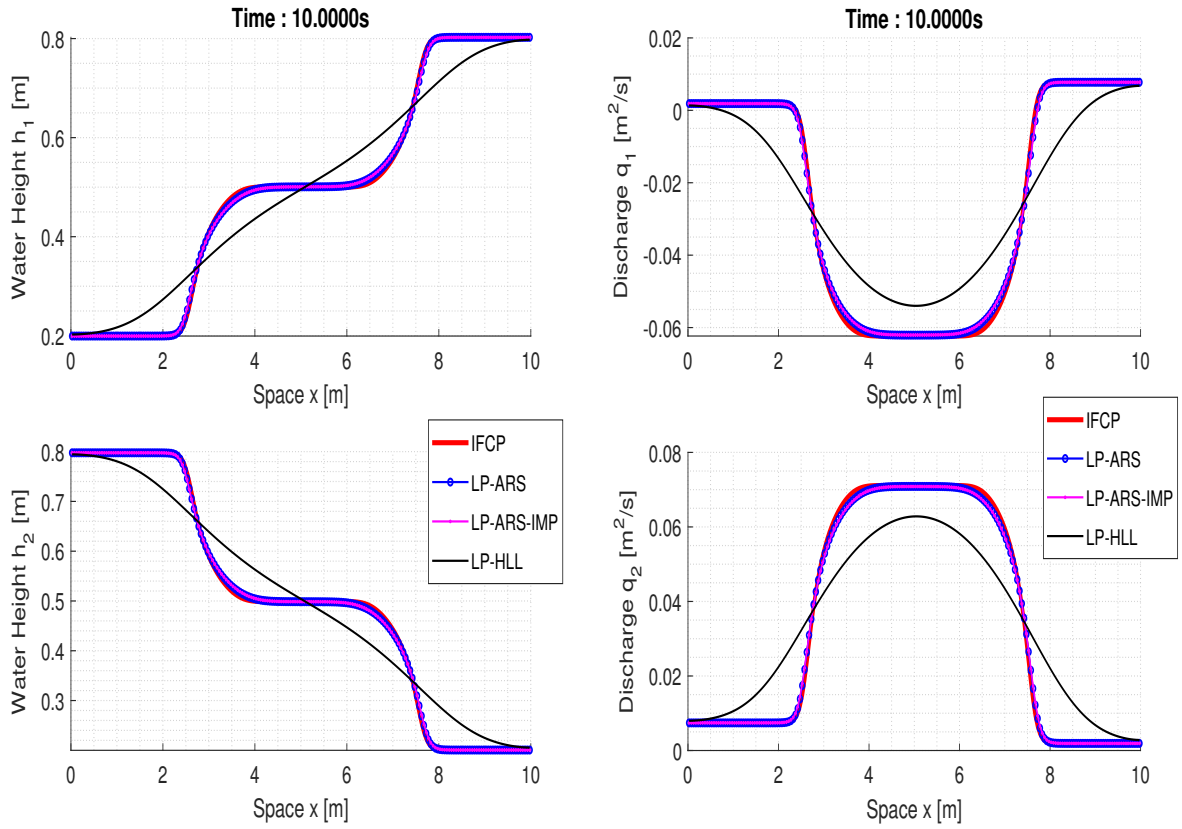


Figure 3.4: RP 1 of section 3.5.1: water height h_1, h_2 (left) and discharge q_1, q_2 (right). IFCP (red), LP-ARS (blue), LP-ARS-IMP (magenta) and LP-HLL (black) outputs obtained with $M = 200$ cells at time $t = 10$ s.

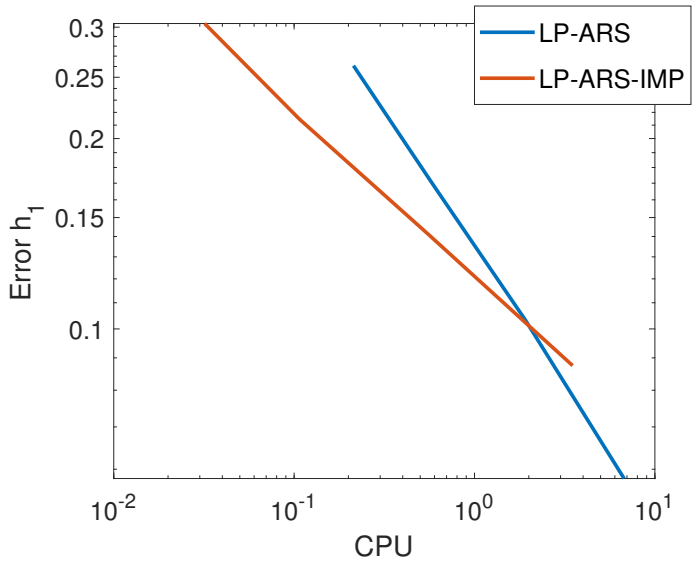


Figure 3.5: RP 1 of section 3.5.1: CPU against error in norm $\mathbf{L}^1$ for the variable h (log scale). Mesh of size $M = (64, 128, 256, 512)$. LP-ARS (blue line) and LP-ARS-IMP (red line) schemes.

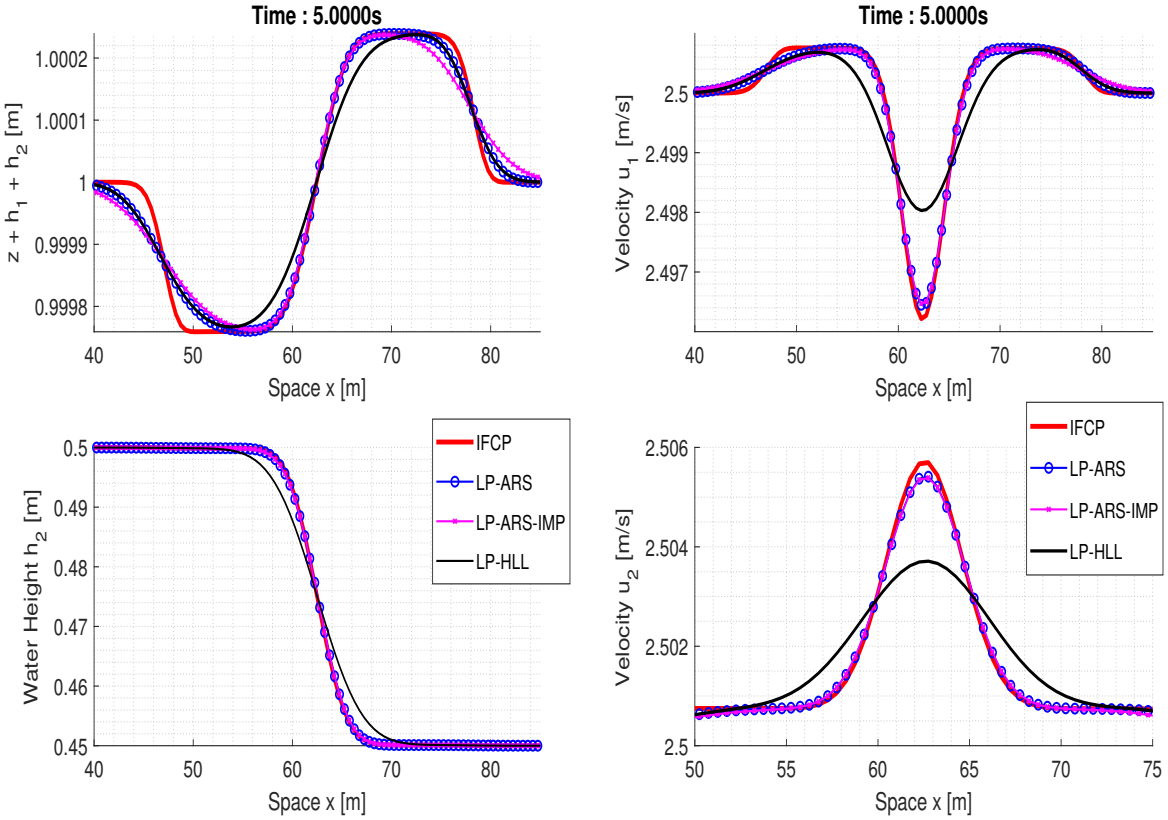


Figure 3.6: RP 2 of section 3.5.1: free surface $h_1 + h_2$, water height h_2 (left) and velocity u_1, u_2 (right). IFCP (red), LP-ARS (blue), LP-ARS-IMP (magenta) and LP-HLL (black) outputs obtained with $M = 200$ cells at time $t = 5$ s.

is it important to highlight that the LP-ARS method gives results very close to the IFCP ones. Moreover, since we have used the same time step, the LP-ARS-IMP outputs are only slightly more diffusive than the LP-ARS ones. Let us conclude observing that we could have used a much larger time step for the implicit method as we are in a sub-critical regime. For this reason, we now consider a series of meshes ($M = 64, 128, 256, 512$ cells) in order to compare the efficiency of the two methods. Thus, for the implicit LP-ARS-IMP method, we only use the transport CFL condition (3.4.2) and neglect the acoustic one (3.4.1). Then, we compute the reference solution with the IFCP method and $M = 2048$ cells. Finally, in table 3.1, we insert the errors in norm $\mathbf{L}^1$ while, in table 3.2, we show the computational times. For the sake of completeness, we also insert the efficiency curve in log scale in figure 3.5. As expected, the errors of the LP-ARS-IMP scheme are slightly greater than those of the LP-ARS method. On the other hand, we immediately see that the LP-ARS-IMP scheme allows faster simulations. Hence, we may conclude that the implicit-explicit method is more efficient than the explicit one. However, this is only true if we do not refine the mesh too much, otherwise the LP-ARS scheme could become the more efficient method between the two.

RP 2. As a second test, we consider a channel of length $L = 100$ with discontinuity in the middle. The initial conditions for the water heights are given by

$$h_{1,L} = 0.5, \quad h_{1,R} = 0.55, \quad h_{2,L} = 0.5, \quad h_{2,R} = 0.45,$$

while for the fluxes we state

$$q_{1,L} = 1.25, \quad q_{1,R} = 1.375, \quad q_{2,L} = 1.25, \quad q_{2,R} = 1.125,$$

	Error ($\mathbf{L}^1$) of h_1		Error ($\mathbf{L}^1$) of q_1		Error ($\mathbf{L}^1$) of h_2		Error ($\mathbf{L}^1$) of q_2	
Mesh	LP-ARS	LP-ARS-IMP	LP-ARS	LP-ARS-IMP	LP-ARS	LP-ARS-IMP	LP-ARS	LP-ARS-IMP
64	0.2630	0.3063	0.0561	0.0661	0.2604	0.3038	0.0556	0.0657
128	0.1714	0.2169	0.0367	0.0467	0.1697	0.2143	0.0364	0.0458
256	0.1059	0.1440	0.0230	0.0312	0.1048	0.1429	0.0227	0.0311
512	0.0605	0.0900	0.0134	0.0204	0.0598	0.0891	0.0132	0.0202

Table 3.1: Errors in norm $\mathbf{L}^1$ of the variables $h_j, q_j = h_j u_j$ with $j = 1, 2$ using LP-ARS and LP-ARS-IMP schemes. Meshes of size $M = (64, 128, 256, 512)$ cells.

Method	$M = 64$	$M = 128$	$M = 256$	$M = 512$
LP-ARS	0.213301	0.588727	1.907348	6.749875
LP-ARS-IMP	0.0319	0.1070	0.5477	3.4986

Table 3.2: Computational times in seconds for LP-ARS and LP-ARS-IMP schemes with meshes of size $M = (64, 128, 256, 512)$ cells.

see again [23]. Then, figure 3.6 shows the different outputs which generally confirms what we have observed for the previous Riemann problem. The solutions are in agreement with the ones presented in [23].

3.5.2 Stationary solution and perturbation

Next, we numerically verify that our numerical strategy is indeed well-balanced in the sense that it preserves the stationary solution (3.1.3). Thus, we take $L = 1$m and as initial condition we consider the following steady state

$$h_1(x, t = 0) = 1, \quad h_2(x, t = 0) + z(x) = 1, \quad u_1(x, t = 0) = 0, \quad u_2(x, t = 0) = 0, \tag{3.5.1}$$

where

$$z(x) = \begin{cases} \frac{1}{4}(1 + \cos(\pi \frac{x-0.5}{0.1})) & \text{if} \quad 0.4 \leq x \leq 0.6 \\ 0 & \text{otherwise.} \end{cases}$$

Our numerical methods are indeed able to preserve this stationary solution up to a machine error of order 10^{-13} computed in the $\mathbf{L}^\infty$ norm using $t = 5$s as ending time and the initial condition as exact solution.

Let us now introduce a small perturbation in the steady state, namely

$$h_1(x, t = 0) = \begin{cases} 1 + 10^{-5} & \text{if} \quad 0.1 \leq x \leq 0.2 \\ 1 & \text{otherwise.} \end{cases}$$

In figure 3.7, we show the LP-ARS results at different times and we observe that the perturbations propagate away, so that we are able to recover the zero-velocity steady state (3.5.1). Indeed, the outputs are in agreement with the ones presented in [3]. We did not include the solutions for the other schemes as they are analogous. However, it is interesting to observe that LP-HLL solution is almost identical to the LP-ARS one, while the LP-ARS-IMP output is more diffusive even if we use the same CFL condition.

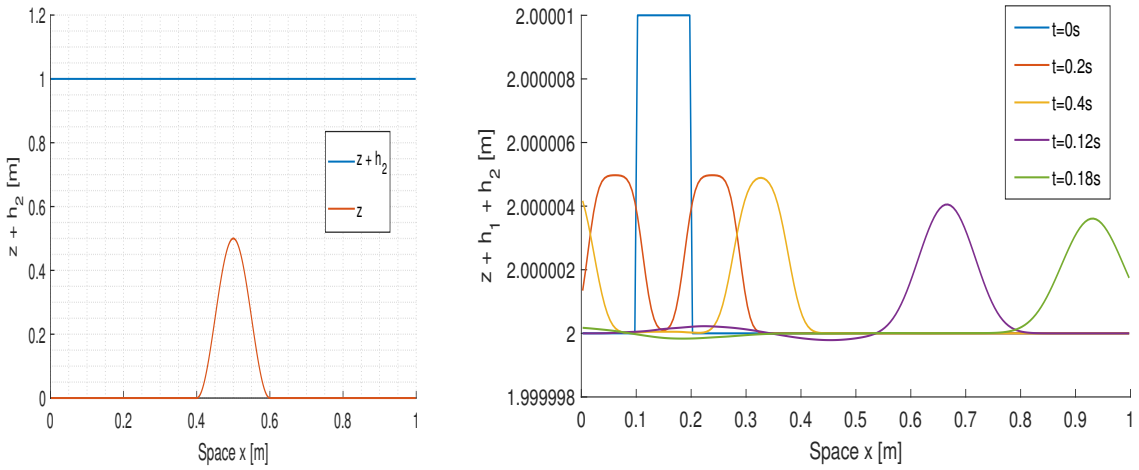


Figure 3.7: Evolution of the perturbation in the lake at rest steady state, section 3.5.2. On the left, LP-ARS solution for h_2+z (blue) and topography z (red). On the right, LP-ARS solution for the free surface h_1+h_2+z for different times: $t=0$ (blue), $t=0.02$ (red), $t=0.04$ (yellow), $t=0.12$ (violet), $t=0.18$ (green). $M=200$ cells.

3.5.3 Transcritical non-smooth stationary solution

In this section, we aim to verify that our schemes are able to recover a transcritical non-smooth stationary solution if proper steady boundary conditions are imposed. We refer the reader to [9] for further details on this simulation. As a second step, we will also introduce some perturbations in the resulting steady state.

Thus, let us consider a channel of length $L=10$ m and the following initial conditions

$$(hu)_j(x,t=0)=0, \qquad h_1(x,t=0)=\begin{cases}0.5 & \text{if} \quad x<5\\ 0.001 & \text{otherwise,}\end{cases}$$

$z(x)=1+0.5x^{-(x-5)^2}$ and finally $h_2(x,t=0)=2-(h_1(x,t=0)+z(x))$. Then, for the boundary conditions we impose $(hu)_2=-(hu)_1$ on both sides and $h_1(x=L,t)+h_2(x=L,t)+z(x=L)=2$ at the end of the channel. Referring to [23], the flow is sub-critical for $x\geq 5$ while for $x<5$ is supercritical at the beginning and then it becomes sub-critical as well. Thus, it is clear that at $x=5$ there is a critical flow.

Then, in figure 3.8, we include the results and compare the different outputs found for h_2+z at time $t=1000$ s. First of all, we generally observe that all the numerical schemes are able to recover the aimed non-smooth stationary solution even if some differences in the shock position and the left state value are present for the same mesh value $M=350$ cells. However, we remarked that the IFCP numerical solution oscillates during the simulation. This is indeed natural as the internal eigenvalues can become complex in a small area of the supercritical region. On the other hand, unphysical oscillations are not observed when using the LP-ARS and LP-ARS-IMP schemes, probably partly related to the fact that the eigenvalues of the acoustic system are always real. Moreover, also the LP-HLL gives a solution without spurious oscillations. However, this is not surprising as the method neglect the middle waves of the mathematical model. Indeed, we can observe that the LP-HLL output is much more diffusive than the others.

Finally, since the regime is partly sub-critical, for the implicit scheme we could use a larger time step. On the right hand side of figure 3.8, we show the LP-ARS-IMP solution computed with a time step taken as the minimum between the transport and the acoustic time steps, where the latter is computed using as CFL values CFL $=0.5,2.5,5$ (observe that in formula (3.4.1) we directly used CFL $=0.5$). Thus, once again, the solutions seem correct with some

Method	$M = 44$	$M = 88$	$M = 175$	$M = 350$
LP-ARS	8.103190	25.550884	100.167893	363.242909
LP-ARS-IMP	1.666153	4.393538	34.577729	219.775353

Table 3.3: Computational times in seconds for LP-ARS and LP-ARS-IMP schemes with meshes of size $M = (44, 88, 175, 350)$ cells.

differences in the left state value. Then, in table 3.3, for different mesh values, we also include the computational times of the LP-ARS and LP-ARS-IMP schemes, where for the latter we use CFL $= 5$ for the acoustic time step condition. Once again, we generally see that the LP-ARS-IMP method is faster, even if the regime is not always sub-critical. Moreover, it is clear that, the more we refine the mesh, the smaller will be the difference of the computational times between the two schemes.

3.5.3.1 Perturbation of the transcritical non-smooth stationary solution

Next, we consider the transcritical non-smooth stationary solution obtained in section 3.5.3 as initial condition and we add a perturbation in the interface. The objective is to verify that the perturbation propagates away and we recover the non-smooth stationary solution. Thus, at initial time we impose

$$h_2(x, t = 0) = \begin{cases} h_2^{eq} + 0.1e^{-100(x-6.5)^2} & \text{if} \quad 6 \leq x \leq 7, \\ h_2^{eq} & \text{otherwise} \end{cases}$$

where the superscript "eq" indicates the transcritical non-smooth steady state. Regarding the boundary conditions, we keep the same as before. Results are shown in figure 3.9 using the LP-ARS-IMP scheme. We do not include the results for the other methods as they are analogous. Indeed, we observe the perturbation propagates away at different times $t = 0, 0.15, 0.5, 1$s.

3.6 Concluding remarks

In this work the classic Lagrange-Projection (LP) approach has been extended to a two-velocities case, namely the two-layer shallow water system. Hence, we started this work by presenting the mathematical model formulated in Lagrangian coordinates. To numerically approximate such a system, we also considered the acoustic-transport splitting, an alternative interpretation to the Lagrange-Projection decomposition. In particular, we were able to build an approximate Riemann solver for the acoustic system and to develop the associated Godunov-type scheme, both explicitly and implicitly. We highlight that such a discretization can also be interpreted as an approximation for the Lagrangian system. Moreover, in the implicit version of the scheme, to find the numerical solution we only need to solve a linear system, which entails a not excessive computational cost. In this way, we were able to obtain a fast implicit-explicit method, as we could use very large time steps, especially in sub-critical regimes.

Numerical simulations were proposed, in which we compared our results against the outputs of the well-known IFCP scheme. We also considered a LP-HLL method, meaning that the HLL approach has been applied to the simplified Lagrangian system (3.2.4), while keeping the same numerical strategy for the projection step. In the numerical tests, the LP-ARS strategy

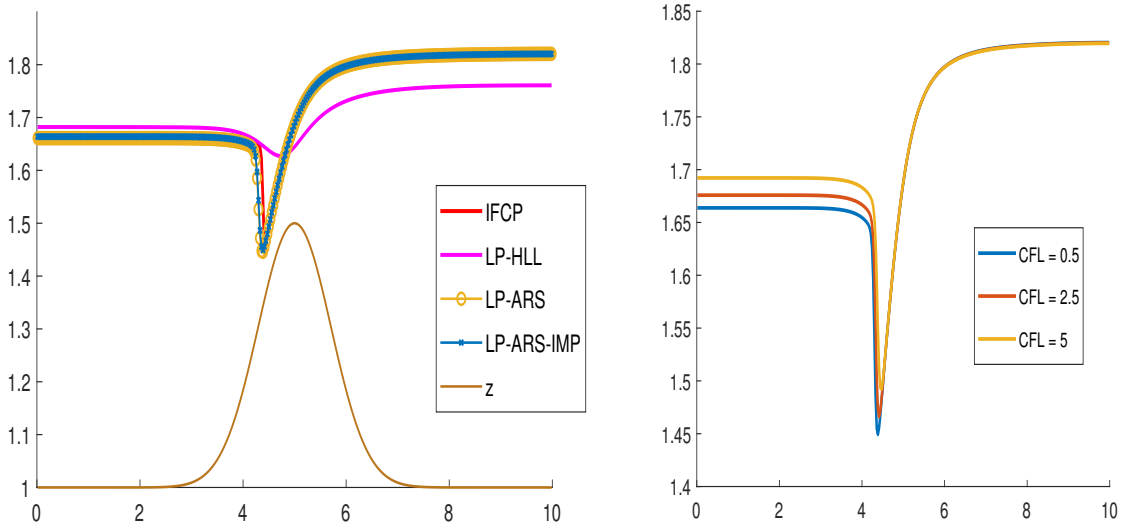


Figure 3.8: Transcritical non smooth stationary solution, section 3.5.3. On the left: IFCP (red), LP-HLL (magenta), LP-ARS (yellow) and LP-ARS-IMP (blue) solution for the interface $h_2 + z$. On the right: LP-ARS-IMP solution for $h_2 + z$ computed with $CFL = 0.5$ (blue), $CFL = 2.5$ (red) and $CFL = 5$ (yellow). $M = 350$ cells.

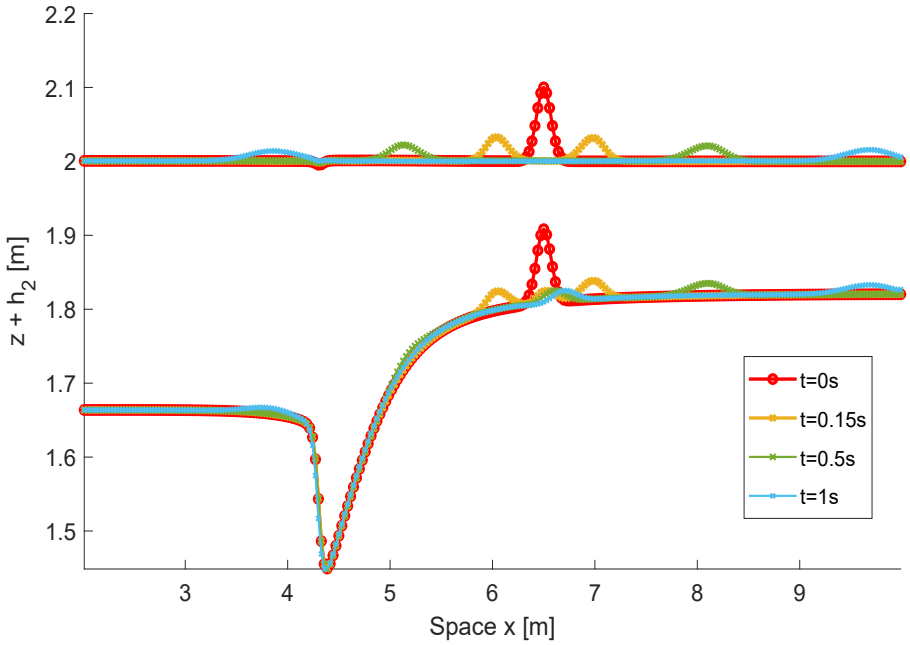


Figure 3.9: Evolution of the perturbation of a transcritical non smooth stationary solution, section 3.5.3.1. LP-ARS-IMP solution for the free surface $h_1 + h_2 + z$ and interface $h_2 + z$ for different times: $t = 0$s (red), $t = 0.15$s (yellow), $t = 0.5$s (green) and $t = 1$s (blue). $M = 350$ cells.

(both explicit and implicit) gave satisfying results, generally slightly more diffusive than the IFCP ones but much more accurate than the LP-HLL outputs.

Furthermore, it is interesting to remark that, in the LP strategy, we considered an approximate version of the Lagrangian system, obtaining a new model for which we are able to explicitly write the eigenvalues and to prove that they are always real. Therefore, the numerical method is able to advance in time even if there are small non-hyperbolic regions with complex internal eigenvalues. Indeed, in the test of section 3.5.3, we acknowledged that our schemes' solutions do not oscillate, contrarily to the IFCP's one.

3.7 Bibliography

[1] R. Abgrall and S. Karni. *Two-Layer Shallow Water System: A Relaxation Approach.* SIAM J. Scientific Computing. 31: 1603-1627, 2009. 10.1137/06067167X.

[2] 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: 2050-2065, 2004. 10.1137/S1064827503431090.

[3] J. Balbás and S. Karni. *A Non-oscillatory Central Scheme for One-Dimensional Two-Layer Shallow Water Flows along Channels with Varying Width.* Journal of Scientific Computing. 55: 499–528, 2013. 10.1007/s10915-012-9642-3.

[4] C. Berthon, F. Foucher. *Hydrostatic Upwind Schemes for Shallow-Water Equations.* Springer Proceedings in Mathematics, 4: 97-105, 2011. 10.1007/978-3-642-20671-9_11

[5] C. Berthon, F. Foucher. *Efficient well–balanced hydrostatic upwind schemes for shallow–water equations.* Journal of Computational Physics. 231(15): 4993–5015, 2012. 10.1016/j.jcp.2012.02.031

[6] C. Berthon, V. Michel-Dansac. *A simple fully well-balanced and entropy preserving scheme for the shallow-water equations.* Applied Mathematics Letters. 86: 284-290, 2018. 10.1016/j.aml.2018.07.013.

[7] F. Bouchut and T. Morales de Luna. *An entropy satisfying scheme for two-layer shallow water equations with uncoupled treatment.* ESAIM: Mathematical Modelling and Numerical Analysis (ESAIM: M2AN), 42: 683–698, 2008. 10.1051/m2an:2008019

[8] M. J. Castro Díaz, C. Chalons and T. Morales De Luna. *A fully well-balanced Lagrange-Projection type scheme for the Shallow-water equations.* SIAM J. Numer. Anal., 56(5): 3071–3098, 2018. 10.1137/17M1156101.

[9] M. Castro, E. Fernández-Nieto, J. González Vida, C. Madroñal. *Numerical Treatment of the Loss of Hyperbolicity of the Two-Layer Shallow-Water System.* Journal of Scientific Computing. 48: 16-40, 2011. 10.1007/s10915-010-9427-5.

[10] M. Castro, J. Frings, S. Noelle, C. Madroñal, G. Puppo. *On the Hyperbolicity of Two- and Three-Layer Shallow Water Equations.* 17, 2012. 10.1142/9789814417099_0030.

[11] M. J. Castro Díaz, J. Macías and C. Parés. *A Q-scheme for a class of systems of coupled conservation laws with source term. Application to a two-layer 1-D shallow water system.* ESAIM M2AN Math. Model. Numer. Anal. 35: 107–127, 2001. 10.1051/m2an:2001108

[12] M. J. Castro, T. Morales de Luna, C. Parés Madroñal. *Well-Balanced Schemes and Path-Conservative Numerical Methods.* Handbook of Numerical Analysis, Elsevier, 18: 131-175, 2017. 10.1016/bs.hna.2016.10.002.

[13] C. Chalons, F. Coquel, S. Kokh, N. Spillane. *Large time-step numerical scheme for the seven-equation model of compressible two-phase flows.* Springer Proceedings in Mathematics, FVCA 6, 4(1): 225-233, 2011. 10.1007/978-3-642-20671-9_24

[14] C. Chalons, A. Del Grosso. *A second-order well-balanced Lagrange-projection scheme for Shallow Water Exner equations in 1D and 2D.* 2021. hal-03251707

[15] C. Chalons, A. Del Grosso. *Exploring different possibilities for second-order well-balanced Lagrange-projection numerical schemes applied to shallow water Exner equations.* International Journal for Numerical Methods in Fluids. 1- 31, 2022. 10.1002/fld.5064

[16] C. Chalons, M. Girardin, S. Kokh. *An All-Regime Lagrange-Projection like scheme for the gas dynamics equations on unstructured meshes.* Communications in Computational Physics. 20(1): 188-233 2016. 10.4208/cicp.260614.061115a.

[17] C. Chalons, M. Girardin, S. Kokh. *An all-regime Lagrange-Projection like scheme for 2D homogeneous models for two-phase flows on unstructured meshes.* Journal of Computational Physics, Elsevier, 335: 885-904, 2017. 10.1016/j.jcp.2017.01.017.

[18] C. Chalons, P. Kestener, S. Kokh, and M. Stauffert. *A large time-step and well-balanced Lagrange-Projection type scheme for the Shallow-water equations.* Communications in Mathematical Sciences. 15(3): 765–788, 2017. 10.4310/CMS.2017.v15.n3.a9.

[19] C. Chalons, S. Kokh, M. Girardin. *Large Time Step and Asymptotic Preserving Numerical Schemes for the Gas Dynamics Equations with Source Terms.* SIAM Journal on Scientific Computing. 35(6): A2874–A2902, 2013. 10.1137/130908671.

[20] A. Chinnayya, A.-Y. LeRoux, N. Seguin. *A well-balanced numerical scheme for the approximation of the shallow-water equations with topography: the resonance phenomenon.* International Journal on Finite Volume (electronic), 1(1), 2004. hal-00017378

[21] G. Dal Maso, P. G. Lefloch, and F. Murat. *Definition and weak stability of nonconservative products.* Journal de Mathématiques Pures et Appliquées, 74(6): 483–548, 1995.

[22] A. Del Grosso and C. Chalons. *Second-order well balanced Lagrange-Projection schemes for Blood Flow Equations.* Calcolo 58, 43, 2021. 10.1007/s10092-021-00434-5

[23] E. Fernández-Nieto, M. Castro, C. Parés. *On an Intermediate Field Capturing Riemann Solver Based on a Parabolic Viscosity Matrix for the Two-Layer Shallow Water System.* Journal of Scientific Computing. 48: 117-140, 2011. 10.1007/s10915-011-9465-7.

[24] J. Lin, B. Mao, X. Lu. *A Two-Layer Hydrostatic-Reconstruction Method for High-Resolution Solving of the Two-Layer Shallow-Water Equations over Uneven Bed Topography.* Mathematical Problems in Engineering. 1-14, 2019. 10.1155/2019/5064171.

[25] V. Michel-Dansac, C. Berthon, S. Clain, F. Foucher. *A well-balanced scheme for the shallow-water equations with topography or Manning friction.* Journal of Computational Physics, Elsevier, 335: 115-154, 2017. 10.1016/j.jcp.2017.01.009.

[26] T. Morales De Luna, M. J. Castro Díaz and C. Chalons. *High order fully well-balanced Lagrange-Projection scheme for Shallow-water.* Commun. Math. Sci., 18(3): 781–807, 2020. 10.4310/CMS.2020.v18.n3.a9

[27] J.B. Schijf, J.C. Schonfeld. *Theoretical considerations on the motion of salt and fresh water.* In: Proc. of the Minn. Int. Hydraulics Conv., 321–333. Joint meeting IAHR and Hyd. Div. ASCE, Sept. 1953.

[28] E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics*, Third Edition. Springer–Verlag, 2009. 10.1007/b79761_5.

Annexe

3.A Système linéaire pour l'approximation implicite du système acoustique

On considére l'approximation implicite du système acoustique présentée dans la section 3.4.2. Dans cette annexe, on veut simplement décrire la matrice A présente dans le système linéaire (3.4.8). En particulier, A est une matrice carrée avec 4M $\times$ 4M entrées où seulement 12 d'entre elles pourraient être non nulles dans chaque ligne, en effet

$$
A = \begin{bmatrix}
d_1^{u_1} & d_1^{u_2} & d_1^{\mathcal{C}_1} & d_1^{\mathcal{C}_2} & f_1^{u_1} & f_1^{u_2} & f_1^{\mathcal{C}_1} & f_1^{\mathcal{C}_2} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
k_1^{u_1} & k_1^{u_2} & k_1^{\mathcal{C}_1} & k_1^{\mathcal{C}_2} & l_1^{u_1} & l_1^{u_2} & l_1^{\mathcal{C}_1} & l_1^{\mathcal{C}_2} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
n_1^{u_1} & n_1^{u_2} & n_1^{\mathcal{C}_1} & n_1^{\mathcal{C}_2} & q_1^{u_1} & q_1^{u_2} & q_1^{\mathcal{C}_1} & q_1^{\mathcal{C}_2} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
v_1^{u_1} & v_1^{u_2} & v_1^{\mathcal{C}_1} & v_1^{\mathcal{C}_2} & w_1^{u_1} & w_1^{u_2} & w_1^{\mathcal{C}_1} & w_1^{\mathcal{C}_2} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
 & & & & & & & \cdots & \cdots & & & & & & & \\
\cdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & \cdots & \cdots & \cdots & \cdots & \vdots \\
\cdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & \cdots & \cdots & \cdots & \cdots & \vdots \\
\cdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & \cdots & \cdots & \cdots & \cdots & \vdots \\
\cdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & \cdots & \cdots & \cdots & \cdots & \vdots \\
\cdots & 0 & b_i^{u_1} & b_i^{u_2} & b_i^{\mathcal{C}_1} & b_i^{\mathcal{C}_2} & d_i^{u_1} & d_i^{u_2} & d_i^{\mathcal{C}_1} & d_i^{\mathcal{C}_2} & f_i^{u_1} & f_i^{u_2} & f_i^{\mathcal{C}_1} & f_i^{\mathcal{C}_2} & 0 & \cdots \\
\cdots & 0 & g_i^{u_1} & g_i^{u_2} & g_i^{\mathcal{C}_1} & g_i^{\mathcal{C}_2} & k_i^{u_1} & k_i^{u_2} & k_i^{\mathcal{C}_1} & k_i^{\mathcal{C}_2} & l_i^{u_1} & l_i^{u_2} & l_i^{\mathcal{C}_1} & l_i^{\mathcal{C}_2} & 0 & \cdots \\
\cdots & 0 & m_i^{u_1} & m_i^{u_2} & m_i^{\mathcal{C}_1} & m_i^{\mathcal{C}_2} & n_i^{u_1} & n_i^{u_2} & n_i^{\mathcal{C}_1} & n_i^{\mathcal{C}_2} & q_i^{u_1} & q_i^{u_2} & q_i^{\mathcal{C}_1} & q_i^{\mathcal{C}_2} & 0 & \cdots \\
\cdots & 0 & s_i^{u_1} & s_i^{u_2} & s_i^{\mathcal{C}_1} & s_i^{\mathcal{C}_2} & v_i^{u_1} & v_i^{u_2} & v_i^{\mathcal{C}_1} & v_i^{\mathcal{C}_2} & w_i^{u_1} & w_i^{u_2} & w_i^{\mathcal{C}_1} & w_i^{\mathcal{C}_2} & 0 & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \cdots \\
 & & & & & & & \cdots & \cdots & & & & & & & \\
\vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & b_M^{u_1} & b_M^{u_2} & b_M^{\mathcal{C}_1} & b_M^{\mathcal{C}_2} & d_M^{u_1} & d_M^{u_2} & d_M^{\mathcal{C}_1} & d_M^{\mathcal{C}_2} \\
\vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & g_M^{u_1} & g_M^{u_2} & g_M^{\mathcal{C}_1} & g_M^{\mathcal{C}_2} & k_M^{u_1} & k_M^{u_2} & k_M^{\mathcal{C}_1} & k_M^{\mathcal{C}_2} \\
\vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & m_M^{u_1} & m_M^{u_2} & m_M^{\mathcal{C}_1} & m_M^{\mathcal{C}_2} & n_M^{u_1} & n_M^{u_2} & n_M^{\mathcal{C}_1} & n_M^{\mathcal{C}_2} \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & s_M^{u_1} & s_M^{u_2} & s_M^{\mathcal{C}_1} & s_M^{\mathcal{C}_2} & v_M^{u_1} & v_M^{u_2} & v_M^{\mathcal{C}_1} & v_M^{\mathcal{C}_2}
\end{bmatrix}
$$

avec

$$b_i^{u_1} = -\frac{\Delta t}{2\Delta x}(\bar{s} + \frac{a_1^2 - \bar{s}^2}{s+\bar{s}})_{i-1/2}, \quad f_i^{u_1} = -\frac{\Delta t}{2\Delta x}(\bar{s} + \frac{a_1^2 - \bar{s}^2}{s+\bar{s}})_{i+1/2}, \quad d_i^{u_1} = 1 - b_i^{u_1} - f_i^{u_1}$$
$$b_i^{u_2} = -\frac{\Delta t}{2\Delta x}(\frac{a_2^2}{s+\bar{s}})_{i-1/2}, \quad f_i^{u_2} = -\frac{\Delta t}{2\Delta x}(\frac{a_2^2}{s+\bar{s}})_{i+1/2}, \quad d_i^{u_2} = -b_i^{u_2} - f_i^{u_2},$$
$$b_i^{\mathcal{C}_1} = -\frac{\Delta t}{2\Delta x}, \quad d_i^{\mathcal{C}_1} = +\frac{\Delta t}{2\Delta x}, \quad f_i^{\mathcal{C}_1} = 0,$$
$$b_i^{\mathcal{C}_2} = -\frac{\Delta t}{2\Delta x}, \quad f_i^{\mathcal{C}_2} = +\frac{\Delta t}{2\Delta x}, \quad d_i^{\mathcal{C}_2} = 0;$$

$$g_i^{u_1} = \frac{\Delta t}{2\Delta x}(\frac{(a_1^2 - s^2)(a_1^2 - \bar{s}^2)}{a_2^2(s+\bar{s})})_{i-1/2},\ l_i^{u_1} = \frac{\Delta t}{2\Delta x}(\frac{(a_1^2 - s^2)(a_1^2 - \bar{s}^2)}{a_2^2(s+\bar{s})})_{i+1/2},\ k_i^{u_1} = -g_i^{u_1} - l_i^{u_1},$$
$$g_i^{u_2} = -\frac{\Delta t}{2\Delta x}(\bar{s} - \frac{a_1^2 - s^2}{s+\bar{s}})_{i-1/2},\ l_i^{u_2} = -\frac{\Delta t}{2\Delta x}(\bar{s} - \frac{a_1^2 - s^2}{s+\bar{s}})_{i+1/2},\ k_i^{u_2} = 1 - g_i^{u_2} - l_i^{u_2},$$
$$g_i^{\mathcal{C}_1} = -r\frac{\Delta t}{2\Delta x}, \quad l_i^{\mathcal{C}_1} = +r\frac{\Delta t}{2\Delta x}, \quad k_i^{\mathcal{C}_1} = 0,$$
$$g_i^{\mathcal{C}_2} = -\frac{\Delta t}{2\Delta x}, \quad l_i^{\mathcal{C}_2} = +\frac{\Delta t}{2\Delta x}, \quad k_i^{\mathcal{C}_2} = 0;$$
$$m_i^{u_1} = -\frac{\Delta t}{2\Delta x}(a_1^2)_{i-1/2}, \quad q_i^{u_1} = +\frac{\Delta t}{2\Delta x}(a_1^2)_{i+1/2}, \quad n_i^{u_1} = -m_i^{u_1} - n_i^{u_1},$$
$$m_i^{u_2} = 0, \quad q_i^{u_2} = 0, \quad n_i^{u_2} = 0,$$
$$m_i^{\mathcal{C}_1} = -\frac{\Delta t}{2\Delta x}(\bar{s} + \frac{a_1^2 - \bar{s}^2}{s+\bar{s}})_{i-1/2}, \quad q_i^{\mathcal{C}_1} = -\frac{\Delta t}{2\Delta x}(\bar{s} + \frac{a_1^2 - \bar{s}^2}{s+\bar{s}})_{i+1/2}, \quad n_i^{\mathcal{C}_1} = 1 - m_i^{\mathcal{C}_1} - q_i^{\mathcal{C}_1},$$
$$m_i^{\mathcal{C}_2} = -\frac{\Delta t}{2\Delta x}(\frac{a_1^2}{s+\bar{s}})_{i-1/2}, \quad q_i^{\mathcal{C}_2} = -\frac{\Delta t}{2\Delta x}(\frac{a_1^2}{s+\bar{s}})_{i+1/2}, \quad n_i^{\mathcal{C}_2} = -m_i^{\mathcal{C}_2} - q_i^{\mathcal{C}_2}$$

et

$$s_i^{u_1} = 0, \quad w_i^{u_1} = 0, \quad v_i^{u_1} = 0,$$
$$s_i^{u_2} = -\frac{\Delta t}{2\Delta x}(a_2^2)_{i-1/2}, \quad w_i^{u_2} = +\frac{\Delta t}{2\Delta x}(a_2^2)_{i+1/2}, \quad v_i^{u_2} = -m_i^{u_1} - n_i^{u_1},$$
$$s_i^{\mathcal{C}_1} = -\frac{\Delta t}{2\Delta x}(\frac{(a_1^2 - s^2)(a_1^2 - \bar{s}^2)}{a_1^2(s+\bar{s})})_{i-1/2},\ w_i^{\mathcal{C}_1} = -\frac{\Delta t}{2\Delta x}(\frac{(a_1^2 - s^2)(a_1^2 - \bar{s}^2)}{a_1^2(s+\bar{s})})_{i+1/2},\ v_i^{\mathcal{C}_1} = -s_i^{\mathcal{C}_1} - w_i^{\mathcal{C}_1},$$
$$s_i^{\mathcal{C}_2} = -\frac{\Delta t}{2\Delta x}(\bar{s} - \frac{a_1^2 - s^2}{s+\bar{s}})_{i-1/2},\ w_i^{\mathcal{C}_2} = -\frac{\Delta t}{2\Delta x}(\bar{s} - \frac{a_1^2 - s^2}{s+\bar{s}})_{i+1/2},\ v_i^{\mathcal{C}_2} = 1 - s_i^{\mathcal{C}_2} - w_i^{\mathcal{C}_2},$$

où nous avons utilisé $s = \lambda_{ext}^+$, $\bar{s} = \lambda_{int}^+$ pour alléger les notations.

Enfin, il est clair que la première et la dernière ligne du système doivent être modifiées en fonction des conditions limites considérées.

3.B Schéma HLL pour le système lagrangien

Dans cette annexe, nous allons donner quelques détails sur la méthode LP-HLL que nous avons considérée dans la section précédente 3.5. En effet, dans le but de disposer d'un schéma numérique de base dans le formalisme Lagrange-projection, nous avons également appliqué la stratégie HLL au système lagrangien (3.2.4).

Afin de pouvoir le décrire et puisque le système lagrangien $\partial_t \mathbf{LQ} + \mathbf{A}(\mathbf{LQ})\partial_\xi \mathbf{LQ} = \mathbf{S}(\mathbf{LQ}, z)$ a une forme non-conservative, nous devons d'abord introduire brièvement le concept de chemin et de schéma numérique préservant le chemin. Ensuite, nous appliquerons une telle théorie à notre cas particulier, afin de trouver l'approximation HLL pour le système lagrangien.

3.B.1 Schéma numérique préservant les chemins

Considérons un système non-conservatif de la forme

$$\partial_t \mathbf{Q} + \mathbf{A}(\mathbf{Q})\partial_x \mathbf{Q} = \mathbf{0} \tag{3.B.1}$$

où $\mathbf{Q}$ est le vecteur des inconnues et $\mathbf{A}(\mathbf{Q})$ la matrice non-conservative. On peut observer qu'un terme source $\mathbf{S}(\mathbf{Q})\partial_x z$, avec z une fonction connue, pourrait être facilement inséré dans le système (3.B.1) et rien ne changerait dans la stratégie numérique. En effet, on pourrait simplement ajouter z au vecteur d'inconnues et l'équation $\partial_t z = 0$ au système, ainsi ici nous ne donnons pas plus de détails.

À la recherche d'une méthode des volumes finis pour approcher le système (3.B.1), nous commençons par l'intégrer sur le volume de contrôle $[x_1, x_2] \times [t_0, t_1]$, en obtenant

$$\int_{x_1}^{x_2} \mathbf{Q}(x, t_1)\mathrm{d}x = \int_{x_1}^{x_2} \mathbf{Q}(x, t_0)\mathrm{d}x - \int_{t_0}^{t_1} \int_{x_1}^{x_2} \mathbf{A}(\mathbf{Q}(x,t))\partial_x \mathbf{Q}(x,t)\mathrm{d}x\mathrm{d}t.$$

Cependant, en raison de la présence du produit non-conservatif $\mathbf{A}(\mathbf{Q})\partial_x \mathbf{Q}$, des masses de Dirac peuvent apparaître en présence de discontinuités et, par conséquent, la notion de solution faible au sens des distributions ne s'applique plus. Pour cette raison, Dal Maso, LeFloch et Murat [21] ont développé une théorie permettant de contourner ce problème. À cette fin, nous devons d'abord définir une famille de chemins continus Lipschitz, $\Phi : [0,1] \times \Omega \times \Omega \to \Omega$ satisfaisant les propriétés suivantes,

$$\Phi(0; \mathbf{Q}_L, \mathbf{Q}_R) = \mathbf{Q}_L, \quad \Phi(1; \mathbf{Q}_L, \mathbf{Q}_R) = \mathbf{Q}_R, \quad \text{et} \quad \Phi(s;\ \mathbf{Q}, \mathbf{Q}) = \mathbf{Q},$$

Puis, nous définissons $\displaystyle\int_{x_1}^{x_2} \mathbf{A}(\mathbf{Q}(x,t))\partial_x \mathbf{Q}(x,t)\mathrm{d}x\mathrm{d}t$ en posant

$$\fint_{x_1}^{x_2} \mathbf{A}(\mathbf{Q}(x))\partial_x \mathbf{Q}(x)\mathrm{d}x = \int_{x_1}^{x_2} \mathbf{A}(\mathbf{Q}(x))\partial_x \mathbf{Q}(x)\mathrm{d}x + \sum_l \int_0^1 \mathbf{A}(\Phi(s; \mathbf{Q}_l^-, \mathbf{Q}_l^+))\frac{\partial \Phi}{\partial s}(s; \mathbf{Q}_l^-, \mathbf{Q}_l^+)\mathrm{d}s$$

où $\mathbf{Q}_l^-$ et $\mathbf{Q}_l^+$ sont respectivement la limite de $\mathbf{Q}$ à gauche et à droite de la l-ième discontinuité. Ainsi, on dit qu'une solution faible est une fonction qui satisfait à

$$\int_{x_1}^{x_2} \mathbf{Q}(x, t_1)\mathrm{d}x = \int_{x_1}^{x_2} \mathbf{Q}(x, t_0)\mathrm{d}x - \int_{t_0}^{t_1} \fint_{x_1}^{x_2} \mathbf{A}(\mathbf{Q}(x,t))\partial_x \mathbf{Q}(x,t)\mathrm{d}x\mathrm{d}t.$$

Pour plus de détails, on peut se référer par exemple à [12, 23].

Cependant, un problème crucial est maintenant donné par le choix du chemin. En effet, la solution pourrait être différente selon le chemin choisi. Comme il s'agit d'un problème épineux, ici nous préférons ne pas nous y attarder et nous exploitons simplement la formulation la plus simple pour la définition du chemin, à savoir le chemin des lignes droites,

$$\Phi(s; \mathbf{Q}_L, \mathbf{Q}_R) = \mathbf{Q}_L + s(\mathbf{Q}_R - \mathbf{Q}_L).$$

De plus, il est important de souligner qu'en raison de la viscosité numérique de la méthode, même si le chemin "correct" est choisi, la solution numérique pourrait converger vers une solution erronée et non vers la solution physique [12]. Après, une fois que la famille de chemins Φ a été choisie, on dit qu'une méthode numérique est Φ-conservative si elle est donnée par la formule suivante

$$\mathbf{Q}_i^{n+1} = \mathbf{Q}_i^n - \frac{\Delta t}{\Delta x}(\mathbf{D}_{i+1/2}^- + \mathbf{D}_{i-1/2}^+) \tag{3.B.2}$$

où $\mathbf{D}^{\pm}_{i+1/2} = \mathbf{D}^{\pm}(\mathbf{Q}_i, \mathbf{Q}_{i+1})$ sont deux fonctions continues satisfaisant

$$\mathbf{D}^{\pm}(\mathbf{Q}, \mathbf{Q}) = 0 \quad \forall\, \mathbf{Q} \in \Omega$$

et

$$\mathbf{D}^{-}(\mathbf{Q}_L, \mathbf{Q}_R) + \mathbf{D}^{+}(\mathbf{Q}_L, \mathbf{Q}_R) = \int_0^1 \mathbf{A}(\Phi(s; \mathbf{Q}_L, \mathbf{Q}_R)) \frac{\partial \Phi}{\partial s}(s; \mathbf{Q}_L, \mathbf{Q}_R) \mathrm{d}s,$$

pour tout ensemble $\{\mathbf{Q}_L, \mathbf{Q}_R\} \subset \Omega$. Une fois de plus, nous ne donnons pas ici plus de détails sur les schémas préservant les chemins, mais nous contentons de faire référence à [23] et aux références qui y figurent. Il est important de souligner que le concept de schéma préservant les chemins peut être considéré comme une généralisation de celui de méthode conservative pour un système conservatif, à savoir le cas où $\mathbf{A}(\mathbf{Q})$ est la matrice jacobienne d'une certaine fonction $\mathbf{F}(\mathbf{Q})$.

Pour conclure cette partie, nous donnons quelques détails sur les schémas de Roe dans ce cadre, puisque cela nous sera utile pour introduire le schéma HLL. Ainsi, étant donné une famille de chemins Φ, une fonction $\mathbf{A}_\Phi : \Omega \times \Omega \to \mathcal{M}_{N\times N}(\mathbb{R})$ est dite être une linéarisation de Roe si elle vérifie les propriétés suivantes :

- $\forall\, \mathbf{Q}_L, \mathbf{Q}_R \in \Omega$, $\mathbf{A}_\Phi(\mathbf{Q}_L, \mathbf{Q}_R)$ a N valeurs propres réelles distinctes;
- $\mathbf{A}_\Phi(\mathbf{Q}, \mathbf{Q}) = \mathbf{A}(\mathbf{Q})\ \forall\, \mathbf{Q} \in \Omega$;
- $\forall\, \mathbf{Q}_L, \mathbf{Q}_R \in \Omega$,

$$\mathbf{A}_\Phi(\mathbf{Q}_L, \mathbf{Q}_R) \cdot (\mathbf{Q}_R - \mathbf{Q}_L) = \int_0^1 \mathbf{A}(\Phi(s; \mathbf{Q}_L, \mathbf{Q}_R)) \frac{\partial \Phi}{\partial s}(s; \mathbf{Q}_L, \mathbf{Q}_R) \mathrm{d}s.$$

Enfin, les fluctuations $\mathbf{D}^{\pm}$ dans (3.B.2) sont définies comme suit

$$\mathbf{D}^{-}(\mathbf{Q}_L, \mathbf{Q}_R) = \mathbf{A}^{-}_\Phi(\mathbf{Q}_L, \mathbf{Q}_R) \cdot (\mathbf{Q}_R - \mathbf{Q}_L), \quad \mathbf{D}^{+}(\mathbf{Q}_L, \mathbf{Q}_R) = \mathbf{A}^{+}_\Phi(\mathbf{Q}_L, \mathbf{Q}_R) \cdot (\mathbf{Q}_R - \mathbf{Q}_L)$$

où

$$\mathbf{A}^{\pm}_\Phi(\mathbf{Q}_L, \mathbf{Q}_R) = \frac{1}{2}(\mathbf{A}_\Phi(\mathbf{Q}_L, \mathbf{Q}_R) \pm |\mathbf{A}_\Phi(\mathbf{Q}_L, \mathbf{Q}_R)|).$$

3.B.2 Schéma HLL

Concentrons-nous maintenant sur le schéma HLL préservant le chemin appliqué au système lagrangien $\partial_t \mathbf{LQ} + \mathbf{A}(\mathbf{LQ})\partial_\xi \mathbf{LQ} = \mathbf{S}(\mathbf{LQ}, z)$ (3.2.4). En particulier, nous montrons ici directement l'interprétation de la méthode HLL comme un schéma PVM (Polynomial Viscosity Matrix) [12], c'est-à-dire que nous remplaçons $\mathbf{A}^{\pm}_\Phi$ par

$$\mathbf{A}^{\pm}_\Phi(\mathbf{Q}_L, \mathbf{Q}_R) = \frac{1}{2}(\mathbf{A}_\Phi(\mathbf{Q}_L, \mathbf{Q}_R) \pm P_r(\mathbf{A}_\Phi(\mathbf{Q}_L, \mathbf{Q}_R)))$$

avec P_r un polynôme de degré r. Dans le cas de la méthode HLL, le polynôme P_r est donné par $P(x) = \alpha_0 + \alpha_1 x$ avec les coefficients

$$\alpha_0 = \frac{\lambda^+_{ext}|\lambda^-_{ext}| - \lambda^-_{ext}|\lambda^+_{ext}|}{\lambda^+_{ext} - \lambda^-_{ext}}, \quad \alpha_1 = \frac{|\lambda^+_{ext}| - |\lambda^-_{ext}|}{\lambda^+_{ext} - \lambda^-_{ext}}.$$

Voir par exemple [12].

Par conséquent, si on considère le système lagrangien $\partial_t \mathbf{LQ} + \mathbf{A}(\mathbf{LQ})\partial_\xi \mathbf{LQ} = \mathbf{0}$ sans terme source et avec les inconnues $\mathbf{LQ}$, nous sommes en mesure d'obtenir la forme finale suivante pour les fluctuations

$$\mathbf{D}^{\pm}_{i+1/2} = \frac{1}{2}\Big(\mathbf{A}_{i+1/2}(\mathbf{LQ}_{i+1} - \mathbf{LQ}_i)\Big) \pm \frac{1}{2} P_{i+1/2}(\mathbf{A}_{i+1/2})\Big(\mathbf{LQ}_{i+1} - \mathbf{LQ}_i\Big)$$

ou de manière équivalente

$$\mathbf{D}^{\pm}_{i+1/2} = \\ \frac{1}{2}\Big(\mathbf{A}_{i+1/2}(\mathbf{LQ}_{i+1} - \mathbf{LQ}_i)\Big) \pm \frac{1}{2}\Big(\alpha_{0,i+1/2}(\mathbf{LQ}_{i+1} - \mathbf{LQ}_i) + \alpha_{1,i+1/2}\mathbf{A}_{i+1/2}(\mathbf{LQ}_{i+1} - \mathbf{LQ}_i)\Big).$$

En rappelant que $\mathbf{A}(\mathbf{LQ}) = \dfrac{\partial \mathbf{F}(\mathbf{LQ})}{\partial \mathbf{LQ}} + \mathbf{B}(\mathbf{LQ})$, nous pouvons également exprimer les fluctuations sous la forme

$$\mathbf{D}^{\pm}_{i+1/2} = \frac{1}{2}\Big(\mathbf{F}(\mathbf{LQ}_{i+1}) - \mathbf{F}(\mathbf{LQ}_i) + \mathbf{B}_{i+1/2}(\mathbf{LQ}_{i+1} - \mathbf{LQ}_i)\Big) \pm \\ \pm \frac{1}{2}\Big(\alpha_{0,i+1/2}(\mathbf{LQ}_{i+1} - \mathbf{LQ}_i) + \alpha_{1,i+1/2}\Big(\mathbf{F}(\mathbf{LQ}_{i+1}) - \mathbf{F}(\mathbf{LQ}_i) + \mathbf{B}_{i+1/2}(\mathbf{LQ}_{i+1} - \mathbf{LQ}_i)\Big)\Big).$$

De plus, puisque dans notre cas $\lambda^-_{ext} = -\lambda^+_{ext}$, il est clair que nous avons toujours $\alpha_1 = 0$.

Enfin, il est simple d'inclure le terme source de la topographie dans la méthode numérique. En effet, quelques calculs nous donnent

$$\mathbf{D}^{\pm}_{i+1/2} = \frac{1}{2}\Big(\mathbf{A}_{i+1/2}(\mathbf{LQ}_{i+1} - \mathbf{LQ}_i) - \mathbf{S}_{i+1/2}(z_{i+1} - z_i)\Big) \pm \\ \pm \frac{1}{2} P_{i+1/2}(\mathbf{A}_{i+1/2})\Big((\mathbf{LQ}_{i+1} - \mathbf{LQ}_i) - \mathbf{A}^{-1}_{i+1/2}\mathbf{S}_{i+1/2}(z_{i+1} - z_i)\Big).$$

En pratique, nous exploitons la formule suivante

$$\mathbf{D}^{\pm}_{i+1/2} = \frac{1}{2}\Big(\mathbf{F}(\mathbf{LQ}_{i+1}) - \mathbf{F}(\mathbf{LQ}_i) + \mathbf{B}_{i+1/2}(\mathbf{LQ}_{i+1} - \mathbf{LQ}_i) - \mathbf{S}_{i+1/2}(z_{i+1} - z_i)\Big) \pm \\ \pm \frac{1}{2}\Big(\alpha_{0,i+1/2}(\mathbf{L\tilde{Q}}_{i+1} - \mathbf{L\tilde{Q}}_i) + \alpha_{1,i+1/2}\Big(\mathbf{F}(\mathbf{LQ}_{i+1}) - \mathbf{F}(\mathbf{LQ}_i) + \mathbf{B}_{i+1/2}(\mathbf{LQ}_{i+1} - \mathbf{LQ}_i) \\ - \mathbf{S}_{i+1/2}(z_{i+1} - z_i)\Big)\Big)$$

avec $\mathbf{L\tilde{Q}} = (L_1, L_1h_1, L_1h_1u_1, L_2, L_2h_2 + z, L_2h_2u_2)^t$.

Chapter 4

Explorant différentes possibilités de schémas de type Lagrange-projection équilibres et de second ordre pour les équations de Saint-Venant-Exner

La première partie de ce chapitre est consacré à l'approximation numérique du système de Saint-Venant-Exner en une dimension d'espace, où l'équation d'Exner exprime l'évolution dans le temps du lit de sédiments. Les formules de Grass et de Meyer-Peter&Müller sont prises en compte pour modéliser les contributions du flux de transport de sédiments. Une fois encore, le schéma numérique est basé sur le formalisme Lagrange-projection qui consiste à décomposer le modèle mathématique en un système acoustique et un système du transport. Dans ce chapitre, nous étudions trois stratégies numériques différentes pour discrétiser l'équation d'Exner. En particulier, cette dernière est prise en compte soit dans les deux étapes acoustique et transport, soit uniquement au niveau acoustique, soit uniquement au niveau du transport. Toutefois, nous verrons que ces trois stratégies numériques sont faiblement couplées. Grâce à cela, les méthodes et leurs extensions du second ordre sont équilibres, notamment ils préservent les états stationnaires "lac au repos" et "lit à pente constante".

Au cours de cette thèse, l'étude des schémas de type Lagrange-projection pour le modèle de Saint-Venant-Exner a été réalisée en deux étapes. D'abord, une stratégie plus simple (équation d'Exner dans l'étape du transport) a été développée en 1D et 2D et présentée dans un premier article. Ensuite, les deux autres approches faiblement couplées ont été proposées dans un deuxième manuscrit. Pour cette raison, dans la deuxième partie de ce chapitre, nous incluons la version 2D du modèle et d'une des stratégies numériques.

En conclusion, dans ce chapitre, nous résumons deux publications dont les références sont:

- C. Chalons. and A. Del Grosso, *A second-order well-balanced Lagrange-projection numerical scheme for Shallow Water Exner equations in 1D and 2D*. 2022. À paraître dans Communications in Mathematical Sciences.

- C. Chalons. and A. Del Grosso *Exploring different possibilities for second-order well-balanced Lagrange-projection numerical schemes applied to shallow water Exner equations.* International Journal for Numerical Methods in Fluids. 1- 31, 2022. https://doi.org/10.1002/fld.5064

Pour conclure l'étude des méthodes de type Lagrange-projection appliquées au système

de Saint-Venant-Exner, nous proposons également une méthode numérique du premier ordre entièrement couplée dans l'annexe 4.A. Toutefois, nous verrons que la propriété équilibre n'est pas facile à préserver dans un tel cas.

Exploring different possibilities for second-order well-balanced Lagrange-projection numerical schemes applied to shallow water Exner system

Abstract

This work is devoted to the numerical approximation of the shallow water Exner system. In the first part of this chapter, we investigate three different numerical strategies to discretize the Exner equation, which expresses the time evolution of bed sediments. The numerical schemes are all based on the Lagrange-projection formalism which consists in splitting the mathematical model into the acoustic and transport systems. In particular, the Exner equation is taken into account either in both the acoustic and transport steps, or only at the acoustic or transport level. The methods and their second-order extensions are designed in such a way to satisfy the well-balanced property, namely the "lake at rest" and the "constant bed slope" steady states.

Then, in the second part of this chapter, we present the 2D extension of the mathematical model and of one numerical strategy (Exner equation in the transport step).

4.1 Introduction and governing equations

This work considers several second-order and well-balanced Lagrange-projection schemes applied to the shallow water system with moving topography. Lagrange-projection schemes consist in splitting the acoustic and transport waves of the model in two different systems (and steps). This expedient reveals itself to be very useful in practice, for instance in subsonic regimes, where the acoustic waves are the reason of the restrictive CFL condition that has to be used to have a stable numerical scheme. Indeed, the Lagrange-Projection (LP) decomposition makes possible to implicitly approximate only the acoustic system and, thus, to circumnavigate the problem of restrictive time steps. Furthermore, this kind of decomposition can also be interpreted as a Lagrange-projection one, in the sense that the system under consideration is first written in Lagrangian coordinates and solved. Then, the Lagrangian solution is projected back into Eulerian coordinates.

Nowadays, the Lagrange-projection approach have been studied in order to satisfy different properties and as applied to several models. Giving few examples, we refer for instance to [18], where all-regime first-order explicit and semi-implicit Lagrange-projection schemes have been applied to the gas dynamics model, or to [19], where the scheme was extended

to the 2D two-phase flows model. Another possible reference is [11], where an asymptotic-preserving LP scheme has been used to approximate low Mach number flows of the barotropic Euler equations. On the other hand, when it comes to the design of LP schemes for the shallow water system, we can refer to [20] for an implicit well-balanced first-order scheme, to [14] for a fully well-balanced first-order explicit method and finally to [45] for high-order fully well-balanced schemes. Last but not least, and without being exhaustive, we refer for instance to [27, 40, 9, 13, 25] for other interesting studies in this framework.

Our main objective is to focus on the topography discretization and to compare several natural LP approaches in which the bed level is taken into account either partially in both steps, or entirely in one of the two steps. Then, in the second part of this chapter, we shall present the 2D extension of the shallow water Exner model and of one numerical strategy.

We highlight that the Saint-Venant equations with moving topography have often been solved numerically by means of splitting methods, meaning that the hydrodynamic and morphologic components are separately treated. Indeed, such methods are easier to implement, although they are known to occasionally produce spurious oscillations in the numerical results. Such oscillations are mainly related to differences in the eigenstructures of the shallow water equations with and without moving topography, for which we directly refer to [23]. However, it is not even necessary to consider a fully coupled scheme in order to avoid this problem; indeed a weak coupling of the equations at the numerical level can lead to satisfying results, see for instance [4]. In this work, a three-waves approximate Riemann solver has been described and then re-interpreted as a splitting strategy. Moreover, they stated that one change in the fluid solver (there, the value of the wave speeds in the approximate Riemann solver) can give place to splitting methods without spurious oscillations. We also refer to [44], where a weakly coupled method based on the HLL scheme is presented. Another possible reference is [43], in which the authors describe two methods based on the Roe approach. In this paper, they present not only a decoupled approach where the oscillations are partly stabilized by controlling the stability region, but also a fully-coupled scheme. Indeed, progresses to obtain fully coupled method for the shallow water Exner system have also been made. See for instance [47], where a Roe-type first-order scheme was used, or [15], in which the authors presented a path-conservative Roe method and its high-order extension equipped with flux limiters. Without being exhaustive, see for instance [29, 42, 28] for further coupled schemes applied to the shallow water Exner system.

Let us now present the model we are interested in, namely the shallow water Exner system. The first two equations (the shallow water system) have been extensively used to describe the time evolution in time of fluid flows, for instance in rivers or coastal areas. As for the third equation, it simulates the bedload sediment transport due to the mechanical action of the fluid. As such, the system reads

$$\begin{cases} \partial_t h + \partial_x(hu) = 0 \\ \partial_t(hu) + \partial_x(hu^2 + p) + gh\partial_x z = -ghS_f \\ \partial_t z + \zeta \partial_x q_b = 0, \end{cases} \tag{4.1.1}$$

where $h(x,t) > 0$ is the water depth, $u(x,t)$ is the averaged velocity, $z(x,t)$ is the bed level and $H = h + z$ is the free surface elevation. Then, $p = \dfrac{gh^2}{2}$ is the pressure term with g the gravitational acceleration, $q_b = q_b(h,u)$ is the solid transport discharge and ζ is a parameter such that $\zeta = \dfrac{1}{1-\rho_0}$ with ρ_0 the porosity of the sediment layer. Moreover, $-ghS_f$ represents the Manning friction term with $S_f = \dfrac{\mu_f^2 |u| u}{R_h^{4/3}}$, μ_f Manning roughness coefficient, $R_h = \dfrac{Lh}{L+2h}$

hydraulic radius and L length of the channel. Observe that this definition of R_h is relevant only for rectangular channels. In compact form, we have

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \mathbf{A}(\mathbf{Q})\partial_x \mathbf{Q} = -ghS_f \mathbf{E}_2$$

where $\mathbf{E}_2 = (0, 1, 0)^t$,

$$\mathbf{Q} = \begin{pmatrix} h \\ hu \\ z \end{pmatrix}, \quad \mathbf{F}(\mathbf{Q}) = \begin{pmatrix} hu \\ hu^2 + p \\ \zeta q_b \end{pmatrix}, \quad \mathbf{A}(\mathbf{Q}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & gh \\ 0 & 0 & 0 \end{pmatrix}.$$

For more details about shallow-water equations with and without moving topography, we refer to [1, 12, 15, 7, 46, 44, 42] and [51, 2, 8, 14] respectively.

Let us know focus on the solid transport discharge q_b, which can be formulated in different ways depending on the characteristics of the sediment and the flow, see for instance [34]. A frequently used formulation is the well-known Grass model, which expresses the instantaneous sediment transport as a power law of the averaged velocity u, namely

$$q_b = A_g u|u|^{m_g - 1}, \quad 1 \leq m_g \leq 4. \tag{4.1.2}$$

In practice we take $m_g = 3$, while the parameter $A_g \in [0, 1]$ is computed using empirical relationships based on local properties (grain size, cinematic viscosity...). Moreover, A_g expresses how strong the interaction between the flow and the sediment is. For instance, the interaction is considered weak if A_g is of order 10^{-3} or smaller, while for values of order 10^{-1} the flow is said to be highly erosive, see for instance [15, 43, 47, 41, 4] for further details. It is important to point out that, when using the Grass formulation, we are implicitly assuming that the bed sediments start moving as soon as the velocity of the water is different from zero. However, in other formulations it is usually assumed that a critical value has to be overcome. Let us see the details considering for instance the Meyer-Peter&Müller (MPM) formulation. The latter is given by

$$q_b = 8Qsgn(u)(\theta^* - \theta_c^*)_+^{\frac{3}{2}} \quad \text{with} \quad \theta^* = \frac{u_*^2}{sgd} \quad \text{and} \quad u_*^2 = \frac{g\mu_f^2 u^2}{h^{\frac{1}{3}}} \tag{4.1.3}$$

where $Q = d\sqrt{gsd}$ is the characteristic discharge with relative density s and sediment diameter d. Then, θ^* represents the non-dimensional shear stress and determines the movement of the sediments. Indeed, only if it is bigger of the critical stress value θ_c^*, q_b is different from zero. Here we do not present further formulations but many have been proposed, see again the previous references. We also highlight that each formulation usually has its own range of applications which depends on the flow and sediment characteristics. For instance, the MPM formula is only used for weakly erosive flow, see directly [15] for details about the range of parameter values.

Finally, let us specify that, depending on the particular form of q_b, the convective part of system (4.1.1) could be strictly hyperbolic or not. In particular, in [15] it has been proved that the use of the Grass formula leads to a strictly hyperbolic system with all real eigenvalues. Indeed, defining the quantities

$$a_1 = -2u, \quad a_2 = u^2 - c^2(1 + \zeta \partial_{hu} q_b) \quad \text{and} \quad a_3 = -\zeta c^2 \partial_h q_b,$$

one can easily see that the eigenvalues are given by the solution of the following equation

$$\lambda^3 + a_1\lambda^2 + a_2\lambda + a_3 = 0.$$

Hence, the three eigenvalues read

$$\lambda_k = 2\sqrt{-p}\cos\left(\frac{\theta + 2k\pi}{3}\right) - \frac{a_1}{3} \quad \text{with} \quad k = 0, 1, 2. \tag{4.1.4}$$

where

$$p = \frac{3a_2 - a_1^2}{9}, \quad r = \frac{9a_1a_2 - 27a_3 + 2a_1^3}{54} \quad \text{and} \quad \theta = \arccos\left(\frac{r}{\sqrt{-p^3}}\right).$$

In order to have real eigenvalues we need $p^3 + r^2 \leq 0$, which can be proved in the case of the Grass model.

Last but not least, we are interested in numerical schemes able to preserve the stationary solutions of the system. This property is generally not trivial to satisfy but, at the same time, it is critical if we want to obtain accurate numerical methods which do not present spurious oscillations in their results (when close to stationary solutions). If such a property is met, the numerical scheme is said to be well-balanced. In our specific case, it means the preservations of the following steady states,

$$q = hu = \text{constant} = q_0, \quad \partial_x\Big(\frac{q_0^2}{2h^2} + g(h+z)\Big) + gS_f = 0 \quad \text{and} \quad q_b = \text{constant}.$$

Here, we are only interested in preserving two particular steady states, namely the so-called "lake at rest" equilibrium with zero-velocity

$$u = 0, \quad h + z = \text{constant}, \tag{4.1.5}$$

and the "constant bed slope" equilibrium

$$\partial_x h = \partial_x u = \partial_{xx} z = 0, \quad \partial_x z + S_f = 0. \tag{4.1.6}$$

Let us observe that, when using the Grass formulation, these are the only two possible steady states. Being the well-balancedness of the scheme a crucial property, many studies have been done in this sense, here we refer for instance to [2, 4, 8, 38] and the references herein. See also [14, 20, 45] for well-balanced methods in the Lagrange-projection formalism.

Chapter outline. We now give the structure of the chapter. In the next section, we briefly present the Lagrange-projection splitting strategy in both Eulerian and Lagrangian variables for the shallow water system, thus we consider the evolution equations of h and hu. Then, in section 4.3, we explain three different strategies to treat the topography equation. In particular, details for the approximate Riemann solver for the acoustic systems are given. Subsequently, we present the numerical schemes both at first and second-order of accuracy in sections 4.4 and 4.5 respectively. Section 4.6 is a summary of the properties of each scheme. We show several numerical evidences to validate our 1D numerical schemes in section 4.7. Then, in section 4.8, we present the 2D extension of the shallow water Exner model and of one of the numerical strategy, namely the one that directly updates the bed elevation in the transport step. Concluding remarks are given in section 4.9.

4.2 Splitting operator for the shallow water system

This section focuses only on the first two equations of system (4.1.1), namely the updating equations for the water height h and discharge hu. Details for the bed level approximation will be given later. Thus, here we explain the decomposition which entails the splitting of the Saint-Venant system into two different ones, the so-called acoustic and transport systems. The former accounts for the acoustic phenomena and topography variations, while the latter focuses on the transport effects. Note that here we neglect the friction term; its contribution will be included directly at the end of the numerical methods, see section 4.4.

Then, we observe that the first two equations of the model can be reformulated as

$$\begin{cases} \partial_t h + h\partial_x u + u\partial_x h = 0 \\ \partial_t(hu) + hu\partial_x u + u\partial_x(hu) + \partial_x p = -gh\partial_x z, \end{cases}$$

where we used the chain rule for space derivatives. Therefore, the so-called acoustic and transport systems are given by

$$\begin{cases} \partial_t h + h\partial_x u = 0 \\ \partial_t(hu) + hu\partial_x u + \partial_x p = -gh\partial_x z \end{cases} \tag{4.2.1}$$

and

$$\begin{cases} \partial_t h + u\partial_x h = 0 \\ \partial_t(hu) + u\partial_x(hu) = 0 \end{cases}$$

respectively, where the latter can be reinterpreted as

$$\partial_t X + u\partial_x X = 0$$

with X either $X = h$ or $X = hu$. We also observe that system (4.2.1) can be expressed as

$$\begin{cases} \partial_t \tau - \partial_m u = 0 \\ \partial_t u + \partial_m p = -\dfrac{g}{\tau}\partial_m z \end{cases}$$

where $\tau = \dfrac{1}{h}$ is the specific volume and the mass variable m is such that $\dfrac{1}{h}\partial_x = \partial_m$. See for instance [14, 20] for additional details about this decomposition applied to the shallow water system. It is then clear that the numerical method would sum up to first solve the acoustic system and then the transport equations, using the acoustic solution as initial data. However, while it is well known how to decompose the shallow water system, this is not true when it comes to the Exner equation. Indeed, one could easily imagine at least three possibilities for numerical treatment. The first one would split the Exner equation inside both steps, as we just did for h and hu. The second one would consider z at the acoustic level, and the last one would directly take into account z in the transport step. The aim of the present contribution is indeed to compare these three approaches, both at first and second order accuracy.

That being said and before going into further details, it is convenient to first introduce the Lagrangian coordinates. We first define the fluid particle ξ and the characteristic curves

$$\begin{cases} \dfrac{\partial x}{\partial t}(\xi, t) = u(x(\xi, t), t) \\ x(\xi, 0) = \xi \end{cases}$$

which define the trajectory : $t \to x(\xi, t)$ of ξ as the time goes on. Therefore, any function : $(x, t) \to \varphi(x, t)$ in Eulerian coordinates can be written in Lagrangian coordinates,

$$\bar{\varphi}(\xi, t) = \varphi(x(\xi, t), t).$$

Introducing now the volume ratio

$$L(\xi, t) = \frac{\partial x}{\partial \xi}(\xi, t)$$

such that

$$\begin{cases} \dfrac{\partial L}{\partial t}(\xi, t) = \partial_\xi u(x(\xi, t), t) \\ L(\xi, 0) = 1, \end{cases} \tag{4.2.2}$$

it clearly follows

$$\partial_t L(\xi,t) = \partial_\xi u(x(\xi,t),t) = \partial_\xi \bar{u}(\xi,t),$$

and thus

$$\partial_\xi \bar{\varphi}(\xi,t) = L(\xi,t)\partial_x\varphi(x,t) \quad \text{and} \quad \partial_t\bar{\varphi}(\xi,t) = \partial_t\varphi(x,t) + u(x,t)\partial_x\varphi(x,t).$$

Focusing first on the governing equations for h and hu, it is easy to show that their counterpart in Lagrangian coordinates reads

$$\begin{cases} \partial_t(L\overline{h}) = 0 \\ \partial_t(L\overline{hu}) + \partial_\xi \bar{p} = -g\bar{h}\partial_\xi \bar{z}. \end{cases} \tag{4.2.3}$$

In the following sections, we shall sometimes omit the bar over the Lagrangian functions to avoid cumbersome notations. Hence, the two-steps (acoustic and transport) numerical method would now consist of solving the Lagrangian system (4.2.3) and then projecting the solution into Eulerian coordinates. For further details about this decomposition, we refer the reader to [45] and the references therein. Let us now discuss the three different strategies proposed for the Exner equation.

4.3 Treatment of the Exner equation

The treatment of the Exner equation is an important issue due to the complexity of the full coupled system. It is known that a fully decoupled scheme, which consists of updating the topography independently from the first two equations of the model, generally produces spurious oscillations in the numerical solutions, see for instance [23]. However and in order to avoid this problem, a weak coupling of the equations can lead to satisfying numerical results, see [4]. In this work, we mainly focus on weakly coupled numerical schemes. In brief, we are going to take into account the flow and sediment interactions in three ways:

1. The usual acoustic-transport splitting is considered for the topography equation and, therefore, the bed level is taken into account in both steps;

2. The topography is accounted for only at the Lagrangian level;

3. The topography is only updated in the transport step. This approach resembles the usual splitting (and therefore decoupled) method for shallow water Exner system but, as we will see, spurious oscillations are often not observed.

4.3.1 Update the bed level in both steps

The first strategy is to split the bed level evolution equation

$$\partial_t z + \zeta \partial_x q_b = 0$$

into the following two equations, namely

$$\partial_t z - u\partial_x z + \zeta\partial_x q_b = 0$$

and

$$\partial_t z + u\partial_x z = 0.$$

Therefore, the complete acoustic system is now given by

$$\begin{cases} \partial_t h + h\partial_x u = 0 \\ \partial_t (hu) + hu\partial_x u + \partial_x p + gh\partial_x z = 0 \\ \partial_t z - u\partial_x z + \zeta \partial_x q_b = 0, \end{cases} \tag{4.3.1}$$

while the transport system is simply formulated as

$$\partial_t X + u\partial_x X = 0, \tag{4.3.2}$$

where now we have not only $X = h, hu$, but also $X = z$. It is thus clear that, with this approach, the evolution of the bed level is taken into account in both the acoustic and transport steps. Then, we note that the acoustic system (4.3.1) can be easily reformulated as follows

$$\begin{cases} \partial_t \tau - \partial_m u = 0 \\ \partial_t u + \partial_m p + \dfrac{g}{\tau}\partial_m z = 0 \\ \partial_t z - \dfrac{u}{\tau}\partial_m z + \zeta \dfrac{1}{\tau}\partial_m q_b = 0, \end{cases} \tag{4.3.3}$$

exploiting once again the notation $\tau = 1/h$ and $\tau\partial_x = \partial_m$. Alternatively, the shallow water Exner system (4.1.1) in Lagrangian coordinates reads

$$\begin{cases} \partial_t (L\overline{h}) = 0 \\ \partial_t (L\overline{hu}) + \partial_\xi \bar{p} + g\bar{h}\partial_\xi \bar{z} = 0 \\ \partial_t (L\overline{z}) - \partial_\xi (\overline{zu}) + \zeta \partial_\xi \overline{q_b} = 0. \end{cases} \tag{4.3.4}$$

4.3.1.1 Approximate Riemann solver

In order to approximate the solutions of system (4.3.4) using a Godunov-type method, in this section we define an approximate Riemann solver for equations (4.3.3) associated with initial data

$$(\tau, u, z)^T (m, t = 0) = \begin{cases} (\tau_L, u_L, z_L)^T & \text{if} \quad m < 0 \\ (\tau_R, u_R, z_R)^T & \text{if} \quad m > 0. \end{cases}$$

The idea is to base the approximate Riemann solver on a relaxation formulation, see for instance [10, 20] for more details. Thus, as a starting point, we considered the following relaxation system of the whole (not only acoustic) Saint-Venant-Exner system (4.1.1),

$$\begin{cases} \partial_t h + \partial_x (hu) = 0 \\ \partial_t (hu) + \partial_x (hu^2 + \Pi) + gh\partial_x z = 0 \\ \partial_t z + \partial_x \Omega = 0 \\ \partial_t \Pi + u\partial_x \Pi + a^2 \partial_x u = 0 \\ \partial_t \Omega + 2u\partial_x \Omega + (b^2 - \dfrac{h^2}{u^2})\partial_x z = 0, \end{cases} \tag{4.3.5}$$

which was proposed in [1]. Observe that we introduced two new variables, Π and Ω, which can be interpreted as a linearization of p and ζq_b respectively. Then, a and b are two constant parameters. However, in order to build a relaxation system for the acoustic system and, at the same time, to define a well-balanced approximate Riemann solver, it was convenient to modify

system (4.3.5) as follows

$$\begin{cases} \partial_t \tau - \partial_m u = 0 \\ \partial_t u + \partial_m \Pi = -\dfrac{g}{\tau}\partial_m z \\ \partial_t z - \dfrac{u}{\tau}\partial_m z + \dfrac{1}{\tau}\partial_m \Omega = 0 \\ \partial_t \Pi + a^2 \partial_m u = 0 \\ \partial_t \Omega + \dfrac{u}{\tau}\partial_m \Omega + u^2(b^2\tau - \dfrac{1}{\tau})\partial_m z = 0. \end{cases} \tag{4.3.6}$$

Thus, the idea would be to take the exact Riemann solution of the latter relaxation system (4.3.6) as the approximate solution of either equations (4.3.3) or system (4.3.4). We also underline that the initial data now are

$$(\tau, u, z, \Pi, \Omega)^T(m, t=0) = \begin{cases} (\tau_L, u_L, z_L, \Pi_L, \Omega_L)^T & \text{if} \quad m < 0 \\ (\tau_R, u_R, z_R, \Pi_R, \Omega_R)^T & \text{if} \quad m > 0 \end{cases}$$

and they are taken at equilibrium, that is to say such that

$$\Pi_{L,R} = \frac{g}{2} h_{L,R}^2 \quad \text{and} \quad \Omega_{L,R} = \zeta(q_b)_{L,R}.$$

Furthermore, the parameters a and b are chosen in such a way to ensure the stability of the relaxation system. In particular, we ask for the validity of the so-called sub-characteristic condition

$$a \geq h\sqrt{\partial_h p} \quad \text{and} \quad u^2 b^2 \geq (hu)^2 + g h^2 \zeta \partial_u q_b, \quad b > 0, \tag{4.3.7}$$

which is clearly different from the one used for the original relaxation system (4.3.5). The relaxation system (4.3.6) can be written in compact form as

$$\partial_t \mathbf{U} + \mathbf{B}(\mathbf{U})\partial_m \mathbf{U} = \mathbf{S}(\mathbf{U})$$

with

$$\mathbf{U} = \begin{pmatrix} \tau \\ u \\ z \\ \Pi \\ \Omega \end{pmatrix} \quad \mathbf{B}(\mathbf{U}) = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{u}{\tau} & 0 & \frac{1}{\tau} \\ 0 & a^2 & 0 & 0 & 0 \\ 0 & 0 & u^2(b^2\tau - \frac{1}{\tau}) & 0 & \frac{u}{\tau} \end{pmatrix} \quad \text{and} \quad \mathbf{S}(\mathbf{U}) = \begin{pmatrix} 0 \\ \tilde{s} \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Observe that the topography term

$$\tilde{s} = -\frac{g}{\tau}\partial_m z$$

is taken into account as a source term and not included in the convective matrix $\mathbf{B}$. This is also a critical point in order to be able to define a well-balanced approximate Riemann solver, even if it is also the reason why the resulting numerical method will only be weakly coupled and not fully coupled. With this in mind, it is easy to show that the matrix $\mathbf{B}$ has five real eigenvalues given by $\lambda_0 = 0, \lambda_a^{\pm} = \pm a, \lambda_b^{\pm} = \pm |u| b$, and that the associated characteristic fields are all linearly degenerate. This property is well-known to make the resolution of the Riemann problem straightforward using the continuity of the Riemann invariants across each wave, see for instance [1] for more details. However, in this case the eigenvalues are not ordered a priori, so that at a continuous level there exist two different cases depending on whether $a < |u| b$ or not (recall that a and b are positive). In practice, we will distinguish between the following two cases: $a < |u_L| b, a < |u_R| b$ and its negation.

4.3.1.2 The case $a < |u_L|b$ and $a < |u_R|b$

In this first case, easy calculations show that the solution of the Riemann problem reads

$$\hat{\mathbf{U}}(\frac{m}{t};\mathbf{U}_L,\mathbf{U}_R) = \begin{cases} \mathbf{U}_L & \text{if} \quad \frac{m}{t} < \lambda_b^- = -|u_L|b \\ \mathbf{U}^*_{b,L} & \text{if} \quad -|u_L|b < \frac{m}{t} < \lambda_a^- = -a \\ \mathbf{U}^*_{a,L} & \text{if} \quad -a < \frac{m}{t} < \lambda_0 = 0 \\ \mathbf{U}^*_{a,R} & \text{if} \quad 0 < \frac{m}{t} < \lambda_a^+ = a \\ \mathbf{U}^*_{b,R} & \text{if} \quad a < \frac{m}{t} < \lambda_b^+ = |u_R|b \\ \mathbf{U}_R & \text{if} \quad \frac{m}{t} > |u_R|b. \end{cases}$$

with

$$\mathbf{U}^*_{b,L} = \begin{pmatrix} \tau_L \\ u_L \\ z^* \\ \Pi_L \\ \Omega^* \end{pmatrix}, \quad \mathbf{U}^*_{a,L} = \begin{pmatrix} \tau_L^* \\ u^* \\ z^* \\ \Pi_L^* \\ \Omega^* \end{pmatrix}, \quad \mathbf{U}^*_{a,R} = \begin{pmatrix} \tau_R^* \\ u^* \\ z^* \\ \Pi_R^* \\ \Omega^* \end{pmatrix}, \quad \text{and} \quad \mathbf{U}^*_{b,R} = \begin{pmatrix} \tau_R \\ u_R \\ z^* \\ \Pi_R \\ \Omega^* \end{pmatrix}. \tag{4.3.8}$$

On one hand we have

$$\begin{cases} \tau_L^* = \tau_L + \dfrac{1}{a}(u^* - u_L) \\ \tau_R^* = \tau_R - \dfrac{1}{a}(u^* - u_R) \\ u^* = \dfrac{1}{2}(u_L + u_R) - \dfrac{1}{2a}(\Pi_R - \Pi_L) - \dfrac{\mathcal{M}}{2a} \\ \Pi_{*L} = \dfrac{1}{2}(\Pi_L + \Pi_R) - \dfrac{a}{2}(u_R - u_L) + \dfrac{\mathcal{M}}{2} \\ \Pi_{*R} = \dfrac{1}{2}(\Pi_L + \Pi_R) - \dfrac{a}{2}(u_R - u_L) - \dfrac{\mathcal{M}}{2} \end{cases} \tag{4.3.9}$$

where the discretization $\mathcal{M}$ of the topography source term reads

$$\mathcal{M} = \frac{g}{2}(\frac{1}{\tau_L} + \frac{1}{\tau_R})(z_R - z_L) \tag{4.3.10}$$

and, on the other hand,

$$\begin{cases} z^* = \dfrac{|u_R|(\text{sign}(u_R) + b\tau_R)z_R - |u_L|(\text{sign}(u_L) - b\tau_L)z_L}{|u_R|(\text{sign}(u_R) + b\tau_R) - |u_L|(\text{sign}(u_L) - b\tau_L)} \\ \qquad - \dfrac{\Omega_R - \Omega_L}{|u_R|(\text{sign}(u_R) + b\tau_R) - |u_L|(\text{sign}(u_L) - b\tau_L)} \\ \Omega^* = \dfrac{\Omega_R + \Omega_L}{2} + \dfrac{|u_R|}{2}(\text{sign}(u_R) + b\tau_R)(z^* - z_R) + \dfrac{|u_L|}{2}(\text{sign}(u_L) - b\tau_L)(z^* - z_L). \end{cases} \tag{4.3.11}$$

4.3.1.3 The case $a \geq |u_L|b$ or $a \geq |u_R|b$

In this case the structure of the Riemann solution is the same but the waves are expected to be ordered in a different way. More precisely, assuming that $a > |u^*|b$, we now have

$$\hat{\mathbf{U}}(\frac{m}{t};\mathbf{U}_L,\mathbf{U}_R) = \begin{cases} \mathbf{U}_L & \text{if} \quad \frac{m}{t} < -a \\ \mathbf{U}^*_{a,L} & \text{if} \quad -a < \frac{m}{t} < -u^*b \\ \mathbf{U}^*_{b,L} & \text{if} \quad -u^*b < \frac{m}{t} < 0 \\ \mathbf{U}^*_{b,R} & \text{if} \quad 0 < \frac{m}{t} < u^*b \\ \mathbf{U}^*_{a,R} & \text{if} \quad u^*b < \frac{m}{t} < a \\ \mathbf{U}_R & \text{if} \quad \frac{m}{t} > a. \end{cases}$$

where

$$\mathbf{U}^*_{a,L} = \begin{pmatrix} \tau^*_L \\ u^* \\ z_L \\ \Pi^*_L \\ \Omega_L \end{pmatrix}, \quad \mathbf{U}^*_{b,L} = \begin{pmatrix} \tau^*_L \\ u^* \\ z^* \\ \Pi^*_L \\ \Omega^* \end{pmatrix}, \quad \mathbf{U}^*_{b,R} = \begin{pmatrix} \tau^*_R \\ u^* \\ z^* \\ \Pi^*_R \\ \Omega^* \end{pmatrix}, \quad \text{and} \quad \mathbf{U}^*_{a,R} = \begin{pmatrix} \tau^*_R \\ u^* \\ z_R \\ \Pi^*_R \\ \Omega_R \end{pmatrix}. \tag{4.3.12}$$

Then, (4.3.9) and (4.3.10) are still valid but (4.3.11) is replaced by

$$\begin{cases} (z|u|)^* = |u^*| \dfrac{(\mathrm{sign}(u^*) + b\tau^*_R)z_R - (\mathrm{sign}(u^*) - b\tau^*_L)z_L}{b(\tau^*_R + \tau^*_L)} - \dfrac{\Omega_R - \Omega_L}{b(\tau^*_R + \tau^*_L)} \\ \Omega^* = \dfrac{\Omega_R + \Omega_L}{2} + \dfrac{1}{2}((\mathrm{sign}(u^*) + b\tau^*_R)((z|u|)^* - |u^*|z_R) + (\mathrm{sign}(u^*) - b\tau^*_L)((z|u|)^* - |u^*|z_L)). \end{cases} \tag{4.3.13}$$

Observe that assuming $a \geq |u_L|b$ or $a \geq |u_R|b$ does not necessarily imply that $a > |u^*|b$. In practice, we proceed as follows. First, we define a and b by

$$a = \max(\Delta x, h_L c_L, h_R c_R) \quad \text{and} \quad b = \max\Big(\Delta x, \sqrt{h_L^2 + g\frac{h_L^2}{u_L^2}\zeta\partial_u(q_b)_L}, \sqrt{h_R^2 + g\frac{h_R^2}{u_R^2}\zeta\partial_u(q_b)_R}\Big), \tag{4.3.14}$$

$c = \sqrt{\partial_h p}$, as a natural approximation of (4.3.7). Then, if a and b are such that $a \geq |u_L|b$ or $a \geq |u_R|b$ but $a \leq |u^*|b$, we increase the value of a and redefine it as $a = (1+\epsilon)|u^*|b$ (with typically $\epsilon = 0.01$). We highlight that, once we have redefined a, we have to recompute the value of u^*, and more generally the quantities in (4.3.8) and (4.3.9). In practice, this iterative process converges in one or two iterations. An easier (and more diffusive) option could be to define a and b such that a is automatically smaller than ub, in this way we only had to use the star values in section 4.3.1.2 and no further details would be required. For instance, this is what has been done in [1] to numerically solve system (4.3.5). However, we will see later in section 4.6.1 that to distinguish between the two cases allows us to obtain a well-balanced numerical scheme.

Notice also that, unlike (4.3.11), we define $(z|u|)^*$ instead of z^* in (4.3.13) in order to avoid any possible ambiguity related to the value of u^* which could be zero. As we will see below, this is sufficient as, in the resulting scheme, the updating formula of the topography z^{n+1-} only requires $(uz)^*$ and not z^*.

Lastly, observe that, if $u = 0$, the relaxation system (4.3.6) is not strictly hyperbolic, unlike the acoustic system (4.3.3). However, the values in the star region result to be well-defined anyway. Indeed, if $u_L = u_R = u^* = 0$, then (4.3.13) gives $(uz)^* = 0$ and $\Omega^* = 0$ as well, as $u_L = u_R = 0$ implies $\Omega_L = \Omega_R = 0$.

4.3.2 Update the bed level in the acoustic step

The second strategy takes into account the Exner equation only in the acoustic step. As a consequence, we will have a different acoustic system and therefore a different approximate Riemann solver. Let us give more details.

The acoustic system is now given by

$$\begin{cases} \partial_t h + h\partial_x u = 0 \\ \partial_t(hu) + hu\partial_x u + \partial_x p + gh\partial_x z = 0 \\ \partial_t z + \zeta\partial_x q_b = 0 \end{cases} \tag{4.3.15}$$

while the transport system is simply formulated as $\partial_t X + u\partial_x X = 0$ with $X = h, hu$, together with $\partial_t z = 0$. Exploiting the variables $\tau = 1/h$ and $\tau\partial_x = \partial_m$, the acoustic system (4.3.15) also reads

$$\begin{cases} \partial_t \tau - \partial_m u = 0 \\ \partial_t u + \partial_m p + \dfrac{g}{\tau}\partial_m z = 0 \\ \partial_t z + \zeta \dfrac{1}{\tau}\partial_m q_b = 0. \end{cases} \tag{4.3.16}$$

4.3.2.1 Relaxation system and approximate Riemann Solver

We proceed as before to approximate the solution of (4.3.16) and to define an approximate Riemann solution based on a relaxation system. The latter is now defined as

$$\begin{cases} \partial_t \tau - \partial_m u = 0 \\ \partial_t u + \partial_m \Pi = -\frac{g}{\tau}\partial_m z \\ \partial_t z + \frac{1}{\tau}\partial_m \Omega = 0 \\ \partial_t \Pi + a^2 \partial_m u = 0 \\ \partial_t \Omega + u^2 b^2 \tau \partial_m z = 0, \end{cases}$$

with once again

$$(\tau, u, z, \Pi, \Omega)^T(m, t=0) = \begin{cases} (\tau_L, u_L, z_L, \Pi_L, \Omega_L)^T & \text{if} \quad m < 0 \\ (\tau_R, u_R, z_R, \Pi_R, \Omega_R)^T & \text{if} \quad m > 0 \end{cases}$$

the initial data taken at equilibrium, that is to say such that

$$\Pi_{L,R} = \frac{g}{2} h_{L,R}^2 \quad \text{and} \quad \Omega_{L,R} = \zeta(q_b)_{L,R}.$$

As far as the parameters a and b are concerned, we now ask for

$$a \geq h\sqrt{\partial_h p} \quad \text{and} \quad u^2 b^2 \geq g h^2 \zeta \partial_u q_b, \quad b > 0. \tag{4.3.17}$$

Considering again the topography term as a source term in the second equation, it is easy to show that the characteristic velocities of the model are still given by $\lambda_0 = 0, \lambda_a^{\pm} = \pm a, \lambda_b^{\pm} = \pm|u|b$, and that the associated characteristic fields are all linearly degenerate again. Let us now discuss the associated Riemann solution.

4.3.2.2 Modification of the well-balanced approximate Riemann solver

The procedure is exactly the same as the one presented in section 4.3.1.1, so that we do not give other details but the differences. In particular, we have to distinguish between the case $a < |u_L|b$ and $a < |u_R|b$ and its negation. In this last case, and even if it means increasing the value of a, once again we will assume that $a > |u^*|b$.

In both cases, (4.3.9) and (4.3.10) are still valid, but definitions of z^* and Ω^* in (4.3.11) and (4.3.13) are modified. More precisely and imposing again the continuity of the Riemann invariants across each wave, easy calculations show that in the case $a < |u_L|b$ and $a < |u_R|b$, formulas (4.3.11) are replaced by

$$\begin{cases} z^* = \dfrac{|u_R|\tau_R z_R + |u_L|\tau_L z_L}{|u_R|\tau_R + |u_L|\tau_L} - \dfrac{\Omega_R - \Omega_L}{b|u_R|\tau_R + b|u_L|\tau_L} \\ \Omega^* = \dfrac{\Omega_R + \Omega_L}{2} + \dfrac{b|u_R|\tau_R}{2}(z^* - z_R) - \dfrac{b|u_L|\tau_L}{2}(z^* - z_L), \end{cases} \tag{4.3.18}$$

while (4.3.13) now reads

$$\begin{cases} (z|u|)^* = |u^*|\dfrac{\tau_R^* z_R + \tau_L^* z_L}{\tau_R^* + \tau_L^*} - \dfrac{\Omega_R - \Omega_L}{b(\tau_R^* + \tau_L^*)} \\ \Omega^* = \dfrac{\Omega_R + \Omega_L}{2} + \dfrac{1}{2}(b\tau_R^*((z|u|)^* - z_R|u^*|) - b\tau_L^*((z|u|)^* - z_L|u^*|)). \end{cases} \tag{4.3.19}$$

Lastly, instead of using (4.3.14), we define a and b as follows

$$a = \max(\Delta x, h_L c_L, h_R c_R) \quad \text{and} \quad b = \max\Big(\Delta x, \sqrt{g\frac{h_L^2}{u_L^2}\zeta\partial_u(q_b)_L}, \sqrt{g\frac{h_R^2}{u_R^2}\zeta\partial_u(q_b)_R}\Big), \tag{4.3.20}$$

according to condition (4.3.17).

4.3.3 Update the bed level in the transport step

In this last strategy, the bottom height is only updated in the transport step. Thus, the associated acoustic system is now simply given by

$$\begin{cases} \partial_t h + h\partial_x u = 0 \\ \partial_t(hu) + hu\partial_x u + \partial_x p + gh\partial_x z = 0 \end{cases}$$

while the transport system has the following form,

$$\begin{cases} \partial_t h + u\partial_x h = 0 \\ \partial_t(hu) + u\partial_x(hu) = 0 \\ \partial_t z + \zeta\partial_x q_b = 0. \end{cases} \tag{4.3.21}$$

Hence, the water height h and flow hu are updated as in the classical shallow-water equations, namely both in the acoustic and transport steps. In other words, we only need to specify the discretization of the bed level, which does not affect the other two variables. For this reason, this strategy resembles a usual splitting method.

Note that in this context, the relaxation system associated with the acoustic step is nothing but

$$\begin{cases} \partial_t \tau - \partial_m u = 0 \\ \partial_t u + \partial_m p = -\dfrac{g}{\tau}\partial_m z \\ \partial_t \Pi + a^2\partial_m u = 0 \end{cases} \tag{4.3.22}$$

and the intermediate states associated with the Riemann problems are given by (4.3.9) and (4.3.10). Finally, let us observe that here the eigenvalues of the relaxation system depend only on the parameter a and not on b, with a still defined as in (4.3.14).

4.4 Numerical method

Before getting into the heart of the matter, we give few details about the time and space discretizations. Given a constant time step Δt, we define the intermediate times by $t^n = n\Delta t$ for $n \in \mathbb{N}$. Then, the mesh interfaces are $x_{j+1/2} = j\Delta x$ for $j \in \mathbb{Z}$, where Δx is the constant space step. Note that, for the Lagrangian variable ξ, we use the same discretization of the one we introduced for x, thus $\Delta x = \Delta\xi, x_{j+\frac{1}{2}} = \xi_{j+\frac{1}{2}}$ and $x_j = \xi_j\ \forall j$, where x_j is the center of the cell $[x_{j-1/2}, x_{j+1/2})$. Hence, given a variable φ, φ_j^n is its piecewise constant approximation

at each point (x_j, t^n) with $n \in \mathbb{N}$ and $j \in \mathbb{Z}$. If φ_j^n is known, we look for its approximation at the next time level t^{n+1}, namely φ_j^{n+1}. The approximate value obtained at the end of the acoustic step will be denoted by φ_j^{n+1-}.

Let us note that our numerical schemes are divided into two different steps. First we have the acoustic (Lagrangian) stage, in which we aim to numerically solve one of the systems (4.3.4), (4.3.16) or (4.2.3) depending on the strategy for the bed level equation. Then, we exploit the acoustic (Lagrangian) solution as initial condition for the transport (projection) step, in which we solve either equations (4.3.2) ($X = h, hu$ and with/without $X = z$) or system (4.3.21), again depending on the chosen strategy. If the friction term is also considered, a third step is needed in order to take into account the friction contribution.

Finally, let us specify the CFL restriction on the time step Δt in order to have stable numerical schemes. Since our methods are composed of two different steps, we obtain two different conditions, namely

$$\Delta t \leq \mathrm{CFL}_l \frac{\Delta x}{\max\limits_j \{\max(\tau_j^n, \tau_{j+1}^n) \max(a_{j+\frac{1}{2}}, (|u|b)_{j+\frac{1}{2}})\}}, \tag{4.4.1}$$

for the acoustic step and

$$\Delta t \leq \mathrm{CFL}_t \frac{\Delta x}{\max\limits_j \{u^+_{j-\frac{1}{2}} - u^-_{j+\frac{1}{2}}\}}, \tag{4.4.2}$$

for the transport step. CFL_l and CFL_t are the CFL number for the Lagrangian and transport systems respectively, and

$$u^+_{j-\frac{1}{2}} = \max(u^*_{j-\frac{1}{2}}, 0) \quad \text{and} \quad u^-_{j+\frac{1}{2}} = \min(u^*_{j+\frac{1}{2}}, 0).$$

Lastly, the final time step is taken as the minimum between the two. It is clear that the value of b in (4.4.1) depends on the numerical treatment of the bottom height z. Clearly, if z is updated in both steps, we consider formula (4.3.14) while, if the bed elevation is completely taken into account at the acoustic level, b is defined according to (4.3.20). In the case of the bed level only updated in the transport step, the acoustic time step definition depends only on a and not on b.

4.4.1 Acoustic step

As for the acoustic systems (4.3.3), (4.3.16) and (4.2.1), we have already defined approximate Riemann solver for them. Hence, we can use a classical Godunov-type method, refer to [30, 31, 37] and the references therein. Such a scheme simply consists in the juxtaposition of the approximate Riemann solutions locally defined at each interface.

Regarding the first two equations which are common to systems (4.3.3), (4.3.16) and (4.2.1), easy calculations (not reported here) give

$$\begin{cases} \tau_j^{n+1-} = \tau_j^n + \dfrac{\Delta t}{\Delta m_j}(u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}) \\ u_j^{n+1-} = u_j^n - \dfrac{\Delta t}{\Delta m_j}(\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}) - \Delta t\{\dfrac{g}{\tau}\partial_m z\}_j^n \end{cases} \tag{4.4.3}$$

where $\Delta m_{j+1/2} = (\Delta m_j + \Delta m_{j+1})/2$, $\Delta m_j = \dfrac{\tau_j^n}{\Delta x}$ and for all j

$$\{\frac{g}{\tau}\partial_m z\}_j^n = \frac{1}{2}\Big(\frac{\Delta m_{j+1/2}}{\Delta m_j}\{\frac{g}{\tau}\partial_m z\}_{j+1/2}^n + \frac{\Delta m_{j-1/2}}{\Delta m_j}\{\frac{g}{\tau}\partial_m z\}_{j-1/2}^n\Big) \tag{4.4.4}$$

with

$$\{\frac{g}{\tau}\partial_m z\}^n_{j+1/2} = \frac{\mathcal{M}^n_{j+1/2}}{\Delta m_{j+1/2}}.$$

The star values $u^*_{j+\frac{1}{2}}$, $\Pi^*_{j+\frac{1}{2}}$ and $\mathcal{M}_{j+1/2}$ are locally defined at each interface $x_{j+1/2}$ and for all j thanks to (4.3.9) and (4.3.10), obtaining

$$\begin{aligned} u^*_{j+\frac{1}{2}} &= u^*_{j+\frac{1}{2}}(\mathbf{Q}^n_j, \mathbf{Q}^n_{j+1}) = \frac{1}{2}(u^n_{j+1} + u^n_j) - \frac{1}{2a^n_{j+\frac{1}{2}}}(\Pi^n_{j+1} - \Pi^n_j) - \frac{\mathcal{M}^n_{j+1/2}}{2a_{j+1/2}} \\ \Pi^*_{j+\frac{1}{2}} &= \Pi^*_{j+\frac{1}{2}}(\mathbf{Q}^n_j, \mathbf{Q}^n_{j+1}) = \frac{1}{2}(\Pi^n_{j+1} + \Pi^n_j) - \frac{a^n_{j+\frac{1}{2}}}{2}(u^n_{j+1} - u^n_j). \end{aligned} \tag{4.4.5}$$

Furthermore, in Lagrangian coordinates, (4.4.3) reveals itself to be strictly equivalent to

$$\begin{cases} L^{n+1-}_j h^{n+1-}_j = L^n_j h^n_j \\ L^{n+1-}_j (hu)^{n+1-}_j = L^n_j (hu)^n_j - \frac{\Delta t}{\Delta x}(\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}) - \Delta t\{gh\partial_x z\}^n_j \end{cases} \tag{4.4.6}$$

where

$$L^{n+1-}_j = L^n_j + \frac{\Delta t}{\Delta x}(u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}) \quad \text{with} \quad L^n_j = 1$$

and

$$s = -gh\partial_x z, \quad s^n_j = \frac{1}{2}\Big(s^n_{j+1/2} + s^n_{j-1/2}\Big), \quad s^n_{j+1/2} = -\frac{\mathcal{M}^n_{j+1/2}}{\Delta x} \quad \forall j.$$

Let us now briefly give the update formulas for the topography.

Bed level in both steps. Considering system (4.3.4), we state

$$L^{n+1-}_j z^{n+1-}_j = L^n_j z^n_j - \frac{\Delta t}{\Delta x}((\Omega - zu)^*_{j+\frac{1}{2}} - (\Omega - zu)^*_{j-\frac{1}{2}})$$

which turns out to be equivalent to

$$z^{n+1-}_j(1 + \frac{\Delta t}{\Delta x}(u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}})) = z^n_j + \frac{\Delta t}{\Delta x}((zu)^*_{j+\frac{1}{2}} - (zu)^*_{j-\frac{1}{2}}) - \frac{\Delta t}{\Delta x}(\Omega^*_{j+\frac{1}{2}} - \Omega^*_{j-\frac{1}{2}}),$$

where the star values are easily defined from (4.3.11) and (4.3.13).

Bed level in the acoustic step. In this case, from system (4.3.15) we find

$$z^{n+1-}_j = z^n_j - \frac{\Delta t}{\Delta x}(\Omega^*_{j+\frac{1}{2}} - \Omega^*_{j-\frac{1}{2}}), \tag{4.4.7}$$

where $\Omega^*_{j\pm\frac{1}{2}}$ are defined from (4.3.18) and (4.3.19).

Bed level in the transport step. Finally, it is clear that here we simply have $z^{n+1-}_j = z^n_j$ as the bed level is completely taken into account in the transport step.

4.4.2 Transport step

Referring to [45], we can approximate $\partial_t X + u\partial_x X = 0$ by

$$X^{n+1}_j = (LX)^{n+1-}_j - \frac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}(LX)^{n+1-}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}(LX)^{n+1-}_{j-\frac{1}{2}}\big), \tag{4.4.8}$$

where

$$(LX)_{j+\frac{1}{2}}^{n+1} = \begin{cases} (LX)_j^{n+1-} & \text{if} \quad u_{j+\frac{1}{2}}^* \geq 0 \\ (LX)_{j+1}^{n+1-} & \text{if} \quad u_{j+\frac{1}{2}}^* < 0 \end{cases}$$

with either $X = h$ or $X = hu$. Notice that this formula can also be explained in an alternative way, using the Lagrangian coordinates. Indeed, defining $\hat{\xi}_{j+\frac{1}{2}}(t)$ such that for all j

$$x(\hat{\xi}_{j+\frac{1}{2}}(T), T) = x_{j+\frac{1}{2}}, \quad \text{with} \quad T \geq 0,$$

it is enough to recall that

$$X_j(t) = \frac{1}{\Delta x}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} X(x,t)dx = \frac{1}{\Delta x}\int_{x(\hat{\xi}_{j-\frac{1}{2}},t)}^{x(\hat{\xi}_{j+\frac{1}{2}},t)} X(x,t)dx = \frac{1}{\Delta x}\int_{\hat{\xi}_{j-\frac{1}{2}}}^{\hat{\xi}_{j+\frac{1}{2}}} L(\xi,t)\bar{X}(\xi,t)d\xi$$

and then split the last integral into three parts to define X_j^{n+1}, namely

$$\begin{aligned} X_j^{n+1} =& \frac{1}{\Delta x}\int_{\hat{\xi}_{j-\frac{1}{2}}}^{\xi_{j-\frac{1}{2}}} L(\xi, t^{n+1-})\bar{X}(\xi, t^{n+1-})d\xi + \\ &+ \frac{1}{\Delta x}\int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} L(\xi, t^{n+1-})\bar{X}(\xi, t^{n+1-})d\xi + \frac{1}{\Delta x}\int_{\xi_{j+\frac{1}{2}}}^{\hat{\xi}_{j+\frac{1}{2}}} L(\xi, t^{n+1-})\bar{X}(\xi, t^{n+1-})d\xi \end{aligned} \tag{4.4.9}$$

where we approximate $x_{j+\frac{1}{2}}$ at first-order,

$$x_{j+\frac{1}{2}} = x(\hat{\xi}_{j+\frac{1}{2}}(T), T) \simeq x(\hat{\xi}_{j+\frac{1}{2}}(T), 0) + T\partial_t x(\hat{\xi}_{j+\frac{1}{2}}(T), 0) \simeq \hat{\xi}_{j+\frac{1}{2}} + Tu_{j+\frac{1}{2}}^*,$$

for a fixed time $T \geq 0$. Then, it is clear that formula (4.4.8) can be seen as a first-order approximation of the integrals in (4.4.9). Note that the second-order scheme will be obtained by approximating the three integrals in (4.4.9) at second-order of accuracy in space. Since this procedure has been explained in details in [45], here we do not provide further information.

As before, let us now give the update formulas for the topography.

Bed level in both steps. In this case, similarly to the formulas for h, hu, we have for all j

$$z_j^{n+1} = (Lz)_j^{n+1-} - \frac{\Delta t}{\Delta x}\big(u_{j+\frac{1}{2}}^* z_{j+\frac{1}{2}}^{n+1-} - u_{j-\frac{1}{2}}^* z_{j-\frac{1}{2}}^{n+1-}\big)$$

with the only difference that we use the values of $z_{j+\frac{1}{2}}^{n+1-}$ and not of its Lagrangian counterpart.

Bed level in the acoustic step. Since in this numerical method the bottom height is completely considered in the acoustic step, here we simply have

$$z_j^{n+1} = z_j^{n+1-}.$$

Bed level in the transport step. In this case, it is a matter of discretizing the full Exner equation, which can be done following at least two options. On one hand, we simply update the topography as follows,

$$z_j^{n+1} = z_j^n - \zeta\frac{\Delta t}{\Delta x}\Big(u_{j+\frac{1}{2}}^*\Big(\frac{q_b(u)}{u}\Big)_{j+\frac{1}{2}}^n - u_{j-\frac{1}{2}}^*\Big(\frac{q_b(u)}{u}\Big)_{j-\frac{1}{2}}^n\Big), \tag{4.4.10}$$

with

$$\Big(\frac{q_b(u)}{u}\Big)^n_{j+1/2} = \begin{cases} \Big(\frac{q_b(u)}{u}\Big)\big(h^n_{j+1},(hu)^n_{j+1}\big) & \text{if} \quad u^*_{j+\frac{1}{2}} \leq 0 \\ \Big(\frac{q_b(u)}{u}\Big)\big(h^n_{j},(hu)^n_{j}\big) & \text{if} \quad u^*_{j+\frac{1}{2}} > 0. \end{cases} \tag{4.4.11}$$

Note that $\Big(\frac{q_b(u)}{u}\Big)^n_{j+1/2}$ can be defined with no ambiguity since we either assume the solid transport discharge to be given by the Grass formula (4.1.2) with $m_g = 3$ (and therefore u simplifies) or by the MPM formula (4.1.3) where q_b is considered null for $u = 0$. Moreover, we see that the numerical fluxes are simply evaluated using the solution at time t^n, and not the one obtained at the end of the Lagrangian step. In some sense, this means that the evolution of the topography is not coupled with the one of the hydrodynamic model from a numerical point of view. This is why we call this strategy *decoupled*.

On the other hand, we propose a *weakly coupled* strategy, in which we exploit the solution obtained at time t^{n+1-} by setting

$$z^{n+1}_j = z^n_j - \zeta\frac{\Delta t}{\Delta x}\Big(u^*_{j+\frac{1}{2}}\Big(\frac{q_b(Lu)}{Lu}\Big)^{n+1-}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}\Big(\frac{q_b(Lu)}{Lu}\Big)^{n+1-}_{j-\frac{1}{2}}\Big). \tag{4.4.12}$$

and $\Big(\frac{q_b(Lu)}{Lu}\Big)^{n+1-}_{j+\frac{1}{2}}$ defined as the corresponding value in the (x_j, t^{n+1-}) or (x_{j+1}, t^{n+1-}) point respectively if $u^*_{j+\frac{1}{2}} > 0$ or $u^*_{j+\frac{1}{2}} \leq 0$ and according to (4.4.11). Since a definition of $(Lu)^{n+1-}$ is needed, we observe that an evolution equation for Lu in the Lagrangian step reads

$$\partial_t(Lu) - \partial_\xi\frac{u^2}{2} = -g\partial_\xi(h+z),$$

and can be discretize as

$$(Lu)^{n+1-}_j = (Lu)^n_j + \frac{\Delta t}{2\Delta x}((u^*_{j+\frac{1}{2}})^2 - (u^*_{j-\frac{1}{2}})^2) + \Delta t\frac{\hat{s}_{j+\frac{1}{2}} + \hat{s}_{j-\frac{1}{2}}}{2} \tag{4.4.13}$$

where $\hat{s}_{j+\frac{1}{2}} = -g((h+z)_{j+1} - (h+z)_j)/\Delta x$, which concludes the definition of the schemes.

As a last remark we observe that (4.4.13) could not be extended in two dimensions as it is (see section 4.8), we propose an alternative approximation of Lu which reads

$$(Lu)^{n+1-}_j = (Lu)^n_j + \frac{u^n_j + u^n_{j+1}}{2}\frac{\Delta t}{\Delta x}(u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}) + \Delta t\frac{\hat{s}_{j+\frac{1}{2}} + \hat{s}_{j-\frac{1}{2}}}{2}.$$

Both formulations turn out to give the same results in 1D.

4.4.3 Friction term approximation

As far as the approximation of the friction term is concerned, here we refer to the work of Audusse et al. [4]. Hence, we briefly recall the discretization they used and refer to their paper for more details. In particular, we exploit an implicit splitting strategy. Once the solution hu^{n+1} from the projection step has been obtained, we state

$$\tilde{hu}^{n+1} = hu^{n+1} - g\Delta t\frac{\mu_f^2|\tilde{hu}^{n+1}|\tilde{hu}^{n+1}}{h^{n+1}R_h^{4/3}} \tag{4.4.14}$$

which gives us the flux hu at the new time level with the friction contributions included, that is $\tilde{hu}^{n+1}$. Note that, imposing $\tilde{hu}^{n+1}$ and hu^{n+1} to have the same sign, it is possible to obtain the explicit solution of equation (4.4.14) so that the computational cost is not high. Moreover, as explained in [4], this discretization will also allow us to preserve the "constant bed slope" equilibrium (4.1.6).

4.5 Increasing the order of accuracy

So far we presented three different first-order numerical schemes for the 1D shallow water Exner system. The discretization for the water height and flow appeared to be the same in all the methods, only the manner in which we updated the bottom height changed. We now aim to develop second-order numerical methods, which could lead to the design of even higher order schemes. As we will see, we do not achieve the second-order of accuracy for all the three schemes in the same way. This is mainly due to the fact that we also ask for the well-balanced property and it cannot always be obtained in the same manner as it depends on the underlying first-order scheme. However, the heart of the method is the same, namely we exploit polynomial reconstructions [49] and Runge-Kutta TVD approach [33] in order to reach the second order of accuracy in space and time respectively. Indeed, what will change is mainly the way we define the slopes for the reconstruction polynomials in space, as the Runge-Kutta procedure does not affect the ability of the scheme of preserving the well-balanced property. We specify that the Runge-Kutta method is used at second order and applied to the overall scheme, namely the acoustic and transport steps together.

4.5.1 Update the bed level in both steps

Let us start with the numerical scheme which entails a splitting of the Exner equation in both steps. Regarding the acoustic step, we first proceed in a very classical way by making use of first-order polynomial reconstructions of the form

$$\mathbf{P}_j(x) = \mathbf{Q}_j^n + \Delta_j(x - x_j), \tag{4.5.1}$$

where $\mathbf{Q} = (h, hu, z)^t$ is the vector of unknowns and $\Delta_j = \big(\Delta_j(h), \Delta_j(hu), \Delta_j(z)\big)^t$ denotes the corresponding slopes. Motivated by the well-balanced property, we compute the slopes using standard ENO or MINMOD limiters applied to the free surface $H = h + z$, hu and z. Then, we simply set $\Delta_j(h) = \Delta_j(H) - \Delta_j(z)$. Indeed, notice that H is constant under the "lake at rest" condition and therefore the slopes $\Delta_j(H)$ automatically reduce themselves to zero in this case. Then, it clearly follows $\Delta_j(h) = -\Delta_j(z)$, which is necessary for the well-balanced property.

Finally, the Lagrangian step reads

$$\begin{cases} L_j^{n+1-} h_j^{n+1-} = L_j^n h_j^n \\ L_j^{n+1-} (hu)_j^{n+1-} = L_j^n (hu)_j^n - \frac{\Delta t}{\Delta x}(\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}) + \Delta t(s_j^n + s_{C,j}^n) \\ L_j^{n+1-} z_j^{n+1-} = L_j^n z_j^n - \frac{\Delta t}{\Delta x}((\Omega - zu)^*_{j+\frac{1}{2}} - (\Omega - zu)^*_{j-\frac{1}{2}}) \end{cases} \tag{4.5.2}$$

where we compute the interface values using the first-order formulas but applied to the left and right traces of the reconstruction polynomials, namely

$$u^*_{j+\frac{1}{2}} = u^*_{j+\frac{1}{2}}(\mathbf{Q}^n_{j+\frac{1}{2}L}, \mathbf{Q}^n_{j+\frac{1}{2}R}) \quad \text{and} \quad \Pi^*_{j+\frac{1}{2}} = \Pi^*_{j+\frac{1}{2}}(\mathbf{Q}^n_{j+\frac{1}{2}L}, \mathbf{Q}^n_{j+\frac{1}{2}R}). \tag{4.5.3}$$

with

$$\mathbf{Q}_{j+\frac{1}{2}L} = \mathbf{P}_j(x_{j+\frac{1}{2}}) \quad \text{and} \quad \mathbf{Q}_{j+\frac{1}{2}R} = \mathbf{P}_{j+1}(x_{j+\frac{1}{2}}). \tag{4.5.4}$$

Note that, in order to have a well-balanced scheme, we introduced an additional term $s^n_{C,j}$, which is defined by

$$s^n_{C,j} = -g\frac{h_{j-\frac{1}{2}R} + h_{j+\frac{1}{2}L}}{2}\frac{z_{j+\frac{1}{2}L} - z_{j-\frac{1}{2}R}}{\Delta x}$$

and which represents the in-cell second-order contribution of the source term. Of course, we note that $s^n_{C,j} = 0$ for all j if the slopes are null.

Regarding the transport step, we exploit again polynomial reconstructions but now we reconstruct the Lagrangian variables $(Lh)_j^{n+1-}$ and $(Lhu)_j^{n+1-}$ obtained at the end of the Lagrangian step. Then, the updating formula for h and hu are given by a second-order approximation of the three integrals that appear in (4.4.9). This is achieved by using a classical mid-point rule, hence we get

$$X_j^{n+1-} = (LX)_j^{n+1-} - \frac{\Delta t}{\Delta x}\Big(u^*_{j+\frac{1}{2}}(LX)^{n+1-}_{j+\frac{1}{2}}\Big(\frac{\xi_{j+\frac{1}{2}}+\hat{\xi}_{j+\frac{1}{2}}}{2}\Big)+ \\ - u^*_{j-\frac{1}{2}}(LX)^{n+1-}_{j-\frac{1}{2}}\Big(\frac{\xi_{j-\frac{1}{2}}+\hat{\xi}_{j-\frac{1}{2}}}{2}\Big)\Big), \tag{4.5.5}$$

where we use the upwind definition

$$(LX)^{n+1-}_{j-\frac{1}{2}}(\xi) = \begin{cases} (LX)^{n+1-}_{j-1}(\xi) & \text{if} \quad u^*_{j-\frac{1}{2}} > 0 \\ (LX)^{n+1-}_{j}(\xi) & \text{if} \quad u^*_{j-\frac{1}{2}} \leq 0. \end{cases} \tag{4.5.6}$$

As far as the topography is concerned, the procedure is similar, the only difference is that we reconstruct its values z_j^{n+1-} instead of $(Lz)_j^{n+1-}$. Hence, the reconstructed polynomial is given by

$$P_j(z,\xi) = z_j^{n+1-} + \Delta_j(\xi - \xi_j),$$

and the second-order updating formula for z simply reads

$$z_j^{n+1} = (Lz)_j^{n+1-} - \frac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}P_{j+\frac{1}{2}}\Big(z, \frac{\xi_{j+\frac{1}{2}}+\hat{\xi}_{j+\frac{1}{2}}}{2}\Big) - u^*_{j-\frac{1}{2}}P_{j-\frac{1}{2}}\Big(z, \frac{\xi_{j-\frac{1}{2}}+\hat{\xi}_{j-\frac{1}{2}}}{2}\Big)\big),$$

with

$$P_{j-\frac{1}{2}}(z,\xi) = \begin{cases} P_{j-1}(z,\xi) & \text{if} \quad \xi_{j-\frac{1}{2}} > \hat{\xi}_{j-\frac{1}{2}} \\ P_j(z,\xi) & \text{if} \quad \xi_{j-\frac{1}{2}} \leq \hat{\xi}_{j-\frac{1}{2}}. \end{cases}$$

4.5.2 Update the bed level in the acoustic step

Let us proceed with the numerical scheme that takes into account the bed level z only at acoustic level. Here, in order to obtain a second-order discretization of the variables h and hu, we exactly follow the same procedure we described above. The only difference is related to the bottom height approximation for which we use formula (4.4.7) and either (4.3.18) or (4.3.19) for the star values. We highlight that, in the acoustic step, the same reconstruction procedure is considered for the variables h, hu and z whereas, in the transport step, nothing has to be done for z.

4.5.3 Update the bed level in the transport step

Let us finally see the third and last approach, which directly updates z in the transport step. Starting with the Lagrangian step, here the main difference, with respect to the previous two LP strategies, concerns the strategy we use to preserve the well-balanced property. We begin by defining at time t^n and for each cell j a stationary solution denoted by $x \mapsto \mathbf{Q}_j^{n,e}(x)$ and defined for all x by

$$\big(h_j^{n,e}\big)(x) = h_j^n + z_j^n - z^n(x), \quad u_j^{n,e}(x) = u_j^n \quad \text{and} \quad z_j^{n,e}(x) = z^n(x), \tag{4.5.7}$$

where $x \mapsto z^n(x)$ is nothing but the piecewise constant approximation of z at time t^n, namely such that $z^n(x) = z_j^n$ for all x in $[x_{j-1/2}, x_{j+1/2})$. Such a reconstructed solution satisfies the in-cell conservation property

$$\frac{1}{\Delta x}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{Q}_j^{n,e}(x)dx = \mathbf{Q}_j^n.$$

Next, we follow [45, 17] and introduce the so-called j-fluctuations defined as

$$\mathbf{D}_{k,j}^n = \mathbf{Q}_k^n - \frac{1}{\Delta x}\int_{x_{k-1/2}}^{x_{k+1/2}} \mathbf{Q}_j^{n,e}(x)dx,$$

for all k. Observe that $\mathbf{D}_{j,j}^n = \mathbf{0}$ by construction, while $\mathbf{D}_{k,j}^n = \mathbf{0}$ for all k if the approximate solution at time t^n satisfies the "lake at rest" condition (4.1.5).

Lastly, for each cell I_j we make use of a reconstructed polynomial vector $\mathbf{P}_j^n(x)$ defined as in (4.5.1), where we compute the slopes using the fluctuations. Namely, we impose $\Delta_j^n = \Delta_j^n(\mathbf{D}_{j-1,j}^n, \mathbf{D}_{j,j}^n, \mathbf{D}_{j+1,j}^n)$ and exploit either ENO [49] or MINMOD [50] limiters.

Then, we define the numerical fluxes $u_{j+\frac{1}{2}}^*$ and $\Pi_{j+\frac{1}{2}}^*$ using the interfaces values (4.5.4) (as in (4.5.3)) and formula (4.4.5). Regarding the source term, once again we exploit formulas (4.4.4). Let us note that, thanks to formula (4.5.7), $z_{j+\frac{1}{2}L} = z_j$ and $z_{j+\frac{1}{2}R} = z_{j+1}$ as the fluctuations related to the topography are null. Finally, the discretization of the Lagrangian system (4.2.3) reads as in the first-order step, namely formulas (4.4.6).

As far as the transport step is concerned, the water height h and the discharge hu are updated like in the two other methods and thus we do not further discuss it. Then, regarding the topography, we consider the weakly coupled scheme (4.4.12) where we naturally set

$$\Big(\frac{q_b}{u}\Big)_{j+1/2}^{n+1-} = \begin{cases} \Big(\frac{q_b}{u}\Big)\Big((Lu)_{j+1}^{n+1-}\Big(\frac{\xi_{j+\frac{1}{2}} + \hat{\xi}_{j+\frac{1}{2}}}{2}\Big)\Big) & \text{if} \quad u_{j+\frac{1}{2}}^* \leq 0 \\ \Big(\frac{q_b}{u}\Big)\Big((Lu)_{j}^{n+1-}\Big(\frac{\xi_{j+\frac{1}{2}} + \hat{\xi}_{j+\frac{1}{2}}}{2}\Big)\Big) & \text{if} \quad u_{j+\frac{1}{2}}^* > 0, \end{cases}$$

and $u_{j\pm\frac{1}{2}}^*$ is given by (4.5.3).

Afterwards, for the decoupled scheme, we first define the reconstructed polynomial for the water height h and the flow hu at time t^n,

$$P(X)_j^n(x) = X_j^n + \Delta_j^n(x - x_j)$$

with $X = h, hu$ and Δ_j^n the slopes (either ENO or MINMOD). Then, we use formula (4.4.10), where we impose

$$\Big(\frac{q_b}{u}\Big)_{j+1/2}^{n} = \begin{cases} \Big(\frac{q_b}{u}\Big)\Big(\frac{P(hu)_{j+1}^n(x_{j+\frac{1}{2}})}{P(h)_{j+1}^n(x_{j+\frac{1}{2}})}\Big) & \text{if} \quad u_{j+\frac{1}{2}}^* \leq 0 \\ \Big(\frac{q_b}{u}\Big)\Big(\frac{P(hu)_{j}^n(x_{j+\frac{1}{2}})}{P(h)_{j}^n(x_{j+\frac{1}{2}})}\Big) & \text{if} \quad u_{j+\frac{1}{2}}^* > 0. \end{cases}$$

4.6 Overall discretization and well-balanced property

This section is devoted to the illustration of the first-order overall discretizations of the previous schemes and their well-balanced property. We also take advantage of this section to observe that the three first-order methods we presented are also positivity-preserving under suitable conditions.

Remark 4. *Is it possible to prove that the three first-order methods are positivity-preserving, meaning that they are able to preserve the strict positivity of the water height under the associated CFL condition 4.4.2 with $CFL_t \leq \frac{1}{2}$. See also [14] for more details.*

4.6.1 Bed level in both steps

Considering the numerical treatment that updates z in both the acoustic and transport steps (sections 4.3.1, 4.4 and 4.5.1), it is easy to see that the whole first-order scheme takes the following final form,

$$\begin{cases} h_j^{n+1} = h_j^n - \dfrac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}(Lh)^{n+1-}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}(Lh)^{n+1-}_{j-\frac{1}{2}}\big) \\ (hu)_j^{n+1} = (hu)_j^n - \dfrac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}(Lhu)^{n+1-}_{j+\frac{1}{2}} + \Pi^*_{j+\frac{1}{2}} - (u^*_{j-\frac{1}{2}}(Lhu)^{n+1-}_{j-\frac{1}{2}} + \Pi^*_{j-\frac{1}{2}})\big) + \Delta t s_j^n, \\ z_j^{n+1} = z_j^n - \dfrac{\Delta t}{\Delta x}\big((\Omega - zu)^*_{j+\frac{1}{2}} - (\Omega - zu)^*_{j-\frac{1}{2}}\big) - \dfrac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}} z^{n+1-}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}} z^{n+1-}_{j-\frac{1}{2}}\big) \end{cases} \tag{4.6.1}$$

with

$$X^{n+1-}_{j+\frac{1}{2}} = \begin{cases} X_j^{n+1-} & \text{if} \quad u^{*,n+1-}_{j+\frac{1}{2}} > 0 \\ X_{j+1}^{n+1-} & \text{if} \quad u^{*,n+1-}_{j+\frac{1}{2}} \leq 0 \end{cases} \tag{4.6.2}$$

and $X = Lh, Lhu, z$. Note that the evolution equation for the topography z can also be reformulated as

$$z_j^{n+1} = z_j^n - \frac{\Delta t}{\Delta x}\big(\Omega^*_{j+\frac{1}{2}} - \Omega^*_{j-\frac{1}{2}}\big) + \frac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}(z^*_{j+\frac{1}{2}} - z^{n+1-}_{j+\frac{1}{2}}) - u^*_{j-\frac{1}{2}}(z^*_{j-\frac{1}{2}} - z^{n+1-}_{j-\frac{1}{2}})\big).$$

It is clear that, without the source term present in the evolution equation for hu, the whole numerical scheme would be conservative. Let us now prove that both the first and second-order schemes are well-balanced.

Theorem 5. *The numerical method with updating formula (4.6.1) and star values given in section 4.3.1.1 is well-balanced under the "lake at rest" condition (4.1.5).*

Proof. Referring to [20], it can be easily seen that $u^*_{j+1/2} = 0$, $h_j^{n+1-} = h_j^n$ and $(hu)_j^{n+1-} = (hu)_j^n \ \forall j$. Let us now consider the topography z. Since we are under the hypothesis that $u_j^n = u^*_{j+1/2} = 0 \ \forall j$, it appears clear that we are in the case $|u^*_{j+1/2}|b_{j+1/2} < a_{j+1/2}$. Consequently, we should use formula (4.3.13) to update z_j^{n+1-}, which leads to $(|u|z)^*_{j+1/2} = 0$ and $\Omega^*_{j+1/2} = 0$ $\forall j$. Thus, the scheme is well-balanced for z as well.

□

Theorem 6. *The second-order numerical method which updates the bed level z in both the acoustic and transport step (formulas in section 4.5.1) preserves the "lake at rest" stationary solution (4.1.5).*

Proof. First of all, we observe that $u^*_{j+\frac{1}{2}} = 0 \ \forall j$ when the variables at time t^n satisfy the "lake at rest" solution. Indeed, we have

$$\begin{aligned} u^*_{j+\frac{1}{2}} = \frac{1}{2}(u_{j+\frac{1}{2}L} + u_{j+\frac{1}{2}R}) &- \frac{1}{2a_{j+\frac{1}{2}}}(\Pi_{j+\frac{1}{2}R} - \Pi_{j+\frac{1}{2}L}) \\ &- \frac{1}{2a_{j+\frac{1}{2}}}(\frac{g}{2}(h_{j+\frac{1}{2}L} + h_{j+\frac{1}{2}R})(z_{j+\frac{1}{2}R} - z_{j+\frac{1}{2}L})) \end{aligned}$$

but $\Delta_j^t(u) = 0 \; \forall j$ and thus $u_{j+\frac{1}{2}L} = u_{j+\frac{1}{2}R} = u_j = u_{j+1} = 0$. Thus, we only have to prove that

$$\Pi_{j+\frac{1}{2}R} - \Pi_{j+\frac{1}{2}L} = -\frac{g}{2}(h_{j+\frac{1}{2}L} + h_{j+\frac{1}{2}R})(z_{j+\frac{1}{2}R} - z_{j+\frac{1}{2}L}) \tag{4.6.3}$$

but

$$\Pi_{j+\frac{1}{2}R} - \Pi_{j+\frac{1}{2}L} = \frac{g}{2}(h_{j+\frac{1}{2}R}{}^2 - h_{j+\frac{1}{2}L}^2) = \frac{g}{2}(h_{j+\frac{1}{2}L} + h_{j+\frac{1}{2}R})(h_{j+\frac{1}{2}R} - h_{j+\frac{1}{2}L})$$

and finally (4.6.3) is equivalent to

$$h_{j+\frac{1}{2}R} - h_{j+\frac{1}{2}L} = -(z_{j+\frac{1}{2}R} - z_{j+\frac{1}{2}L}).$$

Exploiting definitions (4.5.1)-(4.5.4), we write

$$\begin{aligned} h_{j+\frac{1}{2}R} - h_{j+\frac{1}{2}L} &= h_{j+1} - \frac{\Delta_{j+1}^t(h)\Delta x}{2} - (h_j + \frac{\Delta_j^t(h)\Delta x}{2}) = \\ &= -z_{j+1} + z_j + \frac{\Delta_{j+1}^t(z)\Delta x}{2} + \frac{\Delta_j^t(z)\Delta x}{2} = -z_{j+\frac{1}{2}R} + z_{j+\frac{1}{2}L}. \end{aligned}$$

Thus, we proved that $u^*_{j+\frac{1}{2}} = 0 \; \forall j$. Next, if the "lake at rest" condition holds true, from the second equation of system (4.5.2) we observe

$$-(\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}) + \Delta x(s_j^n + s_{C,j}^n) = 0$$

and thus $L_j^{n+1-}(hu)_j^{n+1-} = L_j^n(hu)_j^n = 0$. Let us give the details. Since we know $u_{j\pm\frac{1}{2}L} = u_{j\pm\frac{1}{2}R} = 0 \; \forall j$, we can write

$$\begin{aligned} \Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}} &= \frac{1}{2}(\Pi_{j+\frac{1}{2}R} + \Pi_{j+\frac{1}{2}L}) - \frac{1}{2}(\Pi_{j-\frac{1}{2}R} + \Pi_{j-\frac{1}{2}L}) = \\ &= \frac{1}{2}\Big((\Pi_{j+\frac{1}{2}R} - \Pi_{j+\frac{1}{2}L}) + (\Pi_{j-\frac{1}{2}R} - \Pi_{j-\frac{1}{2}L})\Big) + \Pi_{j+\frac{1}{2}L} - \Pi_{j-\frac{1}{2}R} \\ &\overset{(4.6.3)}{=} \Delta x s_j^n + \Pi_{j+\frac{1}{2}L} - \Pi_{j-\frac{1}{2}R}. \end{aligned}$$

With similar computations, it is straightforward to see that $\Pi_{j+\frac{1}{2}L} - \Pi_{j-\frac{1}{2}R} = \Delta x s^n_{C,j}$. Finally, since $u^*_{j\pm\frac{1}{2}} = 0 \; \forall j$, we also know that $\Omega^*_{j\pm\frac{1}{2}} = 0 \; \forall j$ and thus $z_j^{n+1-} = z_j^n$.

Regarding the transport step, it does not modify the unknowns' values as $u^*_{j\pm\frac{1}{2}} = 0 \; \forall j$. Similarly, since we already know that the Runge-Kutta procedure does not prevent the scheme from being well-balanced, the property is proved. □

Theorem 7. *The numerical method with updating formula (4.6.1) and star values given in section 4.3.1.1 is well-balanced under the "constant bed slope" equilibrium (4.1.6). This statement remains true even if we consider its second-order accurate version presented in section 4.5.1.*

Proof. Assuming to be under the "constant bed slope" equilibrium (4.1.6), namely to have constant velocity $u_i^n = u_{i+1}^n \; \forall i$, constant water height $h_i^n = h_{i+1}^n \; \forall i$ and constant bed slope $z_{i+1}^n - z_i^n = -\Delta x S_{f,i} = K \; \forall i$ with constant K, we want to prove that such a steady state is preserved, namely that $h_i^{n+1} = h_i^n$, $hu_i^{n+1} = hu_i^n$ and $z_i^{n+1} = z_i^n \; \forall i$. Let us start by seeing what are the star values in this case. We consider only the case $a < |u_L|b$, $a < |u_R|b$ as for the other one the procedure is analogous. Thus, it easy to see that we obtain $\Pi^*_{i+1/2} = \Pi_i^n = \Pi^* = constant$ and $u^*_{i+1/2} = u_i^n - \frac{1}{2a}\mathcal{M}_{i+1/2} = u^* = constant$ as also the parameters a and b are clearly constant in all the domain. Moreover, we also have $z^*_{i+1/2} = \frac{z^n_{i+1}+z^n_i}{2} + \frac{u}{2|u|b\tau}(z^n_{i+1} - z^n_i)$ and

$\Omega^*_{i+1/2} = \Omega^* = constant$. Then, it is easy to see that we obtain $h_i^{n+1} = Lh_i^{n+1-} = h_i^n$, while for the discharge we get

$$hu_i^{n+1} = hu_i^n + \Delta t g h_i^n S_{f,i}$$

where the friction has not yet been taken into account and which compensates the presence of the term $\Delta t g h_i^n \Delta x S_f$, leading to $\tilde{hu}_i^{n+1} = hu_i^n$, see also [4]. Finally, as far as the topography is concerned, we have

$$\begin{aligned} z_i^{n+1} &= z_i^n - \frac{\Delta t}{\Delta x}\big(\Omega^*_{i+\frac{1}{2}} - \Omega^*_{i-\frac{1}{2}}\big) + \frac{\Delta t}{\Delta x}\big(u^*_{i+\frac{1}{2}}(z^*_{i+\frac{1}{2}} - z^{n+1-}_{i+\frac{1}{2}}) - u^*_{i-\frac{1}{2}}(z^*_{i-\frac{1}{2}} - z^{n+1-}_{i-\frac{1}{2}})\big) \\ &= z_i^n + u^* \frac{\Delta t}{\Delta x}\big((z^*_{i+\frac{1}{2}} - z^{n+1-}_{i+\frac{1}{2}}) - (z^*_{i-\frac{1}{2}} - z^{n+1-}_{i-\frac{1}{2}})\big). \end{aligned}$$

Then, since u^* is constant, it is clear that we have $z^{n+1-}_{i+\frac{1}{2}} - z^{n+1-}_{i-\frac{1}{2}} = z^{n+1-}_{i+1} - z^{n+1-}_i = z^{n+1-}_i - z^{n+1-}_{i-1} = z^n_{i+1} - z^n_i = z^n_i - z^n_{i-1} = K$ and thus

$$\begin{aligned} z_i^{n+1} &= z_i^n + u^* \frac{\Delta t}{\Delta x}\big(\frac{z^n_{i+1} + z^n_i}{2} - \frac{z^n_i + z^n_{i-1}}{2} - K\big) \\ &= z_i^n + u^* \frac{\Delta t}{\Delta x}\big(\frac{z^n_{i+1} - z^n_i}{2} + \frac{z^n_i - z^n_{i-1}}{2} - K\big) = z_i^n, \end{aligned}$$

which concludes the proof for the first-order method. Concerning the second-order extension, it is enough to check that the slopes of the polynomial reconstruction are either null or constant and similar computations lead to the same result; we therefore do not include further details here. □

4.6.2 Bed level in the acoustic step

The first-order numerical method, which takes into account the bed level z only in the acoustic step (sections 4.3.2, 4.4 and 4.5.2), has an overall form similar to (4.6.1), the only difference related to the updating equation for z. In practice, we write

$$\begin{cases} h_j^{n+1} = h_j^n - \dfrac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}} Lh^{n+1-}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}} Lh^{n+1-}_{j-\frac{1}{2}}\big) \\ hu_j^{n+1} = hu_j^n - \frac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}} Lhu^{n+1-}_{j+\frac{1}{2}} + \Pi^*_{j+\frac{1}{2}} - (u^*_{j-\frac{1}{2}} Lhu^{n+1-}_{j-\frac{1}{2}} + \Pi^*_{j-\frac{1}{2}})\big) + \Delta t s_j^n \\ z_j^{n+1} = z_j^n - \dfrac{\Delta t}{\Delta x}\big(\Omega^*_{j+\frac{1}{2}} - \Omega^*_{j-\frac{1}{2}}\big) \end{cases} \tag{4.6.4}$$

with $X^{n+1-}_{j+\frac{1}{2}}$ given by formula (4.6.2) and $X = Lh, Lhu$.

Theorem 8. *The numerical method with updating formula (4.6.4) and star values given in section 4.3.2.2 is well-balanced under the "lake at rest" condition (4.1.5).*

Proof. Analogous to the one of theorem 5.

□

Theorem 9. *The second-order numerical method which updates the bed level z only in the acoustic step (section 4.5.2) preserves the "lake at rest" stationary solution (4.1.5).*

Proof. Analogous to the one of theorem 6.

□

Theorem 10. *The numerical method with updating formula (4.6.4) and star values given in section 4.3.2.2 is well-balanced under the "constant bed slope" equilibrium (4.1.6). This statement remains true even if we consider its second-order accurate version presented in section 4.5.2.*

Proof. Analogous to the one of theorem 7.

□

4.6.3 Bed level in the transport step

Finally, let us briefly comment on the numerical scheme which resembles to the usual splitting method (sections 4.3.3, 4.4 and 4.5.3). The bed level is updated only in the transport step and, as such, the first-order overall discretization reads

$$\begin{cases} h_j^{n+1} = h_j^n - \dfrac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}} Lh^{n+1-}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}} Lh^{n+1-}_{j-\frac{1}{2}}\big) \\ hu_j^{n+1} = hu_j^n - \frac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}} Lhu^{n+1-}_{j+\frac{1}{2}} + \Pi^*_{j+\frac{1}{2}} - (u^*_{j-\frac{1}{2}} Lhu^{n+1-}_{j-\frac{1}{2}} + \Pi^*_{j-\frac{1}{2}})\big) + \Delta t s_j^n \\ z_j^{n+1} = z_j^n - \zeta \dfrac{\Delta t}{\Delta x}\Big(u^*_{j+\frac{1}{2}} \Big(\dfrac{q_b(Lu)}{Lu}\Big)^{n+1-}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}} \Big(\dfrac{q_b(Lu)}{Lu}\Big)^{n+1-}_{j-\frac{1}{2}}\Big), \end{cases} \tag{4.6.5}$$

where we have used discretization (4.4.12) instead of (4.4.10). See also formulas (4.4.6) and (4.4.8). It is easy to see that both first and second-order accurate versions of this method are well-balanced.

Theorem 11. *The first-order numerical scheme with updating formula (4.6.5) preserves the "lake at rest" stationary solution (4.1.5).*

Proof. Proof analogous to the one of theorem 5. We only observe that under the lake at rest condition, we also find $(Lu)_j^{n+1-} = 0$ as $(h+z)_j^n = (h+z)_{j+1}^n$ for all j. □

Theorem 12. *The second-order numerical method which updates the bed level z only in the transport step (section 4.5.3) preserves the "lake at rest" stationary solution (4.1.5).*

Proof. Since we already proved the well-balanced property for the first-order scheme, it is straightforward to show it for the second-order method as well. Indeed, it is enough to observe that the slopes $\Delta_j^n = \Delta_j^n(\mathbf{D}_{j-1,j}^n, \mathbf{D}_{j,j}^n, \mathbf{D}_{j+1,j}^n)$ are null under the "lake at rest" hypothesis thanks to definition of the fluctuations. Hence, once again we obtain $u^*_{j+\frac{1}{2}} = 0$ and thus $Lhu_j^{n+1-} = hu_j^n = 0$, $Lh_j^{n+1-} = h_j^n$, $hu_j^{n+1} = Lhu_j^{n+1-} = hu_j^n = 0$, $h_j^{n+1} = Lh_j^{n+1-} = h_j^n$ and $z_j^{n+1} = z_j^n$. Finally, it is only worth to specify that the Runge-Kutta procedure automatically preserves the stationary solutions. □

Theorem 13. *The first-order numerical scheme with updating formula (4.6.5) preserves the "constant bed slope" equilibrium (4.1.6). This statement remains true even if we consider its second-order accurate version presented in section 4.5.3.*

Proof. Analogous to the one of theorems 7 and 12. □

4.7 Numerical evidences

Here we test the numerical schemes we presented so far; for the sake of conciseness we distinguish them by calling them as follows:

- "AcTrZ": scheme with bed level z updated in both the ACoustic and the TRansport steps (see sections 4.3.1, 4.4 and 4.5.1);
- "AcZ": scheme with bed level z updated in the ACoustic step (see sections 4.3.2, 4.4 and 4.5.2);
- "TrZ": scheme with bed level z updated in the TRansport step (see sections 4.3.3, 4.4 and 4.5.3). Since the decoupled (4.4.10) and weakly coupled (4.4.12) formulas give similar results, if not otherwise specified, we use the weakly coupled discretization (4.4.12).

Scheme	Mesh M	Var	err $\mathbf{L}^1$	$O(\mathbf{L}^1)$	Var	err $\mathbf{L}^1$	$O(\mathbf{L}^1)$	Var	err $\mathbf{L}^1$	$O(\mathbf{L}^1)$
AcTrZ	64	h	0.0269	–	hu	0.1174	–	z	0.1372×10^{-3}	–
	128		0.0083	1.6927		0.0354	1.7294		0.0639×10^{-3}	1.1030
	256		0.0027	1.6190		0.0115	1.6236		0.0218×10^{-3}	1.5529
	512		0.0007	1.8757		0.0031	1.8773		0.0061×10^{-3}	1.8456
	1024		0.0002	1.9786		0.0008	1.9819		0.0016×10^{-3}	1.9532
AcZ	64	h	0.0269	–	hu	0.1174	–	z	0.1370×10^{-3}	–
	128		0.0083	1.6931		0.0354	1.7301		0.0592×10^{-3}	1.2115
	256		0.0027	1.6186		0.0115	1.6236		0.0198×10^{-3}	1.5780
	512		0.0007	1.8754		0.0031	1.8772		0.0056×10^{-3}	1.8329
	1024		0.0002	1.9785		0.0008	1.9820		0.0014×10^{-3}	1.9427
TrZ	64	h	0.0268	–	hu	0.1175	–	z	0.1824×10^{-3}	–
Weakly	128		0.0083	1.6955		0.0354	1.7320		0.0550×10^{-3}	1.7296
coupled	256		0.0027	1.6182		0.0115	1.6248		0.0183×10^{-3}	1.5890
	512		0.0007	1.8755		0.0031	1.8781		0.0050×10^{-3}	1.8734
	1024		0.0002	1.9782		0.0008	1.9818		0.0013×10^{-3}	1.9792
TrZ	64	h	0.0268	–	hu	0.1175	–	z	0.1792×10^{-3}	–
Decoupled	128		0.0083	1.6953		0.0354	1.7320		0.0544×10^{-3}	1.7190
	256		0.0027	1.6185		0.0115	1.6249		0.0182×10^{-3}	1.5785
	512		0.0007	1.8756		0.0031	1.8782		0.0050×10^{-3}	1.8761
	1024		0.0002	1.9781		0.0008	1.9818		0.0012×10^{-3}	1.9932

Table 4.7.1: Errors and empirical convergence rates for norm $\mathbf{L}^1$. Mesh of size $M = (64, 128, 256, 512, 1024)$, CFL $= 0.25$.

If not otherwise specified, we take $\zeta = 1$, $q_b = A_g u^3$ with $A_g = 0.005$ for the Exner equation, $\mu_f = 0$ and transmissive boundary conditions. For the CFL number we use $\text{CFL}_l = 0.45$ and $\text{CFL}_l = 0.25$ for first and second order schemes respectively, while $\text{CFL}_t = 1$. With the exception of the accuracy test case, when the reference solution is inserted, it is computed exploiting the second-order AcTrZ scheme with $M = 1000$ cells, where $\Delta x = \frac{L}{M}$ with L the length of the channel.

4.7.1 Test of order of accuracy

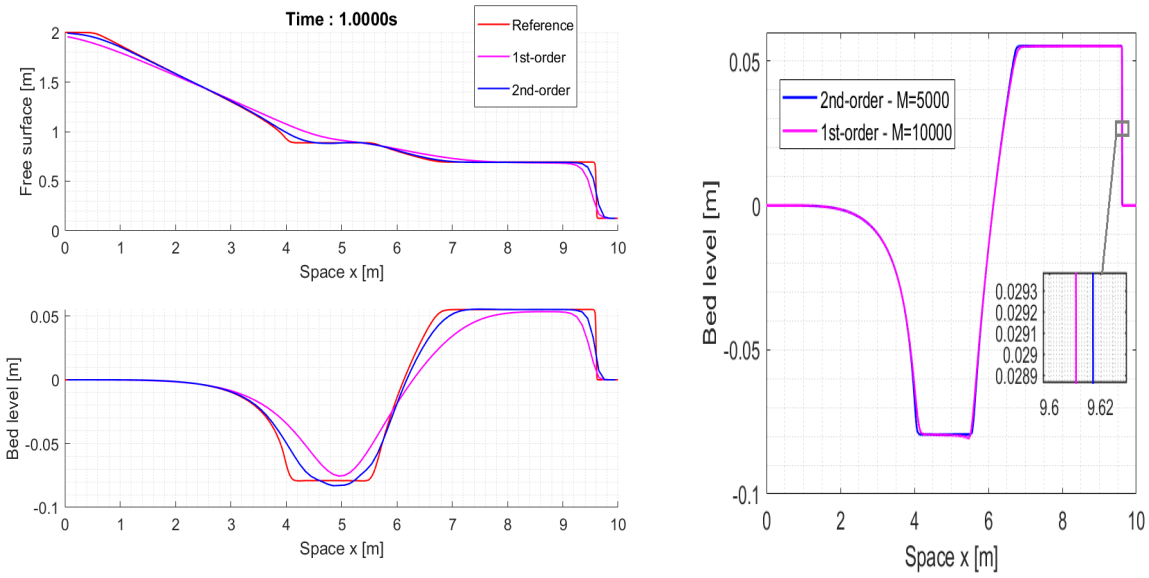
Here we test the order of accuracy of the numerical schemes described previously. Let us consider a channel of length $L = 20m$, $A_g = 0.3$, $m = 3$. The initial condition is given by null velocity and

$$\begin{cases} z_{IC} = 0.1 - 0.01e^{-(x-10)^2} \\ h_{IC} = 2 - 0.1e^{-(x-10)^2}. \end{cases}$$

We refer to paper [15] for this test case. The reference solution is computed using $M = 2048$ cells and TrZ-second-order "decoupled" method (formula (4.4.10)). In table 4.7.1 we insert the errors and the EOA in norm $\mathbf{L}^1$ for all the three schemes. In particular for the TrZ one, we consider both the weakly coupled (formula (4.4.12)) and decoupled (formula (4.4.10)) versions. We can see that the second-order of accuracy is reached in each case.

4.7.2 Riemann problem: dam break on movable bottom

For this Riemann problem, we refer to [1]. The length of the channel is $L = 10m$ and the dam is placed in the middle. The ending time is $t_{end} = 1s$. The initial condition is given by null velocity, flat topography and water height $h_L = 2m$ if $x < L/2$, $h_R = 0.125m$ if $x > L/2$. First of all, in figure 4.7.1, we compare the first and second order method. In particular, we used LP-AcZ scheme but similar results can be found with the other two methods. Clearly,



(a) $M = 100$ cells (b) $M = 2000$ cells

Figure 4.7.1: RP: dam break on movable bottom; free surface (left-top) and bed level (left-bottom and right). Comparison between the first (magenta) and second-order (blue) LP-AcZ scheme with $M = 100$ cells (left) and $M = 2000$ cells (right). Reference solution in red line.

the second-order version of the method gives more accurate results for the same value of the mesh. On the right we used $M = 2000$ cells to show that the first-order solution converges to the second-order one in general, even if we can observe a small difference of order 10^{-2} in the shock position. However, this is not surprising but natural, as the shallow water Exner system is not conservative. Then, in figure 4.7.2, we insert the bed level solution of all the three second-order numerical schemes in order to be able to sum up the merits and flaws of each of them. As expected, the more diffusive numerical scheme is the one in which the bed level is updated in both steps, namely the AcTrZ-scheme. Whereas the AcZ-scheme is slightly more diffusive than the TrZ one but less than the AcTrZ-method. On the other hand, the TrZ-method appears to be the less diffusive scheme but it is the only one which presents some oscillations in the solution. Of course, they could be related to the fact that the TrZ-method entails a decoupled numerical approximation of the shallow-water-Exner system, even if this kind of oscillations are not observed in the first-order version of TrZ-scheme. However, we remark that these oscillations reduce themselves when refining the mesh. Indeed, in figure 4.7.3, we insert the results for both the decoupled and weakly coupled second order TrZ methods for different mesh sizes, $M = 100$, $M = 200$ and $M = 500$ cells. These two schemes give similar results and, in the topography outputs of both of them, we note some instabilities which decrease as we refine the mesh. We also observe that these oscillations are more accentuated in the decoupled scheme outputs.

4.7.3 Transient Riemann problems

Next, we consider three different transient Riemann problems to test the ability of our schemes to reproduce the correct solution in this kind of situation as well. In table 4.7.2, we insert the initial conditions and the value for the coefficient A_g present in the Grass formulation. We highlight that, in the third Riemann problem, A_g is not constant anymore as it depends on the water height h. Moreover, we take $\rho_0 = 0.4$, $M = 200$ cells and $t_{\text{end}} = 0.2$s as ending time. We refer to works [47, 42] for more details and for the analytical solutions reported in figures 4.7.4, 4.7.5, 4.7.6. In particular, we plot the free surface and bed elevation

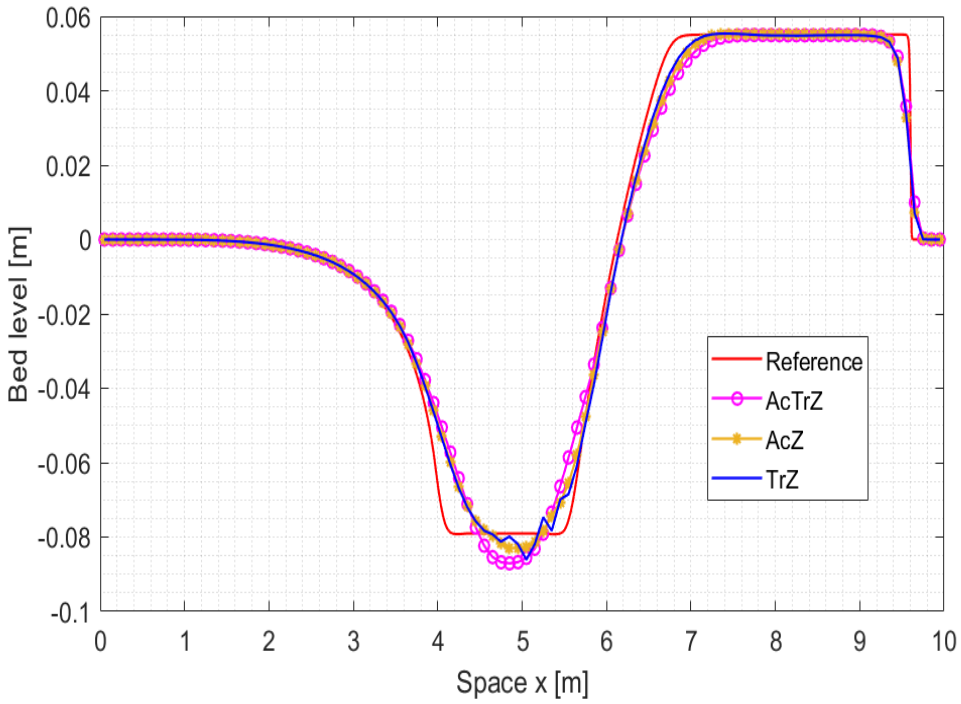


Figure 4.7.2: RP: dam break on movable bottom, bed level. Comparison among the three second-order schemes AcTrZ (magenta), AcZ (yellow) and TrZ (blue) with $M = 100$ cells. Reference solution in red line.

outputs using both first and second order method. Let us see the details. For the first RP (test A) represented in figure 4.7.4, on the left side we insert the AcTrZ solution while the TrZ one is shown on the right. We generally observe that the numerical outputs follow closely the reference one, even if the TrZ solution presents some small oscillations in correspondence of the middle discontinuity. The AcZ output is not inserted as it is very close to the AcTrZ one, even if less diffusive. Similar observations can be inferred for test B, thus we only show the AcTrZ solution on the left side of figure 4.7.5. We remark that neither the AcTrZ nor the AcZ solutions present oscillations for these two RPs at both first and second order of accuracy. Finally, let us see test C outputs. On the right side of figure 4.7.5, we inserted the AcTrZ solution both at first and second order of accuracy. Once again, the numerical solution appear to reproduce correctly the reference one. However, while neither the first-order AcTrZ nor the first-order AcZ produce any oscillations, the latter can be observed when using the second-order schemes. Thus, in figure 4.7.6 we compare the three second-order solutions and we zoom in the areas of interest to show that some small perturbations are present in the numerical outputs, probably due to the fact that less numerical diffusion is present. However, the correct solution is generally reproduced, even if A_g is not constant anymore. We conclude saying that we verified that the numerical outputs converge to the analytical one when refining the mesh.

Remark 5. *Moreover, even if without reporting here the data, even bigger values of A_g have been considered to simulate highly erosive flow for the first dam break problem (section 4.7.2) and these three transient test cases. The interest in trying bigger values of A_g resides in the fact that greater values mean a change in the flow structure of the coupled model and, consequently, more instabilities could appear if the numerical scheme is decoupled. Using the value $A_g = 0.5$, we indeed verified the presence of great oscillations when using the TrZ method. On the other hand, the other two first-order methods AcTrZ and AcZ do not produce any instabilities and remain stable. However, some oscillations have been observed using their second-order extension, even if sufficiently controlled to keep the schemes stable. One could envisage a strategy to remove them, by combining the first-order and second-order version in order to keep the second-order of*

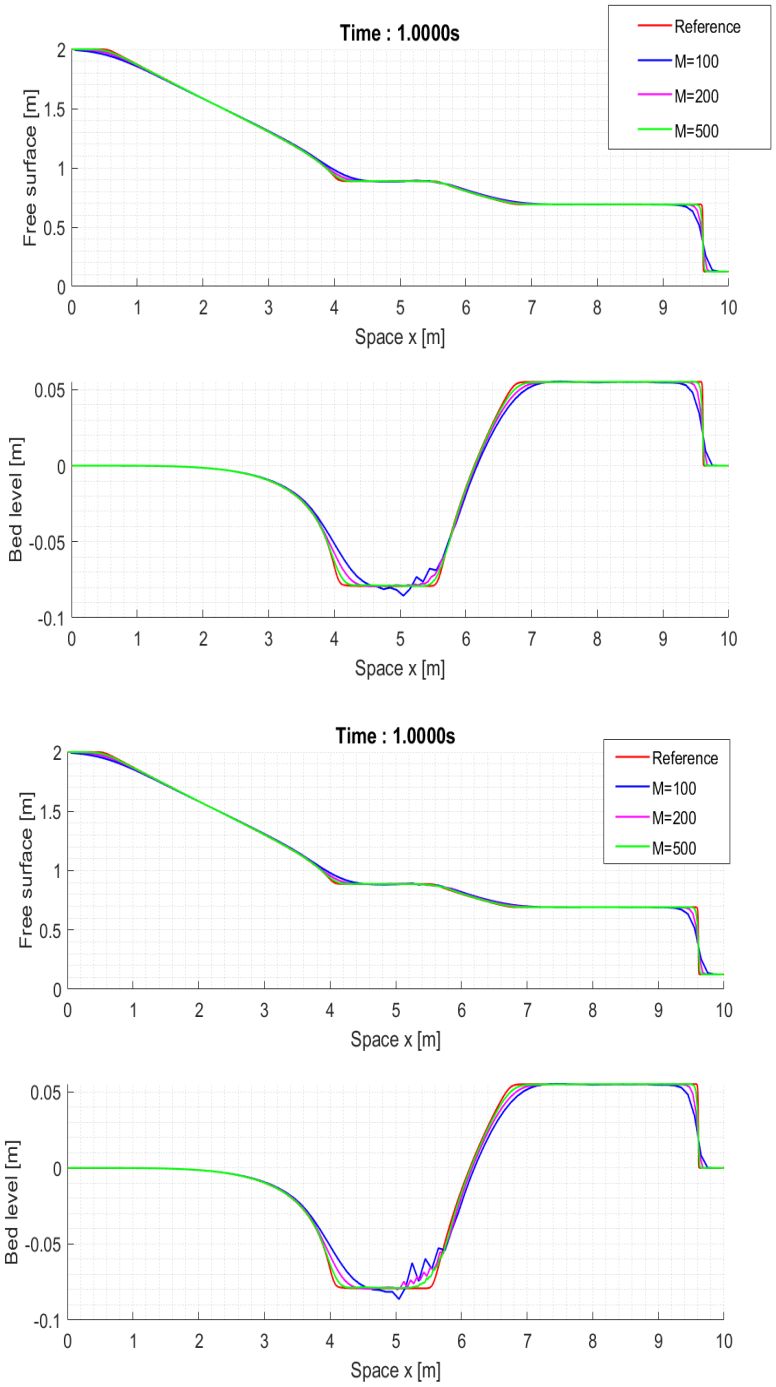
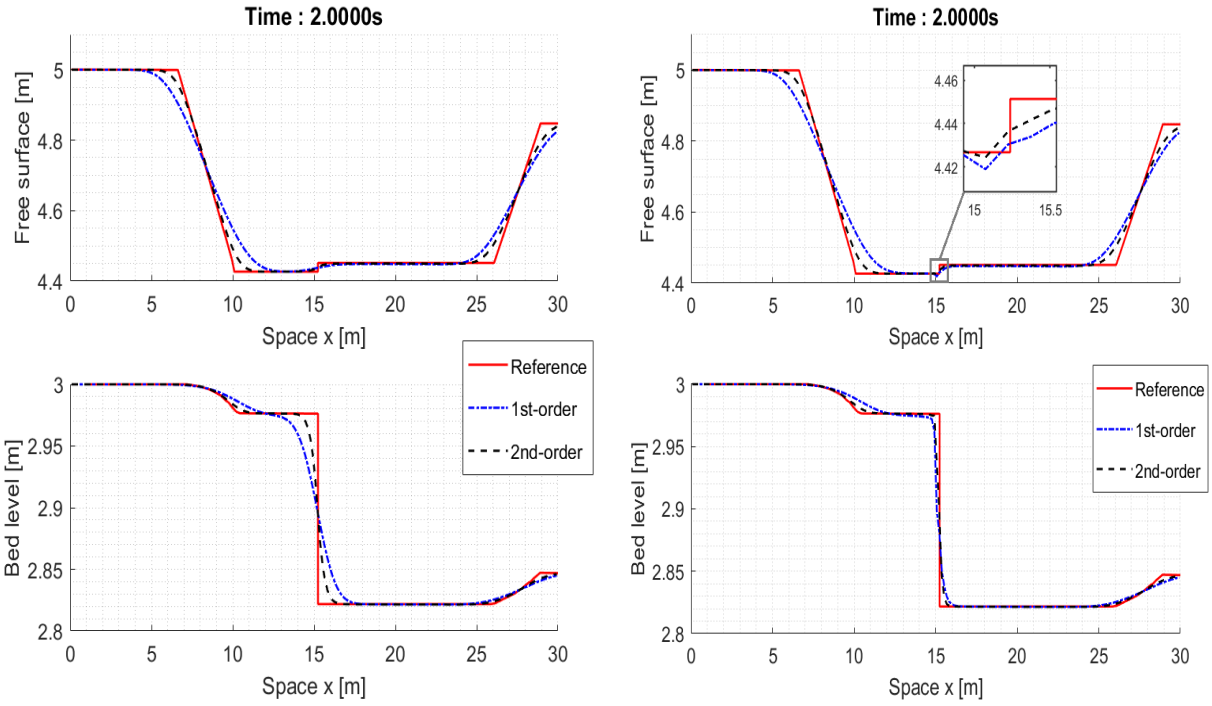


Figure 4.7.3: RP: dam break on movable bottom; free surface (up) and bed level (bottom). "Decoupled" (bottom) and "weakly coupled" (up) solutions. Mesh of size $M = 100$ (blue line), $M = 200$ (magenta line) and $M = 500$ (green line) cells. Red line for reference solution.

Test	h_L	h_R	u_L	u_R	z_L	z_R	A_g
A	2	2	0.25495	2.3247449	3.0	2.846848	0.01
B	2.25	1.18868612	0.2050	2.4321238	5.0	5.124685	0.01
C	6	5.2	0.30037	15.16725	3.0	4.631165	$0.01/h$

Table 4.7.2: Data for the transient Riemann problems.



(a) Method AcTrZ

(b) Method TrZ

Figure 4.7.4: Test A; free surface (up) and bed elevation (bottom) computed with AcTrZ method (left) and TrZ scheme (right). Reference solution (red line), first-order solution (blue symbol -.) and second-order solution (black dashed line). $M = 200$ cells.

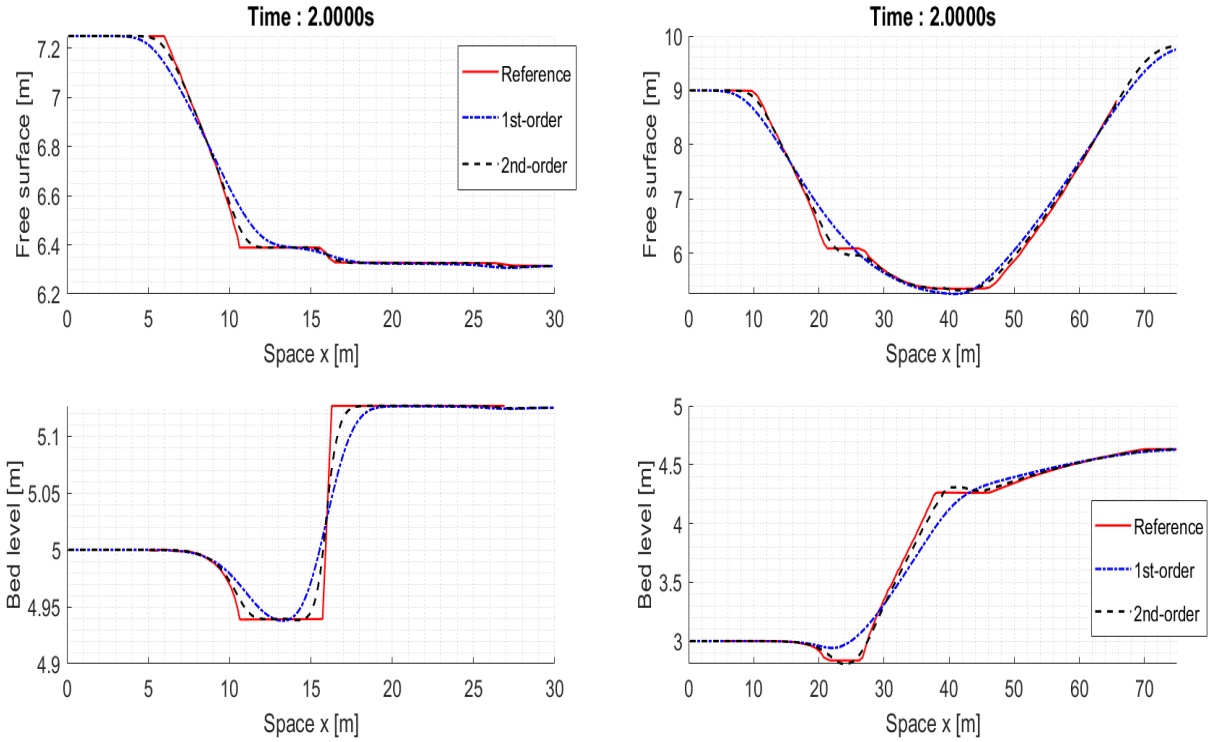
accuracy in the stable parts of the solution, resembling a kind of flux-limiter [49] approach (if a priori*) or a MOOD [22] approach (if* a posteriori*).*

4.7.4 Sub-critical test case

For the following two numerical tests we refer to paper [23]. As initial condition we consider the sub-critical steady state

$$\begin{cases} hu(x, t = 0) = 0.5 \\ z(x, t = 0) = 0.1(1 + e^{-(x-5)^2}) \\ \frac{u^2}{2} + g(h + z) = 6.386, \end{cases}$$

while the length of the channel is $L = 10.0m$. In figure 4.7.7 we insert the results for $A_g = 0.05$ and $A_g = 0.007$. In the latter case, AcZ-solution is not inserted for the sake of clarity as it is very similar to the AcTrZ-one. We observe that the three schemes give similar solutions, which confirm the observations of the previous test case; in order from the least to the most diffusive, we have TrZ, AcZ and AcTrZ-scheme. It is also important to remark that, in these



(a) Test B

(b) Test C

Figure 4.7.5: Test B (left) and C (right); free surface (up) and bed elevation (bottom) computed with AcTrZ method. Reference solution (red line), first-order solution (blue pointed line) and second-order solution (black dashed line). $M = 200$ cells.

two numerical simulations, no oscillations appear even if, in the work [23], it has been explained that decoupled methods could present oscillations in these two test cases.

4.7.5 "Lake at rest" solution and perturbation

Referring to [20], here we test the ability of the schemes to preserve the "lake at rest" steady state. First of all, we consider the following stationary solution where $u = 0$, $h(x, t = 0) + z(x, t = 0) = 3m$ and

$$z(x, t = 0) = \begin{cases} 2 + 0.25(\cos(10\pi(x - 0.5)) + 1) & \text{if} \quad 1.4 < x < 1.6 \\ 2 & \text{otherwise.} \end{cases}$$

The length of the channel is $L = 2.0m$. The numerical schemes were able to preserve this steady state up to an error of order 10^{-15}. Then, we introduce some small perturbations in the initial data, namely we impose

$$h(x, t = 0) = \begin{cases} 3 - z(x, t = 0) + 0.001 & \text{if} \quad 1.1 < x < 1.2 \\ 3 - z(x, t = 0) & \text{otherwise.} \end{cases}$$

In figure 4.7.8, we compare the results of first and second-order AcZ-scheme against the reference solution. We only show the outputs of the AcZ-method as the other two schemes give analogous results. We observe that the outcomes are indeed satisfying as they are in agreement with the ones showed in [20] and no unphysical oscillations appear.

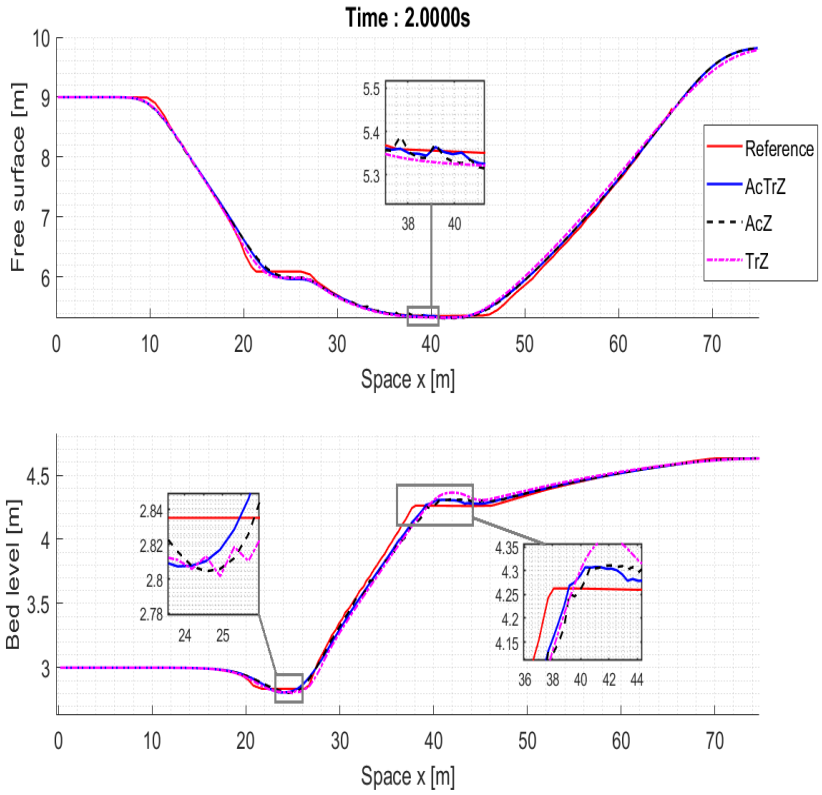
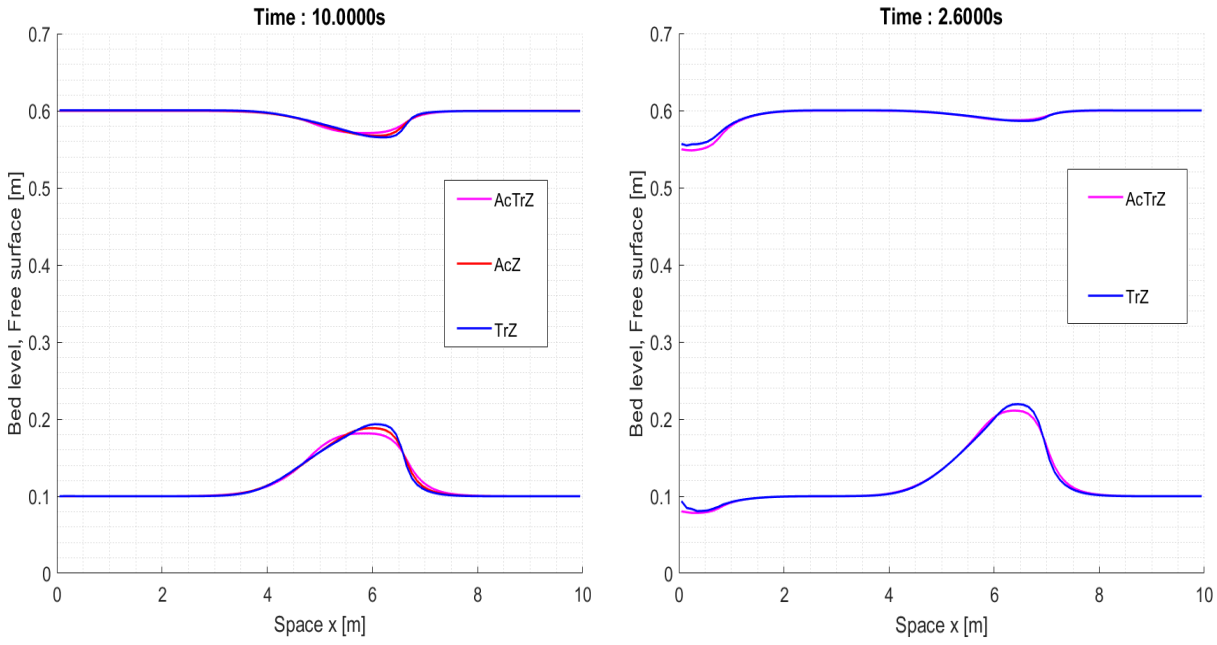


Figure 4.7.6: Test C; free surface (up) and bed level (bottom) computed with the three 2nd order schemes. AcTrZ (blue line), AcZ (black dashed line), TrZ (magenta pointed line) and reference solution (red line). $M = 200$ cells.



(a) $A_g = 0.005$ (b) $A_g = 0.07$

Figure 4.7.7: Flow over a movable bump; free surface and bed level. Values $A_g = 0.005$ (up) and $A_g = 0.07$ (bottom). Schemes AcTrZ (magenta), AcZ (yellow) and TrZ (blue) with $M = 100$ cells.

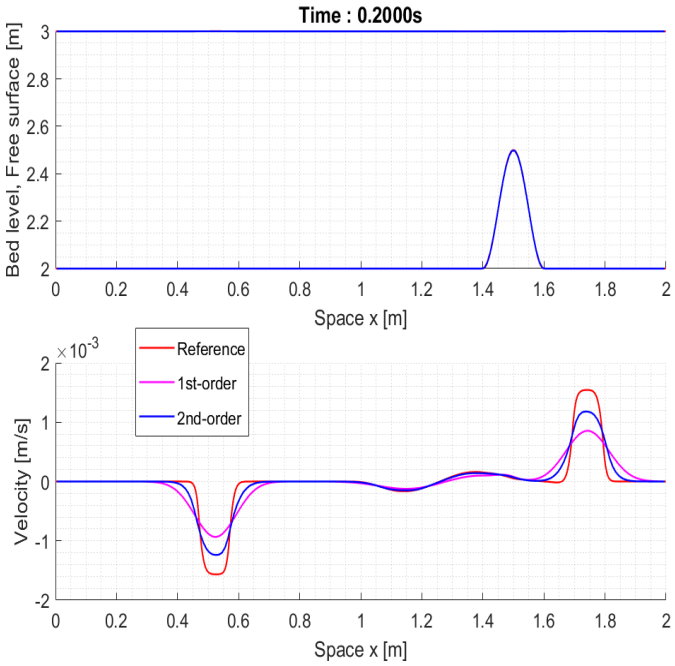


Figure 4.7.8: Propagation of perturbation; bed level z and free surface $z+h$ (top), velocity (bottom). Comparison between the first (magenta) and second-order (blue) LP-AcZ scheme with $M = 200$ cells. Reference solution in red line.

4.7.6 "Constant bed slope" equilibrium for steady flow regimes

In this section, we numerically show that our method is capable of preserving the "constant bed slope" equilibrium and that the solution evolves to the steady state if steady boundary conditions are imposed. We refer to [42] and we take $\mu_f = 0.020006460818026$ s m$^{-1/3}$, $A_g = 0.01$, $L = 100$ m and $M = 100$ cells so that $\Delta x = 1$m. Then, for the equilibrium, we consider the following slope $S_{eq} = -0.002$ for z. Our numerical methods are able to preserve it with an error machine of 10^{-12}. We highlight that, at the left and right boundaries, for the variable z we need to impose constant slope. For instance, for the right boundary, this means that the value in the ghost cell is given by $z_{\text{out}} - z_{\text{end}} = z_{\text{end}} - z_{\text{end}-1}$, where by z_{end} and $z_{\text{end}-1}$ we indicate the value of z in the last and second-last cell.

Then, we move away from the steady state and consider as initial condition $h(x, t = 0) = 0.943$m, $q(x, t = 0) = 1$ m^3/s and either $S_0 = -0.007$ or $S_0 = 0$. As for the boundary conditions, at the inlet we impose $z_{\text{In}} = 2$ m and $q_{\text{In}} = 1$ m^3/s, while at the outlet we use $h_{\text{Out}} = 0.943$ m. Once again, at the right boundary, for the variable z we imposed that the slope is constant. Then, in figure 4.7.9 we insert the results using $S_0 = -0.007$ (left) and $S_0 = 0$ (right). In both cases, we observe that the numerical solution converges towards the exact one, namely the "constant bed slope" equilibrium. The outputs have been computed with the first-order LP-AcTrZ scheme, but all the numerical methods we presented give analogous results.

4.7.7 Dam break with experimental values

Here we present the last numerical test, in which we compare our numerical solution against experimental data. We consider once again a Riemann problem performed at the Université catholique de Louvain with initial condition given by zero-velocity, $h_L = 0.1\ m$, $h_R = 1e-3\ m$ for the water height and flat topography. Then, we take $L = 2.5\ m$, MPM formula (4.1.3) for the solid transport discharge with $d = 3.2\ mm$, $\rho_0 = 0.4$, $s = 0.540$,

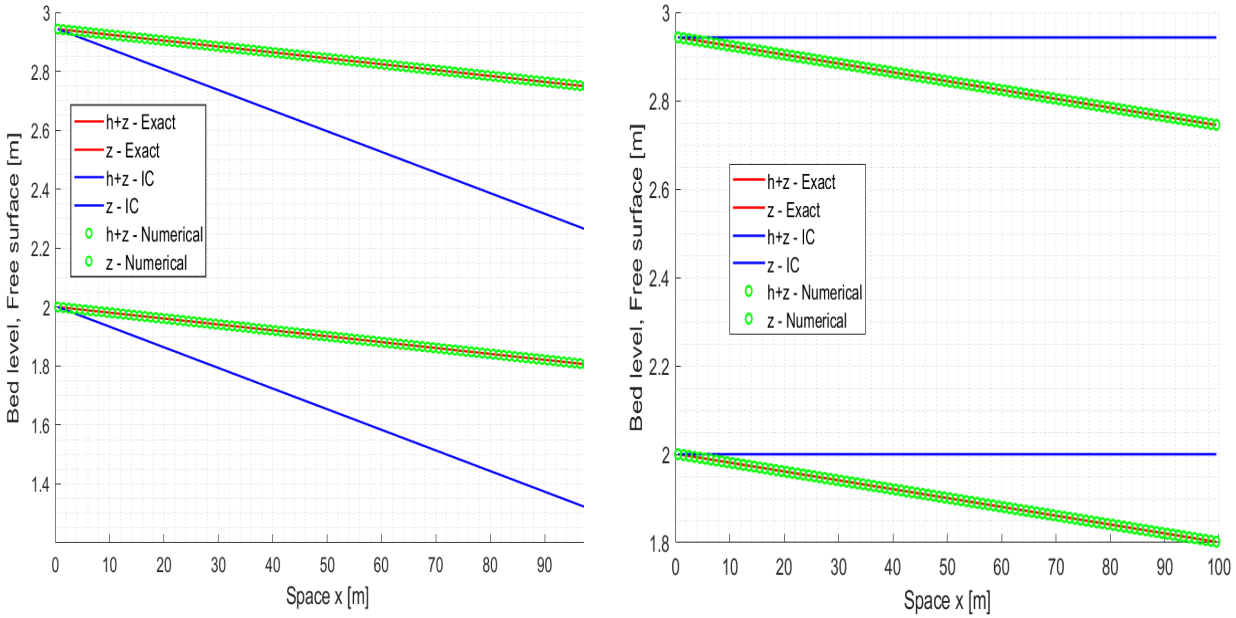


Figure 4.7.9: "Constant bed slope" equilibrium for steady flow regimes; bed level z and free surface $z+h$. Initial condition (blue line), numerical steady state (green symbol) and exact steady state (red line). LP-AcTrZ scheme with $M = 100$ cells.

$\mu_f = 0.03$ and $\theta_c^* = 0.045$. Refer to [34] for more details about this experiment. In figure 4.7.10, we insert the free surface and bed elevation numerical and experimental outputs computed at different times $t_{end} = 5t_0, 7.5t_0, 10t_0\ s$ with $t_0 = \sqrt{gh_0} \approx 0.101$. In particular, we used the AcZ scheme but analogous solution can be found using the AcTrZ or TrZ method, where once again the TrZ solution would be the less diffusive while the AcTrZ output the most diffusive. The results are considered satisfying as they appear to match the experimental data and are comparable to the ones obtained in [34]. Note that the difference between the numerical and experimental output could be related to the fact that we are both neglecting erosion processes (meaning that smaller fractions of sediment could be in suspension into the water) and non-hydrostatic effects, refer also to [32]. Then, when using the second-order schemes we observed the production of some small oscillations which were dumped as time went on. Thus, in figure 4.7.12, we inserted the second-order solutions using the AcTrZ and AcZ schemes at times $t_{end} = 5t_0, 7.5t_0, 10t_0\ s$, while in figure 4.7.11 at time $t_{end} = 1.5t_0$ to better show the presence of oscillations. Note that we did not inserted the second-order TrZ solution as it gave complex values. Concluding, even if small oscillations are present, probably due to the fact that at second-order of accuracy there is less diffusion, the solutions are considered satisfying. Moreover, at this stage we are still neglecting source terms related to the erosion phenomena of the sediment, which could help stabilizing the numerical output, see [32].

4.8 2D extension of the shallow water Exner model

During this thesis, the study of Lagrange-projection schemes for the Saint-Venant-Exner model has been realized in two steps. First, an easier strategy (the one that only updates z in the transport step) has been developed in 1D and 2D and presented in a first paper. Then, two other weakly coupled approaches (which also consider the Exner equation in the acoustic step) have been proposed in a second work. For this reason, as a second part of this chapter, we include the 2D version of the model together with the Lagrange-Projection scheme which updates the bed elevation only in the transport step, see the previous sections 4.3.3, 4.4 and

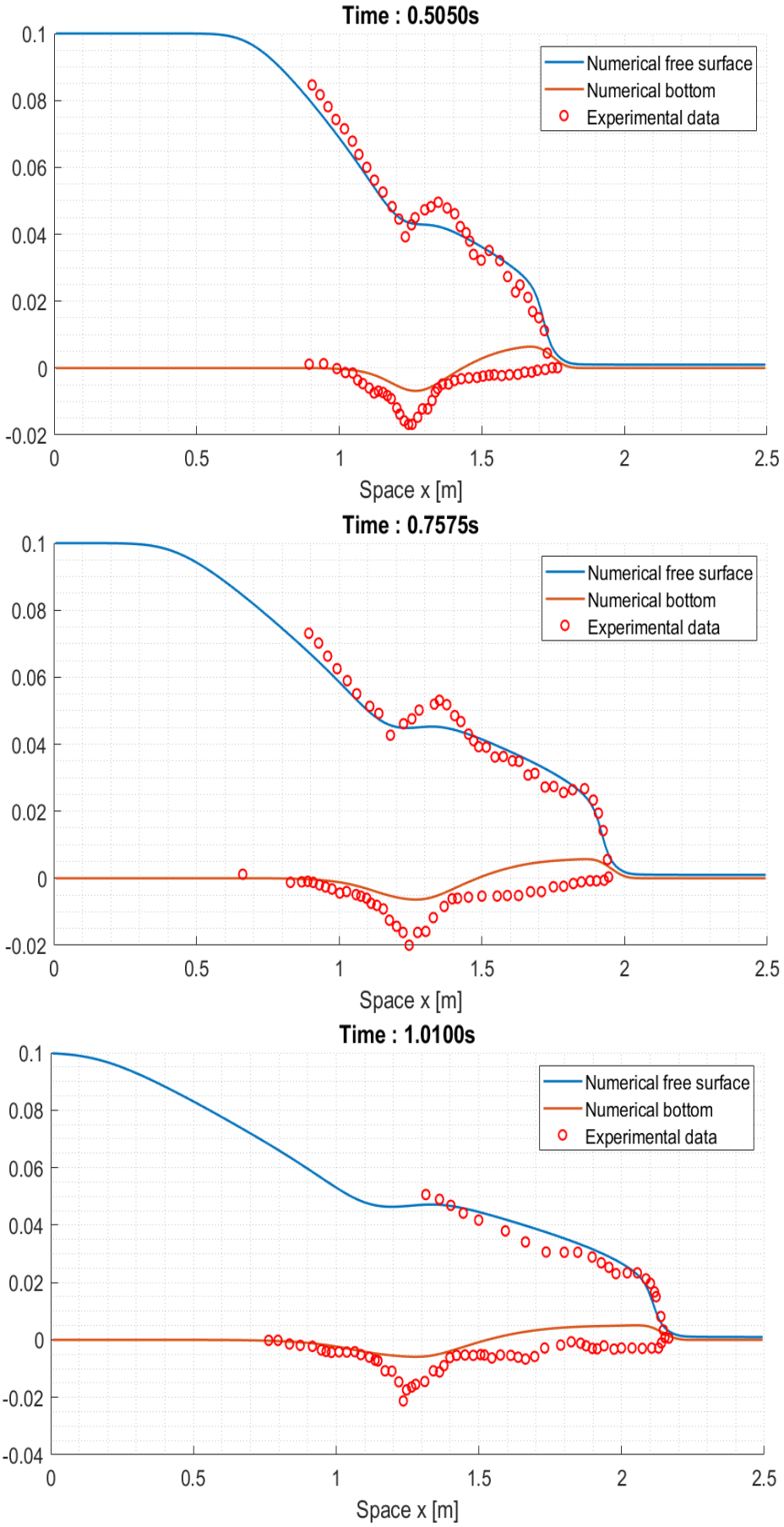


Figure 4.7.10: Dam break with experimental values; numerical bed level z (red line), numerical free surface $z+h$ (blue line) and experimental data (red symbol). Solution at different times: $t = 5t_0\ s$ (up), $t = 7.5t_0\ s$ (middle), $t = 10t_0\ s$ (bottom). LP-AcZ scheme with $M = 200$ cells.

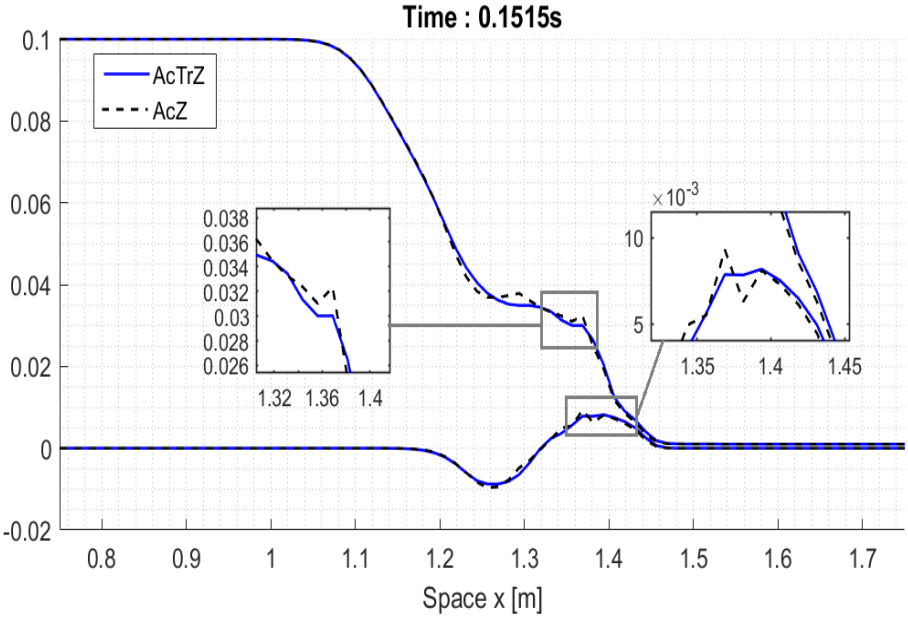


Figure 4.7.11: Dam break with experimental values; AcTrZ scheme (blue line) and AcZ scheme (black dashed line). Solution at time $t = 1.5t_0$ s. $M = 200$ cells.

4.5.3.

Hence, we start by presenting the 2D shallow water Exner system (without friction forces). Denoting $(x, y) \in \mathbb{R}^2$ the space variables in 2D and $\mathbf{u} = (u, v)^T$ the velocity vector, the 2D shallow water system reads

$$\begin{cases} \partial_t h + \nabla \cdot (h\mathbf{u}) = 0 \\ \partial_t (h\mathbf{u}) + \nabla \cdot (h\mathbf{u} \otimes \mathbf{u}) + \nabla p = -gh\nabla z, \end{cases} \tag{4.8.1}$$

while the Exner equation is given by

$$\partial_t z + \zeta \partial_x q_{b,x} + \zeta \partial_y q_{b,y} = 0$$

where $q_{b,x}$ and $q_{b,y}$ are the solid transport discharges in the x and y direction respectively. Exploiting once again the Grass model, their formula are the following,

$$q_{b,x} = A_g u(u^2 + v^2) \quad \text{and} \quad q_{b,y} = A_g v(u^2 + v^2).$$

Notice that the "lake at rest" stationary solution now satisfies

$$u = 0, \quad v = 0 \quad \text{and} \quad \nabla(h + z) = 0.$$

Refer to [36, 43, 44] for details about the shallow water Exner system in 2D.

At this stage, we observe that a Lagrangian formulation of these equations is still possible. More precisely, let us introduce the Lagrangian coordinates, considering a fluid particle in position (ξ_1, ξ_2) and the map : $(\xi_1, \xi_2) \to (x, y)$, with $x = x(\xi_1, \xi_2, t)$, $y = y(\xi_1, \xi_2, t)$ and such that

$$\frac{\partial x}{\partial t} = u(x, y, t) \quad \text{and} \quad \frac{\partial y}{\partial t} = v(x, y, t),$$
$$x(\xi_1, \xi_2, 0) = \xi_1, \quad y(\xi_1, \xi_2, 0) = \xi_2.$$

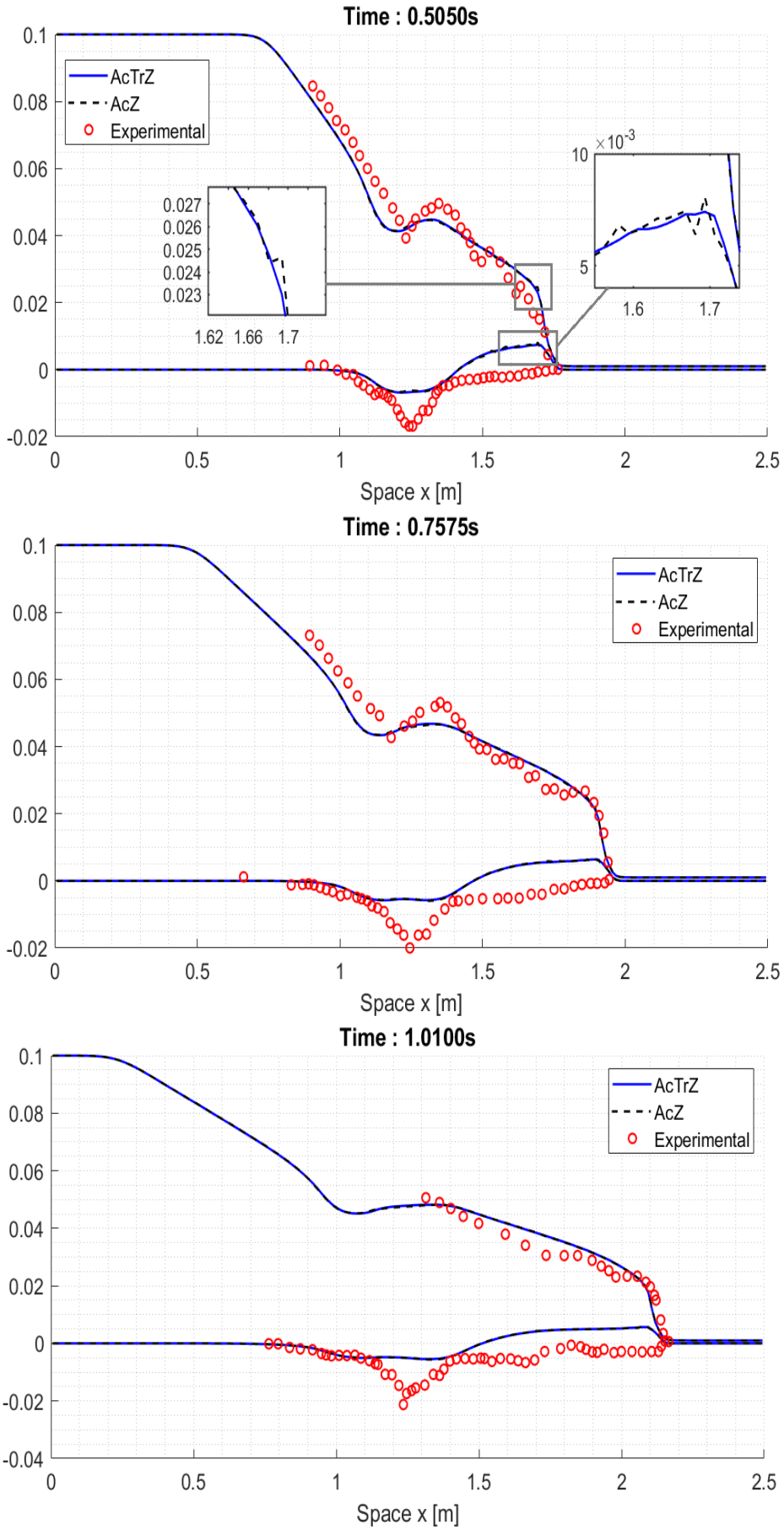


Figure 4.7.12: Dam break with experimental values; AcTrZ scheme (blue line), AcZ scheme (black dashed line) and experimental data (red symbol). Solution at times $t = 5t_0\ s$ (up), $t = 7.5t_0\ s$ (middle), $t = 10t_0\ s$ (bottom). $M = 200$ cells.

We also assume that for each $t > 0$, this map is invertible and its Jacobian (determinant of the Jacobian matrix) is given by

$$L(\xi_1, \xi_2, t) = \begin{vmatrix} \partial_{\xi_1} x & \partial_{\xi_2} x \\ \partial_{\xi_1} y & \partial_{\xi_2} y \end{vmatrix}$$

with $L(\xi_1, \xi_2, 0) = 1$ and, after easy calculations,

$$\frac{\partial L(\xi_1, \xi_2, t)}{\partial t} = L\nabla \cdot \mathbf{u} = L\partial_x u + L\partial_y v. \tag{4.8.2}$$

Then, it can be shown that the Lagrangian formulation of the system reads

$$\begin{cases} \partial_t(Lh) = 0 \\ \partial_t(Lh\mathbf{u}) + L\nabla p = -ghL\nabla z, \end{cases} \tag{4.8.3}$$

where the gradient is still taken with respect to the Eulerian variables (x, y). On the other hand, we also have

$$\partial_t(L\mathbf{u}) - \mathbf{u}L\nabla \cdot \mathbf{u} + gL\nabla(h + z) = 0,$$

which will be useful hereafter. We refer the reader to [39, 40, 26] for more details about Lagrangian coordinates in 2D.

4.8.1 2D Numerical scheme

We now give some basic notations about the discretization of the domain before presenting the 2D extension of the numerical scheme.

First of all, the computational domain $\Omega \subset \mathbb{R}$ is divided into $M_x \times M_y$ rectangular cells with constant space steps Δx and Δy in the x and y directions respectively. Then, the mesh interfaces are given by $x_{i+1/2}$ for $i \in \{0, \dots, M_x\}$ and $y_{j+1/2}$ for $j \in \{0, \dots, M_y\}$. Thus, $\varphi^n_{i,j}$ denotes the piecewise constant approximation of the variable φ in the cell $[x_{i-1/2}, x_{i+1/2}) \times [y_{j-1/2}, y_{j+1/2})$ at time t^n, namely

$$\varphi^n_{i,j} \approx \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \int_{y_{j-1/2}}^{y_{j+1/2}} \varphi(x, y, t^n)\, dx\, dy.$$

Next, we present the 2D extension of the acoustic and transport steps of the LP strategy. We highlight that here the bed elevation z is only updated in the transport step.

Acoustic step

Following the 1D scheme presented in section 4.4.1, the numerical approximation of (4.8.3) reads

$$\begin{cases} (Lh)^{n+1-}_{i,j} = (Lh)^n_{i,j} \\ (Lhu)^{n+1-}_{i,j} = (Lhu)^n_{i,j} - \dfrac{\Delta t}{\Delta x}(\Pi^*_{i+\frac{1}{2},j} - \Pi^*_{i-\frac{1}{2},j}) + \Delta t s^{1,n}_{i,j} \\ (Lhv)^{n+1-}_{i,j} = (Lhv)^n_{i,j} - \dfrac{\Delta t}{\Delta y}(\Pi^*_{i,j+\frac{1}{2}} - \Pi^*_{i,j-\frac{1}{2}}) + \Delta t s^{2,n}_{i,j} \end{cases} \tag{4.8.4}$$

where we imposed

$$L^{n+1-}_{i,j} = L^n_{i,j} + \frac{\Delta t}{\Delta x}(u^*_{i+\frac{1}{2},j} - u^*_{i-\frac{1}{2},j}) + \frac{\Delta t}{\Delta y}(v^*_{i,j+\frac{1}{2}} - v^*_{i,j-\frac{1}{2}}) \tag{4.8.5}$$

with

$$\begin{aligned}
u^*_{i+\frac{1}{2},j} &= \frac{1}{2}(u^n_{i+1,j} + u^n_{i,j}) - \frac{1}{2a^n_{i+\frac{1}{2},j}}(\Pi^n_{i+1,j} - \Pi^n_{i,j}) - \frac{\mathcal{M}^n_{i+1/2,j}}{2a_{i+1/2,j}},\\
v^*_{i,j+\frac{1}{2}} &= \frac{1}{2}(v^n_{i,j+1} + v^n_{i,j}) - \frac{1}{2a^n_{i,j+\frac{1}{2}}}(\Pi^n_{i,j+1} - \Pi^n_{i,j}) - \frac{\mathcal{M}^n_{i,j+1/2}}{2a_{i,j+1/2}},\\
\Pi^*_{i+\frac{1}{2},j} &= \frac{1}{2}(\Pi^n_{i+1,j} + \Pi^n_{i,j}) - \frac{a^n_{i+\frac{1}{2},j}}{2}(u^n_{i+1,j} - u^n_{i,j}),\\
\Pi^*_{i,j+\frac{1}{2}} &= \frac{1}{2}(\Pi^n_{i,j+1} + \Pi^n_{i,j}) - \frac{a^n_{i,j+\frac{1}{2}}}{2}(v^n_{i,j+1} - v^n_{i,j}),
\end{aligned} \tag{4.8.6}$$

with $a^n_{i+1/2,j} = \max((hc)^n_{i,j}, (hc)^n_{i+1,j})$, $a^n_{i,j+1/2} = \max((hc)^n_{i,j}, (hc)^n_{i,j+1})$. Regarding the source term, we have for all j

$$s^{1,n}_{i,j} = \frac{1}{2}\Big(s^n_{i+1/2,j} + s^n_{i-1/2,j}\Big) \quad \text{with} \quad s^n_{i+1/2,j} = -\frac{\mathcal{M}^n_{i+1/2,j}}{\Delta x}$$

$$s^{2,n}_{i,j} = \frac{1}{2}\Big(s^n_{i,j+1/2} + s^n_{i,j-1/2}\Big) \quad \text{with} \quad s^n_{i,j+1/2} = -\frac{\mathcal{M}^n_{i,j+1/2}}{\Delta y}$$

with

$$\mathcal{M}^n_{i+1/2,j} = \frac{g}{2}(\frac{1}{\tau^n_{i,j}} + \frac{1}{\tau^n_{i+1,j}})(z^n_{i+1,j} - z^n_{i,j}), \quad \mathcal{M}^n_{i,j+1/2} = \frac{g}{2}(\frac{1}{\tau^n_{i,j}} + \frac{1}{\tau^n_{i,j+1}})(z^n_{i,j+1} - z^n_{i,j}).$$

It is clear that the numerical formulas (4.8.6) are a natural extension of the ones used for the 1D system, namely (4.3.9). Therefore, if we want to compute h and $\mathbf{u}$ in a cell of indexes (i,j), we consider the flux contributions from cells $(i \pm 1, j)$ and $(i, j \pm 1)$. This also means that we are not considering the values in cells of indexes $(i \pm 1, j \pm 1)$, $(i \pm 1, j \mp 1)$, namely the cells which only share a corner with the cell (i,j). Although this is a standard approach, in future works it might actually be interesting to consider a "truly" multi-dimensional version of the scheme, refer for instance to [6].

Transport step

As before, the second step of the Lagrange-projection scheme consists in either projecting the acoustic solution onto the Eulerian grid or in approximating the transport system

$$\partial_t \varphi + \mathbf{u} \cdot \nabla \varphi = 0$$

or equivalently

$$\partial_t \varphi + \nabla \cdot (\varphi \mathbf{u}) - \varphi \nabla \cdot \mathbf{u} = 0,$$

where we took $\varphi = h, hu, hv$. Here and analogously to the 1D formulation (4.4.8), we simply set

$$\begin{aligned}
\varphi^{n+1}_{i,j} =& (L\varphi)^{n+1-}_{i,j} - \frac{\Delta t}{\Delta x}\Big(u^*_{i+\frac{1}{2},j}(L\varphi)^{n+1-}_{i+\frac{1}{2},j} - u^*_{i-\frac{1}{2},j}(L\varphi)^{n+1-}_{i-\frac{1}{2},j}\Big) +\\
&- \frac{\Delta t}{\Delta y}\Big(v^*_{i,j+\frac{1}{2}}(L\varphi)^{n+1-}_{i,j+\frac{1}{2}} - v^*_{i,j-\frac{1}{2}}(L\varphi)^{n+1-}_{i,j-\frac{1}{2}}\Big),
\end{aligned} \tag{4.8.7}$$

where

$$(L\varphi)^{n+1-}_{i-\frac{1}{2},j} = \begin{cases} (L\varphi)^{n+1-}_{i-1,j} & \text{if} \quad u^*_{i-\frac{1}{2},j} > 0 \\ (L\varphi)^{n+1-}_{i,j} & \text{if} \quad u^*_{i-\frac{1}{2},j} \leq 0, \end{cases}$$

and

$$(L\varphi)^{n+1-}_{i,j-\frac{1}{2}} = \begin{cases} (L\varphi)^{n+1-}_{i,j-1} & \text{if} \quad v^*_{i,j-\frac{1}{2}} > 0 \\ (L\varphi)^{n+1-}_{i,j} & \text{if} \quad v^*_{i,j-\frac{1}{2}} \leq 0. \end{cases}$$

As for the Exner equation, as a natural extension of (4.4.12), we set

$$\begin{aligned} z^{n+1}_{i,j} =& z^n_{i,j} - \zeta \frac{\Delta t}{\Delta x}\Big(u^*_{i+\frac{1}{2},j}\Big(\frac{q_{b,x}}{u}\Big)^{n+1-}_{i+\frac{1}{2},j} - u^*_{i-\frac{1}{2},j}\Big(\frac{q_{b,x}}{u}\Big)^{n+1-}_{i-\frac{1}{2},j}\Big) + \\ & - \zeta \frac{\Delta t}{\Delta y}\Big(v^*_{i,j+\frac{1}{2}}\Big(\frac{q_{b,y}}{v}\Big)^{n+1-}_{i,j+\frac{1}{2}} - v^*_{i,j-\frac{1}{2}}\Big(\frac{q_{b,y}}{v}\Big)^{n+1-}_{i,j-\frac{1}{2}}\Big) \end{aligned}$$

with

$$(\frac{q_{b,x}}{u})^{n+1-}_{i+1/2,j} = \begin{cases} (\frac{q_{b,x}}{u})\Big((L\mathbf{u})^{n+1-}_{i+1,j}\Big) & \text{if} \quad u^*_{i+1/2,j} \leq 0 \\ (\frac{q_{b,x}}{u})\Big((L\mathbf{u})^{n+1-}_{i,j}\Big) & \text{if} \quad u^*_{i+1/2,j} > 0, \end{cases}$$

and

$$(\frac{q_{b,y}}{v})^{n+1-}_{i,j+1/2} = \begin{cases} (\frac{q_{b,y}}{v})\Big((L\mathbf{u})^{n+1-}_{i,j+1}\Big) & \text{if} \quad v^*_{i,j+1/2} \leq 0 \\ (\frac{q_{b,y}}{v})\Big((L\mathbf{u})^{n+1-}_{i,j}\Big) & \text{if} \quad v^*_{i,j+1/2} > 0, \end{cases}$$

where a possible discretization of the evolution equations for Lu and Lv read

$$\begin{aligned} (Lu)^{n+1-}_{i,j} =& (Lu)^n_{i,j} + \Delta t \frac{u_{i+1,j} + u_{i,j}}{2}\big(\frac{1}{\Delta x}(u^*_{i+\frac{1}{2},j} - u^*_{i-\frac{1}{2},j}) + \\ & + \frac{1}{\Delta y}(v^*_{i,j+\frac{1}{2}} - v^*_{i,j-\frac{1}{2}})\big) - \Delta t \frac{\hat{s}_{i+\frac{1}{2},j} + \hat{s}_{i-\frac{1}{2},j}}{2} \end{aligned}$$

and

$$\begin{aligned} (Lv)^{n+1-}_{i,j} =& (Lv)^n_{i,j} + \Delta t \frac{v_{i,j+1} + v_{i,j}}{2}\big(\frac{1}{\Delta x}(u^*_{i+\frac{1}{2},j} - u^*_{i-\frac{1}{2},j}) + \\ & + \frac{1}{\Delta y}(v^*_{i,j+\frac{1}{2}} - v^*_{i,j-\frac{1}{2}})\big) - \Delta t \frac{\hat{s}_{i,j+\frac{1}{2}} + \hat{s}_{i,j-\frac{1}{2}}}{2} \end{aligned}$$

where $\hat{s}_{i+\frac{1}{2},j} = g((h+z)_{i+1,j} - (h+z)_{i,j})/\Delta x$ and $\hat{s}_{i,j+\frac{1}{2}} = g((h+z)_{i,j+1} - (h+z)_{i,j})/\Delta y$.

2D Extension of the second-order scheme

We now briefly discuss the extension of the second-order scheme presented in section 4.5.3, distinguishing among the Exner equation and the Lagrangian and projection steps for the shallow water system. Once again we reach the second order of accuracy in time exploiting the Runge-Kutta procedure, which is applied to the Lagrangian and projection steps together. As expected, the overall strategy is analogous to what we have done for the 1D case.

Regarding the Lagrangian step, we proceed as in the 1D case and compute the numerical fluxes $u^*_{i+\frac{1}{2},j}$, $v^*_{i,j+\frac{1}{2}}$, $\Pi^*_{i+\frac{1}{2},j}$ and $\Pi^*_{i,j+\frac{1}{2}}$, but also the speeds $a_{i+1/2,j}$ and $a_{i,j+1/2}$, using the left and right interfaces values defined by means of reconstructed polynomials, namely

$$\mathbf{V}^n_{i+1/2L,j} = \mathbf{V}^n_{i,j} + \Delta^{x,t}_{i,j}\frac{\Delta x}{2}, \quad \mathbf{V}^n_{i+1/2R,j} = \mathbf{V}^n_{i+1,j} - \Delta^{x,t}_{i+1,j}\frac{\Delta x}{2} \tag{4.8.8}$$

in x direction, and

$$\mathbf{V}^n_{i,j+1/2L} = \mathbf{V}^n_{i,j} + \Delta^{y,t}_{i,j}\frac{\Delta y}{2}, \quad \mathbf{V}^n_{i,j+1/2R} = \mathbf{V}^n_{i,j+1} - \Delta^{y,t}_{i,j+1}\frac{\Delta y}{2} \tag{4.8.9}$$

along the y axis, where we have set $\mathbf{V} = (h, hu, hv)^T$. At this stage, we only need to define the slopes $\Delta^{x,t}_{i,j}$, $\Delta^{y,t}_{i,j}$ which are exactly computed as in the 1D case. We consider the definition of the fluctuations in the x, respectively y, direction, we take $y = y_j$, resp. $x = x_i$, fixed and we use reconstructed stationary solutions direction by direction. The details are left to the reader. In particular, such a strategy guarantees the well-balanced property of the numerical scheme, since the slopes turn out to be null under the "lake at rest" conditions.

Regarding the transport step, we consider a direct 2D extension of (4.5.5), namely

$$\begin{aligned} X^{n+1}_{i,j} = {} & (LX)^{n+1-}_{i,j} + \\ & - \frac{\Delta t}{\Delta x}\Big(u^*_{i+\frac{1}{2},j}(LX)^{n+1-}_{i+\frac{1}{2},j}\Big(\frac{\xi_{1,i+\frac{1}{2}} + \hat{\xi}_{1,i+\frac{1}{2}}}{2}\Big) - u^*_{i-\frac{1}{2},j}(LX)^{n+1-}_{i-\frac{1}{2},j}\Big(\frac{\xi_{1,i-\frac{1}{2}} + \hat{\xi}_{1,i-\frac{1}{2}}}{2}\Big)\Big) \\ & - \frac{\Delta t}{\Delta y}\Big(v^*_{i,j+\frac{1}{2}}(LX)^{n+1-}_{i,j+\frac{1}{2}}\Big(\frac{\xi_{2,j+\frac{1}{2}} + \hat{\xi}_{2,j+\frac{1}{2}}}{2}\Big) - v^*_{i,j-\frac{1}{2}}(LX)^{n+1-}_{i,j-\frac{1}{2}}\Big(\frac{\xi_{2,j-\frac{1}{2}} + \hat{\xi}_{2,j-\frac{1}{2}}}{2}\Big)\Big), \end{aligned}$$

where we have used clear notations based on classical first-order polynomial reconstructions of the Lagrangian unknowns (LX) in each direction as in the 1D case, clearly with $X = h, hu, hv$.

Finally, exploiting the reconstructed values for Lu and Lv and using again classical notations, we suggest a direct 2D extension of (4.4.12) namely

$$\begin{aligned} z^{n+1}_{i,j} = {} & z^n_{i,j} - \zeta\frac{\Delta t}{\Delta x}\Big(u^*_{i+1/2,j}\Big(\frac{q_{b,x}}{u}\Big)^{n+1-}_{i+1/2,j} - u^*_{i-1/2,j}\Big(\frac{q_{b,x}}{u}\Big)^{n+1-}_{i-1/2,j}\Big) + \\ & - \zeta\frac{\Delta t}{\Delta y}\Big(v^*_{i,j+1/2,}\Big(\frac{q_{b,y}}{v}\Big)^{n+1-}_{i,j+1/2} - u^*_{i,j-1/2}\Big(\frac{q_{b,y}}{v}\Big)^{n+1-}_{i,j-1/2}\Big) \end{aligned}$$

with a natural definition for the numerical fluxes $\frac{q_{b,x}}{u}$ and $\frac{q_{b,y}}{v}$. Again, the details are left to the reader since there is no ambiguity.

To conclude this 2D section, let us mention that both schemes described here preserve the "lake at rest" stationary solution. The proof is analogous to the one seen in 1D.

4.8.2 Numerical results

Finally, this section is devoted to the presentation of the simulations and outputs of the 2D numerical scheme we described so far. Regarding the 2D time step value, at each time t^n we compute two different time steps, one for the acoustic system and the other for the transport part. Then, the final time step is taken as the minimum between the two. The acoustic time step is automatically extended, see formula (4.4.1). Indeed, it is enough to consider both directions. Then, for the transport time step, we state

$$\Delta t \leq \mathrm{CFL}_t \min_{i,j}\Big\{\Big\{\frac{u^+_{i-\frac{1}{2},j} - u^-_{i+\frac{1}{2},j}}{\Delta x} + \frac{v^+_{i,j-\frac{1}{2}} - v^-_{i,j+\frac{1}{2}}}{\Delta y}\Big\}^{-1}\Big\}.$$

Once again, if not otherwise specified, we impose transmissive boundary conditions, $\mathrm{CFL}_t = 1$, while $\mathrm{CFL}_l = 0.45$ and $\mathrm{CFL}_l = 0.25$ for the first and second order schemes respectively.

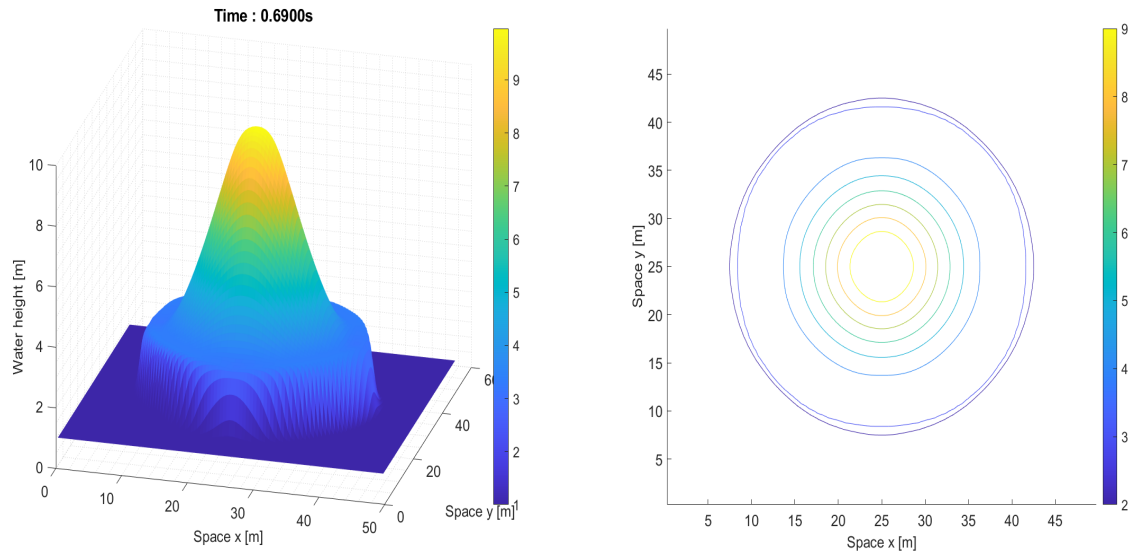


Figure 4.8.1: 2D circular dam break on wet bed; water height (left) and its contour plot (right). 2D extension of first-order scheme solution. $M = 100$ cells, $t_{\text{End}} = 0.69$s and $\text{CFL}_l = 0.45$.

4.8.3 Circular dam break on wet bed

In this first test, we do not take into account the Exner equation and we refer to [48]. The domain is a $L \times L$ square with $L = 50$m. Here as initial condition we consider a flat topography, null velocities in both the x and y directions and water height

$$h(x, y, t = 0) = \begin{cases} 10 & \text{if} \quad r \leq 11\text{m} \\ 1 & \text{if} \quad r > 11\text{m}, \end{cases}$$

with $r = \sqrt{(x - 25)^2 + (y - 25)^2}$. Thus, we are considering a cylindrical dam that instantaneously breaks at initial time $t = 0$s. We take ending time $t_{\text{End}} = 0.69$s. Satisfying results of the first order scheme are reported in figure 4.8.1.

4.8.4 Water drop in a basin

For this numerical test, we refer to [3]. Here we simulate a water drop in a basin and consequently reflective boundary conditions are used. A L-side square domain with $L = 20$m is considered. The topography is still taken flat and constant in time. At initial time we assume $\mathbf{u} = (0, 0)^t$ and

$$h(x, y, t = 0) = 2.4(1 + e^{-0.25((x-10.05)^2+(y-10.05)^2)})$$

The outputs are shown in figure 4.8.2 at time $t = 1$s, $t = 2$s, $t = 3$s and $t = 4$s respectively. The results agree with the ones reported in [3]. In particular, in the same picture 4.8.2, we compare the results at time $t = 4$s obtained using the 2D extensions of the first-order and second-order schemes. We can clearly see that the latter scheme gives less diffusive solutions for the same mesh value $M = 100$ cells.

4.8.5 2D flow over a smooth bump

The following test problem is useful to check the well-balanced property of the scheme, see [48]. The domain is a square of side $L = 1$m and we consider the Grass formulation for the Exner equation with $A_g = 1$ and $\zeta = 1$. At initial time we assume null velocities,

$$z(x, y, t = 0) = \max(0, 0.25 - 5((x - 0.5)^2 + (y - 0.5)^2)) \text{ and } h(x, y, t = 0) = 0.5 - z(x, y).$$

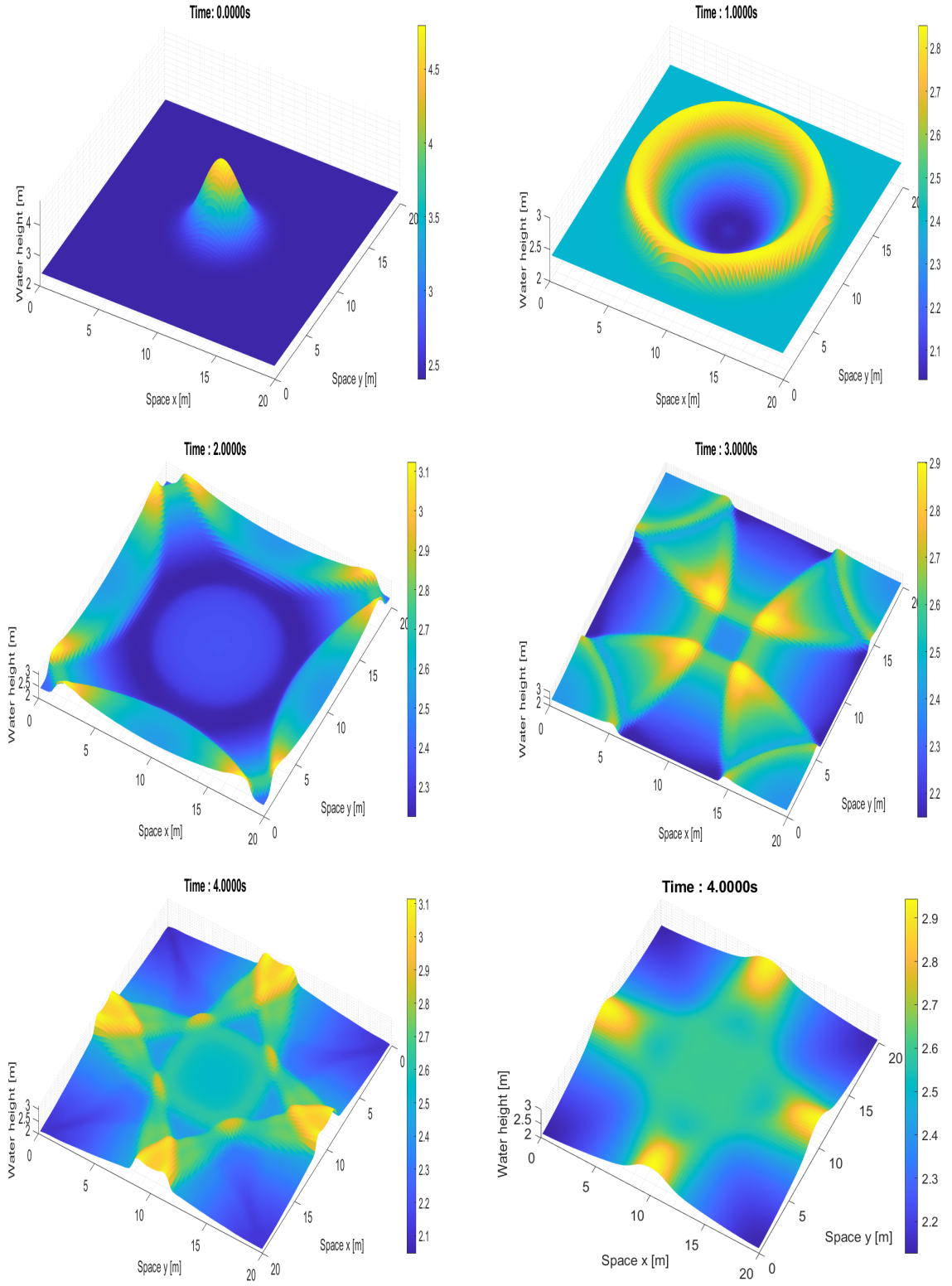


Figure 4.8.2: Water drop in a basin; water height at time $t = 0$s (up, left), $t = 1$s (up, right), $t = 2$s (middle, left), $t = 3$s (middle, right) and $t = 4$s (bottom). $M = 100$ cells, $\text{CFL}_l = 0.25$ and $\text{CFL}_l = 0.45$ for the 2D extension of second and first order scheme respectively. 2D-extension of the 1D first-order scheme used only in the image on the bottom-right.

Thus, the initial solution satisfies the "lake at rest" condition. Indeed, our 2D numerical schemes are able to preserve this kind of stationary solutions up to an error of order 10^{-15}.

4.8.6 Conical dune of sand

This test case has been vastly used to validate numerical schemes for shallow water Exner system, here we do refer for instance to [5, 35]. When considering the Grass formulation for the sediment discharge, we take porosity $\rho_0 = 0.4$, where we recall that $\zeta = \frac{1}{1-\rho_0}$. The domain is a $L \times L$ square with $L = 1000$m. At time $t = 0$, we impose

$$z(x,y,t=0) = \begin{cases} 0.1 + (\sin(\frac{\pi(x-300)}{200}))^2(\sin(\frac{\pi(y-400)}{200}))^2 & \text{if} \quad 300 \leq x \leq 500,\ 400 \leq y \leq 600 \\ 0.1 & \text{otherwise} \end{cases},$$

$$h(x,y,t=0) = 10 - z(x,y,t=0),$$

$$u(x,y,t=0) = \frac{10}{h(x,y,0)} \quad \text{and} \quad v(x,y,t=0) = 0.$$

As for the boundary conditions, at the upstream we impose $u(x,y,t) = \dfrac{10}{h(x,y,t)}$, while we use transmissive conditions at the other boundaries.

Then, we start considering two different cases: in the first one we take $A_g = 1$ and ending time $t_{\text{End}} = 500$s, thus we are assuming a fast interaction between the flow and the sediments. As second case, we diminish the value of A_g, namely we impose $A_g = 0.1$, thus the strength of the interaction decreases. The outputs for these two test cases can be found in pictures 4.8.3 and 4.8.4 respectively. The result are in agreement with the ones reported in [5] even if more diffusive due to a coarser mesh size ($M = 100$ cells).

Finally, we consider a slow interaction case with $A_g = 0.001$ and ending time $t = 100$h. The aim of this simulation is to observe what is the spreading angle of the dune. Indeed, for slow water-sediment interaction cases ($A_g < 0.01$), the sediment bottom should expand according to a star-shaped. In particular, an approximation of the spreading angle α has been proposed in the work of De Vriend [24]. Considering the Grass formulation, the angle is given by

$$\tan(\alpha) = \frac{3\sqrt{3}(m-1)}{9m-1}$$

and thus, we should obtain $\alpha = 21.786789°$ for $m = 3$. Afterwards, in figure 4.8.5 we insert the bed elevation, while its contour plot is shown at different times in figure 4.8.6. In particular, we consider the iso-levels $z = 0.12$ (left) and $z = 0.15$ (right) at times $t = 0$s, $t = 25$h, $t = 50$h, $t = 75$h, $t = 100$h. We also insert the analytical approximation of the spreading angle α in order to show that our second-order method is capable to compute the solution with enough accuracy. Indeed, for the iso-level $z = 0.12$, our results are comparable with the ones shown for instance in [21, 35] as the majority of the spread of the dune is inside the cone. Then, since on the right we use a greater value for the iso-level, namely $z = 0.15$, it is not surprising to see that now almost all the spread is contained in the theoretical cone. See again the following works [16, 5, 35, 21] for similar considerations.

4.9 Conclusions

In this work, we introduced three well-balanced second-order Lagrange-projection schemes for the shallow water Exner model. In particular, we described three different ways of discretizing the bed level, by considering z either in both the acoustic and transport steps or only in

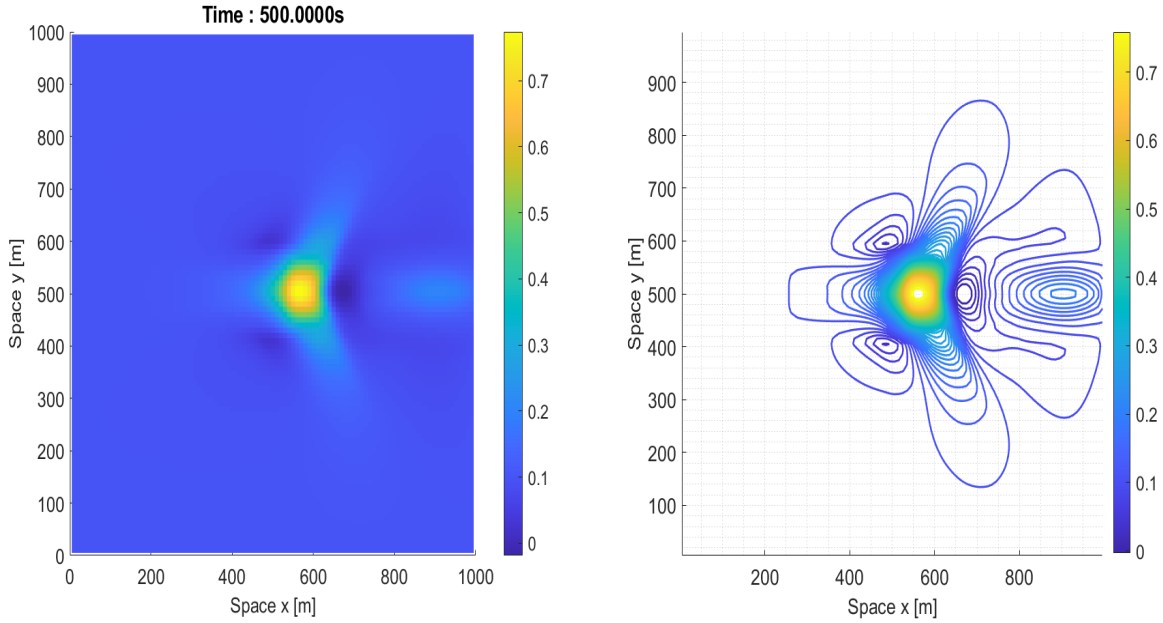


Figure 4.8.3: Conical dune of sand; fast interaction $A_g = 1$. Bed level (left) and contour plot (right) at time $t = 500$s. 2D extension of second order scheme with $M = 100$ cells and $\text{CFL}_l = 0.25$.

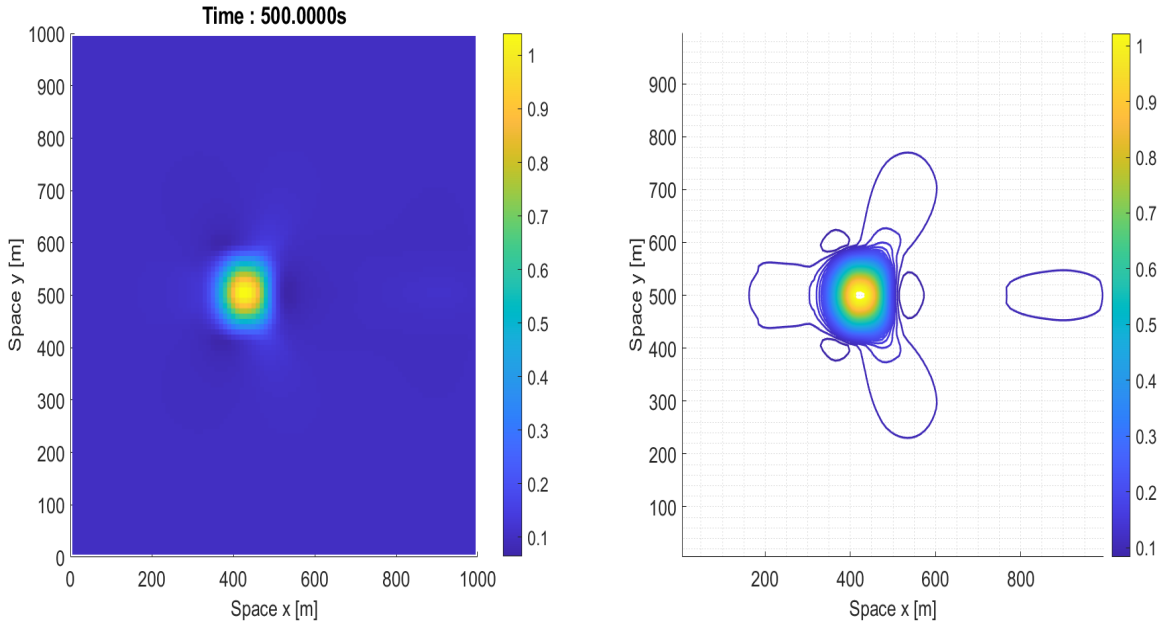


Figure 4.8.4: Conical dune of sand; medium interaction $A_g = 0.1$. Bed level (left) and contour plot (right) at time $t = 500$s. 2D extension of second order scheme with $M = 100$ cells and $\text{CFL}_l = 0.25$.

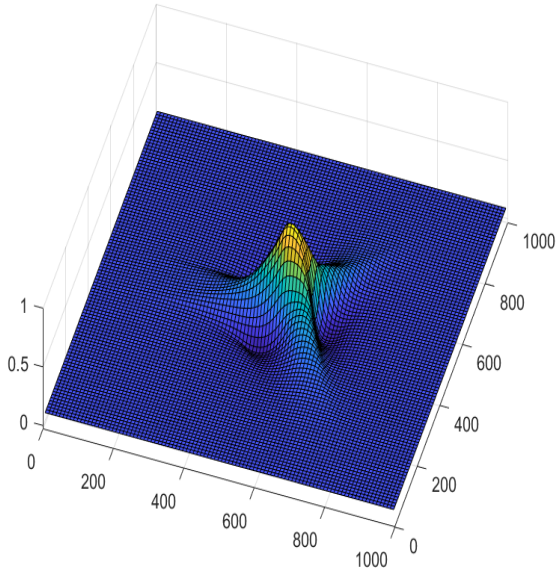


Figure 4.8.5: Conical dune of sand; slow interaction $A_g = 0.001$. Bed elevation computed with 2D extension of second order scheme with $M = 100$ cells and $\text{CFL}_l = 0.25$.

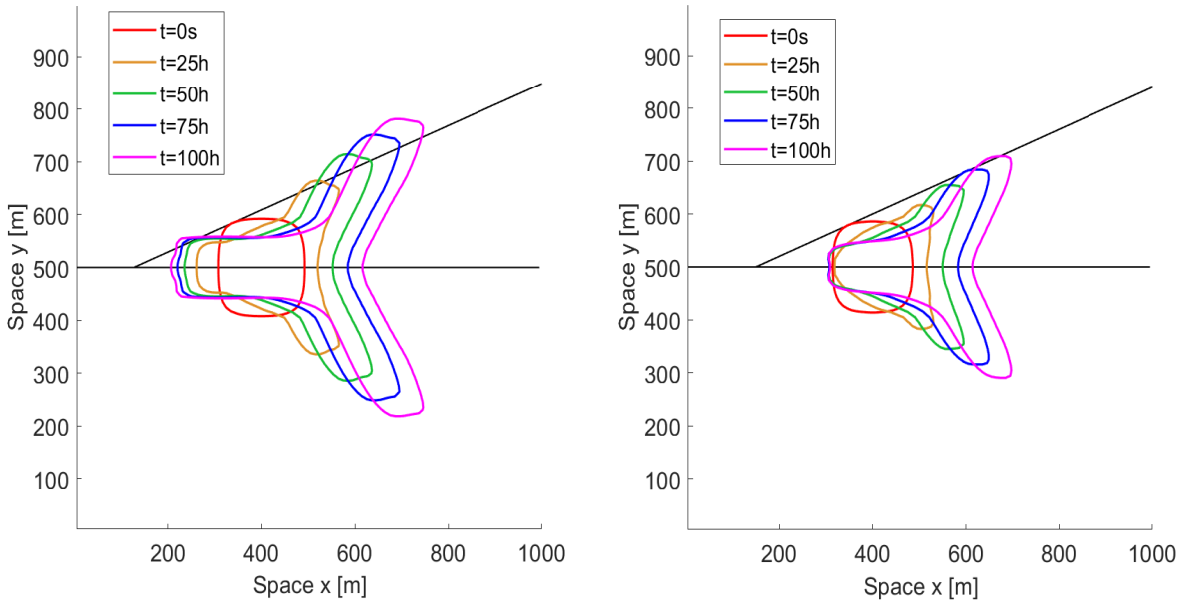


Figure 4.8.6: Conical dune of sand; slow interaction $A_g = 0.001$. Reference angle (black line) and iso-level at times $t = 0$s (red), $t = 25$h (orange), $t = 50$h (green), $t = 75$h (blue), $t = 100$h (magenta). Iso-level $z = 0.12$ (left) and $z = 0.15$ (right). 2D extension of second order scheme with $M = 100$ cells and $\text{CFL}_l = 0.25$.

one of them. All the methods have been designed in such a way to be well-balanced and at the same time second-order accurate. Here, by well-balanced, we mean that the schemes preserve the "lake at rest" and the "constant bed slope" equilibrium steady states. Moreover, both Grass and MPM formulation have been considered. We tested these methods considering several numerical experiments and we generally observed that no oscillations are present in the numerical results when using their first-order version. However, when moving to second-order of accuracy, this is not always true. In particular, the TrZ scheme (z only in the transport step) revealed itself to be the worst method among the three in this sense. On the other hand, the other two methods AcTrZ and AcZ performed a lot better, showing only minor and controlled oscillations in some of the numerical tests. Moreover, we also presented the 2D extension of the model and of one numerical strategy (Exner equation in the transport strategy). A benchmark test (conical dune of sand) has been considered to test the good behavior of the scheme.

Further improvements could be related to the extension of all these numerical schemes to two dimensions, higher order of accuracy and the use of other formulations for the solid transport discharge. Finally, it could be interesting to consider the 1D and 2D implicit version of such numerical methods, especially in situations of weak interaction between the flow and the sediments, where long time simulations are needed.

4.10 Bibliography

[1] E. Audusse, C. Berthon, C. Chalons, O. Delestre, N. Goutal, M. Jodeau, J. Sainte-Marie, J. Gieselmann, G. Sadaka. *Sediment transport modelling : Relaxation schemes for Saint-Venant – Exner and three layer models.* ESAIM: Proc. 38: 78-98, 2012. 10.1051/proc/201238005

[2] 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: 2050-2065, 2004. 10.1137/S1064827503431090.

[3] E. Audusse, M. O. Bristeau. *A 2d Well-balanced Positivity Preserving Second Order Scheme for Shallow Water Flows on Unstructured Meshes*, [Research Report] RR-5260, INRIA, 31, 2004. inria-00070738

[4] E. Audusse, C. Chalons, and P. Ung. *A simple three-wave approximate riemann solver for the Saint-Venant-Exner equations.* International Journal for Numerical Methods in Fluids, 87(10): 508-528, 2018. 10.1002/fld.4500

[5] E. Audusse, O. Delestre, M. H. Le, M. Masson-Fauchier, P. Navaro, and R. Serra. *Parallelization of a relaxation scheme modelling the bedload transport of sediments in shallow water flow.* ESAIM: Proceedings, 43: 80-94, 2013. 10.1051/proc/201343005

[6] W. Barsukow, *Truly multi-dimensional all-speed schemes for the Euler equations on Cartesian grids.* Journal of Computational Physics, 435: 110216, 2021. 10.1016/j.jcp.2021.110216.

[7] C. Berthon, S. Cordier, O. Delestre, and M. H. Le. *An analytical solution of the shallow water system coupled to the exner equation.* Comptes Rendus Mathematique, 350(3-4): 183-186, 2012. 10.1016/j.crma.2012.01.007

[8] C. Berthon, F. Foucher. *Efficient well–balanced hydrostatic upwind schemes for shallow–water equations.* Journal of Computational Physics. 231(15): 4993–5015, 2012. 10.1016/j.jcp.2012.02.031.

[9] M. Billaud Friess, B. Boutin, F. Caetano, G. Faccanoni, S. Kokh, F. Lagoutière, L. Navoret. *A second order anti-diffusive Lagrange-remap scheme for two-component flows.* ESAIM: Proceedings, EDP Sciences, 32: 149-162, 2011. 10.1051/proc/2011018.

[10] F. Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources.* Frontiers in mathematics, 2004. 10.1007/b93802.

[11] F. Bouchut, C. Chalons, S. Guisset. *An entropy satisfying two-speed relaxation system for the barotropic Euler equations. Application to the numerical approximation of low Mach number flows.* Numerische Mathematik, Springer Verlag, 145: 35-76, 2020. 10.1007/s00211-020-01111-5.

[12] R. Briganti, N. Dodd, D. Kelly, and D. Pokrajac. *An efficient and flexible solver for the simulation of the morphodynamics of fast evolving flows on coarse sediment beaches.* International Journal for Numerical Methods in Fluids, 69(4): 859-877, 2012. 10.1002/fld.2618.

[13] R. Bürger, C. Chalons and L. Villada. *Antidiffusive and Random-Sampling Lagrangian-Remap Schemes for the Multiclass Lighthill-Whitham-Richards Traffic Model.* SIAM Journal on Scientific Computing, 35, 2013. 10.1137/130923877

[14] M. J. Castro Díaz, C. Chalons and T. Morales De Luna. *A fully well-balanced Lagrange-Projection type scheme for the Shallow-water equations.* SIAM J. Numer. Anal., 56(5): 3071–3098, 2018. 10.1137/17M1156101.

[15] M. J. Castro Díaz, E. D. Fernández-Nieto and A.M. Ferreiro. *Sediment transport models in Shallow Water equations and numerical approach by high order finite volume methods.* Computers & Fluids, 37(3): 299-316, 2008. 10.1016/j.compfluid.2007.07.017

[16] M. J. Castro Díaz, E. D. Fernández-Nieto, A.M. Ferreiro and C. Parés. *Two-dimensional sediment transport models in shallow water equations. A second order finite volume approach on unstructured meshes*, Computer Methods in Applied Mechanics and Engineering, 198: 2520-2538, 2009. 10.1016/j.cma.2009.03.001

[17] M. J. Castro, T. Morales de Luna, C. Parés Madroñal. *Well-Balanced Schemes and Path-Conservative Numerical Methods.* Handbook of Numerical Analysis, Elsevier, 18: 131-175, 2017. 10.1016/bs.hna.2016.10.002.

[18] C. Chalons, M. Girardin, S. Kokh. *An All-Regime Lagrange-Projection like scheme for the gas dynamics equations on unstructured meshes.* Communications in Computational Physics. 20(1): 188-233 2016. 10.4208/cicp.260614.061115a.

[19] C. Chalons, M. Girardin, S. Kokh. *An all-regime Lagrange-Projection like scheme for 2D homogeneous models for two-phase flows on unstructured meshes.* Journal of Computational Physics, Elsevier, 335: 885-904, 2017. 10.1016/j.jcp.2017.01.017.

[20] C. Chalons, P. Kestener, S. Kokh, and M. Stauffert. *A large time-step and well-balanced Lagrange-Projection type scheme for the Shallow-water equations.* Communications in Mathematical Sciences. 15(3): 765–788, 2017. 10.4310/CMS.2017.v15.n3.a9.

[21] A. Chertock, A. Kurganov and T. Wu. *Operator splitting based central-upwind schemes for shallow water equations with moving bottom topography*, Communications in Mathematical Sciences, 18: 2149-2168, 2020. 10.4310/CMS.2020.v18.n8.a3

[22] S. Clain, S. Diot, R. Loubère. *A high-order finite volume method for hyperbolic systems: Multi-dimensional Optimal Order Detection (MOOD).* Journal of Computational Physics, Elsevier, 230, 2011. ffhal-00518478v2f

[23] S. Cordier, M.H. Le, and T. Morales de Luna. *Bedload transport in shallow water models: Why splitting (may) fail, how hyperbolicity (can) help.* Advances in Water Resources, 34(8): 980-989, 2011. 10.1016/j.advwatres.2011.05.002

[24] H.J. De Vriend, *2DH mathematical modelling of morphological evolutions in shallow water*, Coast. Engrg., 11:1–27, 1987. 10.1016/0378-3839(87)90037-8

[25] A. Del Grosso and C. Chalons. *Second-order well balanced Lagrange-Projection schemes for Blood Flow Equations.* Calcolo 58, 43, 2021. 10.1007/s10092-021-00434-5

[26] B. Després. *Numerical Methods for Eulerian and Lagrangian Conservation Laws.* 2017. 10.1007/978-3-319-50355-4.

[27] F. Duboc, C. Enaux, S. Jaouen, H. Jourdren, M. Wolff. *High-order dimensionally split Lagrange-remap schemes for compressible hydrodynamics.* Comptes Rendus Mathematique, 348(1–2):105-110, 2010. 10.1016/j.crma.2009.12.008.

[28] E. Fernández-Nieto, C. Lucas, T. Morales de Luna, S. Cordier. *On the influence of the thickness of the sediment moving layer in the definition of the bedload transport formula in Exner systems.* Computers & Fluids 91, 87–106, 2014. 10.1016/j.compfluid.2013.11.031

[29] F. Franzini, S. Soares-Frazão. *Coupled finite-volume scheme with adapted Augmented Roe scheme for simulating morphological evolution in arbitrary cross-sections.* J. Hydroinform. 20, 1111–1130, 2018. 10.2166/hydro.2018.109

[30] G. Gallice. *Solveurs simples positifs et entropiques pour les systèmes hyperboliques avec terme source.* C. R. Math. Acad. Sci. Paris 334(8): 713-716, 2002. 10.1016/S1631-073X(02)02307-5.

[31] G. Gallice. *Positive and entropy stable Godunov-type schemes for gas dynamics and MHD equations in Lagrangian or Eulerian coordinates.* Numer. Math. 94(4): 673-713, 2003. 10.1007/s00211-002-0430-0

[32] J.C. González-Aguirre, M.J. Castro, and T. Morales de Luna. *A robust model for rapidly varying flows over movable bottom with suspended and bedload transport: Modelling and numerical approach.* Advances in Water Resources, 140: 103575, 2020. 10.1016/j.advwatres.2020.103575.

[33] S. Gottlieb and C.-W. Shu. *Total variation diminishing RUNGE-KUTTA schemes.* 1996. Mathematics of Computation. 67. 10.1090/S0025-5718-98-00913-2.

[34] P. H. Gunawan. *Numerical simulation of shallow water equations and related models.* General Mathematics [math.GM]. Université Paris-Est, 2015. English. ffNNT : 2015PEST1010ff. tel01216642v2

[35] J. Hudson. *Numerical techniques for morphodynamic modelling.* PhD thesis, University of Reading, October 2001.

[36] C. Juez, J. Murillo, P. García-Navarro. *A 2D weakly-coupled and efficient numerical model for transient shallow flow and movable bed.* Advances in Water Resources, 71: 93-109, 2014. 10.1016/j.advwatres.2014.05.014

[37] R.J. LeVeque. *Numerical Methods for Conservation Laws.* 1992 - Springer. 10.1007/978-3-0348-5116-9

[38] X. Liu, A. Mohammadian, A. Kurganov, J. A. Infante Sedano. *Well-balanced central-upwind scheme for a fully coupled shallow water system modeling flows over erodible bed.* Journal of Computational Physics, 300: 202-218, 2015. 10.1016/j.jcp.2015.07.043

[39] R. Loubère. *Une Méthode Particulaire Lagrangienne de type Galerkin Discontinu : Application à la Mécanique des Fluides et l'Interaction Laser/Plasma*, These doctorale, 2002.

[40] P. H. Maire, R. Abgrall, J. Breil, J. Ovadia. *A Cell-Centered Lagrangian Scheme for Two-Dimensional Compressible Flow Problems.* SIAM Journal on Scientific Computing, 2007. 10.1137/050633019

[41] S. Martínez-Aranda, J. Murillo, P. García-Navarro. *A comparative analysis of capacity and non-capacity formulations for the simulation of unsteady flows over finite-depth erodible beds.* Advances in Water Resources 130: 91-112, 2019. 10.1016/j.advwatres.2019.06.001

[42] S. Martínez-Aranda, J. Murillo, P. García-Navarro. *A 1D numerical model for the simulation of unsteady and highly erosive flows in rivers.* Computers & Fluids 181: 8–34, 2019. 10.1016/j.compfluid.2019.01.011

[43] S. Martínez-Aranda, J. Murillo, P. García-Navarro. *Comparison of new efficient 2D models for the simulation of bedload transport using the augmented Roe approach.* Advances in Water Resources 153(3-4): 103931, 2021. 10.1016/j.advwatres.2021.103931

[44] R. Meurice, S. Soares-Frazão. *A 2D HLL-based weakly coupled model for transient flows on mobile beds.* J. Hydroinform. 22: 1351–1369, 2020. 10.2166/hydro.2020.033

[45] T. Morales De Luna, M. J. Castro Díaz and C. Chalons. *High order fully well-balanced Lagrange-Projection scheme for Shallow-water.* Commun. Math. Sci., 18(3): 781–807, 2020. 10.4310/CMS.2020.v18.n3.a9

[46] T. Morales de Luna, M. J. Castro Díaz, C. Parés Madroñal, and E. D. Fernández Nieto. *On a shallow water model for the simulation of turbidity currents.* Communications in Computational Physics, 6(4): 848-882, 2009. 10.4208/cicp.2009.v6.p848

[47] J. Murillo and P. García-Navarro. *An exner-based coupled model for two-dimensional transient flow over erodible bed.* Journal of Computational Physics, 229(23): 8704-8732, 2010. 10.1016/j.jcp.2010.08.006.

[48] M. Seaïd. *Non-oscillatory relaxation methods for the shallow-water equations in one and two space dimensions*, International Journal for Numerical Methods in Fluids, 46(5): 457-484, 2004. 10.1002/fld.766

[49] E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics*, Third Edition. Springer–Verlag, 2009. 10.1007/b79761_5.

[50] E. F. Toro and A. Siviglia. *PRICE: Primitive Centred Schemes for Hyperbolic Systems.* Int. J. Numer. Meth. in Fluids, 42: 1263–1291, 2003. 10.1002/fld.491.

[51] C. B. Vreugdenhil. *Numerical Methods for Shallow Water Flow.* Dordrecht; Boston: Kluwer Academic Publishers, 1994. Part of the Water Science and Technology Library book series (WSTL, volume 13).

Annexe

4.A Sur un schéma de type Lagrange-projection entièrement couplé pour le système de Saint-Venant-Exner

Dans cette annexe, on presente une première tentative de Solveur de Riemann Approché (SRA) que nous avons conçu pour la partie acoustique du système de Saint-Venant-Exner. On estime intéressant de montrer ce SRA car il nous paraît tout à fait naturel. En effet, contrairement aux solveurs précédents, il conduirait à un schéma numérique entièrement couplé avec l'élévation du lit z mise à jour dans les deux étapes : acoustique et transport. Malheureusement ce SRA s'est avéré ne pas être automatiquement équilibre, c'est-à-dire qu'il ne préserve pas a priori la solution stationnaire "lac au repos" (4.1.5). De plus, des problèmes liés à son caractère bien posé pourraient survenir si la vitesse u est nulle partout.

On considère à nouveau les équations acoustiques (4.3.3),

$$\begin{cases} \partial_t \tau - \partial_m u = 0 \\ \partial_t u + \partial_m p + \dfrac{g}{\tau}\partial_m z = 0 \\ \partial_t z - \dfrac{u}{\tau}\partial_m z + \zeta \dfrac{1}{\tau}\partial_m q_b = 0 \end{cases} \tag{4.A.1}$$

où nous rappelons que $\tau = 1/h$ et $\tau \partial_x = \partial_m$. Une fois encore, on introduit deux nouvelles variables de relaxation Π et Ω, à savoir la linéarisation de p et ζq_b respectivement. Par contre, le système de relaxation résultant est maintenant donné par

$$\begin{cases} \partial_t \tau - \partial_m u = 0 \\ \partial_t u + \partial_m \Pi + \dfrac{g}{\tau}\partial_m z = 0 \\ \partial_t z - \dfrac{u}{\tau}\partial_m z + \dfrac{1}{\tau}\partial_m \Omega = 0 \\ \partial_t \Pi + a^2 \partial_m u = 0 \\ \partial_t \Omega + \dfrac{u}{\tau}\partial_m \Omega + (b^2\tau - \dfrac{u^2}{\tau})\partial_m z = 0 \end{cases} \tag{4.A.2}$$

avec a et b paramètres constants. Les différences par rapport au système de relaxation (4.3.6) sont les suivantes. Tout d'abord, le terme de couplage $\dfrac{g}{\tau}\partial_m z$ est considéré dans le membre de gauche de la deuxième équation et non comme un terme source. Deuxièmement, dans l'équation d'évolution sur Ω, le terme lié à la variation spatiale de z est légèrement différent, c'est-à-dire que le terme $(b^2\tau - \frac{u^2}{\tau})\partial_m z$ a ensuite été remplacé par $u^2(b^2\tau - \frac{1}{\tau})\partial_m z$. Ce choix a été motivé par le fait que nous voulions obtenir les valeurs propres $\pm|u|b$ plutôt que $\pm b$, afin

d'avoir une méthode équilibre. Il est donc clair que la condition de stabilité pour b change et s'écrit maintenant $b \geq \sqrt{(hu)^2 + gh^2\zeta\partial_u q_b}$ tandis que la condition pour a est toujours donnée par (4.3.7).

On introduit le système (4.A.2) sous forme compacte comme suit

$$\partial_t \mathbf{U} + \mathbf{B}(\mathbf{U})\partial_m \mathbf{U} = \mathbf{0}$$

avec

$$\mathbf{U} = \begin{pmatrix} \tau \\ u \\ z \\ \Pi \\ \Omega \end{pmatrix} \quad \text{et} \quad \mathbf{B}(\mathbf{U}) = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & \frac{g}{\tau} & 1 & 0 \\ 0 & 0 & -\frac{u}{\tau} & 0 & \frac{1}{\tau} \\ 0 & a^2 & 0 & 0 & 0 \\ 0 & 0 & b^2\tau - \frac{u^2}{\tau} & 0 & \frac{u}{\tau} \end{pmatrix}.$$

La matrice non-conservative du système a des valeurs propres constantes $0, \pm a, \pm b$ et par conséquent les champs caractéristiques associés sont linéairement dégénérés. Ensuite, les vecteurs propres de droite correspondants sont

$$\mathbf{R}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{R}_a^{\pm} = \begin{pmatrix} 1 \\ \mp a \\ 0 \\ -a^2 \\ 0 \end{pmatrix} \quad \text{et} \quad \mathbf{R}_b^{\pm} = \begin{pmatrix} 1 \\ \mp b \\ -\frac{\tau}{g}(b^2 - a^2) \\ -a^2 \\ -\frac{\tau}{g}(b^2 - a^2)(u \pm \tau b) \end{pmatrix}. \tag{4.A.3}$$

Ainsi, nous pouvons enfin trouver les invariants de Riemann. En commençant par la valeur propre λ_0, on obtient

$$RI_{0,1} = u, \quad RI_{0,2} = z, \quad RI_{0,3} = \Pi \quad \text{et} \quad RI_{0,4} = \Omega. \tag{4.A.4}$$

Après, en considérant $\lambda_a^{\pm}$, on a

$$RI_{\pm a,1} = u \pm a\tau, \quad RI_{\pm a,2} = au \mp \Pi, \quad RI_{\pm a,3} = z \quad \text{et} \quad RI_{\pm a,4} = \Omega. \tag{4.A.5}$$

tandis que pour $\lambda_b^{\pm}$ on trouve les invariants de Riemann suivants

$$RI_{\pm b,1} = u \pm b\tau, \quad RI_{\pm b,2} = \frac{a^2}{b}u \mp \Pi, \quad RI_{\pm b,3} = \tau^2 - \frac{2g}{a^2 - b^2}z \quad \text{et} \quad RI_{\pm b,4} = (u \pm b\tau)z - \Omega. \tag{4.A.6}$$

Nous avons maintenant tous les ingrédients pour résoudre le problème de Riemann associé au système (4.A.2).

4.A.1 Solveur de Riemann approché

On souhaite donc résoudre le problème de Riemann suivant,

$$\begin{cases} \partial_t \mathbf{U} + \mathbf{B}(\mathbf{U})\partial_m \mathbf{U} = 0 \\ \mathbf{U}(m, t = 0) = \begin{cases} \mathbf{U}_L & \text{if} \quad m < 0 \\ \mathbf{U}_R & \text{if} \quad m \geq 0 \end{cases} \end{cases} \tag{4.A.7}$$

où en particulier

$$\mathbf{U}_L = \begin{pmatrix} \tau_L \\ u_L \\ z_L \\ \Pi_L \\ \Omega_L \end{pmatrix} \quad \text{et} \quad \mathbf{U}_R = \begin{pmatrix} \tau_R \\ u_R \\ z_R \\ \Pi_R \\ \Omega_R \end{pmatrix}.$$

Comme il y a cinq ondes, la solution du problème de Riemann est composée par six états différents séparés par les cinq discontinuités. Cependant, l'ordre des ondes n'est pas fixé car il dépend de la valeur relative des paramètres a et b. Ainsi, deux situations différentes peuvent apparaître, à savoir soit $a < b$, soit $b < a$. Par conséquent, la solution sera donnée respectivement soit par

$$\hat{\mathbf{U}}(\frac{m}{t};\mathbf{U}_L,\mathbf{U}_R) = \begin{cases} \mathbf{U}_L & \text{if} \quad \frac{m}{t} < \lambda_b^- = -b \\ \mathbf{U}^*_{b,L} & \text{if} \quad \lambda_b^- = -b < \frac{m}{t} < \lambda_a^- = -a \\ \mathbf{U}^*_{a,L} & \text{if} \quad \lambda_a^- < \frac{m}{t} < \lambda_0 = 0 \\ \mathbf{U}^*_{a,R} & \text{if} \quad 0 < \frac{m}{t} < \lambda_a^+ = a \\ \mathbf{U}^*_{b,R} & \text{if} \quad \lambda_a^+ < \frac{m}{t} < \lambda_b^+ = b \\ \mathbf{U}_R & \text{if} \quad \frac{m}{t} > \lambda_b^+ \end{cases} \tag{4.A.8}$$

soit par

$$\hat{\mathbf{U}}(\frac{m}{t};\mathbf{U}_L,\mathbf{U}_R) = \begin{cases} \mathbf{U}_L & \text{if} \quad \frac{m}{t} < \lambda_a^- = -a \\ \mathbf{U}^*_{a,L} & \text{if} \quad \lambda_a^- = -a < \frac{m}{t} < \lambda_b^- = -b \\ \mathbf{U}^*_{b,L} & \text{if} \quad \lambda_b^- < \frac{m}{t} < \lambda_0 = 0 \\ \mathbf{U}^*_{b,R} & \text{if} \quad 0 < \frac{m}{t} < \lambda_b^+ = b \\ \mathbf{U}^*_{a,R} & \text{if} \quad \lambda_b^+ < \frac{m}{t} < \lambda_a^+ = a \\ \mathbf{U}_R & \text{if} \quad \frac{m}{t} > \lambda_a^+ \end{cases} \tag{4.A.9}$$

avec

$$\mathbf{U}^*_{b,L} = \begin{pmatrix} \tau^*_{b,L} \\ u^*_{b,L} \\ z^*_{b,L} \\ \Pi^*_{b,L} \\ \Omega^*_{b,L} \end{pmatrix}, \quad \mathbf{U}^*_{b,R} = \begin{pmatrix} \tau^*_{b,R} \\ u^*_{b,R} \\ z^*_{b,R} \\ \Pi^*_{b,R} \\ \Omega^*_{b,R} \end{pmatrix}, \quad \mathbf{U}^*_{a,L} = \begin{pmatrix} \tau^*_{a,L} \\ u^*_{a,L} \\ z^*_{a,L} \\ \Pi^*_{a,L} \\ \Omega^*_{a,L} \end{pmatrix} \quad \text{et} \quad \mathbf{U}^*_{a,R} = \begin{pmatrix} \tau^*_{a,R} \\ u^*_{a,R} \\ z^*_{a,R} \\ \Pi^*_{a,R} \\ \Omega^*_{a,R} \end{pmatrix}. \tag{4.A.10}$$

Il est intéressant de souligner que les deux cas $a < b$ et $b < a$ correspondent à des flux supercritiques et sous-critiques respectivement.

Ensuite, en utilisant les invariants de Riemann, il est clair que les états étoilés (4.A.10) se réduisent à

$$\mathbf{U}^*_{b,L} = \begin{pmatrix} \tau^*_{b,L} \\ u^*_{b,L} \\ z^* \\ \Pi^*_{b,L} \\ \Omega^* \end{pmatrix}, \quad \mathbf{U}^*_{a,L} = \begin{pmatrix} \tau^*_{a,L} \\ u^*_{a} \\ z^* \\ \Pi^*_{a} \\ \Omega^* \end{pmatrix}, \quad \mathbf{U}^*_{a,R} = \begin{pmatrix} \tau^*_{a,R} \\ u^*_{a} \\ z^* \\ \Pi^*_{a} \\ \Omega^* \end{pmatrix} \quad \text{et} \quad \mathbf{U}^*_{b,R} = \begin{pmatrix} \tau^*_{b,R} \\ u^*_{b,R} \\ z^* \\ \Pi^*_{b,R} \\ \Omega^* \end{pmatrix} \tag{4.A.11}$$

si $a < b$, et à

$$\mathbf{U}^*_{a,L} = \begin{pmatrix} \tau^*_{a,L} \\ u^*_{a,L} \\ z_L \\ \Pi^*_{a,L} \\ \Omega_L \end{pmatrix}, \quad \mathbf{U}^*_{b,L} = \begin{pmatrix} \tau^*_{b,L} \\ u^*_{b} \\ z^* \\ \Pi^*_{b} \\ \Omega^* \end{pmatrix}, \quad \mathbf{U}^*_{b,R} = \begin{pmatrix} \tau^*_{b,R} \\ u^*_{b} \\ z^* \\ \Pi^*_{b} \\ \Omega^* \end{pmatrix} \quad \text{et} \quad \mathbf{U}^*_{a,R} = \begin{pmatrix} \tau^*_{a,R} \\ u^*_{a,R} \\ z_R \\ \Pi^*_{a,R} \\ \Omega_R \end{pmatrix}, \tag{4.A.12}$$

si $b < a$. Sans montrer tous les calculs, dans le cas $a < b$, on obtient les états étoilés suivants

$$\begin{cases} z^* = \dfrac{(u_R + b\tau_R)z_R - (u_L - b\tau_L)z_L}{(u_R + b\tau_R) - (u_L - b\tau_L)} - \dfrac{\Omega_R - \Omega_L}{(u_R + b\tau_R) - (u_L - b\tau_L)} \\ \Omega^* = \dfrac{(u_R + b\tau_R)\Omega_L - (u_L - b\tau_L)\Omega_R}{(u_R + b\tau_R) - (u_L - b\tau_L)} + \dfrac{(u_R + b\tau_R)(u_L - b\tau_L)}{(u_R + b\tau_R) - (u_L - b\tau_L)}(z_R - z_L), \end{cases}$$

$$\begin{cases} \tau^*_{b,L} = \left(\tau_L^2 - \dfrac{2g}{b^2-a^2}(z^* - z_L)\right)^{\frac{1}{2}} \\ \tau^*_{b,R} = \left(\tau_R^2 + \dfrac{2g}{b^2-a^2}(z_R - z^*)\right)^{\frac{1}{2}} \\ u^*_{b,L} = u_L + b(\tau^*_{b,L} - \tau_L) \\ u^*_{b,R} = u_R + b(\tau_R - \tau^*_{b,R}) \\ \Pi^*_{b,L} = \Pi_L - a^2(\tau^*_{b,L} - \tau_L) \\ \Pi^*_{b,R} = \Pi_R + a^2(\tau_R - \tau^*_{b,R}) \end{cases}$$

et

$$\begin{cases} u^*_a = \dfrac{1}{2}(u^*_{b,L} + u^*_{b,R}) - \dfrac{1}{2a}(\Pi^*_{b,R} - \Pi^*_{b,L}) \\ \Pi^*_a = \dfrac{1}{2}(\Pi^*_{b,L} + \Pi^*_{b,R}) - \dfrac{a}{2}(u^*_{b,R} - u^*_{b,L}) \\ \tau^*_{a,L} = \tau^*_{b,L} + \dfrac{1}{a^2}(\Pi^*_{b,L} - \Pi^*_a) \\ \tau^*_{a,R} = \tau^*_{b,R} + \dfrac{1}{a^2}(\Pi^*_{b,R} - \Pi^*_a). \end{cases}$$

Passons au cas $b < a$. Maintenant, la solution du problème de Riemann n'est pas aussi facilement obtenue que précédemment en raison de la présence de conditions non linéaires. En effet, nous avons

$$\begin{cases} z^* = \dfrac{(u^*_{a,R} + b\tau^*_{a,R})z_R - (u^*_{a,L} - b\tau^*_{a,L})z_L}{(u^*_{a,R} + b\tau^*_{a,R}) - (u^*_{a,L} - b\tau^*_{a,L})} - \dfrac{\Omega_R - \Omega_L}{(u^*_{a,R} + b\tau^*_{a,R}) - (u^*_{a,L} - b\tau^*_{a,L})} \\ \Omega^* = \dfrac{(u^*_{a,R} + b\tau^*_{a,R})\Omega_L - (u^*_{a,L} - b\tau_{a,L})\Omega_R}{(u^*_{a,R} + b\tau^*_{a,R}) - (u^*_{a,L} - b\tau^*_{a,L})} + \dfrac{(u^*_{a,R} + b\tau^*_{a,R})(u^*_{a,L} - b\tau^*_{a,L})}{(u^*_{a,R} + b\tau^*_{a,R}) - (u^*_{a,L} - b\tau^*_{a,L})}(z_R - z_L), \\ u^*_b = \dfrac{1}{2}(u^*_{a,L} + u^*_{a,R}) - \dfrac{b}{2a^2}(\Pi^*_{a,R} - \Pi^*_{a,L}) \\ \Pi^*_b = \dfrac{1}{2}(\Pi^*_{a,L} + \Pi^*_{a,R}) - \dfrac{a^2}{2b}(u^*_{a,R} - u^*_{a,L}) \\ u^*_{a,L} = u_L + a(\tau^*_{a,L} - \tau_L) \\ u^*_{a,R} = u_R + a(\tau_R - \tau^*_{a,R}) \\ \Pi^*_{a,L} = \Pi_L - a^2(\tau^*_{a,L} - \tau_L) \\ \Pi^*_{a,R} = \Pi_R + a^2(\tau_R - \tau^*_{a,R}) \\ \tau^*_{b,L} = \tau_L + \dfrac{1}{a^2}(\Pi_L - \Pi^*_b) \\ \tau^*_{b,R} = \tau_R + \dfrac{1}{a^2}(\Pi_R - \Pi^*_b) \\ \dfrac{a^2-b^2}{2g}((\tau^*_{a,L})^2 - (\tau^*_{b,L})^2) = z_L - z^* \\ \dfrac{a^2-b^2}{2g}((\tau^*_{a,R})^2 - (\tau^*_{b,R})^2) = z_R - z^* \end{cases} \tag{4.A.13}$$

Ainsi, il est clair que ce dernier système n'est pas linéaire et nous devons le résoudre pour obtenir les états étoilés $\tau^*_{a,R}, \tau^*_{a,L}, \tau^*_{b,R}$ and $\tau^*_{b,L}$.

Cependant, même sans le résoudre, nous sommes en mesure de prouver que le solveur de Riemann approché résultant n'est pas équilibre, ce qui signifie qu'il ne préserve pas la solution stationnaire "lac au repos" (4.1.5). Regardons les détails. Supposons que nous soyons dans la condition stationnaire "lac au repos" (4.1.5), c'est-à-dire que nous avons $u_L = u_R = 0$ et $h_L + z_L = h_R + z_R$, alors nous disons que le solveur de Riemann est équilibre si nous avons également $\mathbf{U}^*_{a,L} = \mathbf{U}^*_{b,L} = \mathbf{U}_L$ and $\mathbf{U}^*_{a,R} = \mathbf{U}^*_{b,R} = \mathbf{U}_R$. Tout d'abord, puisque la vitesse est nulle, nous avons clairement $b < a$, c'est-à-dire que nous sommes dans le deuxième cas avec les états étoilés données par (4.A.13). Nous remarquons également que nous ne pouvons pas

prendre exactement $b = 0$ sinon le solveur de Riemann pourrait ne pas être bien posé, voir les états étoilés de z^* et Ω^* (4.A.13). Ensuite, comme en général $z_L \neq z_R$, il est clair que les conditions $z^* = z_L$ et $z^* = z_R$ ne peuvent pas être satisfaites en même temps. Cependant, même si nous ne demandons qu'à la variable u de satisfaire une telle condition, nous obtenons une contradiction. En effet, imposons que $u^*_{a,L} = u^*_{a,R} = u^*_b = u_L = u_R = 0$. Par conséquent, nous obtenons $\tau^*_{a,L} = \tau_L$ et $\tau^*_{a,R} = \tau_R$, et après $\Pi^*_{a,L} = \Pi_L$ et $\Pi^*_{a,R} = \Pi_R$. Ainsi, de la définition de u^*_b il découle aussi $\Pi^*_{a,R} = \Pi^*_{a,L}$ ce qui va clairement en contradiction avec le fait que nous avons généralement $\Pi_L \neq \Pi_R$.

Chapter 5

Méthodes de Lagrange-projection pour les équations de Saint-Venant avec fond mobile et processus de dépôt et d'érosion des sédiments

Ce chapitre a été soumis pour publication dans le Journal "Communications in Mathematics and Applications" sous la référence: A. Del Grosso, M. J. Castro Dìaz, C. Chalons and T. Morales de Luna. *Lagrange-projection methods for shallow water equations with movable bottom and erosion-deposition processes.*

Dans le chapitre précédent, nous avons considéré le système de Saint-Venant-Exner afin de simuler l'écoulement des fluides ainsi que le transport de sédiments au long du lit. Dans ce chapitre, nous cherchons à simuler numériquement le transport de sédiments en général, ce qui signifie que les processus d'érosion et de dépôt des sédiments sont également pris en compte. Ainsi, la particule de sédiment peut aller en suspension dans l'eau ou être déposée sur le fond. A cette fin, la densité du mélange eau-sédiment est considérée.

De plus, le transport de sédiments étant généralement un processus lent, nous cherchons à développer des schémas implicites-explicites afin d'obtenir des simulations rapides. Dans ce but, nous savons que la stratégie de type Lagrange-projection est bien adaptée car elle implique une décomposition des ondes acoustiques (rapides) et des ondes matérielles (lentes) du modèle. Ainsi, dans les régimes subsoniques, une approximation implicite des équations acoustiques nous permet de négliger la condition CFL correspondante et d'obtenir des schémas numériques rapides avec un grand pas de temps.

D'autre part, en ce qui concerne les régimes supercritiques, nous savons que l'utilisation d'une approche découplée pour le système de Saint-Venant-Exner implique généralement la présence d'oscillations non physiques dans les résultats numériques. Ceci est principalement lié à la différence entre les structures propres du système de Saint-Venant avec et sans équation d'Exner. Par conséquent, dans de tels cas, nous proposons une approximation différente pour la topographie, que nous savons être plus stable dans ce type de situations (voir le chapitre précédent 4).

À la fin de ce chapitre, nous insérons également deux annexes. Dans la première 5.A, nous donnons quelques détails sur la discretisation du système acoustique implicite, en rappelant que l'approximation numérique de l'élévation du lit z reste toujours explicite. Dans la deuxième annexe 5.B, nous proposons d'approcher implicitement le système acoustique qui

actualise également l'élévation du lit, en résolvant deux systèmes linéaires différents.

Lagrange-projection methods for shallow water equations with movable bottom and erosion-deposition processes

Abstract

This work concerns the numerical simulation of shallow water flows with sediment transport. On one hand, erosion and deposition processes are taken into account, meaning that the sediment particle could go in suspension into the water (fine fraction) or being deposited on the bottom (coarse fraction). To this end, the density of the water-sediment mixture is considered. On the other hand, we take into account bedload sediment transport by including the Exner equation.

Since sediment transport is generally a slow process, we aim to develop semi-implicit schemes in order to obtain fast simulations. For this purpose, well-balanced Lagrange-projection methods are applied to the resulting mathematical model. Indeed, the Lagrange-projection splitting entails a decomposition of the (fast) acoustic waves and the (slow) material waves of the model. Hence, in subsonic regimes, an implicit approximation of the acoustic equations allows us to neglect the corresponding CFL condition and to obtain fast numerical schemes with large time step.

On the other hand, when it comes to supercritical regimes, we know that using a decoupled approach for the shallow water Exner system usually implies the presence of unphysical oscillations in the numerical results. This is mainly related to the difference in the eigenstructures of the shallow water systems with and without Exner equation. Hence, in such cases, we propose a different approximation for the topography, which we know to be more stable in this kind of situations (see the previous chapter 4).

5.1 Introduction and mathematical model

Sediment transport is an interesting and active topic in the field of geophysical flows. Sediments are transported by the action of a river current or due to currents near coastal areas mainly in two ways: a suspended load (fine fractions carried by the flow) and bedload (coarse fractions which move close to the bottom rolling, jumping and sliding), see [48].

Knowledge of sediment transport has different practical applications. For instance, in civil engineering, to plan the extended life of a dam forming a reservoir. Moreover, sediment deposition downstream reduces river capacity in that area, which may be a potential problem

in flood situations. Sediments also play an important role in some environmental problems as well. For example, suspended sediments have a direct impact on fish habitat in river or estuaries [38].

A first common approach to model sediment transport by a fluid is to couple the shallow water equations with the so-called Exner equation [30]. Many works have been proposed to study such a problem, which depends on an empirical definition of the solid transport flux for bedload transport (see [43, 29, 39, 50, 10, 12, 46] among many others). This first approach is then completed by including some transport equations for suspended sediment, that is, sediment particles which have been eroded from the bottom and remain floating in the current for some time until subsequent sedimentation (see [26, 42, 41, 47, 40, 45, 38] among many others).

One of the key points of such problems is that the characteristic time associated with sediment transport dynamics is much larger than the one corresponding to fluids. Hence, studying sediment transport usually requires long time simulations to see sediment's evolution. As such, numerical simulations will run for long wall-clock times, which are carried out in small time steps by the numerical scheme. These small time steps are mainly dominated by the characteristic fluid speed, which is much faster than that of the sediment.

To overcome this difficulty, different strategies have been proposed. The most common approach is to use semi-implicit schemes (see [13, 15, 14, 16] among others). In particular, this approach is exploited in [7, 34], where bedload transport with the simple Grass formula is considered as well as variable density. Moreover, in [35], the authors propose a semi-implicit scheme based on the theta method for sediment bedload transport models with gravitational effects under subcritical regimes. Another approach is the use of the Lagrange-Projection strategy (see [28, 20, 21, 22, 11, 44] and references therein). This framework allows us to naturally decouple the acoustic terms of the model from the transport ones. Such a decomposition is useful and very efficient to deal with subsonic or near low-Froude number flows. In such cases, the usual CFL time step limitation of Godunov-type schemes is driven by the acoustic waves and can thus be very restrictive. The Lagrange-projection strategy allows us to design a very natural implicit-explicit and large time step scheme, with a CFL restriction based on the (slow) transport waves and not on the (fast) acoustic waves. Therefore, in this work, we consider the Lagrange-projection technique adapted to the problem of sediment transport. In particular, we aim to define a semi-implicit scheme for sediment transport problems.

Hence, let us briefly present the corresponding mathematical model. It is deduced from the Navier-Stokes equations under the hypothesis that the horizontal scale is much greater than the vertical one, assuming hydrostatic pressure and incompressibility of the fluid. For more details about its derivation, we directly refer to [38]. As such, the system is composed of four different equations, which express the evolution in time of the variables $h(x,t) \geq 0$, $h\rho(x,t) \geq 0$, $h\rho u(x,t)$ and $z(x,t)$. Here, h and u stand for water's height and averaged velocity respectively, where the bottom elevation is represented by z. Then, ρ is the density of the mixture water-sediments, where the latter are transported by the currents and can either move along the bottom (bed-load) or being finer fractions carried by the flow (suspended-load). Finally, using $t > 0$ and x to represent the time and the axial coordinate respectively, the resulting model reads

$$\begin{cases} \partial_t h + \partial_x(hu) = \zeta\phi_z \\ \partial_t(h\rho) + \partial_x(h\rho u) = \zeta\phi_z\rho_z \\ \partial_t(h\rho u) + \partial_x(h\rho u^2 + p) + gh\rho\partial_x z = \rho_z \dfrac{u}{2}\zeta\phi_z - \tau_f(u) \\ \partial_t z + \zeta\partial_x q_z = -\zeta\phi_z. \end{cases} \tag{5.1.1}$$

In particular, in the third equation, the pressure term is given by $p = g\frac{h^2}{2}\rho$. Then, the evolution in time of the topography is described by the Exner equation, where $q_z = q_z(h, h\rho, h\rho u)$ represents the solid transport discharge. For the latter, there exist many different empirical laws for the solid transport discharge. Classically, it only depends on the hydrodynamical variables, $q_z = q_z(h, u)$, and on different parameters that are calibrated depending on the type of the considered sediments. Among many others, we refer for instance to the works [46, 3, 1, 9, 12, 6] for more information. For the sake of simplicity, here we only consider two of the most used formulations: the first one is the simplest one, namely the Grass model [39]. It expresses q_z as a power law of the velocity,

$$q_z = A_g u|u|^{m_g-1}, \quad m_g = 3 \tag{5.1.2}$$

with $A_g \in [0, 1]$ a constant which represents the strength of the interaction between the sediment and the flow. Then, another (more realistic) possibility is given by the Meyer-Peter&Müller (MPM) formula [43], which reads

$$q_z = 8Qsgn(u)(\theta^* - \theta_c^*)_+^{\frac{3}{2}} \quad \text{with} \quad \theta^* = \frac{u_*^2}{sgd} \quad \text{and} \quad u_*^2 = \frac{g\mu_f^2 u^2}{h^{\frac{1}{3}}}. \tag{5.1.3}$$

Moreover, $Q = d\sqrt{gsd}$ is the characteristic discharge with $s = \frac{\rho_s}{\rho_w} - 1$ the relative density, ρ_s is the density of the sediment and ρ_w is the density of water. Finally, d represents the sediment diameter, μ_f is the dimensionless Manning's coefficient and θ_c^* is the critical Shield's stress for incipient motion. Let us observe that, depending on the particular form of q_z, the resulting system could be strictly hyperbolic or not. Indeed, we already know that the shallow water Exner system is strictly hyperbolic, with all real eigenvalues, in the case of the Grass model (5.1.2). Then, regarding the MPM formula (5.1.3), it has been proved that a sufficient condition for the resulting model to be strictly hyperbolic is $|u| < 6gh$, which is generally true in physical situations, see [25, 38]. In any case, we specify that our numerical strategy can be applied whatever the formulation for q_z is.

Next, let us focus on the source terms. Here, $\phi_z = F_e - F_d$ expresses the sediment exchange between the bottom and the water-sediment mixture, where F_e and F_d are the erosion and deposition rate respectively. In particular, we state $F_e = v_s \mathcal{P} E_s$ where

$$v_s = \sqrt{(\frac{13.95\nu}{d})^2 + 1.09sgd} - \frac{13.95}{d}\nu$$

is the settling velocity of sediment where ν is the kinematic viscosity of the water and d is the sediment diameter. Then, the constant $\mathcal{P}$ stands for the volume fraction of the sediment in the bottom or, equivalently, $1 - \mathcal{P} = \Psi$ is the porosity of the bottom (see [38, 37]), with $\zeta = (1 - \Psi)^{-1}$. Subsequently, the sediment entrainment coefficient is given by

$$E_s = \frac{1.3 \times 10^{-7}\mathcal{Z}^5}{1 + 4.3 \times 10^{-7}\mathcal{Z}^5} \quad \text{with} \quad \mathcal{Z} = \frac{\alpha_1\sqrt{c_D}|u|}{v_s}\mathcal{R}_p^{\alpha_2}, \tag{5.1.4}$$

where $\mathcal{R}_p = \frac{\sqrt{sgd}d}{\nu}$ is the Reynold number and c_D is the bed drag coefficient. Finally, α_1, α_2 are two parameters depending on $\mathcal{R}_p$ and, for which, there exist different value choices. Here we refer to [38] and take

$$(\alpha_1, \alpha_2) = \begin{cases} (1, 0.6) & \text{if} \quad \mathcal{R}_p > 2.36 \\ (0.586, 1.23) & \text{if} \quad \mathcal{R}_p \leq 2.36. \end{cases}$$

Continuing with the deposition rate, we assume $F_d = v_s c_z$, where c_z is the fractional concentration of suspension near by the bed, namely

$$c_z = c\Big(0.4\Big(\frac{d}{D_{sg}}\Big)^{1.64} + 1.64\Big)$$

where $c(x,t)$ is the volumetric sediment concentration such that $\rho(x,t) = \rho_w + c(x,t)(\rho_s - \rho_w)$ and D_{sg} is the geometric mean size of the suspended sediment mixture. In particular, in this work we take $D_{sg} = d$ as all the particles are assumed to be of equal size. Lastly, $\rho_z = \rho_w \Psi + \rho_s(1-\Psi)$ is the density of the saturated bottom.

Lastly and for the friction term, we state $\tau_f(u) = \rho u_*^2(1 + r_w)$ where r_w is the ratio of upper-interface resistance to bed resistance. For the sake of clarity, in tables 5.1.1-5.1.2, we include the description of the parameters and symbols used in this work.

To complete the presentation of the mathematical model, let us observe that system (5.1.1) can be reformulated in compact form as follows:

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \mathbf{B}(\mathbf{Q})\partial_x \mathbf{Q} = \mathbf{S}(\mathbf{Q})$$

where $\mathbf{Q}$ is the vector of unknowns, $\mathbf{F}(\mathbf{Q})$ is the physical flux, $\mathbf{B}(\mathbf{Q})$ is the non-conservative product matrix and $\mathbf{S}(\mathbf{Q})$ is the source term. More explicitly,

$$\mathbf{Q} = \begin{pmatrix} h \\ h\rho \\ h\rho u \\ z \end{pmatrix}, \quad \mathbf{F}(\mathbf{Q}) = \begin{pmatrix} hu \\ h\rho u \\ h\rho u^2 + g\frac{h^2}{2}\rho \\ \zeta q_z \end{pmatrix}, \quad \mathbf{B}(\mathbf{Q}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & gh\rho \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{5.1.5}$$

and

$$\mathbf{S}(\mathbf{Q}) = \begin{pmatrix} \zeta\phi_z \\ \zeta\phi_z\rho_z \\ \rho\frac{u}{2}\phi_z - \tau_f(u) \\ -\zeta\phi_z \end{pmatrix}.$$

Then, let us recall that classic Saint Venant Exner system does not satisfy a global entropy equation. Nevertheless, as shown in [31], a modified version of Saint Venant Exner system could be introduced so that the model satisfies a global entropy. Therefore, we do not expect to find an entropy inequality for system (5.1.1), unless similar modifications are performed for Exner's equation, which is out of the scope of this paper. Nevertheless, we may prove a partial result which is given in the next theorem.

Theorem 2. *Consider system (5.1.1) without bedload transport, that is, $q_z = 0$. Then, smooth solutions of the system satisfy the following relation*

$$\partial_t\left(\rho h\frac{u^2}{2} + \frac{g}{2}\rho h^2 + g\rho hz\right) + \partial_x\left(\rho hu\left(\frac{u^2}{2} + \frac{g}{2}h\right) + up + g\rho huz\right) = \frac{g}{2}h\zeta\phi_z(\rho_z - \rho) + gz\zeta\phi_z\rho_z - u\tau_f.$$

Proof. Combining second and third equations in (5.1.1), we get

$$\partial_t u + \partial_x\frac{u^2}{2} + \frac{1}{\rho h}\partial_x p + g\partial_x z = -\frac{\rho_z}{\rho h}\zeta\phi_z\frac{u}{2} - \frac{1}{\rho h}\tau_f.$$

Multiplying this equation by ρhu and adding the second equation in (5.1.1) times $\frac{u^2}{2}$, we obtain:

$$\partial_t\left(\rho h\frac{u^2}{2}\right) + \partial_x\left(\rho hu\frac{u^2}{2}\right) + u\partial_x p + g\rho hu\partial_x z = -u\tau_f. \tag{5.1.6}$$

Now, taking into account that $p = \frac{g}{2}\rho h^2$, from the first equation in (5.1.1) we get

$$\partial_t h + u\partial_x h + \frac{p}{\frac{g}{2}\rho h}\partial_x u = \zeta\phi_z,$$

which, combined with second equation in (5.1.1), gives

$$\partial_t\left(\frac{g}{2}\rho h^2\right) + \partial_x\left(\frac{g}{2}\rho h^2 u\right) + p\partial_x u = \frac{g}{2}h\zeta\phi_z(\rho + \rho_z). \tag{5.1.7}$$

Adding (5.1.6) and (5.1.7) gives us

$$\partial_t\left(\rho h\frac{u^2}{2} + \frac{g}{2}\rho h^2\right) + \partial_x\left(\rho hu\left(\frac{u^2}{2} + \frac{g}{2}h\right) + up\right) + g\rho hu\partial_x z = -u\tau_f + \frac{g}{2}h\zeta\phi_z(\rho + \rho_z).$$

Finally, we have

$$\begin{aligned} g\rho hu\partial_x z = g\rho hu\partial_x z + g\rho h\left(\partial_t z + \zeta\phi_z\right) + gz\left(\partial_t(\rho h) + \partial_x(\rho hu) - \zeta\phi_z\rho_z\right) \\ = \partial_t(g\rho hz) + \partial_x(g\rho huz) + g\rho h\zeta\phi_z - gz\zeta\phi_z\rho_z \end{aligned}$$

and the result follows.

□

As a last remark and referring to [38], from a physical point of view it is interesting to consider the solutions of the model when $u = 0$. Indeed, in this case the solution should satisfy

$$\begin{cases} \partial_t h = \zeta\phi_z \\ \partial_t(h\rho) = \zeta\phi_z\rho_z \\ \partial_x(\frac{gh^2\rho}{2}) = -gh\rho\partial_x z \\ \partial_t z = -\zeta\phi_z. \end{cases} \tag{5.1.8}$$

Moreover, we observe that the free surface $H = h + z$ is constant in time but

$$\partial_x(h + z) = -\frac{h}{2\rho}\partial_x\rho$$

and, in particular,

$$\partial_t\rho = \frac{\zeta\phi_z}{h}(\rho_s - \rho_w)(1 - \Psi - c).$$

Thus, if $u = 0$, we expect the free surface H to remain constant in time but the water height to decrease and the bed level to increase, as the sediments start to accumulate on the bottom. As a consequence, the density ρ will remain constant in time if $1 - \Psi = c$, or either increase or decrease if $1 - \Psi < c$ or $1 - \Psi > c$ respectively.

Chapter outline. To describe the numerical approach to approximate model (5.1.1), we will proceed step by step. We first consider a simplified version of the system with no source terms and constant bed level in time. The Lagrangian formulation of the resulting system is proposed. An approximate Riemann solver for the acoustic system is also described, see section 5.2. Then, in section 5.3, the strategy is extended in order to include the Exner equation. The numerical approximation is finally presented in section 5.4, the source terms related to the friction and the erosion-deposition processes are considered as well. Section 5.5 is devoted to the numerical simulations, while in section 5.6 we draw the conclusions.

Symbol	Description	Value or formula
a	Parameter in syst. (5.2.12), (5.3.4)	$a \geq h^2\rho^2\bar{c}^2$
b	Parameter in syst. (5.3.4)	$u^2b^2 \geq (h\rho u)^2 + g(h\rho)^2\partial_u q_z,\, b > 0$
c	Volumetric sediment concentration	$c(x,t) = (\rho - \rho_w)(\rho_s - \rho_w)^{-1}$
$\bar{c}$	Sound speed	$\bar{c} = \sqrt{gh}$
c_D	Bed drag coefficient	-
c_z	Fractional concentration of suspension near by the bed	$c_z = c\Big(0.4\Big(\frac{d}{D_{sg}}\Big)^{1.64} + 1.64\Big)$
d	Sediment diameter	[mm]
D_{sg}	Geometric mean size of the suspended sediment mixture	[mm]
E_s	Sediment entrainment coefficient	$E_s = \dfrac{1.3\times 10^{-7}\mathcal{Z}^5}{1 + 4.3\times 10^{-7}\mathcal{Z}^5}$
F_d	Deposition rate	$F_d = v_s c_z$
F_e	Erosion rate	$F_e = v_s \mathcal{P} E_s$
g	Gravitational acceleration	$9.81\ [m\ s^{-2}]$
h	water height	[m]
H	Free surface elevation	$H = h + z$
L	Volume ratio	$L(\xi,t) = \partial_\xi x(\xi,t)$
m	Mass variable	$\dfrac{1}{h_0\rho_0}\partial_x = \partial_m$
p	Pressure term	$p = g\dfrac{h^2}{2}\rho$
$\mathcal{P}$	Volume fraction of the sediment in the bottom	$1 - \mathcal{P} = \Psi$
q_z	Solid transport discharge	-
Q	Characteristic discharge	$Q = d\sqrt{gsd}$
r_w	Ratio of upper-interface resistance to bed resistance	-
$\mathcal{R}_p$	Reynold number	$\dfrac{d\sqrt{gsd}}{\nu}$
s	Relative density	$s = \dfrac{\rho_s}{\rho_w} - 1$
t	Time	[s]
u	Averaged velocity of water	$[m\ s^{-1}]$
u_*^2	Bed shear velocity	$u_*^2 = \dfrac{g\mu_f^2 u^2}{h^{\frac{1}{3}}}$
v_s	Settling velocity of the sediment	$v_s = \sqrt{(\dfrac{13.95\nu}{d})^2 + 1.09sgd} - \dfrac{13.95}{d}\nu$
x	Axial coordinate	-
z	Bed level	[m]
$\mathcal{Z}$	Parameter in eq. (5.1.4)	$\mathcal{Z} = \dfrac{\alpha_1\sqrt{c_D}\lvert u\rvert}{v_s}\mathcal{R}_p^{\alpha_2}$

Table 5.1.1: Symbols with description and formula, part one.

Symbol	Description	Value or formula
α_1	Parameter in eq. (5.1.4)	1 if $\mathcal{R}_p > 2.36$, 0.6 otherwise
α_2	Parameter in eq. (5.1.4)	0.586 if $\mathcal{R}_p > 2.36$, 1.23 otherwise
ζ	Parameter in model (5.1.1)	$\zeta = \dfrac{1}{1-\Psi}$
θ	Variable for the inverse of the density	$\theta = 1/\rho$
θ^*	Shield's parameter	$\theta^* = \dfrac{u_*^2}{sgd}$
θ_c^*	Critical Shield's stress for incipient motion	-
λ	Relaxation parameter	-
μ_f	Dimensionless Manning's coefficient	$[sm^{-1/3}]$
ν	Kinematic viscosity of the water	1×10^{-6} $[m\ s^{-2}]$
ξ	Lagrangian coordinate	-
Π	Relaxation linearization of $g\frac{h^2}{2}\rho$	-
ρ	Density of mixture of water and sediment	$[g\ cm^{-3}]$
ρ_w	Density of water	1 $[g\ cm^{-3}]$
ρ_s	Density of sediment	$[g\ cm^{-3}]$
ρ_z	Density of the saturated bottom	$\rho_z = \rho_w\Psi + \rho_s(1-\Psi)$
τ	Variable for the inverse of the water height	$\tau = 1/h$
τ_f	Friction term	$\tau_f(u) = \rho u_*^2(1+r_w)$
ϕ_z	Sediment exchange between the bottom and the water	$\phi_z = F_e - F_d$
Ψ	Porosity	-
Ω	Relaxation linearization of ζq_b	-

Table 5.1.2: Symbols with description and formula, part two.

5.2 Splitting strategy for shallow water equations with non-constant density

For the sake of clarity, let us first apply the the Lagrange-Projection (LP) strategy to system (5.1.1) without Exner equation or any source terms, namely

$$\begin{cases} \partial_t h + \partial_x(hu) = 0 \\ \partial_t(h\rho) + \partial_x(h\rho u) = 0 \\ \partial_t(h\rho u) + \partial_x(h\rho u^2 + p) = -gh\rho\partial_x z \end{cases} \tag{5.2.1}$$

or alternatively in compact form

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) = \mathbf{S}(\mathbf{Q})$$

where $\mathbf{Q}$, $\mathbf{F}(\mathbf{Q})$ and $\mathbf{S}(\mathbf{Q})$ reduce themselves to

$$\mathbf{Q} = \begin{pmatrix} h \\ h\rho \\ h\rho u \end{pmatrix}, \quad \mathbf{F}(\mathbf{Q}) = \begin{pmatrix} hu \\ h\rho u \\ h\rho u^2 + p \end{pmatrix}, \quad \mathbf{S}(\mathbf{Q}) = \begin{pmatrix} 0 \\ 0 \\ -gh\rho\partial_x z \end{pmatrix}.$$

Afterwards, in section 5.3, we will describe the general case with the solid transport flux q_z.

Then, it is easy to prove that system (5.2.1) is hyperbolic. Indeed, its Jacobian matrix reads

$$\mathbf{J}(\mathbf{Q}) = \begin{pmatrix} u & -\dfrac{u}{\rho} & \dfrac{1}{\rho} \\ 0 & 0 & 1 \\ \dfrac{gh\rho}{2} & -u^2 + \dfrac{gh}{2} & 2u \end{pmatrix},$$

where the eigenvalues are given by $u, u \pm \bar{c}$ with $\bar{c} = \sqrt{gh}$ sound speed. Note that, in this particular case without solid transport flux, z is constant in time. Therefore, the quantity $gh\rho\partial_x z$ in the third equation is treated as a source term.

It is interesting to consider stationary solutions of this reduced model. Indeed, in practical applications, we could find such steady states or perturbations of them. In the particular case of stationary solutions with zero-velocity, we get the family

$$u = 0, \quad \partial_x(h+z) = -\frac{1}{2\rho}\partial_x(h\rho) + \frac{1}{2}\partial_x h \tag{5.2.2}$$

or, written in an alternative way,

$$u = 0, \quad \partial_x(h+z) = -\frac{h}{2\rho}\partial_x \rho. \tag{5.2.3}$$

Moreover, among all the stationary solutions (5.2.2), it is interesting to show two particular families: the one with constant bed level

$$u = 0, \quad \frac{h^2\rho}{2} = \text{constant} \quad \text{and} \quad z = \text{constant}, \tag{5.2.4}$$

and the usual "lake at rest" solution

$$u = 0, \quad \rho = \text{constant} \quad \text{and} \quad H = h + z = \text{constant}. \tag{5.2.5}$$

When the friction and erosion-deposition source terms are neglected, we design our numerical scheme in such a way that it exactly preserves the stationary solutions (5.2.4) and (5.2.5). That is, we want our numerical method to be exactly well-balanced for those stationary solutions (see for instance [2, 36]). Note that (5.2.4) and (5.2.5) are two particular families of the more general case (5.2.3). In sections 5.2.2 and 5.4.1, we will see that the proposed schemes do not exactly preserve (5.2.3), but a discrete version of it. In such a case, we say that the numerical scheme is well-balanced with order 2 for (5.2.3), according to the definition introduced in [36]. Namely, the numerical scheme preserves a discrete stationary solution that is a second order approximation of (5.2.3).

5.2.1 Lagrange-projection decomposition

As mentioned in the introduction, the idea of the Lagrange-projection approach is to split the acoustic and transport terms of the model. In practice, this strategy can be explained by using Lagrangian coordinates. The approach then results in first considering the mathematical model formulated in Lagrangian coordinates and then perform the projection of the Lagrangian solution onto Eulerian coordinates. We shall detail both steps in what follows.

Let us first briefly recall the corresponding formalism. We consider a fluid particle located at position ξ at time $t = 0$. Then, its trajectory through time or its characteristic curve $t \mapsto x(\xi, t)$ is given by

$$\begin{cases} \partial_t x(\xi, t) = u(x(\xi, t), t), \\ x(\xi, 0) = \xi. \end{cases} \tag{5.2.6}$$

Then, any function $(x, t) \mapsto \varphi(x, t)$ in Eulerian coordinates can be expressed in Lagrangian coordinates as follows,

$$(\xi, t) \mapsto \overline{\varphi}(\xi, t) = \varphi(x(\xi, t), t).$$

In particular, using the volume ratio $L(\xi, t)$,

$$L(\xi, t) = \frac{\partial x}{\partial \xi}(\xi, t), \tag{5.2.7}$$

which satisfies

$$\begin{cases} \partial_t L(\xi,t) = \partial_\xi u(x(\xi,t),t), \\ L(\xi,0) = 1, \end{cases} \tag{5.2.8}$$

we can easily write the original system (5.2.1) in Lagrangian coordinates, namely

$$\begin{cases} \partial_t(L\bar{h}) = 0 \\ \partial_t(L\overline{h\rho}) = 0 \\ \partial_t(L\overline{h\rho u}) + \partial_\xi\left(g\bar{\rho}\dfrac{\bar{h}^2}{2}\right) = -g\bar{h}\bar{\rho}\partial_\xi\bar{z}. \end{cases} \tag{5.2.9}$$

More details about the Lagrange-projection decomposition applied to the shallow water system can be found for instance in [44, 11]. Let us remark that the Lagrangian formulation (5.2.9) will reveal itself to be very useful and convenient when trying to include the Exner equation in the model, see section 5.3. Indeed, we will remark that it is easier to consider the Lagrangian formulation with variable Lz rather than z, see [18].

Let us point out that system (5.2.9) may also be formulated in a different way. Indeed, observing that both $L\bar{h}$ and $L\bar{h}\bar{\rho}$ do not depend on time, we get

$$L\bar{h}(\xi,t) = L\bar{h}(\xi,0) = h(\xi,0) = h_0 \quad \text{and consequently} \quad L = \frac{h_0}{\bar{h}}.$$

Therefore $0 = \partial_t(L\bar{h}\bar{\rho}) = L\bar{h}\partial_t\bar{\rho} = h_0\partial_t\bar{\rho}$, which means that

$$\partial_t\bar{\rho} = 0 \quad \text{and in particular} \quad \bar{\rho}(\xi,t) = \rho(\xi,0) = \rho_0.$$

Defining now the variables $\bar{\tau} = \dfrac{1}{\bar{h}}$ and $\bar{\theta} = \dfrac{1}{\bar{\rho}}$, we find the equivalent form of system (5.2.9),

$$\begin{cases} \partial_t\bar{\theta} = 0 \\ \partial_t(h_0\rho_0\bar{\tau}\bar{\theta}) - \partial_\xi\bar{u} = 0 \\ \partial_t(h_0\rho_0\bar{u}) + \partial_\xi\left(\dfrac{g}{2\bar{\tau}^2\bar{\theta}}\right) = -\dfrac{g}{\bar{\tau}\bar{\theta}}\partial_\xi\bar{z} \end{cases}$$

and, alternatively, neglecting the bar for the sake of simplicity,

$$\begin{cases} \partial_t\theta = 0 \\ \partial_t(h_0\rho_0\tau\theta) - \partial_\xi u = 0 \\ \partial_t(h_0\rho_0 u) + \partial_\xi\left(\dfrac{g}{2\tau^2\theta}\right) = -\dfrac{g}{\tau\theta}\partial_\xi z. \end{cases} \tag{5.2.10}$$

In this framework, the numerical strategy could again be summarized in two steps. First, we need to numerically solve the Lagrangian-acoustic system (5.2.9)-(5.2.10). Then, we project its solution into Eulerian coordinates. We will see that the most problematic part of this strategy will not be the approximation of the projection step but that of the Lagrangian system, especially when trying to satisfy the well-balanced property. In particular, for these equations (5.2.10), we describe an approximate Riemann solver which will be used to define the associated Godunov-type scheme. For this reason, it is convenient to reformulate system (5.2.10) exploiting the so-called mass variable m, which is given by $\dfrac{1}{h_0\rho_0}\partial_\xi = \partial_m$. Thus, it is easy to show that equations (5.2.10) are equivalent to the following system,

$$\begin{cases} \partial_t\theta = 0 \\ \partial_t(\tau\theta) - \partial_m u = 0 \\ \partial_t u + \partial_m p = -\dfrac{g}{\tau\theta}\partial_m z, \end{cases} \tag{5.2.11}$$

where we recall that $p = \dfrac{g\theta}{2(\tau\theta)^2}$. We will refer to equations (5.2.11) as the acoustic system, as it can be obtained from the starting system (5.2.1) by considering only the acoustic phenomena and the topography variations. See [22, 18, 27] for more details about the acoustic-transport interpretation. Moreover, the eigenvalues of system (5.2.11) are given by 0 and $\pm h\rho\bar{c}$, where the latter are the speed of propagation of the acoustic waves; the material (transport) waves being related to the projection step. Thus, in situations in which the acoustic waves are much faster than the material ones, it can be very convenient to exploit an implicit approximation for the acoustic equations, obtaining in this way a very fast implicit-explicit method. Further details about the implicit formulation for the acoustic system are given in section 5.4.1. See also [22, 23, 19] for implicit-explicit Lagrange-projection numerical methods.

Looking for an approximate Riemann solver associated with system (5.2.11), we follow the Suliciu relaxation approach [49] and we introduce the following approximate relaxation system,

$$\begin{cases} \partial_t \theta = 0 \\ \partial_t(\tau\theta) - \partial_m u = 0 \\ \partial_t u + \partial_m \Pi = -\dfrac{g}{\tau\theta}\partial_m z \\ \partial_t \Pi + a^2 \partial_m u = 0 \end{cases} \tag{5.2.12}$$

where Π is a new variable such that $\Pi = p$ at time $t = 0$. Our approximate Riemann solver will consist of an exact Riemann solver associated with system (5.2.12). Moreover, a^2 is a constant which linearizes $h^2\rho^2\bar{c}^2$ and which is taken as $a^2 \geq h^2\rho^2\bar{c}^2$ according to the sub-characteristic condition. Then, easy computations show that the eigenvalues of (5.2.12) are given by $\lambda = 0$, $\lambda_{\pm a} = \pm a$ and that the associated characteristic fields are all linearly degenerate. This property is well-known to provide an exact and easy solution of the Riemann problem. Indeed, we will obtain three waves that correspond to contact discontinuities. Then, exploiting the Rankine-Hugoniot relations across each wave, we are able to exactly define the solution of the Riemann problem associated to system (5.2.12) (see for instance [32, 33] for more details). For applications related to the Suliciu relaxation approach, see for instance [8, 5, 23, 22, 24].

5.2.2 Approximate Riemann solver

In this section, we aim to briefly describe the approximate Riemann problem solution for system (5.2.11), which is found solving the Riemann problem associated with the relaxation system (5.2.12). The initial data of the Riemann problem are given by

$$(\theta, \tau\theta, u, \Pi)^T(m, t=0) = \begin{cases} (\theta_L, \tau_L\theta_L, u_L, \Pi_L)^T & \text{if} \quad m < 0 \\ (\theta_R, \tau_R\theta_R, u_R, \Pi_R)^T & \text{if} \quad m \geq 0, \end{cases}$$

where $\Pi_{L,R} = \dfrac{g}{2}(h^2\rho)_{L,R}$. Then, its solution would be composed of four different states separated by the three discontinuities,

$$\hat{\mathbf{U}}(\frac{m}{t}; \mathbf{U}_L, \mathbf{U}_R) = \begin{cases} \mathbf{U}_L & \text{if} \quad \frac{m}{t} < \lambda_a^- = -a \\ \mathbf{U}_L^* & \text{if} \quad -a < \frac{m}{t} < \lambda_0 = 0 \\ \mathbf{U}_R^* & \text{if} \quad 0 < \frac{m}{t} < \lambda_a^+ = a \\ \mathbf{U}_R & \text{if} \quad \frac{m}{t} > a, \end{cases}$$

where $\mathbf{U} = (\theta, \tau\theta, u, \Pi)^T$. The definition of $\mathbf{U}_L^*$ and $\mathbf{U}_R^*$ relies on the validity of the Rankine Hugoniot relations across each wave (recall that the characteristic fields are linearly degener-

ate) and, in particular, on a consistent approximation

$$\mathcal{M} = \mathcal{M}(\mathbf{U}_L, \mathbf{U}_R) = g\{h\rho\}_{ST}(z_R - z_L)$$

of the source term in (5.2.12), such that across the stationary wave one has

$$\begin{cases} u^* = u_L^* = u_R^* \\ \Pi_R^* - \Pi_L^* + \mathcal{M} = 0. \end{cases} \tag{5.2.13}$$

Here, $\{h\rho\}_{ST}$ needs to be specified in a consistent way, namely

$$\lim_{\substack{h_L, h_R \to h \\ \rho_L, \rho_R \to \rho}} \{h\rho\}_{ST} = h\rho.$$

For the sake of brevity, we shall not give all the details here and we refer the reader to [22] for further information. The star values $\mathbf{U}_L^*, \mathbf{U}_R^*$ are then given by

$$\begin{cases} \theta_L^* = \theta_L \\ \theta_R^* = \theta_R \\ (\tau\theta)_L^* = \tau_L\theta_L + \dfrac{1}{a}(u^* - u_L) \\ (\tau\theta)_R^* = \tau_R\theta_R - \dfrac{1}{a}(u^* - u_R) \\ u^* = \dfrac{1}{2}(u_L + u_R) - \dfrac{1}{2a}(\Pi_R - \Pi_L) - \dfrac{\mathcal{M}}{2a} \\ \Pi_L^* = \dfrac{1}{2}(\Pi_L + \Pi_R) - \dfrac{a}{2}(u_R - u_L) + \dfrac{\mathcal{M}}{2} \\ \Pi_R^* = \dfrac{1}{2}(\Pi_L + \Pi_R) - \dfrac{a}{2}(u_R - u_L) - \dfrac{\mathcal{M}}{2}. \end{cases} \tag{5.2.14}$$

5.2.2.1 Well-balanced property and definition of $\mathcal{M}$

The definition of $\mathcal{M}$ is driven by the well-balanced property. More precisely, assume that we want to preserve a discrete approximation of the stationary solutions with zero-velocity, namely the ones defined in formula (5.2.2), and consider the following discretization of such stationary solutions:

$$(z_R - z_L) + \frac{h_R - h_L}{2} = -\frac{1}{2}\left\{\frac{1}{\rho}\right\}_{StS}(h_R\rho_R - h_L\rho_L) \tag{5.2.15}$$

where the term $\left\{\dfrac{1}{\rho}\right\}_{StS}$ needs to be specified in a consistent way, namely

$$\lim_{\rho_L, \rho_R \to \rho} \left\{\frac{1}{\rho}\right\}_{StS} = \frac{1}{\rho}.$$

Hence, when $\mathbf{U}_L$ and $\mathbf{U}_R$ satisfy (5.2.15), we require $\mathbf{U}_L^* = \mathbf{U}_L$ and $\mathbf{U}_R^* = \mathbf{U}_R$ in order to have a well-balanced approximate Riemann solver. On one hand, (5.2.13) gives

$$\Pi_R - \Pi_L + \mathcal{M} = 0 \quad \text{with} \quad \mathcal{M} = g\{h\rho\}_{ST}(z_R - z_L), \tag{5.2.16}$$

so that, inserting (5.2.15) into (5.2.16), we find

$$\begin{aligned} \frac{h_R^2\rho_R}{2} - \frac{h_L^2\rho_L}{2} &= -\{h\rho\}_{ST}(z_R - z_L) \\ &= \frac{1}{2}\{h\rho\}_{ST}\Big(h_R - h_L + \left\{\frac{1}{\rho}\right\}_{StS}(h_R\rho_R - h_L\rho_L)\Big). \end{aligned}$$

On the other hand, we have

$$\frac{h_R^2\rho_R}{2} - \frac{h_L^2\rho_L}{2} = \frac{1}{2}\Big((h_R\rho_R - h_L\rho_L)\frac{h_R+h_L}{2} + (h_R-h_L)\Big(\frac{h_R\rho_R+h_L\rho_L}{2}\Big)\Big).$$

Hence, a possibility is to set

$$\begin{cases} \{h\rho\}_{ST} = \dfrac{h_R\rho_R+h_L\rho_L}{2} \\ \Big\{\dfrac{1}{\rho}\Big\}_{StS} = \dfrac{1}{\{h\rho\}_{ST}}\dfrac{h_R+h_L}{2}, \end{cases}$$

which is clearly consistent. In particular, we would define

$$\mathcal{M} = g\Big(\frac{h_R\rho_R+h_L\rho_L}{2}\Big)(z_R - z_L), \tag{5.2.17}$$

which allows us to preserve the stationary solutions which, according to (5.2.15), satisfy the following discretization

$$(z_R - z_L) + \frac{h_R-h_L}{2} = -\frac{1}{2}\frac{h_R+h_L}{h_R\rho_R+h_L\rho_L}(h_R\rho_R - h_L\rho_L) \tag{5.2.18}$$

together with zero-velocity. Note that, if $\rho_L = \rho_R$ and $u_L = u_R = 0$, we recover the well-known lake at rest stationary solution (5.2.5) as (5.2.18) gives $h_L+z_L = h_R+z_R$. Moreover, the stationary solution (5.2.4) (defined by $z_L = z_R$, $u_L = u_R = 0$ and $\rho_L h_L^2 = \rho_R h_R^2$) gives $\mathcal{M} = 0$ and $u_L^* = u_L$, $u_R^* = u_R$ from (5.2.14). In other words, our scheme is well-balanced with order 2 for the stationary solutions (5.2.2) and it is exactly well-balanced for the stationary solutions (5.2.4) and (5.2.5).

Finally, observe that in practice the constant a is defined as follows,

$$a = \max(\varepsilon, h_L\rho_L\bar{c}_L, h_R\rho_R\bar{c}_R),$$

where ε is a tolerance value so that a will not be zero. In practice, we take $\varepsilon = \Delta x$.

Remark 6. *The proposed definition of $\mathcal{M}$ is not unique. Indeed, let us consider the following discretization of the stationary solutions based on (5.2.3) (instead of (5.2.2) above):*

$$z_R + h_R - (z_L + h_L) = -\frac{1}{2}\Big\{\frac{h}{\rho}\Big\}_{StS}(\rho_R - \rho_L) \tag{5.2.19}$$

where the term $\Big\{\dfrac{h}{\rho}\Big\}_{StS}$ needs to be specified. Similar calculations lead to

$$\mathcal{M} = \frac{g}{2}\Big(\frac{\rho_R+\rho_L}{2}\frac{h_R+h_L}{2} + \frac{h_R\rho_R+h_L\rho_L}{2}\Big)(z_R - z_L), \tag{5.2.20}$$

which allows to preserve stationary solutions defined by (5.2.19) with

$$\Big\{\frac{h}{\rho}\Big\}_{StS} = \Big(\frac{h_R+h_L}{2}\Big)^2\frac{8}{(\rho_R+\rho_L)(h_R+h_L)+2(h_R\rho_R+h_L\rho_L)}.$$

Hence, in this section we have considered system (5.2.1) as a starting point to show how to include the evolution equation for the density variable in the Lagrange-projection approach. We have also defined an approximate Riemann solver for the resulting Lagrangian system. Let us now extend this strategy by including the Exner equation in the model.

5.3 Including the Exner equation in the splitting strategy

Moving to the next step, we now aim to include the Exner equation in the splitting strategy. Thus, we consider system (5.1.1) without friction and erosion-deposition source terms, namely

$$\begin{cases} \partial_t h + \partial_x(hu) = 0 \\ \partial_t(h\rho) + \partial_x(h\rho u) = 0 \\ \partial_t(h\rho u) + \partial_x(h\rho u^2 + g\frac{h^2}{2}\rho) + gh\rho\partial_x z = 0 \\ \partial_t z + \zeta\partial_x q_z = 0. \end{cases} \tag{5.3.1}$$

In the following section, we show its Lagrangian-acoustic formulation with the aim of describing an associated approximate Riemann solver.

Referring to the previous works [17] and [18] (chapter 4), we consider two different strategies to take into account the Exner equation. In the first one, we update it directly in the projection step (section 5.4.2), resembling an usual splitting strategy: we first approximate the shallow water equations with variable density and then also the topography is updated. Thus, here we do not have to give further details about its Lagrangian formulation. Moreover, we will see that this strategy could be particularly useful when considering the implicit-explicit version of the scheme. As for the second strategy, we propose to decouple the Exner equation in both steps, which would be the most natural approach in the Lagrange-projection framework, see next section 5.3.1. Observe that even the latter strategy is only weakly-coupled and not fully-coupled in order to preserve the well-balanced property. Let us recall that decoupled approaches applied to the Exner system could produce unphysical oscillations in the numerical results due to the different eigenstructure of the shallow water model with and without the Exner equation (systems (5.2.1) and (5.3.1) respectively), see [25]. In this sense, the method that decouples the Exner equation in both steps proved to be more stable than the other one, especially at second-order of accuracy (refer to [18]). For this reason, here we consider and extend both strategies.

5.3.1 Updating the bed level in both steps

Exploiting once again the Lagrangian formalism introduced in section 5.2.1, few computations allow us to write system (5.3.1) in Lagrangian coordinates, namely

$$\begin{cases} \partial_t(L\bar{h}) = 0 \\ \partial_t(L\overline{h\rho}) = 0 \\ \partial_t(L\overline{h\rho u}) + \partial_\xi(g\dfrac{\bar{h}^2}{2}\bar{\rho}) = -g\bar{h}\bar{\rho}\partial_\xi z \\ \partial_t(L\overline{z}) - \partial_\xi(\overline{zu}) + \zeta\partial_\xi\overline{q_z} = 0 \end{cases} \tag{5.3.2}$$

or equivalently, using once again the change of variables $\bar{\tau} = \dfrac{1}{\bar{h}}$ and $\bar{\theta} = \dfrac{1}{\bar{\rho}}$,

$$\begin{cases} \partial_t\bar{\theta} = 0 \\ \partial_t(h_0\rho_0\bar{\tau}\bar{\theta}) - \partial_\xi\bar{u} = 0 \\ \partial_t(h_0\rho_0\bar{u}) + \partial_\xi\left(\dfrac{g}{2\bar{\tau}^2\bar{\theta}}\right) = -\dfrac{g}{\bar{\tau}\bar{\theta}}\partial_\xi\bar{z} \\ \partial_t(h_0\rho_0\overline{z}) - \dfrac{\bar{u}}{\bar{\tau}\bar{\theta}}\partial_\xi\overline{z} + \dfrac{\zeta}{\bar{\tau}\bar{\theta}}\partial_\xi\overline{q_z} = 0. \end{cases} \tag{5.3.3}$$

Let us observe that the evolution equation for the bed elevation z is written as a conservation law in system (5.3.2), contrarily to the one in system (5.3.3). Then, neglecting the bars and

exploiting again the mass variable m, system (5.3.3) can also be reformulated as

$$\begin{cases} \partial_t \theta = 0 \\ \partial_t(\tau\theta) - \partial_m u = 0 \\ \partial_t u + \partial_m p = -\dfrac{g}{\tau\theta}\partial_m z \\ \partial_t z - \dfrac{u}{\tau\theta}\partial_m z + \dfrac{\zeta}{\tau\theta}\partial_m q_z = 0. \end{cases}$$

Proceeding with the relaxation system, we consider the same variable Π to linearize the pressure term $p = \dfrac{gh^2\rho}{2}$, while for the solid transport discharge q_z we introduce Ω such that $\Omega = \zeta q_z$ at time $t = 0$. Hence, we consider the following linearized system

$$\begin{cases} \partial_t \theta = 0 \\ \partial_t(\tau\theta) - \partial_m u = 0 \\ \partial_t u + \partial_m \Pi = -\dfrac{g}{\tau\theta}\partial_m z \\ \partial_t z - \dfrac{u}{\tau\theta}\partial_m z + \dfrac{1}{\tau\theta}\partial_m \Omega = 0 \\ \partial_t \Pi + a^2 \partial_m u = 0 \\ \partial_t \Omega + u^2(b^2\tau\theta - \dfrac{1}{\tau\theta})\partial_m z + \dfrac{u}{\tau\theta}\partial_m \Omega = 0 \end{cases} \tag{5.3.4}$$

where the sub-characteristic condition is now given by $a^2 \geq h^2\rho^2\bar{c}^2$ and $u^2b^2 \geq (h\rho u)^2 + g(h\rho)^2\partial_u q_z$, $b > 0$, see also [18] (chapter 4). In compact form, system (5.3.4) is equivalent to

$$\partial_t \mathbf{U} + \mathbf{A}(\mathbf{U})\partial_x \mathbf{U} = \mathbf{S}(\mathbf{U})$$

with

$$\mathbf{U} = \begin{pmatrix} \theta \\ \tau\theta \\ u \\ z \\ \Pi \\ \Omega \end{pmatrix}, \quad \mathbf{A}(\mathbf{U}) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -\dfrac{u}{\tau\theta} & 0 & \dfrac{1}{\tau\theta} \\ 0 & 0 & a^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & u^2(b^2\tau\theta - \dfrac{1}{\tau\theta}) & 0 & \dfrac{u}{\tau\theta} \end{pmatrix}, \quad \mathbf{S}(\mathbf{U}) = \begin{pmatrix} 0 \\ 0 \\ -\dfrac{g}{\tau\theta}\rho\partial_m z \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{5.3.5}$$

Considering only the convective part of system (5.3.4), namely neglecting the term related to z in the third equation, we find that the eigenvalues are given by $\lambda = 0, \lambda_{\pm a} = \pm a, \lambda_{\pm b} = \pm|u|b$. Hence, once again the associated characteristic fields are all linearly degenerate. We remark that system (5.3.4) is not strictly hyperbolic anymore for $u = 0$ and that the eigenvalues are not ordered a priori, see again [18].

5.3.1.1 Approximate Riemann solver

Here we solve the Riemann problem associated with system (5.3.4) with initial data

$$(\theta, \tau\theta, u, z, \Pi, \Omega)^T(m, t = 0) = \begin{cases} (\theta_L, \tau_L\theta_L, u_L, z_L, \Pi_L, \Omega_L)^T & \text{if} \quad m < 0 \\ (\theta_R, \tau_R\theta_R, u_R, z_R, \Pi_R, \Omega_R)^T & \text{if} \quad m > 0 \end{cases}$$

with $\Pi_{L,R} = p_{L,R}$ and $\Omega_{L,R} = (\zeta q_z)_{L,R}$. Since the eigenvalues are not ordered a priori, at a continuous level there exists two different cases depending on whether $a < |u|b$ or not (recall

that a and b are positive). In practice, we will distinguish between the following two cases $a < |u_L|b, a < |u_R|b$ and $a > |u^*|b$. As a consequence, if $a < |u_L|b, a < |u_R|b$, the solution reads

$$\hat{\mathbf{U}}(\frac{m}{t};\mathbf{U}_L,\mathbf{U}_R) = \begin{cases} \mathbf{U}_L & \text{if} \quad \frac{m}{t} < \lambda_b^- = -|u_L|b \\ \mathbf{U}^*_{b,L} & \text{if} \quad -|u_L|b < \frac{m}{t} < \lambda_a^- = -a \\ \mathbf{U}^*_{a,L} & \text{if} \quad -a < \frac{m}{t} < \lambda_0 = 0 \\ \mathbf{U}^*_{a,R} & \text{if} \quad 0 < \frac{m}{t} < \lambda_a^+ = a \\ \mathbf{U}^*_{b,R} & \text{if} \quad a < \frac{m}{t} < \lambda_b^+ = |u_R|b \\ \mathbf{U}_R & \text{if} \quad \frac{m}{t} > |u_R|b. \end{cases}$$

with

$$\mathbf{U}^*_{b,L} = \begin{pmatrix} \theta_L \\ \tau_L\theta_L \\ u_L \\ z^* \\ \Pi_L \\ \Omega^* \end{pmatrix}, \quad \mathbf{U}^*_{a,L} = \begin{pmatrix} \theta_L \\ \tau^*_L\theta_L \\ u^* \\ z^* \\ \Pi^*_L \\ \Omega^* \end{pmatrix}, \quad \mathbf{U}^*_{a,R} = \begin{pmatrix} \theta_R \\ \tau^*_R\theta_R \\ u^* \\ z^* \\ \Pi^*_R \\ \Omega^* \end{pmatrix}, \quad \text{and} \quad \mathbf{U}^*_{b,R} = \begin{pmatrix} \theta_R \\ \tau_R\theta_R \\ u_R \\ z^* \\ \Pi_R \\ \Omega^* \end{pmatrix}. \tag{5.3.6}$$

On the other hand, if $a > |u^*|b$, the solution is given by

$$\hat{\mathbf{U}}(\frac{m}{t};\mathbf{U}_L,\mathbf{U}_R) = \begin{cases} \mathbf{U}_L & \text{if} \quad \frac{m}{t} < -a \\ \mathbf{U}^*_{a,L} & \text{if} \quad -a < \frac{m}{t} < -|u^*|b \\ \mathbf{U}^*_{b,L} & \text{if} \quad -|u^*|b < \frac{m}{t} < 0 \\ \mathbf{U}^*_{b,R} & \text{if} \quad 0 < \frac{m}{t} < |u^*|b \\ \mathbf{U}^*_{a,R} & \text{if} \quad |u^*|b < \frac{m}{t} < a \\ \mathbf{U}_R & \text{if} \quad \frac{m}{t} > a. \end{cases}$$

with

$$\mathbf{U}^*_{a,L} = \begin{pmatrix} \theta_L \\ \tau^*_L\theta_L \\ u^* \\ z_L \\ \Pi^*_L \\ \Omega_L \end{pmatrix}, \quad \mathbf{U}^*_{b,L} = \begin{pmatrix} \theta_L \\ \tau^*_L\theta_L \\ u^* \\ z^* \\ \Pi^*_L \\ \Omega^* \end{pmatrix}, \quad \mathbf{U}^*_{b,R} = \begin{pmatrix} \theta_R \\ \tau^*_R\theta_R \\ u^* \\ z^* \\ \Pi^*_R \\ \Omega^* \end{pmatrix}, \quad \text{and} \quad \mathbf{U}^*_{a,R} = \begin{pmatrix} \theta_R \\ \tau^*_R\theta_R \\ u^* \\ z_R \\ \Pi^*_R \\ \Omega_R \end{pmatrix} \tag{5.3.7}$$

respectively. In order to define the star states in (5.3.6) and (5.3.7), we follow the exact same lines as in section 5.2.2. In particular, it is worth noticing that z and Ω are constant through the waves with $\pm a$-velocity and zero-velocity, thus we only need to find a single star value z^*, Ω^* for these two variables. Similarly for the variables θ, τ, u and Π, since they are constant through the waves with $\pm|u|b$-velocity, at most two star values are necessary. This property is related to the fact that in the third equation we treat the coupling term $-\dfrac{g}{\tau\theta}\partial_m z$ as a source term. Moreover, in both cases we exploit the same star values for the variables $\theta, \tau\theta, u$, given by (5.2.14) and (5.2.17). Regarding z^* and Ω^*, we need to separate the two cases and, in partic-

ular, if $a < |u_L|b$ and $a < |u_R|b$, we get

$$\begin{cases} z^* = \dfrac{|u_R|(\text{sign}(u_R) + b\tau_R\theta_R)z_R - |u_L|(\text{sign}(u_L) - b\tau_L\theta_L)z_L}{|u_R|(\text{sign}(u_R) + b\tau_R\theta_R) - |u_L|(\text{sign}(u_L) - b\tau_L\theta_L)} \\ \qquad - \dfrac{\Omega_R - \Omega_L}{|u_R|(\text{sign}(u_R) + b\tau_R\theta_R) - |u_L|(\text{sign}(u_L) - b\tau_L\theta_L)} \\ \Omega^* = \dfrac{\Omega_R + \Omega_L}{2} + \dfrac{|u_R|}{2}(\text{sign}(u_R) + b\tau_R\theta_R)(z^* - z_R) + \dfrac{|u_L|}{2}(\text{sign}(u_L) - b\tau_L\theta_L)(z^* - z_L), \end{cases} \tag{5.3.8}$$

otherwise we state

$$\begin{cases} (z|u|)^* = |u^*|\dfrac{(\text{sign}(u^*) + b\tau_R^*\theta_R)z_R - (\text{sign}(u^*) - b\tau_L^*\theta_L)z_L}{b(\tau_R^*\theta_R + \tau_L^*\theta_L)} - \dfrac{\Omega_R - \Omega_L}{b(\tau_R^*\theta_R + \tau_L^*\theta_L)} \\ \Omega^* = \dfrac{\Omega_R + \Omega_L}{2} + \\ \quad + \dfrac{1}{2}((\text{sign}(u^*) + b\tau_R^*\theta_R)((z|u|)^* - |u^*|z_R) + (\text{sign}(u^*) - b\tau_L^*\theta_L)((z|u|)^* - |u^*|z_L)). \end{cases} \tag{5.3.9}$$

Finally, the parameters a and b are defined as $a = \max(\Delta x, h_L\rho_L\bar{c}_L, h_R\rho_R\bar{c}_R)$ and

$$b = \max\Big(\varepsilon, \sqrt{(h_L\rho_L)^2 + g\frac{(h_L\rho_L)^2}{u_L^2}\partial_u(q_b)_L}, \sqrt{(h_R\rho_R)^2 + g\frac{(h_R\rho_R)^2}{u_R^2}\partial_u(q_b)_R}\Big) \tag{5.3.10}$$

respectively. However, since assuming $a \geq |u_L|b$ or $a \geq |u_R|b$ does not necessarily imply that $a > |u^*|b$, in practice we need to do the following. If a and b are such that $a \geq |u_L|b$ or $a \geq |u_R|b$ but $a \leq |u^*|b$, we need to increase the value of a and redefine it as $a = (1+\epsilon)|u^*|b$ (with typically $\epsilon = 0.01$). We highlight that, once we have redefined a, we have to recompute the value of u^*, and more generally the quantities in (5.2.14). In practice, this iterative process usually converges in one or two iterations. See again the previous work [18] for more details about this approximate Riemann solver applied to system (5.3.2) with constant density ρ in time and space.

Finally, let us remark that the choice not to use a fully coupled approximation for this system also contributes to the fulfillment of the well-balanced property, namely to the preservation of the stationary solutions (5.2.2). Indeed, uz^* and Ω^* are automatically equal to zero when the steady state condition (5.2.2) is satisfied.

5.4 Numerical method

Notations. First, we define the constant space step Δx and constant time step Δt. The mesh interfaces are given by $x_{j+1/2} = j\Delta x$ for $j \in \mathbb{Z}$ and the intermediate times by $t^n = n\Delta t$ for $n \in \mathbb{N}$. At each time t^n, we seek for an approximation $\mathbf{Q}_j^n$ of the solution in the interval $[x_{j-1/2}, x_{j+1/2})$, $j \in \mathbb{Z}$. Therefore, a piecewise constant approximate solution $x \to \mathbf{Q}_{\Delta t,\Delta x}(x, t^n)$ of the solution $\mathbf{Q}$ is given by

$$\mathbf{Q}_{\Delta t,\Delta x}(x, t^n) = \mathbf{Q}_j^n \text{ for all } x \in C_j = [x_{j-1/2}; x_{j+1/2}), \quad j \in \mathbb{Z}, \quad n \in \mathbb{N}.$$

Numerical strategy. As it has already been explained in the previous sections, the numerical scheme consists of three steps: first, one has to solve the acoustic-Lagrangian step, then the transport-projection one and, finally, we need to include the source terms. Thus, using the above notations and the Lagrangian coordinates we have:

1. Update $\mathbf{Q}^n$ to $L\mathbf{Q}^{n+}$ by solving the Lagrangian system (section 5.4.1);

2. Project $L\mathbf{Q}^{n+}$ into Eulerian coordinates, finding $\mathbf{Q}^{n+1-}$ (section 5.4.2);
3. Consider the erosion, deposition and friction source terms and update $\mathbf{Q}^{n+1-}$ to $\mathbf{Q}^{n+1}$ (section 5.4.3).

5.4.1 Lagrangian step

As we have already anticipated, in order to approximate the acoustic or Lagrangian step, we exploit a first-order Godunov-type scheme associated with the approximate Riemann solver for the acoustic system that we have built in the previous sections. Since the approximate Riemann problem solution appears to be the same for the variables τ, θ and u (both with and without the contribution of the Exner equation), we can immediately write their numerical approximation. Indeed, we have

$$\begin{cases} L_j^{n+} h_j^{n+} = L_j^n h_j^n \\ L_j^{n+} h_j^{n+} \rho_j^{n+} = L_j^n h_j^n \rho_j^n \\ L_j^{n+} (h\rho u)_j^{n+} = L_j^n (h\rho u)_j^n - \dfrac{\Delta t}{\Delta x} (\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}) - \Delta t \{ gh\rho \partial_x z \}_j^n \end{cases} \tag{5.4.1}$$

where

$$L_j^{n+} = L_j^n + \frac{\Delta t}{\Delta x} (u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}) \quad \text{with} \quad L_j^n = 1 \tag{5.4.2}$$

with star values $u^*_{j+1/2}, \Pi^*_{j+1/2}$ being locally defined at each interface $x_{j+1/2}$ using formulas (5.2.14). For the source term, we simply state

$$s_j^n = \{ gh\rho \partial_x z \}_j^n = \frac{1}{2} \Big(s^n_{j+1/2} + s^n_{j-1/2} \Big) \quad \text{with} \quad s^n_{j+1/2} = - \frac{\mathcal{M}^n_{j+1/2}}{\Delta x} \tag{5.4.3}$$

and $\mathcal{M}^n_{j+1/2} = \mathcal{M}((h_j^n, h_j^n \rho_j^n; z_j^n); (h_{j+1}^n, h_{j+1}^n \rho_{j+1}^n; z_{j+1}^n))$ given by (5.2.20) for all j. Few algebraic computations give a completely equivalent numerical approximation for the relaxation acoustic system (5.2.12), namely

$$\begin{cases} \theta_j^{n+} = \theta_j^n \\ \tau \theta_j^{n+} = \tau \theta_j^n + \dfrac{\Delta t}{\Delta m_j} (u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}) \\ u_j^{n+} = u_j^n - \dfrac{\Delta t}{\Delta m_j} (\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}) - \Delta t \{ \dfrac{g}{\tau\theta} \partial_m z \}_j^n \\ \Pi_j^{n+} = \Pi_j^n - (a^n_{j+\frac{1}{2}})^2 \dfrac{\Delta t}{\Delta m_j} (u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}). \end{cases} \tag{5.4.4}$$

Remark 7. *It is particularly useful to show the latter formulation as it can be interpreted in an implicit way by defining the star values as follows,*

$$\begin{aligned} u^*_{j+\frac{1}{2}} &= \frac{1}{2}(u_{j+1}^{n+} + u_j^{n+}) - \frac{1}{2a^n_{j+\frac{1}{2}}} (\Pi_{j+1}^{n+} - \Pi_j^{n+}) - \frac{\mathcal{M}^n_{j+1/2}}{2a^n_{j+\frac{1}{2}}} \\ \Pi^*_{j+\frac{1}{2}} &= \frac{1}{2}(\Pi_{j+1}^{n+} + \Pi_j^{n+}) - \frac{a^n_{j+\frac{1}{2}}}{2} (u_{j+1}^{n+} - u_j^{n+}). \end{aligned} \tag{5.4.5}$$

Hence, for an implicit approximation of the acoustic step, we highlight that first we solve the evolution equations for u and Π and then we use the obtained solution to compute $\tau\theta^{n+}$. In particular, the third and fourth equations of system (5.4.4) can be reformulated as a linear system and, as such, their resolution is not computationally expensive. Here we do not provide further details, see either appendix 5.A or refer to [22] and [23] for this approach applied to the shallow water equations and the gas dynamics equations respectively.

Then, let us focus on the numerical approximation of the topography by considering the two different possibilities already described.

Exner equation in the projection step. It is evident that in this case we simply have $z^{n+} = z^n$ as the Exner equation is completely taken into account in the transport system. As such, for the stability of the numerical scheme, we ask for the following CFL condition for the time step,

$$\Delta t \leq \text{CFL}_l \frac{\Delta x}{\max\limits_j \{\max(\tau_j^n \theta_j^n, \tau_{j+1}^n \theta_{j+1}^n) \max(a_{j+\frac{1}{2}})\}} \tag{5.4.6}$$

with CFL_l a constant value. Clearly, we do not need anymore condition (5.4.6) if we use the implicit version of the numerical approximation.

Exner equation in both steps. Using the approximate Riemann solver presented in section 5.3.1.1, here we state

$$L_j^{n+} z_j^{n+} = L_j^n z_j^n - \frac{\Delta t}{\Delta x}((\Omega - zu)^*_{j+\frac{1}{2}} - (\Omega - zu)^*_{j-\frac{1}{2}}). \tag{5.4.7}$$

In this case, the CFL condition on the time step for the explicit approximation is given by

$$\Delta t \leq \text{CFL}_l \frac{\Delta x}{\max\limits_j \{\max(\tau_j^n \theta_j^n, \tau_{j+1}^n \theta_{j+1}^n) \max(a_{j+\frac{1}{2}}, (|u|b)_{j+\frac{1}{2}})\}}. \tag{5.4.8}$$

Let us remark that we could envisage to compute implicitly the variables Lz and $L\Omega$ as well without an excessive computational cost. Indeed, once we have found u^{n+}, Π^{n+} and τ^{n+} by solving the implicit linear system given by formula (5.4.5), the star values $z^{*,n+}$ and $\Omega^{*,n+}$ lead to another linear system thanks to the fact that u^{n+}, Π^{n+} and τ^{n+} are now fixed values.

5.4.2 Projection step

In this section we present the numerical approximation for the projection step. Since it has already been presented in different papers for other systems, here we give few details about it (see for instance [44, 18, 27]).

Aiming to project the Lagrangian solution into Eulerian coordinates, we consider the following identity

$$\int_{\xi_1}^{\xi_2} L(\xi, t) X(\xi, t) d\xi = \int_{x(\xi_1,t)}^{x(\xi_2,t)} X(x, t) dx$$

with $X = h, h\rho, h\rho u$ and, if needed, $X = z$. Then, we define $\hat{\xi}_{j+1/2}$ such that $x(\hat{\xi}_{j+1/2}, t^{n+1}) = x_{j+1/2}$ and $x(\hat{\xi}_{j+1/2}, t^n) = \hat{\xi}_{j+1/2}$ for all j. Subsequently, we can start by writing

$$\begin{aligned} X_j^{n+1} &= \frac{1}{\Delta x} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} X(x, t^{n+1}) dx = \frac{1}{\Delta x} \int_{x(\hat{\xi}_{j-\frac{1}{2}}, t^{n+1})}^{x(\hat{\xi}_{j+\frac{1}{2}}, t^{n+1})} X(x, t^{n+1}) dx \\ &= \frac{1}{\Delta x} \int_{\hat{\xi}_{j-\frac{1}{2}}}^{\hat{\xi}_{j+\frac{1}{2}}} L(\xi, t^{n+1-}) X(\xi, t^{n+1-}) d\xi \end{aligned}$$

and splitting the last integral into three parts

$$\frac{1}{\Delta x}\int_{\hat{\xi}_{j-\frac{1}{2}}}^{\hat{\xi}_{j+\frac{1}{2}}} L(\xi,t^{n+1-})X(\xi,t^{n+1-})d\xi = \frac{1}{\Delta x}\int_{\hat{\xi}_{j-\frac{1}{2}}}^{\xi_{j-\frac{1}{2}}} L(\xi,t^{n+1-})X(\xi,t^{n+1-})d\xi + \\ + \frac{1}{\Delta x}\int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} L(\xi,t^{n+1-})X(\xi,t^{n+1-})d\xi + \frac{1}{\Delta x}\int_{\xi_{j+\frac{1}{2}}}^{\hat{\xi}_{j+\frac{1}{2}}} L(\xi,t^{n+1-})X(\xi,t^{n+1-})d\xi. \tag{5.4.9}$$

Then, estimating $\hat{\xi}_{j+1/2}$ such that

$$x_{j+1/2} = x(\hat{\xi}_{j+1/2},t^{n+1}) \simeq x(\hat{\xi}_{j+1/2},t^n) + \Delta t\partial_t x(\hat{\xi}_{j+1/2},t^n) \simeq \hat{\xi}_{j+1/2} + \Delta t u^*_{j+1/2},$$

we approximate the last three integrals in (5.4.9) and obtain

$$X_j^{n+1-} = (LX)_j^{n+} - \frac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}(LX)^{n+}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}(LX)^{n+}_{j-\frac{1}{2}}\big) \tag{5.4.10}$$

where for all j

$$(LX)^{n+}_{j+1/2} = \begin{cases} (LX)^{n+}_j & \text{if} \quad u^*_{j+1/2} \geq 0 \\ (LX)^{n+}_{j+1} & \text{if} \quad u^*_{j+1/2} < 0. \end{cases}$$

For the projection step, the CFL condition on the time step is the following,

$$\Delta t \leq \mathrm{CFL}_t \frac{\Delta x}{\max\limits_j\{u^+_{j-\frac{1}{2}} - u^-_{j+\frac{1}{2}}\}} \tag{5.4.11}$$

with CFL_t a constant value and

$$u^+_{j-\frac{1}{2}} = \max(u^*_{j-\frac{1}{2}},0) \quad \text{and} \quad u^-_{j+\frac{1}{2}} = \min(u^*_{j+\frac{1}{2}},0).$$

As for the final time step, we take the minimum between the Lagrangian and projection ones. Then, we only need to specify the numerical approximation for the topography.

Exner equation in the transport step. Referring to the previous work [17] (chapter 4), for the numerical approximation of the Exner equation we state

$$z_j^{n+1-} = z_j^n - \zeta\frac{\Delta t}{\Delta x}\Big(u^*_{j+\frac{1}{2}}\Big(\frac{q_b}{u}\Big)^{n+}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}\Big(\frac{q_b}{u}\Big)^{n+}_{j-\frac{1}{2}}\Big) \tag{5.4.12}$$

where

$$(\frac{q_b}{u})^{n+}_{j+1/2} = \begin{cases} (\frac{q_b}{u})\Big((Lu)^{n+}_{j+1}\Big) & \text{if} \quad u^*_{j+\frac{1}{2}} \leq 0 \\ \\ (\frac{q_b}{u})\Big((Lu)^{n+}_{j}\Big) & \text{if} \quad u^*_{j+\frac{1}{2}} > 0. \end{cases}$$

Exner equation in both steps. In this case the numerical approximation for z is similar to the one of the other variables with the only difference that, in the numerical flux, we use the values z_j^{n+} instead of its Lagrangian counterpart $(Lz)_j^{n+}$. Namely, we state

$$z_j^{n+1-} = (Lz)_j^{n+} - \frac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}z^{n+}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}z^{n+}_{j-\frac{1}{2}}\big). \tag{5.4.13}$$

This choice is actually related to the second-order extension of this scheme, for which we refer to [18] (chapter 4).

For the reader's sake, let us summarize which formulas we use for each scheme:

- Explicit scheme with topography only updated at the end of the transport step: formulas (5.4.4), (5.4.10), (5.4.12).
- Implicit-explicit scheme with topography only updated at the end of the transport step: formulas (5.4.4), (5.4.5), (5.4.10), (5.4.12).
- Explicit scheme with topography updated in both the acoustic and transport step: formulas (5.4.4), (5.4.7), (5.4.10), (5.4.13).

Finally, let us sum up the properties of the numerical schemes we presented so far in the following theorem.

Theorem 3. *Consider the three LP numerical schemes we presented so far, namely the explicit method with Exner equation in the transport step (formulas (5.4.4), (5.4.10), (5.4.12)), its implicit-explicit version (formulas (5.4.4) with (5.4.5), (5.4.10), (5.4.12)) and finally the scheme with Exner equation in both steps (formulas (5.4.4) with (5.4.7), (5.4.10), (5.4.13)).*

1. *The above-mentioned numerical schemes are well-balanced in the following sense. They are able to exactly preserve the stationary solutions (5.2.4)-(5.2.5). Whereas, concerning the steady state (5.2.2), the methods are well-balanced with order two: they preserve its second-order discretization (5.2.15).*
2. *Under the CFL condition (5.4.11) with $\text{CFL}_t = 0.5$, the above-mentioned numerical schemes preserve the positivity of the water height h.*
3. *Under the CFL condition (5.4.11) with $\text{CFL}_t = 0.5$, the assumption of the positivity of the water height (point b) and of $h_j^{n+1-} \neq 0 \ \forall j$, if $\rho_w \leq \rho_j^n \leq \rho_s \ \forall j$, then it follows that $\rho_w \leq \rho_j^{n+1-} \leq \rho_s \ \forall j$. This statement is true for any of the above-mentioned numerical schemes.*

Proof.

1. The well-balanced property comes down from the definition of the approximate Riemann solver presented in section 5.2.2 (section 5.3.1.1 with the Exner equation). For this reason we do not insert further details.
2. Assume $h_j^n \geq 0 \ \forall j$. It is straightforward to see that $L_j^{n+} h_j^{n+} \geq 0 \ \forall j$ from discretization (5.4.1). Furthermore, given $h_j^n > 0$ we can also show that $h_j^{n+} > 0$ under the CFL condition (5.4.11) with $\text{CFL}_t = 1$.

   Then, considering the transport approximation (5.4.10) with $X = h$, we have

$$\begin{aligned} h_j^{n+1-} &= L_j^{n+} h_j^{n+} - \frac{\Delta t}{\Delta x}(u^*_{j+\frac{1}{2}}(Lh)^{n+}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}(Lh)^{n+}_{j-\frac{1}{2}}) \\ &= L_j^{n+} h_j^{n+} - \frac{\Delta t}{\Delta x}((u^*_{j+\frac{1}{2}})^+(Lh)^{n+}_j + (u^*_{j+\frac{1}{2}})^-(Lh)^{n+}_{j+1} \\ &\qquad - (u^*_{j-\frac{1}{2}})^+(Lh)^{n+}_{j-1} - (u^*_{j-\frac{1}{2}})^-(Lh)^{n+}_j) \\ &= L_j^{n+} h_j^{n+}(1 - \frac{\Delta t}{\Delta x}((u^*_{j+\frac{1}{2}})^+ - (u^*_{j-\frac{1}{2}})^-)) \\ &\qquad - \frac{\Delta t}{\Delta x}((u^*_{j+\frac{1}{2}})^-(Lh)^{n+}_{j+1} - (u^*_{j-\frac{1}{2}})^+(Lh)^{n+}_{j-1}). \end{aligned}$$

   Then, it is clear that

$$-\frac{\Delta t}{\Delta x}((u^*_{j+\frac{1}{2}})^-(Lh)^{n+}_{j+1} - (u^*_{j-\frac{1}{2}})^+(Lh)^{n+}_{j-1}) \geq 0.$$

Hence, if

$$\frac{\Delta t}{\Delta x}((u^*_{j+\frac{1}{2}})^+ - u^*_{j-\frac{1}{2}})^-) \leq 1, \tag{5.4.14}$$

we have $h_j^{n+1-} \geq 0$. Finally, let us observe that condition (5.4.11) with $\text{CFL}_t = 0.5$ implies condition (5.4.14).

3. Let us assume that $\rho_w \leq \rho_j^n \leq \rho_s \; \forall j$. To prove that $\rho_w \leq \rho_j^{n+1-} \leq \rho_s \; \forall j$, we start by showing that $\rho_w \leq \rho_j^{n+1-} \; \forall j$ and then that $\rho_j^{n+1-} \leq \rho_s \; \forall j$. We also assume $h_j^n \geq 0$ and $h_j^{n+1-} > 0$.

   First of all, we define the new variable $\tilde{\rho} = \rho - \rho_w$. As such, it is clear that its evolution equation is given by $\partial_t(h\tilde{\rho}) + \partial_x(h\tilde{\rho}u) = 0$ and its approximation reads

$$\begin{aligned}(h\tilde{\rho})_j^{n+1-} &= (Lh\tilde{\rho})_j^{n+} - \frac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}(Lh\rho)^{n+}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}(Lh\rho)^{n+}_{j-\frac{1}{2}}\big) \\ &\quad + \rho_w\frac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}(Lh)^{n+}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}(Lh)^{n+}_{j-\frac{1}{2}}\big) \\ &= (Lh\tilde{\rho})_j^{n+} - \frac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}(Lh\tilde{\rho})^{n+}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}(Lh\tilde{\rho})^{n+}_{j-\frac{1}{2}}\big).\end{aligned}$$

   Now, we aim to prove that $\tilde{\rho}_j^{n+1-} \geq 0$, namely that $\rho_j^{n+1-} \geq \rho_w$. Hence, we use the same strategy as in point (b). We have

$$\begin{aligned}(h\tilde{\rho})_j^{n+1-} &= (Lh\tilde{\rho})_j^{n+} - \frac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}(Lh\tilde{\rho})^{n+}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}(Lh\tilde{\rho})^{n+}_{j-\frac{1}{2}}\big) = \\ &= (Lh\tilde{\rho})_j^{n+} - \frac{\Delta t}{\Delta x}\big((u^*_{j+\frac{1}{2}})^+(Lh\tilde{\rho})_j^{n+} + (u^*_{j+\frac{1}{2}})^-(Lh\tilde{\rho})_{j+1}^{n+} \\ &\qquad - (u^*_{j-\frac{1}{2}})^+(Lh\tilde{\rho})_{j-1}^{n+} - (u^*_{j-\frac{1}{2}})^-(Lh)_j^{n+}\big) \\ &= (Lh\tilde{\rho})_j^{n+}(1 - \frac{\Delta t}{\Delta x}((u^*_{j+\frac{1}{2}})^+ - (u^*_{j-\frac{1}{2}})^-)) \\ &\qquad - \frac{\Delta t}{\Delta x}((u^*_{j+\frac{1}{2}})^-(Lh\tilde{\rho})_{j+1}^{n+} - (u^*_{j-\frac{1}{2}})^+(Lh\tilde{\rho})_{j-1}^{n+}),\end{aligned}$$

   and then, as in point (b), it follows that $(h\tilde{\rho})_j^{n+1-} \geq 0$ under condition (5.4.11) with $\text{CFL}_t = 0.5$. Namely, we found $\rho_j^{n+1-} \geq \rho_w$ as we already know that $h_j^{n+1-} > 0$.

   As a second step, we want to prove that $\rho_j^{n+1-} \leq \rho_s$. This time we define the variable $\tilde{\rho} = \rho_s - \rho$, whose approximation reads

$$(h\tilde{\rho})_j^{n+1-} = (Lh\tilde{\rho})_j^{n+} - \frac{\Delta t}{\Delta x}\big(u^*_{j+\frac{1}{2}}(Lh\tilde{\rho})^{n+}_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}(Lh\tilde{\rho})^{n+}_{j-\frac{1}{2}}\big)$$

   once again. Then, we just follow the same strategy as before to prove that $\tilde{\rho}_j^{n+1-} \geq 0$ and thus $\rho_s \geq \rho_j^{n+1-}$.

□

5.4.3 Including the source terms for friction and erosion-deposition fluxes

In the last step, we include the erosion, deposition and friction terms in the mathematical model. Following the lines of [38], we aim to exploit a semi-implicit approximation for the

source terms. Hence, considering first the variables $h\rho$ and z, their updating formulas read

$$\begin{cases} h_j^{n+1}\rho_j^{n+1} = h_j^{n+1-}\rho_j^{n+1-} + \Delta t\zeta\rho_z\Big((1-\Psi)v_s E_{s,j}^{n+1-}\dfrac{z_j^{n+1}}{z_j^{n+1-}} - v_s c_{z,j}^{n+1-}\dfrac{h_j^{n+1}\rho_j^{n+1}}{h_j^{n+1-}\rho_j^{n+1-}}\Big) \\ z_j^{n+1} = z_j^{n+1-} - \Delta t\zeta\Big((1-\Psi)v_s E_{s,j}^{n+1-}\dfrac{z_j^{n+1}}{z_j^{n+1-}} - v_s c_{z,j}^{n+1-}\dfrac{h_j^{n+1}\rho_j^{n+1}}{h_j^{n+1-}\rho_j^{n+1-}}\Big) \end{cases} \tag{5.4.15}$$

where we recall that

$$\phi_z = F_e - F_d, \quad F_e = (1-\Psi)v_s E_s, \quad F_d = v_s c_z.$$

Then, the solution of this linear system (5.4.15) can be explicitly found, we just refer to [38] for more details about it. Subsequently, as we have found $h_j^{n+1}\rho_j^{n+1}$ and z_j^{n+1}, we can automatically define ϕ_z as

$$\phi_z = \frac{1-\Psi}{\rho_z\Delta t}(h_j^{n+1}\rho_j^{n+1} - h_j^{n+1-}\rho_j^{n+1-})$$

and thus update h_j^{n+1} and $(h\rho u)_j^{n+1}$, for which the friction term is treated semi-implicitly as well, see [38] for more details.

Remark 8. *In paper [38], it has been proved that if $h\rho_j^{n+1-}$ and z_j^{n+1-} are positive, then $h\rho_j^{n+1}$ and z_j^{n+1} will remain positive using this approximation.*

Let us also observe that asking for the positivity of the topography is not a restrictive condition. Indeed, from a numerical point of view, we can simply consider its translation $\tilde{z} = z + K$ with K a constant so that $\tilde{z}$ is always positive. At the end of the simulation, we can obtain z by imposing $z = \tilde{z} - K$. Such an expedient does not modify the final solution.

5.5 Numerical simulations

This section is devoted to the presentation of the numerical results. Here we consider the three different numerical methods we have presented so far. For the sake of brevity, we call them as follows:

- "LP-TrZ" if the topography is updated at the end of the transport step and the approximation of the acoustic step is explicit;
- "LP-TrZ-Imp" if the topography is updated at the end of the transport step and the approximation of the acoustic step is implicit;
- "LP-AcTrZ" if the topography is updated both in the acoustic and transport steps.

If not otherwise specified, in the numerical simulations we take $\text{CFL}_l = 0.45$ and $\text{CFL}_t = 0.99$. Finally, we point out that for all the numerical simulations, we exploited MATLAB language with a single Intel Core i7 CPU.

5.5.1 Lake at rest solution with suspended sediment

In [38], it has been presented the following numerical test which is useful to check if our numerical scheme produces indeed the physical solution described in section 5.1. We consider a closed channel of length $L = 2\ m$, where the water is still $u = 0$, the free surface is constant in space ($h + z = 3\ m$) as well as the bed height $z = 1.05\ m$. Then, the sediment diameter is given by $d = 3.9\ mm$ and the sediment concentration is $c(x,t) = 0.2$. Finally, we also

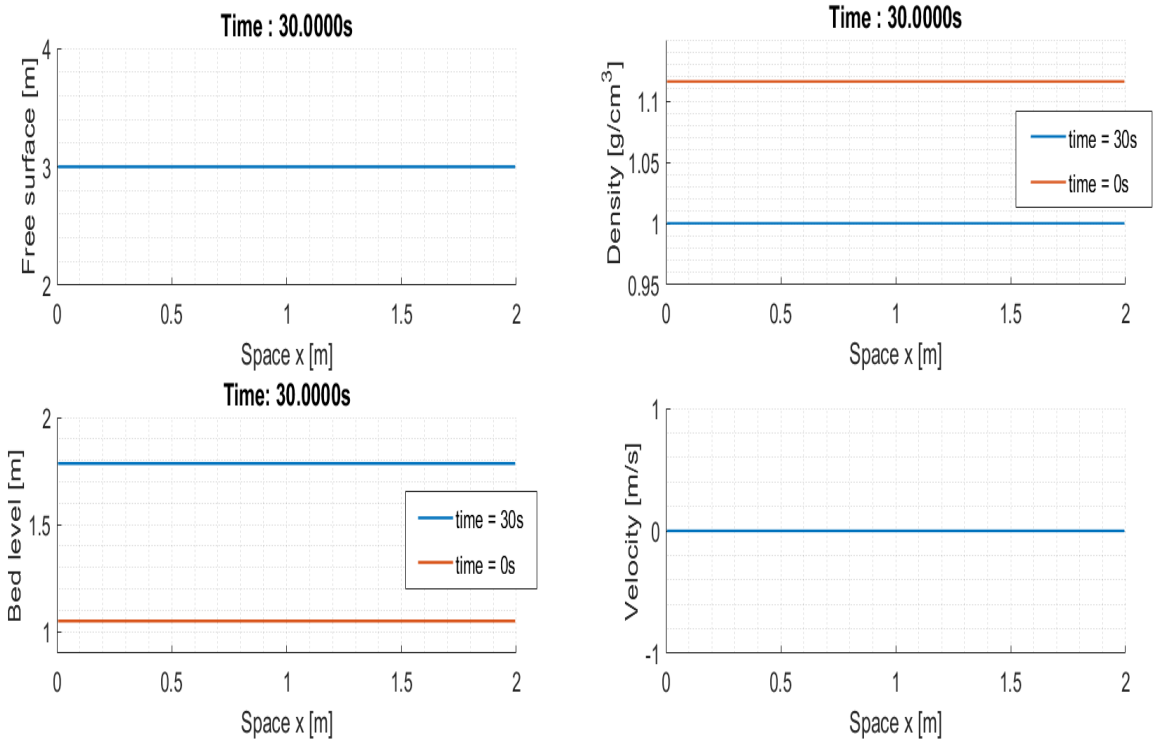


Figure 5.5.1: Lake at rest solution with suspended sediment in section 5.5.1: $M = 200$ cells, $t_{end} = 30s$ (blue line), $t_0 = 0s$ (red line). Free surface and bed level (left), density and velocity (right).

impose the following parameters values, $\rho_s = 1.580$, $\Psi = 0.47$, $\theta_c^* = 0.045$ and $\mu_f = 0.03$. The solution is computed at final time $t_{End} = 30\ s$. As explained in section 5.1, we expect the bed level to increase due to sediment deposition and, consequently, the water height to decrease as the free surface should remain constant in time. Moreover, the velocity should remain null as the channel is closed and the pressure term is constant in the domain. Finally, since $c < 1 - \Psi$, the density of the mixture water sediment is expected to decrease. All these remarks are indeed verified by our numerical outputs, reported in figure 5.5.1. The solution is computed using the LP-AcTrZ method. We do not report here the results obtained with the other methods as they give analogous solution.

5.5.2 Turbidity currents

In this section we want to simulate how the bed elevation evolves in time when imposing some turbidity currents into a channel with clear water. Thus, as initial condition we take $h(x, t = 0) = 4$m, $\rho(x, t = 0) = 1$, $z(x, t = 0) = 1$m and $q(x, t = 0) = 0.001m^2/s$. Moreover, we consider the erosion-deposition source terms together with the friction one but we neglect the solid transport discharge q_z in the Exner equation. Then, we impose $\rho_s = 2.650$ and $r_w = 2.5$ while, for the other parameters values, we use the ones of the previous section 5.5.1. Then, for the left boundary condition we do the following,

$$
\begin{aligned}
q(x=0,t) &= \begin{cases} \frac{3}{2}\left(1+\sin\left(\frac{\pi t}{2}\right)\right) & \text{if} \quad (t \le 20s \ \vee \ 60s \le t \le 90s) \\ q_1(t) & \text{otherwise} \end{cases} \\
\rho(x=0,t) &= \begin{cases} 1+0.2\max\left(\sin\left(\frac{\pi t}{2}\right),0\right) & \text{if} \quad (t \le 20s \ \vee \ 60s \le t \le 90s) \\ \rho_1(t) & \text{otherwise} \end{cases}
\end{aligned}
\tag{5.5.1}
$$

where the index 1 indicates the value of the variable in the first cell of the mesh. Otherwise, we ask for transmissive boundary conditions. Then, in Figure 5.5.2, we insert the bed elevation and density outputs at times $t = 30s$, $t = 50s$, $t = 100s$ and $t = 120s$ using both explicit and implicit LP-TrZ schemes. Notice that, for the latter, we used a time step based only on the transport CFL condition (5.4.11). Hence, in this case the implicit time step is at least 14 times larger than the explicit time step, where the latter is about $\Delta t \approx 0.0026$. Notice that we also inserted a reference solution computed with $M = 1024$ cells and the explicit LP-TrZ scheme.

The two schemes give similar results even if some differences can be observed, mainly in the density outputs. This is probably due to the different time steps used for the two simulations. Indeed, differences are reduced if we consider the same time step for the two schemes. Moreover, we verified that the two schemes give analogous solutions when refining the mesh. We stress, anyway, that in this type of test we are mainly interested in understanding how the topography evolves over long periods of time rather than density, whose values change several times due to boundary conditions. We can observe that the two schemes give similar pattern for the topography outputs. Moreover, even if the implicit scheme overestimate the topography values, it seems to capture the waveform better, in the sense that the fluctuations are more damped when using the explicit scheme.

Then, to better highlight how the bed elevation changes in time, we insert the outputs at different times in the same figure for the two schemes, see Figure 5.5.3. In particular, we observe that, when inserting the flow with sediment in the channel, the bed elevation increases. On the other hand, when no sediments are imposed, the bed elevation decreases while the density of sediment augments. This is probably due to the fact that here erosion is greater than deposition and, as such, there is an increase of suspended sediments in the water.

Finally, it is interesting to show the errors in norm $\mathbf{L}^1$ and the computational times for both the explicit and implicit schemes, see Table 5.5.1. We can conclude that, even if errors are greater when using the implicit scheme, it allows much faster simulation. This explains why the implicit version of the method is useful.

	Error $\mathbf{L}^1$ of z		Error $\mathbf{L}^1$ of ρ		CPU [s]	
Time	LP-TrZ	LP-TrZ-Imp	LP-TrZ	LP-TrZ-Imp	LP-TrZ	LP-TrZ-Imp
30s	0.1669	0.5107	0.0216	0.0988	72.4549	4.1708
50s	0.1175	1.4431	0.0103	0.2778	124.0635	6.1695
100s	1.4623	1.6950	0.2042	0.1127	245.2995	15.2274
120s	1.6630	3.4737	0.1265	0.3918	287.0587	17.9937

Table 5.5.1: CPU time in seconds and errors in norm $\mathbf{L}^1$ for the variables z and ρ computed at times $t = 30s$, $t = 50s$, $t = 100s$ and $t = 120s$ using LP-TrZ and LP-TrZ-Imp schemes.

5.5.3 Dune evolution test case

For this test case we refer to [4] and we use the following parameters values: $L = 1000$m, $\zeta = \dfrac{1}{1-0.47}$, Grass formula with $A_g = 1$ and $\text{CFL}_t = 0.5$. Then, the Initial Conditions (IC) are given by

$$z_{\text{IC}} = \begin{cases} 0.1 + \left(\sin\left(\dfrac{\Pi(x-300)}{200}\right)\right)^2 & \text{if } 300 \le x \le 500 \\ 0.1 & \text{otherwise} \end{cases},$$

$h_{\text{IC}} = 10 - z_{\text{IC}}$ and $q_{\text{IC}} = 10$. We show the results in figure 5.5.4, obtained using the LP-TrZ and LP-TrZ-Imp methods. Here the reference solution is computed with $M = 1024$ cells and

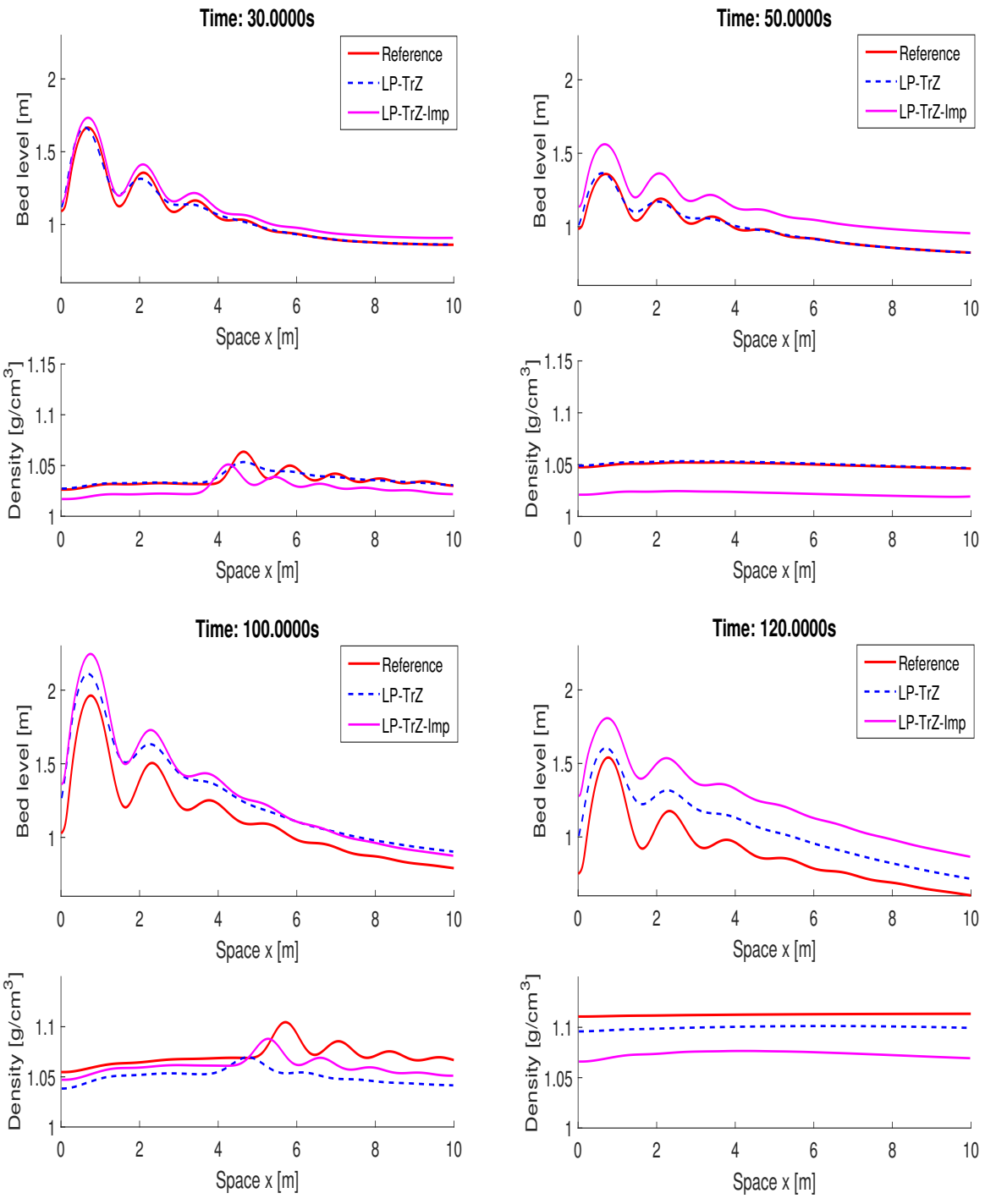
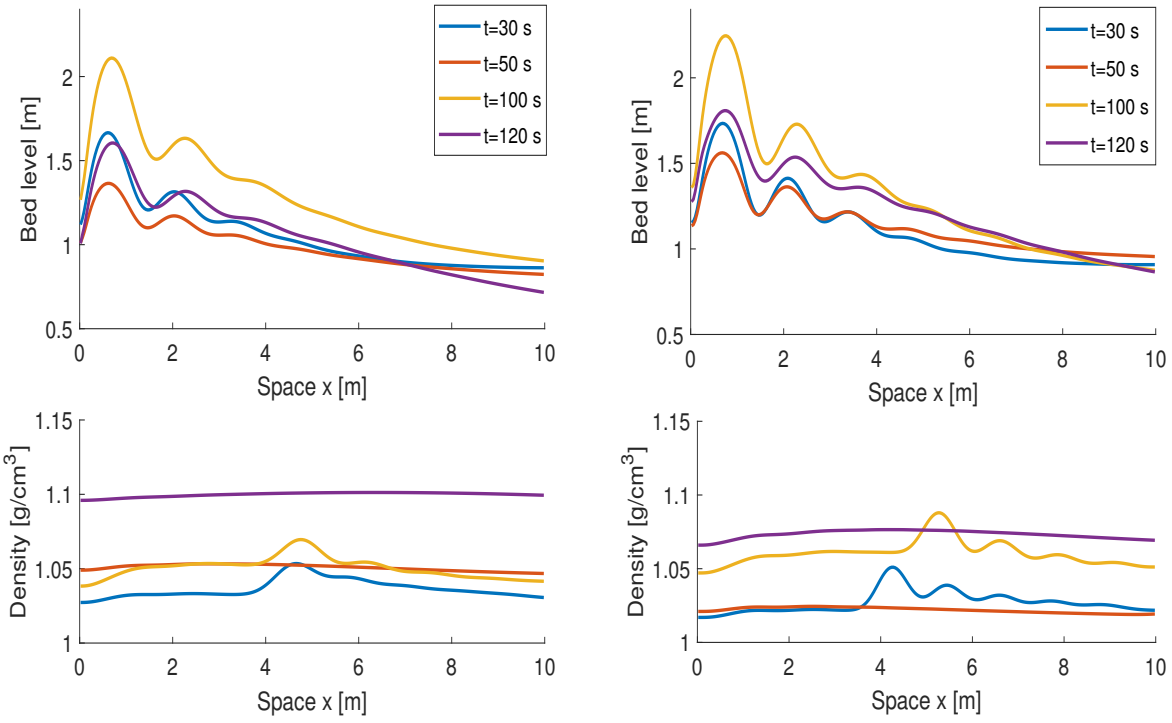


Figure 5.5.2: Turbidity test in section 5.5.2: bed level and density computed with explicit (blue dashed line) and implicit (magenta line) methods with $M = 256$ cells. Reference solution (red line) computed with the explicit scheme and $M = 1024$ cells. Solutions at times $t = 30s$ (up - left), $t = 50s$ (up - right), $t = 100s$ (bottom - left) and $t = 120s$ (bottom - right).



(a) LP-TrZ scheme

(b) LP-TrZ-Imp scheme

Figure 5.5.3: Turbidity test in section 5.5.2: bed level (up) and density (bottom) computed with explicit (left) and implicit (right) methods and $M = 256$ cells. Solution at times $t = 30s$ (blue), $t = 50s$ (red), $t = 100s$ (yellow) and $t = 120s$ (purple).

an implicit-explicit second-order scheme for which we refer to [19]. The numerical solution seems to be in agreement with the one reported in [4]. Let us point out that here we used the implicit-explicit version of the scheme as the dune evolution problem is in general a slow test case. Indeed, here the regime is already sub-critical with a Froude number Fr ≈ 0.1. Here we used a time step of about $\Delta t \approx 2.3s$ for the implicit-explicit scheme, whereas for the explicit version of the method, the time step is limited to $\Delta t \approx 0.23s$ approximately. For the sake of completeness, in tables 5.5.2-5.5.3 we show the errors in norm $\mathbf{L}^1$ and the computational times computed with LP-TrZ and LP-TrZ-Imp methods. For the variable z, we also insert the comparison between the CPU and the errors in log scale in Figure 5.5.5. We highlight that, for this test, we considered a long channel ($L = 1000$m) with relatively coarse meshes and large space steps Δx. As a consequence, also the errors appear to be big, which is normal as we assume them to be of the same order of Δx. Then, we observe that the errors of the explicit scheme are a bit smaller than the implicit ones when using the same mesh values. This is expected as the implicit method is more diffusive than its explicit version. On the other hand, it is immediately evident that the LP-TrZ-Imp scheme is much faster than the explicit one. Therefore, it is clear that the LP-TrZ method is more accurate, but the the difference on CPU time when compared to the implicit scheme makes the LP-TrZ-Imp method a very good alternative for long time simulations. This fact is made even clearer by figure 5.5.5, where we observe that the line for the LP-TrZ-Imp scheme is always under the LP-TrZ one, confirming our previous statement.

Moreover, here we have taken $A_g = 1$, meaning that the interaction between the fluid and the sediments of the bottom is strong. In this way we are able to see in a relative short time substantial changes in the bed elevation. However, if we consider a weaker interaction, for instance $A_g < 0.1$, the evolution of the bottom height is slower and we need a greater ending

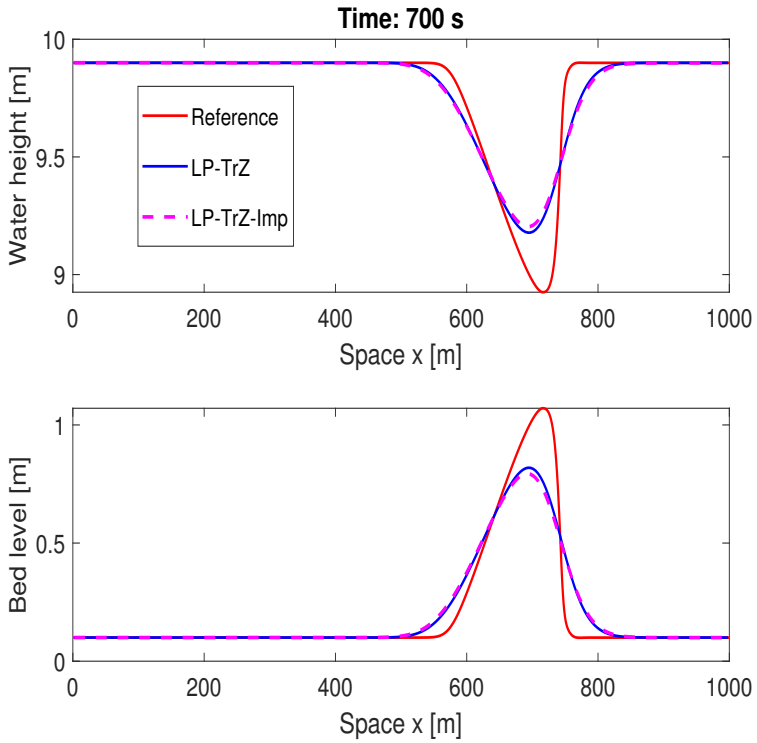


Figure 5.5.4: Dune evolution test case in section 5.5.3, water height (up) and bed level (bottom). Ending time $t_{end} = 700s$. LP-TrZ (blue line) and LP-TrZ-Imp (magenta dashed line) with $M = 256$ cells and reference solution with $M = 1024$ cells (red line).

time to be able to see significant changes in z, hence the usefulness of the implicit-explicit version of the method. Furthermore, we highlight that, if we are in a low-Froude number regime (Fr $< 10^{-2}$), the CFL condition for the acoustic time step is even more limited and difference in value between the acoustic and transport time steps could be more remarkable.

Finally, for the sake of completeness, in table 5.5.4 we insert the empirical convergence rates for the LP-TrZ and LP-TrZ-Imp schemes in order to show that their order of accuracy tends to one.

		Error $\mathbf{L}^1$ of h		Error $\mathbf{L}^1$ of hu		Error $\mathbf{L}^1$ of z	
Mesh	Δx	LP-TrZ	LP-TrZ-Imp	LP-TrZ	LP-TrZ-Imp	LP-TrZ	LP-TrZ-Imp
64	15.625	74.7145	80.0825	30.7352	36.3154	74.2672	79.8077
128	7.8125	52.8432	57.6222	21.6851	26.6577	52.6857	57.4422
256	3.90625	33.3327	37.0244	13.8278	18.3581	33.2159	36.8888
512	1.953125	19.0688	21.6013	8.0738	11.1259	18.9822	21.4826

Table 5.5.2: Errors in norm $\mathbf{L}^1$ of h, hu, z and computational cost using LP-TrZ and LP-TrZ-Imp methods. Mesh of size $M = (64, 128, 256, 512)$ and space step $\Delta x = (15.625, 7.8125, 3.90625, 1.953125)$.

5.5.4 Dam break problems

Here we present two different dam break problems for which the experimental results are available, see also [38]. We consider the MPM formulation (5.1.3) for q_z and describe the parameters values in table 5.5.5. We also take $M = 500$ cells. Let us remark that for these dam break problems, we need to use a smaller time step value for the LP-TrZ method in order to

	CPU [s]	
Mesh	LP-TrZ	LP-TrZ-Imp
64	1.2478	0.1833
128	4.2625	0.4024
256	15.4166	1.9843
512	61.2181	9.5504

Table 5.5.3: Errors in norm $\mathbf{L}^1$ of h, hu, z and computational cost using LP-TrZ and LP-TrZ-Imp methods. Mesh of size $M = (64, 128, 256, 512)$ and space step $\Delta x = (15.625, 7.8125, 3.90625, 1.953125)$.

	$\mathcal{O}(\mathbf{L}^1)$ of h		$\mathcal{O}(\mathbf{L}^1)$ of hu		$\mathcal{O}(\mathbf{L}^1)$ of z	
Mesh	LP-TrZ	LP-TrZ-Imp	LP-TrZ	LP-TrZ-Imp	LP-TrZ	LP-TrZ-Imp
64	–	–	–	–	–	–
128	0.4997	0.4749	0.5032	0.4460	0.4953	0.4744
256	0.6648	0.6381	0.6491	0.5381	0.6655	0.6389
512	0.8057	0.7774	0.7763	0.7225	0.8072	0.7800

Table 5.5.4: Empirical convergence rates in norm $\mathbf{L}^1$ of h, hu, z using LP-TrZ and LP-TrZ-Imp methods. Mesh of size $M = (64, 128, 256, 512)$.

avoid unphysical oscillations.

Experiment 1. This experiment has been carried on at the Université catholique de Louvain. We report the LP-AcTrZ and LP-TrZ results for this first dam break experiment in figure 5.5.6. In particular, on the left we insert the LP-AcTrZ solution against the experimental results, while we compare the two LP-AcTrZ and LP-TrZ methods on the right. We considered three different ending times, namely $t_{end} = 5t_0, 7.5t_0, 10t_0\ s$ with $t_0 = \sqrt{gh_0} \approx 0.101$. The numerical outputs seem to be in agreement with the ones reported in [38] and they are also close to the experimental solution. We can also observe that the two LP-AcTrZ and LP-TrZ methods give similar results, even if the LP-TrZ method is less diffusive, see again Figure 5.5.6. Finally, we do not report here the implicit LP-TrZ outputs as they are analogous to the ones of the explicit version of the scheme.

Experiment 2. In figure 5.5.7, once again we show the LP-AcTrZ and LP-TrZ outputs for the second dam break problem. Analogous comments to what we have made for the first experiment may be said: our numerical solution correctly describe the solution in the sense that it is close to the experimental one. The position of the front is correct as well. Finally, the LP-TrZ is slightly less diffusive than the LP-AcTrZ one but it may require the use of a smaller CFL number to avoid possible spurious oscillations.

Remark 9. *Let us observe that for these two dam break problems, the regime is supercritical. As such, to exploit the implicit version of the method does not lead to an actual improvement of the numerical simulation in the sense that CPU times are not necessarily better. On the other hand,*

Experiment	$h_L\ [m]$	$h_R\ [m]$	$L\ [m]$	$d\ [mm]$	Ψ	$\rho_s\ [g\ cm^{-3}]$	$\mu_f\ [\mathrm{s\ m}^{-1/3}]$	θ_c^*	r_w
1	0.1	0	2.5	3.2	0.4	1.540	0.02	0.045	1
2	0.35	0	6	3.9	0.47	1.580	0.0165	0.047	2.5

Table 5.5.5: Parameters values for dam break problems in section 5.5.4, two different experiments.

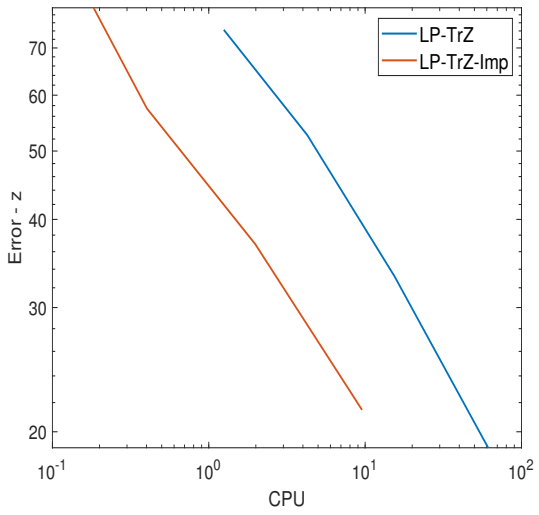


Figure 5.5.5: Dune evolution test case in section 5.5.3: CPU against error in norm $\mathbf{L}^1$ for the variable z (log scale). Mesh of size $M = (64, 128, 256, 512)$. LP-TrZ (blue line) and LP-TrZ-Imp (red line) schemes.

while the first-order numerical scheme LP-TrZ does not produce spurious oscillations, its second (or higher) order extension could, see [18] (chapter 4). Hence the usefulness of the LP-AcTrZ method, which is more stable.

5.6 Conclusion and perspectives

In this chapter, we have presented both explicit and implicit-explicit well-balanced Lagrange-projection schemes applied to the shallow water system with moving topography and variable density of the mixture water-sediment. In particular, we assumed that sediments could move along the bottom (bed-load) or being finer fraction that could be carried in suspension into the water or be deposited on the bed.

Since sediment transport is generally a slow process, which requires long-in-time simulation in order to see appreciable changes in the bed elevation, implicit-explicit method can be very useful. We considered academic problems to show that the implicit-explicit version of the Lagrange-projection numerical scheme allows very fast simulations, especially when we are in subsonic or low-Froude number regimes. Indeed, the LP approach entails a decomposition of the acoustic and transport waves of the model, leading to the possibility of implicitly approximating only the acoustic equations. As a consequence, the CFL condition on the time step can be based only on the transport waves. Finally, we also provided comparison between experimental and numerical results.

We highlight that two explicit LP methods have been described, which differ only in the approximation of the bed elevation z, namely the LP-TrZ and the LP-AcTrZ schemes. Indeed, while the former is easier, the latter proved to be more stable in situations in which we could expect unphysical oscillations in the numerical outputs of fully decoupled methods.

Improvements could be related to the design and implementation of high order extension of implicit-explicit well-balanced Lagrange-projection schemes.

5.7 Bibliography

[1] E. Audusse, C. Berthon, C. Chalons, O. Delestre, N. Goutal, M. Jodeau, J. Sainte-Marie, J. Gieselmann, and G. Sadaka. *Sediment transport modelling : Relaxation schemes for Saint-Venant – Exner*

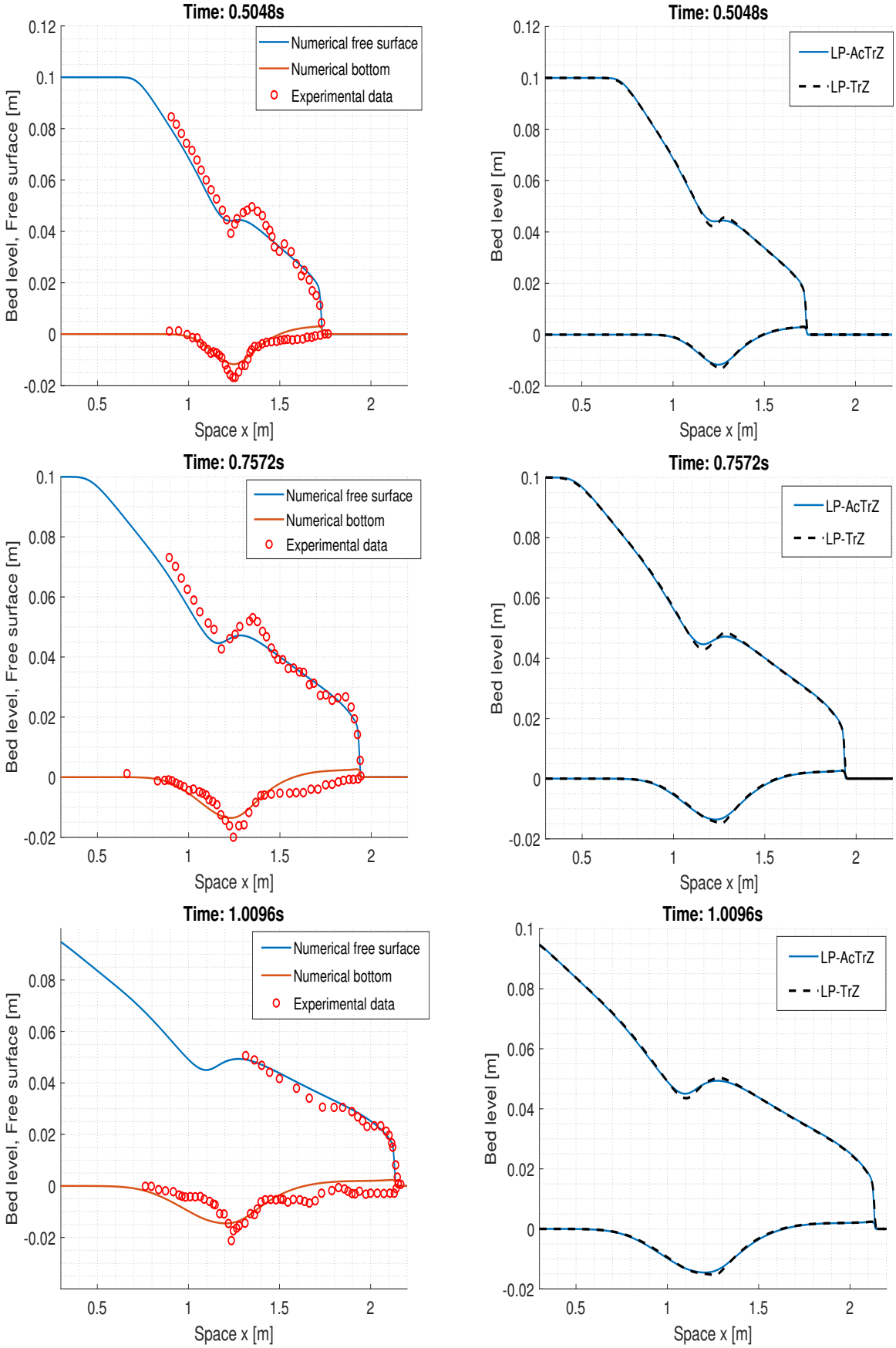


Figure 5.5.6: Dam break problem, experiment 1, Section 5.5.4: free surface and bed level, $M = 500$ cells. Ending time $t_{end} = 5t_0 s$ (up), $t_{end} = 7.5t_0 s$ (middle), $t_{end} = 10t_0 s$ (bottom). On the left, LP-AcTrZ free surface (blue line), LP-AcTrZ bottom (red line) and experimental values (red symbol). On the right, LP-AcTrZ (continuous blue line) and LP-TrZ (black dashed line) numerical methods.

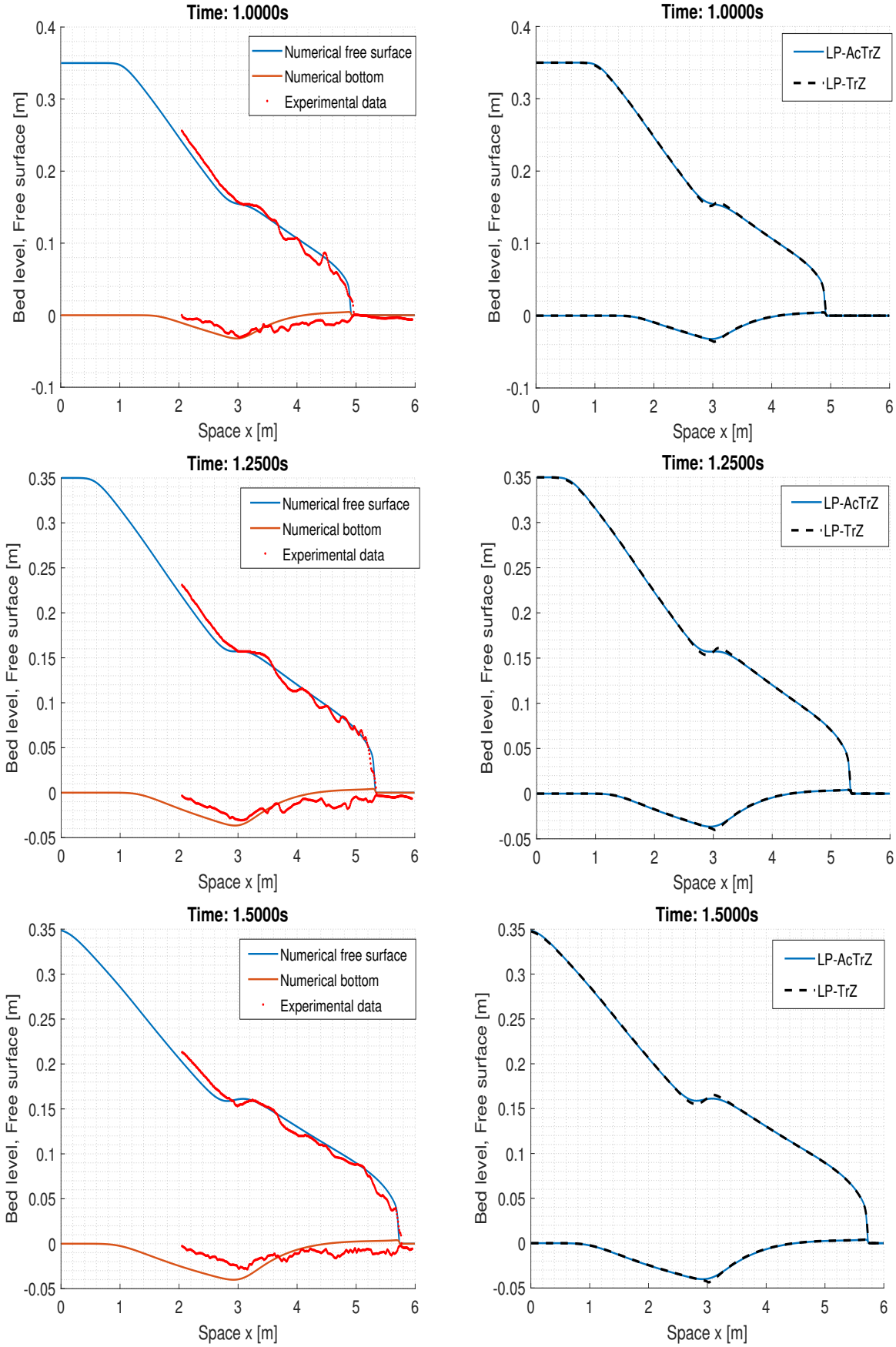


Figure 5.5.7: Dam break problem, experiment 2, Section 5.5.4: free surface and bed level, $M = 500$ cells. Ending time $t_{end} = 1s$ (up), $t_{end} = 1.25s$ (middle), $t_{end} = 1.5s$ (bottom). On the left, LP-AcTrZ free surface (blue line), LP-AcTrZ bottom (red line) and experimental values (red symbol). On the right, LP-AcTrZ (continuous blue line) and LP-TrZ (black dashed line) numerical methods.

and three layer models. ESAIM: Proceedings, 38: 78-98, 2012. 10.1051/proc/201238005

[2] 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: 2050-2065, 2004. 10.1137/S1064827503431090.

[3] E. Audusse, C. Chalons, and P. Ung. *A simple three-wave approximate riemann solver for the Saint-Venant-Exner equations*. International Journal for Numerical Methods in Fluids, 87(10): 508-528, 2018. 10.1002/fld.4500

[4] E. Audusse, O. Delestre, M. H. Le, M. Masson-Fauchier, P. Navaro, and R. Serra. *Parallelization of a relaxation scheme modelling the bedload transport of sediments in shallow water flow*. ESAIM: Proceedings, 43: 80-94, 2013. 10.1051/proc/201343005

[5] M. Baudin, C. Berthon, F. Coquel, R. Masson, Q. HuyTran. *A relaxation method for two-phase flow models with hydrodynamic closure law*. Numerische Mathematik. 99: 411-440, 2005. 10.1007/s00211-004-0558-1.

[6] C. Berthon, S. Cordier, O. Delestre, and M. H. Le. *An analytical solution of the shallow water system coupled to the exner equation*. Comptes Rendus Mathematique, 350(3-4): 183-186, 2012. 10.1016/j.crma.2012.01.007

[7] L. Bonaventura, E. D. Fernández-Nieto, J. Garres-Díaz, and G. Narbona-Reina. *Multilayer shallow water models with locally variable number of layers and semi-implicit time discretization*. Journal of Computational Physics, 364: 209-234, 2018. 10.1016/j.jcp.2018.03.017

[8] F. Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources*. Frontiers in mathematics, 2004. 10.1007/b93802.

[9] R. Briganti, N. Dodd, D. Kelly, and D. Pokrajac. *An efficient and flexible solver for the simulation of the morphodynamics of fast evolving flows on coarse sediment beaches*. International Journal for Numerical Methods in Fluids, 69(4): 859-877, 2012. 10.1002/fld.2618.

[10] V. Caleffi, A. Valiani, and A. Bernini. *High-order balanced cweno scheme for movable bed shallow water equations*. Advances in Water Resources, 30(4): 730-741, 2007. 10.1016/j.advwatres.2006.06.003

[11] M. J. Castro Díaz, C. Chalons and T. Morales De Luna. *A fully well-balanced Lagrange-Projection type scheme for the Shallow-water equations*. SIAM J. Numer. Anal., 56(5): 3071–3098, 2018. 10.1137/17M1156101.

[12] M. J. Castro-Díaz, E. D. Fernández-Nieto, and A. M Ferreiro. *Sediment transport models in shallow water equations and numerical approach by high order finite volume methods*. Computers & Fluids, 37(3): 299-316, 2008. 10.1016/j.compfluid.2007.07.017

[13] V. Casulli. *Semi-implicit finite difference methods for the two-dimensional shallow water equations*. Journal of Computational Physics, 86(1): 56-74, 1990. 10.1016/0021-9991(90)90091-E

[14] V. Casulli. *Numerical simulation of three-dimensional free surface flow in isopycnal co-ordinates*. International Journal for Numerical Methods in Fluids, 25(6): 645-658, 1997. 10.1002/(SICI)1097-0363(19970930)25:63.0.CO;2-L

[15] V. Casulli and E. Cattani. *Stability, accuracy and efficiency of a semi-implicit method for three-dimensional shallow water flow*. Computers & Mathematics with Applications, 27(4): 99-112, 1994. 10.1016/0898-1221(94)90059-0

[16] V. Casulli and P. Zanolli. *Semi-implicit numerical modeling of nonhydrostatic free-surface flows for environmental problems.* Mathematical and Computer Modelling, 36(9-10): 1131-1149, 2002. 10.1016/S0895-7177(02)00264-9

[17] C. Chalons, A. Del Grosso. *A second-order well-balanced Lagrange-projection scheme for Shallow Water Exner equations in 1D and 2D.* 2021. hal-03251707

[18] C. Chalons and A. Del Grosso. *Exploring different possibilities for second-order well-balanced Lagrange-projection numerical schemes applied to shallow water Exner equations.* International Journal for Numerical Methods in Fluids. 1- 31, 2022. 10.1002/fld.5064

[19] C. Chalons and A. Del Grosso. *Second-order well-balanced implicit-explicit scheme for systems of balance laws.* 2022.

[20] C. Chalons, M. Girardin, S. Kokh. *An All-Regime Lagrange-Projection like scheme for the gas dynamics equations on unstructured meshes.* Communications in Computational Physics. 20(1): 188-233 2016. 10.4208/cicp.260614.061115a.

[21] C. Chalons, M. Girardin, S. Kokh. *An all-regime Lagrange-Projection like scheme for 2D homogeneous models for two-phase flows on unstructured meshes.* Journal of Computational Physics, Elsevier, 335: 885-904, 2017. 10.1016/j.jcp.2017.01.017.

[22] C. Chalons, P. Kestener, S. Kokh, and M. Stauffert. *A large time-step and well-balanced Lagrange-Projection type scheme for the Shallow-water equations.* Communications in Mathematical Sciences. 15(3): 765–788, 2017. 10.4310/CMS.2017.v15.n3.a9.

[23] C. Chalons, S. Kokh, M. Girardin. *Large Time Step and Asymptotic Preserving Numerical Schemes for the Gas Dynamics Equations with Source Terms.* SIAM Journal on Scientific Computing. 35(6): A2874–A2902, 2013. 10.1137/130908671.

[24] F. Coquel, E. Godlewski, B. Perthame, A. In, and P. Rascle. *Some new Godunov and relaxation methods for two-phase flow problems.* In: Toro E.F. (eds) Godunov Methods. Springer, New York, NY, pages 179–188, 2001. 10.1007/978-1-4615-0663-8_18

[25] S. Cordier, M.H. Le, and T. Morales de Luna. *Bedload transport in shallow water models: Why splitting (may) fail, how hyperbolicity (can) help.* Advances in Water Resources, 34(8): 980-989, 2011. 10.1016/j.advwatres.2011.05.002

[26] L. Cozzolino, L. Cimorelli, C. Covelli, R. Della Morte, and D. Pianese. *Novel Numerical Approach for 1D Variable Density Shallow Flows over Uneven Rigid and Erodible Beds.* Journal of Hydraulic Engineering, 140(3): 254-268, 2014. 10.1061/(ASCE)HY.1943-7900.0000821.

[27] A. Del Grosso and C. Chalons. *Second-order well balanced Lagrange-Projection schemes for Blood Flow Equations.* Calcolo 58, 43, 2021. 10.1007/s10092-021-00434-5

[28] F. Duboc, C. Enaux, S. Jaouen, H. Jourdren, M. Wolff. *High-order dimensionally split Lagrange-remap schemes for compressible hydrodynamics.* Comptes Rendus Mathematique, 348(1–2):105-110, 2010. 10.1016/j.crma.2009.12.008.

[29] H. A Einstein. *The bed load function for sediment transportation in open channel flows.* Bulletin 1026, Washington, D.C., 1950.

[30] F. M Exner. *Über die wechselwirkung zwischen wasser und geschiebe in flüsen.* Akad. Wiss. Wien Math. Naturwiss, Klasse, 134: 165-2014, 1925.

[31] E. D. Fernández-Nieto, T. Morales de Luna, G. Narbona-Reina, and J. de Dieu Zabsonré. *Formal deduction of the Saint-Venant-Exner model including arbitrarily sloping sediment beds and associated energy.* ESAIM: Mathematical Modelling and Numerical Analysis, 51(1): 115–145, 2017. 10.1051/m2an/2016018

[32] G. Gallice. *Solveurs simples positifs et entropiques pour les systèmes hyperboliques avec terme source.* C. R. Math. Acad. Sci. Paris 334(8): 713-716, 2002. 10.1016/S1631-073X(02)02307-5.

[33] G. Gallice. *Positive and entropy stable Godunov-type schemes for gas dynamics and MHD equations in Lagrangian or Eulerian coordinates.* Numer. Math. 94(4): 673-713, 2003. 10.1007/s00211-002-0430-0

[34] J. Garres-Díaz and L. Bonaventura. *Flexible and efficient discretizations of multilayer models with variable density.* Applied Mathematics and Computation, 402: 126097, 2021. 10.1016/j.amc.2021.126097

[35] J. Garres-Díaz, E.D. Fernández-Nieto, and G. Narbona-Reina. *A semi-implicit approach for sediment transport models with gravitational effects.* Applied Mathematics and Computation. 421(206): 126938, 2022. 10.1016/j.amc.2022.126938.

[36] I. Gómez-Bueno, M. J. Castro, C. Parés, and G. Russo. *Collocation methods for high-order well-balanced methods for systems of balance laws.* Mathematics, 9(15), 2021. 10.3390/math9151799.

[37] C. Goñi, D. Celi, and F. Concha. *Determination of the volumetric solids fraction of saturated polydisperse ore tailing sediments.* Powder Technology, 305: 528-537, 2017. 10.1016/j.powtec.2016.10.001.

[38] J.C. González-Aguirre, M.J. Castro, and T. Morales de Luna. *A robust model for rapidly varying flows over movable bottom with suspended and bedload transport: Modelling and numerical approach.* Advances in Water Resources, 140: 103575, 2020. 10.1016/j.advwatres.2020.103575.

[39] A.J. Grass. *Sediment transport by waves and currents.* SERC London Cent. Mar. Technol, Report No. FL29, 1981.

[40] S. M. Khan, J. Imran, S. Bradford, and J. Syvitski. *Numerical modeling of hyperpycnal plume.* Marine Geology, 222-223: 193-211, 2005. 10.1016/j.margeo.2005.06.025.

[41] Y. Kubo. *Experimental and numerical study of topographic effects on deposition from two-dimensional, particle-driven density currents.* Sedimentary Geology, 164(3-4): 311-326, 2004. 10.1016/j.sedgeo.2003.11.002.

[42] Y. Kubo and T. Nakajima. *Laboratory experiments and numerical simulation of sediment-wave formation by turbidity currents.* Marine Geology, 192(1-3): 105-121, 2002. 10.1016/S0025-3227(02)00551-0.

[43] E. Meyer-Peter and R. Müller. *Formulas for bed-load transport.* In 2nd meeting IAHSR, Stockholm, Sweden, 1-26, 1948.

[44] T. Morales De Luna, M. J. Castro Díaz and C. Chalons. *High order fully well-balanced Lagrange-Projection scheme for Shallow-water.* Commun. Math. Sci., 18(3): 781–807, 2020. 10.4310/CMS.2020.v18.n3.a9

[45] T. Morales de Luna, M. J. Castro Díaz, C. Parés Madroñal, and E. D. Fernández Nieto. *On a shallow water model for the simulation of turbidity currents.* Communications in Computational Physics, 6(4): 848-882, 2009. 10.4208/cicp.2009.v6.p848

[46] J. Murillo and P. García-Navarro. *An exner-based coupled model for two-dimensional transient flow over erodible bed.* Journal of Computational Physics, 229(23): 8704-8732, 2010. 10.1016/j.jcp.2010.08.006.

[47] G. Parker, Y. Fukushima, and H. M. Pantin. *Self-accelerating turbidity currents.* Journal of Fluid Mechanics, 171: 145–181, 1986. 10.1017/S0022112086001404.

[48] D. Subhasish. *Fluvial Hydrodynamics, Hydrodynamic and Sediment Transport Phenomena.* GeoPlanet: Earth and Planetary Sciences. Springer-Verlag Berlin Heidelberg, 2014.

[49] I. Suliciu. *On the thermodynamics of fluids with relaxation and phase transitions. Fluids with relaxation.* Int. J. Engag. Sci. 36: 921-947, 1998.

[50] L. C Van Rijn. *Sediment transport, part I: Bed load transport.* Journal of Hydraulic Engineering, 110(10): 1431–1456, 1984. 10.1061/(ASCE)0733-9429(1984)110:10(1431)

Annexe

5.A Sur l'approximation implicite du système acoustique

Dans cette section, nous montrons brièvement comment formuler la version implicite-explicite du schéma numérique qui approche l'équation d'Exner directement dans l'étape de transport. À cette fin, nous reformulons les troisième et quatrième équations du système (5.4.4) sous forme de système linéaire: $A\overrightarrow{X} = \overrightarrow{b}$, où

$$\overrightarrow{X} = \begin{pmatrix} u_1^{n+} \\ \Pi_1^{n+} \\ \vdots \\ u_j^{n+} \\ \Pi_j^{n+} \\ \vdots \\ u_M^{n+} \\ \Pi_M^{n+} \end{pmatrix} \quad \text{et} \quad \overrightarrow{b} = \begin{pmatrix} u_1^n + \dfrac{\Delta t}{\Delta m_1} s_1^n \\ \Pi_1^n - a_{\frac{3}{2}}^2 \dfrac{\Delta t}{2\Delta m_1} \left(\dfrac{s_{3/2}^n}{a_{\frac{3}{2}}} - \dfrac{s_{1/2}^n}{a_{\frac{1}{2}}} \right) \\ \vdots \\ \vdots \\ u_j^n + \dfrac{\Delta t}{\Delta m_j} s_j^n \\ \Pi_j^n - a_{j+\frac{1}{2}}^2 \dfrac{\Delta t}{2\Delta m_j} \left(\dfrac{s_{j+1/2}^n}{a_{j+\frac{1}{2}}} - \dfrac{s_{j-1/2}^n}{a_{j-\frac{1}{2}}} \right) \\ \vdots \\ \vdots \\ u_M^n + \dfrac{\Delta t}{\Delta m_M} s_M^n \\ \Pi_M^n - a_{M+\frac{1}{2}}^2 \dfrac{\Delta t}{2\Delta m_M} \left(\dfrac{s_{M+1/2}^n}{a_{M+\frac{1}{2}}} - \dfrac{s_{M-1/2}^n}{a_{M-\frac{1}{2}}} \right) \end{pmatrix} \tag{5.A.1}$$

et

$$A = \begin{pmatrix} c_1 & f_1 & d_1 & g_1 & 0 & \dots & \dots & \dots & \dots & 0 \\ \tilde{c}_1 & \tilde{f}_1 & \tilde{d}_1 & \tilde{g}_1 & 0 & \dots & \dots & \dots & \dots & 0 \\ b_2 & e_2 & c_2 & f_2 & d_2 & g_2 & 0 & \dots & \dots & 0 \\ \tilde{b}_2 & \tilde{e}_2 & \tilde{c}_2 & \tilde{f}_2 & \tilde{d}_2 & \tilde{g}_2 & 0 & \dots & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & \dots & 0 \\ \dots & 0 & b_j & e_j & c_j & f_j & d_j & g_j & 0 & \dots \\ \dots & 0 & \tilde{b}_j & \tilde{e}_j & \tilde{c}_j & \tilde{f}_j & \tilde{d}_j & \tilde{g}_j & 0 & \dots \\ 0 & \dots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & \dots & \dots & \dots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & \dots & \dots & \dots & \dots & 0 & b_M & e_M & c_M & f_M \\ 0 & \dots & \dots & \dots & \dots & 0 & \tilde{b}_M & \tilde{e}_M & \tilde{c}_M & \tilde{f}_M \end{pmatrix} \tag{5.A.2}$$

avec

$$b_j = -\frac{\Delta t}{2\Delta m_j} a^n_{j-1/2}, \quad c_j = 1 + \frac{\Delta t}{\Delta m_j}\left(\frac{a^n_{j+1/2} + a^n_{j-1/2}}{2}\right), \quad d_j = -\frac{\Delta t}{2\Delta m_j} a^n_{j+1/2},$$
$$e_j = -\frac{\Delta t}{2\Delta m_j}, \quad f_j = 0, \quad g_j = +\frac{\Delta t}{2\Delta m_j},$$
$$\tilde{b}_j = -\frac{\Delta t}{2\Delta m_j}(a^n_{j+1/2})^2, \quad \tilde{c}_j = 0, \quad \tilde{d}_j = +\frac{\Delta t}{2\Delta m_j}(a^n_{j+1/2})^2$$
$$\tilde{e}_j = -\frac{(a^n_{j+1/2})^2 \Delta t}{2a^n_{j-1/2}\Delta m_j}, \quad \tilde{f}_j = 1 + \frac{(a^n_{j+1/2})^2 \Delta t}{2\Delta m_j}\left(\frac{1}{a^n_{j+1/2}} + \frac{1}{a^n_{j-1/2}}\right) \quad \text{et} \quad \tilde{g}_j = -\frac{(a^n_{j+1/2})^2 \Delta t}{2a^n_{j+1/2}\Delta m_j}.$$

Par les indices 1 et M, nous indiquons la valeur de la variable dans la première et la dernière cellule respectivement, où M est le nombre de cellules. Enfin, il est clair que la première et la dernière ligne du système doivent être modifiées en fonction de la condition aux limites considérée.

Ensuite, une fois que u^{n+} et Π^{n+} ont été été trouvés, il est simple de mettre à jour les variables Lh^{n+}, $Lh\rho^{n+}$ et $Lh\rho u^{n+}$ en considérant l'approximation lagrangienne 5.4.1 avec les valeurs des étoiles au temps t^{n+}. L'étape de transport suit sans aucun changement, comme décrit dans la section 5.4.2.

5.B Sur l'approximation implicite du système acoustique avec topographie non-constante

Dans les sections précédentes, nous avons proposé une approximation implicite du système acoustique (5.2.12) qui ne tient pas compte de l'évolution dans le temps de l'élévation du lit z. Ainsi, l'approximation numérique de z résulte être complètement explicite.

Dans cette annexe, nous proposons d'approcher implicitement le système acoustique (5.3.4), en résolvant deux systèmes linéaires. En effet, grâce à cela, nous pourrions à nouveau obtenir un schéma numérique rapide qui a un Δt basé uniquement sur la condition CFL de l'étape de transport (5.4.11). D'autre part, cette nouvelle méthode serait également plus stable dans les situations où des oscillations non-physiques pourraient apparaître en raison des propriétés hyperboliques du système couplé (5.1.1), voir encore [25] pour plus de détails sur cette problématique. Cependant, les instabilités dans les données numériques apparaissent surtout à un ordre de précision plus élevé. Il serait donc particulièrement intéressant d'étendre ces deux

différentes méthodes implicites-explicites au second ordre de précision afin de mettre en évidence les avantages et les inconvénients de chaque approche.

Ainsi, nous décrivons ici une méthode implicite-explicite du premier ordre alternative pour le système (5.1.1) (sans effectuer de tests numériques) dans la perspective d'étendre ces méthodes au second ordre de précision dans des travaux futurs. Voyons brièvement les détails.

Considérons le système (5.3.4). Nous avons déjà montré que les équations d'évolution pour u et Π peuvent être approximée implicitement en résolvant un système linéaire, voir aussi l'annexe précédente 5.A. En particulier, à cette fin, les valeurs de l'étoile u^*, Π^* sont évaluées au prochain pas de temps t^{n+}. Ainsi, une fois que nous avons trouvé la vitesse u et la pression linéarisée Π au nouveau niveau temporel t^{n+}, nous pouvons également mettre à jour les variables τ et θ sans avoir besoin de résoudre les équations d'évolution pour z et Ω. Donc, à ce stade, nous avons seulement besoin d'évaluer ces dernières. En ce qui concerne l'élévation du lit z, nous rappelons que son approximation explicite est définie par (5.4.7), il est donc simple d'observer que la discrétisation implicite est donnée par

$$\big(1+\frac{\Delta t}{\Delta x}(u^{*,n+}_{j+\frac{1}{2}}-u^{*,n+}_{j-\frac{1}{2}})\big)z^{n+}_j = z^n_j - \frac{\Delta t}{\Delta x}\big((\Omega-zu)^{*,n+}_{j+\frac{1}{2}}-(\Omega-zu)^{*,n+}_{j-\frac{1}{2}}\big). \tag{5.B.1}$$

D'autre part, pour la variable Ω, nous observons que son équation d'évolution peut également être formulée comme suit,

$$\partial_t(L\Omega)-\partial_\xi(\Omega u)+2u\partial_\xi\Omega+u^2(b^2\tau^2\theta^2-1)\partial_\xi z=0$$

ou encore

$$\partial_t(L\Omega)-\partial_\xi(\Omega u)+2u\partial_\xi\Omega+u(b^2\tau^2\theta^2-1)(\partial_\xi(zu)-z\partial_\xi u)=0,$$

où nous avons utilisé le fait que $u^2\partial_\xi z = u\partial_\xi(zu)-zu\partial_\xi u$. Ainsi, nous proposons la discrétisation suivante,

$$\begin{aligned}\big(1+\frac{\Delta t}{\Delta x}(u^{*,n+}_{j+\frac{1}{2}}-u^{*,n+}_{j-\frac{1}{2}})\big)\Omega^{n+}_j &= \Omega^n_j+\frac{\Delta t}{\Delta x}\big((\Omega u)^{*,n+}_{j+\frac{1}{2}}-(\Omega u)^{*,n+}_{j-\frac{1}{2}}\big)-2u^{n+}_j\frac{\Delta t}{\Delta x}(\Omega^{*,n+}_{j+\frac{1}{2}}-\Omega^{*,n+}_{j-\frac{1}{2}})\\ &-u^{n+}_j(b^2(\tau^{n+}_j\theta^{n+}_j)^2-1)\frac{\Delta t}{\Delta x}\Big(\big((zu)^{*,n+}_{j+\frac{1}{2}}-(zu)^{*,n+}_{j-\frac{1}{2}}\big)-z^{n+}_j\big(u^{*,n+}_{j+\frac{1}{2}}-u^{*,n+}_{j-\frac{1}{2}}\big)\Big).\end{aligned} \tag{5.B.2}$$

Donc, en considérant ensemble les approximations (5.B.1) et (5.B.2), nous obtenons un autre système linéaire grâce au fait que nous avons déjà calculé les variables τ, θ, u^*, Π^* au nouveau niveau de temps t^{n+}. Par conséquent, nous n'avons que deux systèmes linéaires à résoudre et le coût de calcul ne devrait pas être élevé.

Chapter 6

Schéma implicite-explicite équilibre du second ordre pour les systèmes de lois de conservation avec terme source

Ce chapitre a été soumis pour publication dans le Journal "Numerische Mathematik" sous la référence: C. Chalons and A. Del Grosso. *Second-order Well-Balanced Implicit-Explicit Scheme for Systems of Balance Laws.*

Dans ce chapitre, nous proposons un schéma de type Lagrange-projection implicite-explicite équilibre du second ordre pour les équations de Saint-Venant. Dans tous les chapitres précédents, nous avons atteint le second ordre de précision en exploitant la stratégie de reconstruction polynomiale et les schémas Runge-Kutta. Alors que, dans l'étape lagrangienne implicite de ce chapitre, le second ordre de précision est atteint en utilisant des développements de Taylor et la procédure de Cauchy-Kovalevskaya. Nous soulignons que, dans cette étape, il nous suffit de résoudre deux systèmes linéaires pour trouver la solution mise à jour. Une fois de plus, l'approximation implicite de l'étape lagrangienne permet d'utiliser de grandes valeurs de pas de temps, en particulier dans le cas des régimes subsoniques où la condition CFL acoustique est beaucoup plus restrictive que celle de l'étape de transport. De plus, le schéma est équilibre dans le sens où la solution "lac au repos" avec vitesse nulle est préservée.

D'autre part, il est bien connu que des oscillations parasites peuvent apparaître en présence de discontinuités lorsqu'on utilise des schémas du second ordre. Pour remédier à ce problème, nous exploitons une approche de limitation *a posteriori* de type MOOD. Cependant, nous soulignons que l'utilisation d'une telle stratégie conduit à la perte de la précision du second ordre.

Second-order well-balanced implicit-explicit scheme for systems of balance laws

Abstract

In this work, we propose a second-order well-balanced implicit-explicit Lagrange-projection scheme for the shallow water equations. In the implicit Lagrangian step, the second-order of accuracy is reached by means of Taylor expansions and Cauchy-Kovalevskaya's procedure. In this step, we only need to solve two linear systems to find the updated solution. Thanks to the implicit approximation of the Lagrangian step, we can use large time step values in the case of subsonic regimes, where the acoustic CFL condition is much more restrictive than the transport one. The scheme is well-balanced in the sense that it preserves the "lake at rest" stationary solution with zero-velocity.

In order to avoid the well-known spurious oscillations in presence of discontinuities when employing second-order schemes, we exploit an *a posteriori* limiting approach. We highlight that the use of such a strategy leads to the loss of the second-order of accuracy.

6.1 Introduction

In this work, we aim to design a second-order well-balanced implicit-explicit Lagrange-projection scheme for hyperbolic systems of balance laws. In particular, as far as the mathematical model is concerned, we take into considerations the Saint-Venant (shallow water) system, which is derived from the Navier-Stokes equations under the hypothesis that the vertical scale is much smaller than the horizontal one. As such, it has been vastly used to simulate fluid flows, for instance in rivers, channels and coastal areas. See for instance [28, 1, 5, 10, 22, 25] for more details about this model and its numerical approximation.

In a Lagrange-Projection (LP) formalism, we are meant to split the acoustic and transport terms of the mathematical model, leading to two different set of equations: the acoustic (or Lagrangian) system and the transport (projection) system. We refer either to [9, 14, 25] for LP schemes applied to the shallow water system, or to [7, 8, 12, 13, 19, 21, 11] for LP methods employed for other sets of hyperbolic partial differential equations. We remark that the presence of two distinct systems leads to two different Courant–Friedrichs–Lewy (CFL) conditions on the time step, the acoustic and transport ones. Then, the time step value should be taken as the minimum between the two. It is worth to note that, in particular situations as subsonic regimes, the acoustic waves result to be faster than the transport ones [14]. As a

consequence, the acoustic time step condition turns out to be much more restrictive than the transport one. This explains why it could be interesting to exploit an implicit approximation for the Lagrangian step, while keeping an explicit discretization for the projection step. Indeed, a very fast numerical scheme could be derived and this is indeed our goal. On the other hand, the use of a large time step could also mean a less accurate solution. For this reason, we also aim to use a numerical method which is at least second-order accurate for smooth solutions.

In order to be able to design such a scheme, we start by considering the scalar equation. In particular, we follow the Jin and Xin relaxation procedure [24] to enlarge the model to a system of two equations. Then, we derive a second-order implicit scheme by exploiting Taylor expansions and the Cauchy-Kovalevskaya procedure. Subsequently, the Saint-Venant system is taken into account and the implicit method is extended to numerically approximate the Lagrangian solution. For a very similar explicit (and high order) version of this scheme applied to the 1D Euler equations, we refer to the work of F. Duboc et al. [21]. Finally, the transport system's approximation is kept explicit and a simple upwind approach is considered. A second-order accurate approximation of the projection step is obtained by using the polynomial reconstruction technique [25].

That been said, another problem could appear. It is known that the use of a second (or higher) order scheme in presence of discontinuities could lead to the presence of unphysical oscillations in the numerical results. In order to solve this issue, many techniques have been proposed, for which we refer to [27] and the references therein. However, the great majority of these approaches lead to the presence of non-linear systems in the numerical approach. This would mean a high computational cost when using the implicit scheme and, thus, it would be in contradiction with the fact that we aim to obtain a fast numerical method. Hence, here we propose to use an *a posteriori* limiting approach resembling the MOOD strategy [17, 18]. The idea is the following. Once a spurious oscillation is detected, we simply substitute the second-order numerical approximation with a convex combination between the original second-order method and another implicit first-order scheme. It is clear that such a strategy would lead to the loss of the second-order of accuracy. This kind of approach has already been proposed in the work of Dimarco et al. [20] for the Euler system.

Last but not least, we are interested in the well-balanced property of a method, which means the ability of a scheme to preserve the zero-velocity stationary solutions of a mathematical model [2]. In the case of the shallow water system, we aim to preserve the so-called "lake at rest" steady state. This property is linked to the accuracy of the numerical method as well, as it is known that a non-well-balanced scheme could produce non-physical oscillations in the numerical solution. As such, numerical schemes with the well-balanced property have been intensely studied and designed. Here, we refer to [1, 5, 4, 6, 16, 3] for well-balanced schemes applied to the Saint-Venant system and to [14, 19, 9, 25, 11] for well-balanced Lagrange-projection schemes.

Chapter outline. In the following section 6.2, we present the implicit scheme applied to the scalar equation and we prove that the scheme is second-order accurate. Moreover, when considering the linear advection equation, we show that the numerical scheme is unconditionally stable. Then, in section 6.3, we introduce the shallow water system for which we explain both the explicit and the implicit-explicit Lagrange-projection numerical scheme. The source term related to a non constant topography in space is included as well. In section 6.4, we present an a posteriori limiting approach which allows us to remove the spurious oscillations in presence of discontinuities. Numerical results are presented in section 6.5 to validate our numerical method. Finally, conclusions are drawn in section 6.6.

6.2 Scalar conservation law

Before getting into the heart of the matter, that is to say the numerical approximation of a system of partial differential equations, let us consider the following scalar equation,

$$\partial_t u + \partial_x f(u) = 0. \tag{6.2.1}$$

Here, $u = u(x,t) \in \mathbb{R}$ is the unknown and the flux $f : \mathbb{R} \to \mathbb{R}$ is a nonlinear function assumed to be smooth and strictly convex. Then, $t > 0$ is the time, x is the spatial coordinate and, as usual, Δt and Δx are the time and space steps respectively. We also specify that we consider Δt and Δx to be of the same order of magnitude. We define the intermediate times as $t^n = n\Delta t$ for $n \in \mathbb{N}$, while the mesh interfaces are $x_{j+1/2} = j\Delta x$ for $j \in \mathbb{Z}$.

Referring to the Jin and Xin relaxation procedure [24], let us introduce the relaxation parameter ε and the unknown v such that $\varepsilon > 0$ and $v \to f(u)$ at least formally when $\varepsilon \to 0$. In this sense, v is understood as a linearization of the nonlinear flux $f(u)$ and equation (6.2.1) is approximated as

$$\left\{ \begin{array}{rcl} \partial_t u + \partial_x v & = & 0, \\ \partial_t v + a^2 \partial_x u & = & -\dfrac{1}{\varepsilon}\left(v - f(u)\right), \end{array} \right. \tag{6.2.2}$$

where a is a positive constant which is assumed to obey to

$$a^2 \geq (f'(u))^2, \tag{6.2.3}$$

according to the sub-characteristic condition [29]. Moreover, as initial condition, we consider $u(x, t = 0) = u_0(x)$ and $v(x, t = 0) = v_0(x)\ \forall x,$ where $v_0(x) = f(u_0(x))$. This allows us to simply look for the numerical solutions of system

$$\begin{cases} \partial_t u + \partial_x v = 0 \\ \partial_t v + a^2 \partial_x u = 0. \end{cases} \tag{6.2.4}$$

As usual, $\forall j \in \mathbb{Z}$ and $\forall n \in \mathbb{N}$, on each cell we compute the constant average approximation of u, namely

$$u_j^n \approx \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} u(x, t^n)\, \mathrm{d}x \quad \text{with} \quad u_j^0 = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} u_0(x).$$

Thus, given the sequence $\{u_j^n\}_{n\in\mathbb{Z}}$, we look for $\{u_j^{n+1}\}_{j\in\mathbb{Z}}$ at the next time level t^{n+1}. Working in the finite volume framework, the numerical schemes read

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \frac{v_{j+1/2}^* - v_{j-1/2}^*}{\Delta x} = 0,$$

where the numerical fluxes $v_{j+1/2}^*$ have to be defined such that

$$v_{j+1/2}^* \approx \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} v(x_{j+1/2}, \theta)\, \mathrm{d}\theta. \tag{6.2.5}$$

Depending on which kind of numerical scheme we are looking for, namely either explicit or implicit, $v_{j+1/2}^*$ could depend on the values of the variables computed both at time t^n and time t^{n+1}.

As far as the explicit scheme is concerned, we take inspiration from the work of Duboc et al. [21], where they built explicit schemes of arbitrary order of accuracy for the Euler equations in the Lagrange-projection formalism. Here, we only explain how to reach the second order of accuracy and for more details we refer to [21]. Then, starting from the explicit scheme, we deduce how to obtain the second-order implicit version of the method.

6.2.1 Explicit scheme

Let us start with the explicit scheme; since the numerical fluxes will only depend on values of u and v at time t^n, we state $v^*_{j+1/2} = v^{*,n}_{j+1/2}$ for all $j \in \mathbb{Z}$. Then, considering definition (6.2.5), the second-order mid-point rule of integration gives

$$v^*_{j+1/2} = v^n_{j+1/2} + \frac{\Delta t}{2}\partial_t v^n_{j+1/2} + \mathcal{O}(\Delta t^2). \tag{6.2.6}$$

At this stage, we only need to find a way to approximate the two terms $v^n_{j+1/2}$ and $\partial_t v^n_{j+1/2}$. For the former, since we only aim to reach the second order of accuracy, we simply define

$$v^n_{j+1/2} = \frac{v^n_{j+1} + v^n_j}{2}$$

where $v^n_j = f(u^n_j)$. For the time derivative of v, $\partial_t v^n_{j+1/2}$, it is sufficient to obtain a first-order approximation as it already multiplies the term $\frac{\Delta t}{2}$. Thus, first we exploit the Cauchy-Kovalevskaya procedure and find

$$\partial_t v^n_{j+1/2} = -(a^n_{j+1/2})^2 \partial_x u^n_{j+1/2}$$

and, then, we impose

$$\partial_t v^n_{j+1/2} = -(a^n_{j+1/2})^2 \frac{u^n_{j+1} - u^n_j}{\Delta x}.$$

Concluding, the numerical flux is approximated as

$$v^*_{j+1/2} = \frac{v^n_{j+1} + v^n_j}{2} - \frac{(a^n_{j+1/2})^2 \Delta t}{2\Delta x}(u^n_{j+1} - u^n_j) \tag{6.2.7}$$

with

$$a^n_{j+1/2} = \max(f'(u)^n_j, f'(u)^n_{j+1}). \tag{6.2.8}$$

In section 6.2.3, we will see that this definition of a allows us to obtain a second-order accurate scheme. Indeed, this flux (6.2.7) clearly evokes the Lax-Wendroff flux [27] as applied to system (6.2.4).

In an analogous manner, it is easy to see that the star values for u are given by

$$u^*_{j+1/2} = \frac{u^n_{j+1} + u^n_j}{2} - \frac{\Delta t}{2\Delta x}(v^n_{j+1} - v^n_j). \tag{6.2.9}$$

6.2.2 Implicit scheme

Following the lines of the explicit scheme, we are now ready to present the second-order implicit method whose updating formula reads

$$\frac{u^{n+1}_j - u^n_j}{\Delta t} + \frac{v^*_{j+1/2} - v^*_{j-1/2}}{\Delta x} = 0 \tag{6.2.10}$$

where $v^*_{j+1/2} = v^*_{j+1/2}(u^n, u^{n+1})$. Then, as before, we start by computing a Taylor expansion in time of $v^*_{j+1/2}$ and we write

$$v^*_{j+1/2} = v^n_{j+1/2} + \frac{\Delta t}{2}\partial_t v^{n+1}_{j+1/2} + \mathcal{O}(\Delta t^2), \tag{6.2.11}$$

which is of course as relevant as formula (6.2.6) since $\partial_t v^n_{j+1/2} = \partial_t v^{n+1}_{j+1/2} + \mathcal{O}(\Delta t)$. While the term $v^n_{j+1/2}$ in (6.2.11) is computed as an average as in the explicit scheme, for the time derivative $\partial_t v^{n+1}_{j+1/2}$ we use the Cauchy-Kovalevskaya procedure at time t^{n+1} and we find

$$\partial_t v^{n+1}_{j+1/2} = -(a^n_{j+1/2})^2 \partial_x u^{n+1}_{j+1/2}.$$

Approximating the space derivative at first-order, that is to say

$$\partial_t v^{n+1}_{j+1/2} = -(a^n_{j+1/2})^2 \frac{u^{n+1}_{j+1} - u^{n+1}_j}{\Delta x},$$

we finally obtain

$$v^*_{j+1/2} = \frac{v^n_{j+1} + v^n_j}{2} - \frac{(a^n_{j+1/2})^2 \Delta t}{2\Delta x}(u^{n+1}_{j+1} - u^{n+1}_j). \tag{6.2.12}$$

As a consequence, in order to be able to find u^{n+1}, we need to solve the linear system $A\overrightarrow{X} = \overrightarrow{R}$, where we define

$$\overrightarrow{X} = \begin{pmatrix} u^{n+1}_1 \\ u^{n+1}_2 \\ \vdots \\ u^{n+1}_j \\ \vdots \\ u^{n+1}_M \end{pmatrix} \quad \text{and} \quad \overrightarrow{R} = \begin{pmatrix} u^n_1 - \dfrac{\Delta t}{2\Delta x}(v^n_2 - v^n_0) \\ u^n_2 - \dfrac{\Delta t}{2\Delta x}(v^n_3 - v^n_1) \\ \vdots \\ u^n_j - \dfrac{\Delta t}{2\Delta x}(v^n_{j+1} - v^n_{j-1}) \\ \vdots \\ u^n_M - \dfrac{\Delta t}{2\Delta x}(v^n_{M+1} - v^n_{M-1}) \end{pmatrix} \tag{6.2.13}$$

and

$$A = \begin{pmatrix} c_1 & d_1 & 0 & \dots & \dots & \dots & 0 \\ b_2 & c_2 & d_2 & 0 & \dots & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & \dots & 0 \\ \dots & 0 & b_j & c_j & d_j & 0 & \dots \\ 0 & \dots & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \ddots & \ddots & \ddots \\ 0 & \dots & \dots & \dots & 0 & b_M & c_M \end{pmatrix} \tag{6.2.14}$$

where

$$b_j = -(a^n_{j-1/2})^2 \frac{\Delta t^2}{2\Delta x^2}, \quad c_j = 1 + \frac{\Delta t^2}{2\Delta x^2}((a^n_{j+1/2})^2 + (a^n_{j-1/2})^2), \quad d_j = -(a^n_{j+1/2})^2 \frac{\Delta t^2}{2\Delta x^2}$$

$\forall j = 1, \dots, M$ and M the number of cells. We observe that boundary conditions are needed in order to be able to solve the system and that the values of a are always computed at time t^n in order to simplify the scheme. Furthermore, matrix A (6.2.14) is clearly diagonally dominant, so that system $A\overrightarrow{X} = \overrightarrow{R}$ is well-posed.

For the sake of completeness, we specify that the star values for u can be found in a similar way and read

$$u^*_{j+1/2} = \frac{u^n_{j+1} + u^n_j}{2} - \frac{\Delta t}{2\Delta x}(v^{n+1}_{j+1} - v^{n+1}_j). \tag{6.2.15}$$

6.2.3 Truncation error and order of accuracy of the implicit scheme

In this section, we prove that the implicit scheme is indeed second order accurate. As far as the explicit method is concerned, the reader can refer to the work of Duboc et al. [21]. Here, the Taylor expansions are always performed around the point (x_j, t^n) and, thus, sometimes we neglect its reference for the sake of conciseness in the notations.

Theorem 14. *The semi-implicit numerical scheme (6.2.10) with fluxes (6.2.12)-(6.2.15) is second-order accurate in both space and time under conditions*

$$\frac{a^2_{j+1/2} + a^2_{j-1/2}}{2} = f'(u)^2 + \mathcal{O}(\Delta x) \quad \textit{and} \quad \frac{a^2_{j+1/2} - a^2_{j-1/2}}{\Delta x} = \partial_x (f'(u)^2) + \mathcal{O}(\Delta x). \tag{6.2.16}$$

Proof. Let us first recall that the Cauchy-Kowalewski procedure for (6.2.1) gives $\partial_t u = -f'(u)\partial_x u$ and therefore

$$\partial_t^2 u = \left(f'(u)\right)^2 \partial_x^2 u + 2f'(u)\partial_x \left(f'(u)\right) \partial_x u.$$

Then, the time increments satisfy the following relation

$$\begin{aligned}
\frac{u_j^{n+1} - u_j^n}{\Delta t} &= \partial_t u + \frac{\Delta t}{2}\partial_t^2 u + \mathcal{O}(\Delta t^2) \\
&= \partial_t u + \frac{\Delta t}{2}\left[\left(f'(u)\right)^2 \partial_x^2 u + 2f'(u)\partial_x \left(f'(u)\right) \partial_x u\right]_j^n + \mathcal{O}(\Delta t^2).
\end{aligned}$$

Regarding the space increments, given a function α, it is useful and easy to prove that

$$\frac{\alpha_j + \alpha_{j+1}}{\Delta x} = \frac{2\alpha}{\Delta x} + \partial_x \alpha + \frac{\Delta x}{2}\partial_x^2 \alpha + \frac{\Delta x^2}{6}\partial_x^3 \alpha + \mathcal{O}(\Delta x^3),$$

$$\frac{\alpha_{j-1} + \alpha_j}{\Delta x} = \frac{2\alpha}{\Delta x} - \partial_x \alpha + \frac{\Delta x}{2}\partial_x^2 \alpha - \frac{\Delta x^2}{6}\partial_x^3 \alpha + \mathcal{O}(\Delta x^3),$$

$$\frac{\alpha_{j+1} - \alpha_j}{\Delta x} = \partial_x \alpha + \frac{\Delta x}{2}\partial_x^2 \alpha + \mathcal{O}(\Delta x^2),$$

and

$$\frac{\alpha_j - \alpha_{j-1}}{\Delta x} = \partial_x \alpha - \frac{\Delta x}{2}\partial_x^2 \alpha + \mathcal{O}(\Delta x^2).$$

Next, we know that the numerical flux reads as in (6.2.12). Then, observing that

$$u_j^{n+1} = u_j^n + \Delta t \partial_t u + \frac{\Delta t^2}{2}\partial_t^2 u + \mathcal{O}(\Delta t^3) = u_j^n - \Delta t \partial_x v + \frac{\Delta t^2}{2}\partial_t^2 u + \mathcal{O}(\Delta t^3),$$

we can write

$$\begin{aligned}
v^*_{j+1/2} &= \frac{v^n_{j+1} + v^n_j}{2} - \frac{(a^n_{j+1/2})^2 \Delta t}{2\Delta x}(u^{n+1}_{j+1} - u^{n+1}_j) \\
&= \frac{v^n_{j+1} + v^n_j}{2} - \frac{(a^n_{j+1/2})^2 \Delta t}{2\Delta x}(u^n_{j+1} - u^n_j) + \frac{(a^n_{j+1/2})^2 \Delta t^2}{2\Delta x}(\partial_x v^n_{j+1} - \partial_x v^n_j) + \mathcal{O}(\Delta t^3)
\end{aligned}$$

and

$$v^*_{j-1/2} = \frac{v^n_j + v^n_{j-1}}{2} - \frac{(a^n_{j-1/2})^2 \Delta t}{2\Delta x}(u^n_j - u^n_{j-1}) + \frac{(a^n_{j-1/2})^2 \Delta t^2}{2\Delta x}(\partial_x v^n_j - \partial_x v^n_{j-1}) + \mathcal{O}(\Delta t^3).$$

Consequently, we get

$$\begin{aligned}\frac{v^*_{j+1/2}-v^*_{j-1/2}}{\Delta x} &= \frac{v^n_{j+1}+v^n_j-(v^n_j+v^n_{j-1})}{2\Delta x}\\ &\quad-\frac{\Delta t}{2\Delta x^2}\big((a^n_{j+1/2})^2(u^n_{j+1}-u^n_j)-(a^n_{j-1/2})^2(u^n_j-u^n_{j-1})\big)+\\ &\quad+\frac{\Delta t^2}{2\Delta x^2}\big((a^n_{j+1/2})^2(\partial_x v^n_{j+1}-\partial_x v^n_j)-(a^n_{j-1/2})^2(\partial_x v^n_j-\partial_x v^n_{j-1})\big)+\mathcal{O}(\Delta t^2),\end{aligned}$$

and thus

$$\begin{aligned}\frac{v^*_{j+1/2}-v^*_{j-1/2}}{\Delta x} &= \partial_x v-\frac{\Delta t}{2\Delta x}\big((a^n_{j+1/2})^2(\partial_x u^n_j+\frac{\Delta x}{2}\partial_x^2 u^n_j)-(a^n_{j-1/2})^2(\partial_x u^n_j-\frac{\Delta x}{2}\partial_x^2 u^n_j)\big)+\\ &\quad+\frac{\Delta t^2}{2\Delta x}\big((a^n_{j+1/2})^2(\partial_x^2 v^n_j)-(a^n_{j-1/2})^2(\partial_x^2 v^n_j)\big)+\mathcal{O}(\Delta x^2)+\mathcal{O}(\Delta t^2)\\ &= \partial_x v-\frac{\Delta t}{2}\Big(\partial_x u^n_j\frac{(a^n_{j+1/2})^2-(a^n_{j-1/2})^2}{\Delta x}+\partial_x^2 u^n_j\frac{(a^n_{j+1/2})^2+(a^n_{j-1/2})^2}{2}\Big)\\ &\quad+\mathcal{O}(\Delta x^2)+\mathcal{O}(\Delta t^2).\end{aligned}$$

Lastly, we find

$$\begin{aligned}\frac{u^{n+1}_j-u^n_j}{\Delta t}+\frac{v^*_{j+1/2}-v^*_{j-1/2}}{\Delta x} &= \partial_t u+\partial_x f(u)+\mathcal{O}(\Delta x^2)+\mathcal{O}(\Delta t^2)+\\ &\quad+\frac{\Delta t}{2}\left[\left(f'(u)\right)^2\partial_x^2 u+2f'(u)\partial_x\left(f'(u)\right)\partial_x u\right]\\ &\quad-\frac{\Delta t}{2}\left[\frac{a^2_{j+1/2}+a^2_{j-1/2}}{2}\partial_x^2 u+\frac{a^2_{j+1/2}-a^2_{j-1/2}}{\Delta x}\partial_x u\right],\end{aligned}$$

and, therefore, in order to obtain a second-order scheme, it is enough to impose condition (6.2.16).

At this stage, let us do the analogous for the evolution equation for v. When considering the scalar equation, it is not strictly necessary to prove that also the evolution equation for v is approximated at second order of accuracy as, at the end of the day, we only need the values of u at time t^{n+1}. However, when expanding this numerical scheme to a system of conservation laws, it will be necessary to consider the "equivalent values of v" at time t^{n+1} (in the shallow water system, it will correspond to Π, the linearization of the pressure term).

The time increments for v satisfy the following relation

$$\frac{v^{n+1}_j-v^n_j}{\Delta t}=\partial_t v+\frac{\Delta t}{2}\partial_t^2 v+\mathcal{O}(\Delta t^2)=\partial_t v+a_j^2\frac{\Delta t}{2}\partial_x^2 v+\mathcal{O}(\Delta t^2).$$

Regarding the numerical fluxes, we have

$$u^*_{j+1/2}=\frac{u^n_{j+1}+u^n_j}{2}-\frac{\Delta t}{2\Delta x}(v^{n+1}_{j+1}-v^{n+1}_j)$$

and, using expansions in time and relation $\partial_t v=-a^2\partial_x u$, we get

$$u^*_{j+1/2}=\frac{u^n_{j+1}+u^n_j}{2}-\frac{\Delta t}{2\Delta x}(v^n_{j+1}-v^n_j)+\frac{\Delta t^2}{2\Delta x}((a^n_{j+1})^2\partial_x u^n_{j+1}-(a^n_j)^2\partial_x u^n_j)+\mathcal{O}(\Delta t^3).$$

Consequently, we find

$$\begin{aligned}
&\frac{u^*_{j+1/2}-u^*_{j-1/2}}{\Delta x} = \frac{u^n_{j+1}+u^n_j-(u^n_j+u^n_{j-1})}{2} - \frac{\Delta t}{2\Delta x^2}(v^n_{j+1}-v^n_j-(v^n_j-v^n_{j-1}))+\\
&+\frac{\Delta t^2}{2\Delta x^2}((a^n_{j+1})^2\partial_x u^n_{j+1}-(a^n_j)^2\partial_x u^n_j-((a^n_j)^2\partial_x u^n_j-(a^n_{j-1})^2\partial_x u^n_{j-1}))+\mathcal{O}(\Delta x^2)+\mathcal{O}(\Delta t^2)\\
&=\partial_x u-\frac{\Delta t}{2\Delta x}(\partial_x v+\frac{\Delta x}{2}\partial_x^2 v-(\partial_x v-\frac{\Delta x}{2}\partial_x^2 v))+\\
&+\frac{\Delta t^2}{2\Delta x}(\partial_x(a^2\partial_x u)+\frac{\Delta x}{2}\partial_x^2(a^2\partial_x u)-(\partial_x(a^2\partial_x u)-\frac{\Delta x}{2}\partial_x^2(a^2\partial_x u)))+\mathcal{O}(\Delta x^2)+\mathcal{O}(\Delta t^2)\\
&=\partial_x u-\frac{\Delta t}{2}\partial_x^2 v+\mathcal{O}(\Delta x^2)+\mathcal{O}(\Delta t^2).
\end{aligned}$$

Finally, we get

$$\begin{aligned}
\frac{v^{n+1}_j-v^n_j}{\Delta t}+a_j^2\frac{u^*_{j+1/2}-u^*_{j-1/2}}{\Delta x} &= \partial_t v+a_j^2\frac{\Delta t}{2}\partial_x^2 v+a_j^2(\partial_x u-\frac{\Delta t}{2}\partial_x^2 v)+\mathcal{O}(\Delta t^2)+\mathcal{O}(\Delta x^2)\\
&=\partial_t v+a_j^2\partial_x u+\mathcal{O}(\Delta t^2)+\mathcal{O}(\Delta x^2)
\end{aligned}$$

which concludes our proof. □

Thus, we proved that the implicit scheme is second-order accurate as well. However, let us observe that, in order to be able prove the right order of accuracy for the second equation, we do not need to satisfy a particular definition for a as long as the value we use is coherent in all the equations. Indeed, we could decide to use either

$$\partial_t v^n_j+(a^n_j)^2\partial_x u^n_j=0 \quad \forall j$$

with a defined at each cell or, for instance,

$$\partial_t v^n_j+(a^n_{j+1/2})^2\partial_x u^n_j=0 \quad \forall j$$

where a is computed at each interface.

6.2.4 Study of the stability of the implicit scheme applied to the linear advection equation

Let us consider the Linear Advection Equation (LAE) $\partial_t u+\alpha\partial_x u=0$ with constant α. The objective of this section is to show that, in the case of the LAE, the implicit numerical scheme is unconditionally stable.

Since $f(u)=v=\alpha u$, the updating equation reads

$$\frac{u^{n+1}_j-u^n_j}{\Delta t}+\frac{v^*_{j+1/2}-v^*_{j-1/2}}{\Delta x}=0$$

and in particular

$$u^{n+1}_j-\frac{\mu^2}{2}(u^{n+1}_{j+1}-2u^{n+1}_j+u^{n+1}_{j-1})=u^n_j-\frac{\mu}{2}(u^n_{j+1}-u^n_{j-1})$$

where we have set $\mu=\alpha\dfrac{\Delta t}{\Delta x}$. Next, we consider the Von Neumann method to analyze the stability of the scheme, see [27]. We introduce a test function

$$u^n_j=A^n e^{I\theta j}$$

where A is the amplitude, θ the angle and I the imaginary unit. As such, the numerical scheme is stable if and only if $||A|| \leq 1$. Hence, we write

$$A^{n+1}e^{I\theta j}\Big(1-\frac{\mu^2}{2}(e^{I\theta}-2+e^{-I\theta})\Big) = A^n e^{I\theta j}\Big(1-\frac{\mu}{2}(e^{I\theta}-e^{-I\theta})\Big).$$

By simplifying $A^n e^{I\theta j}$ and recalling that $e^{I\theta} = \cos\theta + I\sin\theta$, we get

$$A\Big(1-\frac{\mu^2}{2}(\cos\theta + I\sin\theta - 2 + \cos\theta - I\sin\theta)\Big) = 1-\frac{\mu}{2}(\cos\theta + I\sin\theta - \cos\theta + I\sin\theta)$$

and thus

$$A\Big(1+\mu^2(1-\cos\theta)\Big) = 1-\mu I\sin\theta.$$

Therefore, $||A||^2 \leq 1$ if and only if

$$\Big|\Big|\frac{1-\mu I\sin\theta}{1+\mu^2(1-\cos\theta)}\Big|\Big|^2 \leq 1,$$

and equivalently,

$$\frac{1+\mu^2\sin^2\theta}{1+\mu^4(1-\cos\theta)^2+2\mu^2(1-\cos\theta)} \leq 1,$$

and thus we get

$$1+\mu^2\sin^2\theta \leq 1+\mu^4(1-\cos\theta)^2+2\mu^2(1-\cos\theta)$$

as the denominator is always positive and the norm of a complex number is given by $||z|| = \sqrt{x^2+y^2}$ if $z = x+Iy$. Finally, we find that

$$1+\mu^2\sin^2\theta \leq 1+\mu^4(1-\cos\theta)^2+2\mu^2(1-\cos\theta)$$

which is true if and only if

$$1-\cos^2\theta \leq 2(1-\cos\theta)+\mu^2(1-\cos\theta)^2$$

or, equivalently,

$$-1-\cos^2\theta+2\cos\theta = -(1-\cos\theta)^2 \leq \mu^2(1-\cos\theta)^2,$$

which is always true. Thus, we proved that $||A|| \leq 1$ always and, as such, that the scheme is unconditionally stable.

6.3 Shallow water equations and numerical method

We finally arrived at the heart of this work, which is the numerical approximation of an hyperbolic system of partial differential equations. In particular, the mathematical model we are going to consider is the well-known shallow water system. It is composed of the continuity and momentum equations, and it reads

$$\begin{cases} \partial_t h + \partial_x(hu) = 0 \\ \partial_t(hu) + \partial_x(hu^2 + \dfrac{gh^2}{2}) = -gh\partial_x z \end{cases} \tag{6.3.1}$$

where $h(x,t)$ is the total depth of water column, $u(x,t)$ is the averaged velocity and $z(x)$ is the bed level. Then, the pressure is given by $p = \dfrac{gh^2}{2}$ where g stays for the gravitational

acceleration. Observe that the zero-velocity stationary solution is called "lake at rest" solution and reads

$$u = 0, \quad h + z = \text{constant}.$$

For more details about this system, see for instance [28, 1, 5, 9, 22].

As mentioned in the introduction, we are particularly interested in the Lagrange-projection approach, which can lead to the design of efficient implicit-explicit numerical schemes [14, 9, 25]. The Lagrange-projection decomposition consists in splitting of the mathematical model into the acoustic and transport systems. In the case of the shallow-water system, they are respectively given by

$$\begin{cases} \partial_t h + h\partial_x u = 0 \\ \partial_t(hu) + hu\partial_x u + \partial_x p = -gh\partial_x z, \end{cases} \tag{6.3.2}$$

and

$$\begin{cases} \partial_t h + u\partial_x h = 0 \\ \partial_t(hu) + u\partial_x(hu) = 0. \end{cases}$$

The latter system equivalently reads

$$\partial_t X + u\partial_x X = 0, \tag{6.3.3}$$

with $X = h$ and $X = hu$. While the transport system is numerically solved by simply using a standard time-explicit upwind discretization, for the acoustic system we extend the scheme described in section 6.2.1-6.2.2 when considering the scalar equation. In such a way, we will have the possibility to find the approximate solution both through an explicit scheme and an implicit-explicit one.

Lagrangian coordinates. The acoustic-transport splitting can also be interpreted as a Lagrange-projection decomposition. This means that you have to first solve the system formulated in Lagrangian coordinates and then project the solution into Eulerian coordinates. In order to be able to do this, we first define the fluid particle in the position ξ and the characteristic curves

$$\begin{cases} \dfrac{\partial x}{\partial t}(\xi, t) = u(x(\xi, t), t) \\ x(\xi, 0) = \xi \end{cases} \tag{6.3.4}$$

which define the trajectory : $t \to x(\xi, t)$, of ξ as the time goes on. Therefore, any function : $(x, t) \to \varphi(x, t)$ in Eulerian coordinates can be written in Lagrangian coordinates,

$$\overline{\varphi}(\xi, t) = \varphi(x(\xi, t), t).$$

Then, we introduce the volume ratio $L(\xi, t)$ which is defined such that

$$L(\xi, t) = \frac{\partial x}{\partial \xi}(\xi, t) \tag{6.3.5}$$

and which satisfies

$$\begin{cases} \dfrac{\partial L}{\partial t}(\xi, t) = \partial_\xi u(x(\xi, t), t) \\ L(\xi, 0) = 1. \end{cases} \tag{6.3.6}$$

Since it has already been described in different works [9, 25, 11], no additional details are given here and we directly present the shallow water equations (6.3.1) in Lagrangian coordinates, which reads

$$\begin{cases} \partial_t(L\bar{h}) = 0 \\ \partial_t(L\overline{hu}) + \partial_\xi \bar{p} = -g\bar{h}\partial_\xi \bar{z}. \end{cases} \tag{6.3.7}$$

At this stage, since we aim to define its implicit approximation, it is convenient to reformulate this system. Hence, we note that $L\bar{h}$ does not depend on time and thus we have

$$(L\bar{h})(\xi,t) = (L\bar{h})(\xi,0) = h(\xi,0) = h_0 \quad \text{and} \quad L(\xi,t) = \frac{h_0}{\bar{h}(\xi,t)}.$$

Then, exploiting this relation, system (6.3.7) also reads

$$\begin{cases} \partial_t h_0 = 0 \\ \partial_t L - \partial_\xi \bar{u} = 0 \\ \partial_t(h_0\bar{u}) + \partial_\xi \bar{p} = -g\bar{h}\partial_\xi \bar{z} \end{cases}$$

and, introducing the variable $\tau = 1/h$, we finally get

$$\begin{cases} \partial_t h_0 = 0 \\ \partial_t(h_0\bar{\tau}) - \partial_\xi \bar{u} = 0 \\ \partial_t(h_0\bar{u}) + \partial_\xi \bar{p} = -\dfrac{g}{\bar{\tau}}\partial_\xi \bar{z}. \end{cases}$$

From now on, we neglect the bar over the unknowns for the sake of simplicity. Hence, we rewrite the last system as

$$\begin{cases} \partial_t h_0 = 0 \\ \partial_t(h_0\tau) - \partial_\xi u = 0 \\ \partial_t(h_0 u) + \partial_\xi p = -\dfrac{g}{\tau}\partial_\xi z. \end{cases} \tag{6.3.8}$$

Lastly, following the Suliciu relaxation approach [26], we introduce the relaxation parameter λ and the new variable Π such that at least formally

$$\lim_{\lambda\to\infty} \Pi = p.$$

Π can be interpreted as a linearization of the pressure p, whose evolution equation reads $\partial_t(h_0 p) + h^2c^2\partial_\xi u = 0$ and justifies the following system

$$\begin{cases} \partial_t h_0 = 0 \\ \partial_t(h_0\tau) - \partial_\xi u = 0 \\ \partial_t(h_0 u) + \partial_\xi \Pi = -\dfrac{g}{\tau}\partial_\xi z \\ \partial_t(h_0\Pi) + a^2\partial_\xi u = \lambda(p(\tau) - \Pi). \end{cases} \tag{6.3.9}$$

Here, a^2 is a constant which linearizes h^2c^2 and that should be taken as $a^2 \geq h^2c^2$ according to the sub-characteristic condition. We also observe that the source term in the evolution equation for Π can be formally neglected by simply assuming $\Pi = p$ at initial time.

At this stage, we look for both the explicit and implicit approximations of the Lagrangian system (6.3.9). In the following section 6.3.1, we start by describing the numerical method for the Lagrangian system without topography contributions. Considerations to include the source term into the numerical scheme are given in section 6.3.2. Then, in section 6.3.4, we briefly explain how to take into account the projection step.

6.3.1 Lagrangian step with no source term

The aim of this section is to extend the implicit second-order scheme (that we have seen in section 6.2) to the acoustic system (6.3.9) with no source term. System (6.3.9) can also be

reformulated in a more compact way as $\partial_t \mathbf{Q} + \partial_\xi \mathbf{F} = 0$, with $\mathbf{Q} = (h_0, h_0\tau, h_0 u, h_0\Pi)^t$ and the numerical flux given by $\mathbf{F} = (0, -u, \Pi, a^2 u)^t$. Along the lines of section 6.2, let us find the numerical approximation for the numerical fluxes. Assuming that

$$\mathbf{F}^*_{j+1/2} \approx \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \mathbf{F}(\xi_{j+1/2}, \theta)\, \mathrm{d}\theta, \tag{6.3.10}$$

its Taylor expansion with respect to time and at 2nd-order of accuracy reads either

$$\mathbf{F}^*_{j+\frac{1}{2}} = \mathbf{F}(\xi_{j+\frac{1}{2}}, t^n) + \frac{\Delta t}{2} \partial_t \mathbf{F}(\xi_{j+\frac{1}{2}}, t^n) + O(\Delta t^2), \tag{6.3.11}$$

if the scheme is explicit, or

$$\mathbf{F}^*_{j+\frac{1}{2}} = \mathbf{F}(\xi_{j+\frac{1}{2}}, t^n) + \frac{\Delta t}{2} \partial_t \mathbf{F}(\xi_{j+\frac{1}{2}}, t^{n+1}) + O(\Delta t^2) \tag{6.3.12}$$

for an implicit method. Thus, we only need to evaluate $\mathbf{F}(\xi_{j+\frac{1}{2}}, t^n)$, $\partial_t \mathbf{F}(\xi_{j+\frac{1}{2}}, t^n)$ and $\partial_t \mathbf{F}(\xi_{j+\frac{1}{2}}, t^{n+1})$. For $\mathbf{F}(\xi_{j+\frac{1}{2}}, t^n)$, we simply approximate it by averaging the values in the adjacent cells and we state

$$\mathbf{F}^n_{j+\frac{1}{2}} = \frac{\mathbf{F}^n_{j+1} + \mathbf{F}^n_j}{2} \quad \text{where} \quad \mathbf{F}^n_j = \mathbf{F}(\mathbf{Q}^n_j).$$

Then, for the second term $\partial_t \mathbf{F}(\xi_{j+\frac{1}{2}}, t^n)$, the idea is to use the Cauchy-Kovalevskaya procedure and to write the time derivative of the flux in term of spatial derivatives of another quantity ψ, which is then approximated at first-order by

$$\partial_x \psi^n_{j+\frac{1}{2}} \approx \frac{1}{\Delta x}(\psi^n_{j+1} - \psi^n_j). \tag{6.3.13}$$

For instance, to compute $\partial_t u(\xi_{j+\frac{1}{2}}, t^n)$ we just note that

$$\partial_t u = \frac{1}{h_0} \partial_t (h_0 u) = -\frac{1}{h_0} \partial_\xi \Pi \tag{6.3.14}$$

and thus,

$$\partial_t u(\xi_{j+\frac{1}{2}}, t^n) = -\frac{1}{h_0} \partial_\xi \Pi(\xi_{j+\frac{1}{2}}, t^n) \approx -\frac{1}{h_{0,j+1/2}\Delta\xi}(\Pi^n_{j+1} - \Pi^n_j).$$

Lastly, $u^*_{j+\frac{1}{2}}$ and $\Pi^*_{j+\frac{1}{2}}$ read either

$$\begin{cases} u^*_{j+\frac{1}{2}} = \dfrac{u^n_{j+1} + u^n_j}{2} - \dfrac{\Delta t}{2h_{0,j+1/2}\Delta\xi}(\Pi^n_{j+1} - \Pi^n_j), \\ \Pi^*_{j+\frac{1}{2}} = \dfrac{\Pi^n_{j+1} + \Pi^n_j}{2} - \dfrac{(a^n_{j+\frac{1}{2}})^2 \Delta t}{2h_{0,j+1/2}\Delta\xi}(u^n_{j+1} - u^n_j) \end{cases} \tag{6.3.15}$$

if the scheme is explicit, or

$$\begin{cases} u^*_{j+\frac{1}{2}} = \dfrac{u^n_{j+1} + u^n_j}{2} - \dfrac{\Delta t}{2h_{0,j+1/2}\Delta\xi}(\Pi^{n+1-}_{j+1} - \Pi^{n+1-}_j), \\ \Pi^*_{j+\frac{1}{2}} = \dfrac{\Pi^n_{j+1} + \Pi^n_j}{2} - \dfrac{(a^n_{j+\frac{1}{2}})^2 \Delta t}{2h_{0,j+1/2}\Delta\xi}(u^{n+1-}_{j+1} - u^{n+1-}_j), \end{cases} \tag{6.3.16}$$

for an implicit approximation. We also define $h^n_{0,j+1/2} = \dfrac{h^n_{0,j} + h^n_{0,j+1}}{2}$ and $a^n_{j+\frac{1}{2}} = \max(h^n_j c^n_j, h^n_{j+1} c^n_{j+1})$. As for the scalar equation, in the case of the implicit scheme,

we need to solve a linear system in order to be able to find the unknowns at time t^{n+1-}. In particular, we observe that in the updating formula for u, only u has to be computed at time t^{n+1-}, and similarly for Π. Hence, we only need to solve the two following separated system $M_u\overrightarrow{X_u} = \overrightarrow{R_u}$ and $M_\Pi\overrightarrow{X_\Pi} = \overrightarrow{R_\Pi}$, where

$$\overrightarrow{X_u} = \begin{pmatrix} u_1^{n+1-} \\ u_2^{n+1-} \\ \vdots \\ u_j^{n+1-} \\ \vdots \\ u_M^{n+1-} \end{pmatrix}, \quad \overrightarrow{R_u} = \begin{pmatrix} h_{0,1}u_1^n - \dfrac{\Delta t}{2\Delta\xi}(\Pi_2^n - \Pi_0^n) \\ h_{0,2}u_2^n - \dfrac{\Delta t}{2\Delta\xi}(\Pi_3^n - \Pi_1^n) \\ \vdots \\ h_{0,j}u_j^n - \dfrac{\Delta t}{2\Delta\xi}(\Pi_{j+1}^n - \Pi_{j-1}^n) \\ \vdots \\ h_{0,M}u_M^n - \dfrac{\Delta t}{2\Delta\xi}(\Pi_{M+1}^n - \Pi_{M-1}^n) \end{pmatrix} \tag{6.3.17}$$

and

$$M_u = \begin{pmatrix} c_1 & d_1 & 0 & \dots & \dots & \dots & 0 \\ b_2 & c_2 & d_2 & 0 & \dots & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & \dots & 0 \\ \dots & 0 & b_j & c_j & d_j & 0 & \dots \\ 0 & \dots & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \ddots & \ddots & \ddots \\ 0 & \dots & \dots & \dots & 0 & b_M & c_M \end{pmatrix} \tag{6.3.18}$$

with

$$b_j = -(a_{j-1/2}^n)^2\frac{\Delta t^2}{2\Delta\xi^2 h_{0,j-1/2}}, \quad c_j = h_{0,j} + \frac{\Delta t^2}{2\Delta\xi^2}\Big(\frac{(a_{j+1/2}^n)^2}{h_{0,j+1/2}} + \frac{(a_{j-1/2}^n)^2)}{h_{0,j-1/2}}\Big)$$

and

$$d_j = -(a_{j+1/2}^n)^2\frac{\Delta t^2}{2\Delta\xi^2 h_{0,j+1/2}}.$$

Similarly, for Π we have

$$\overrightarrow{X_\Pi} = \begin{pmatrix} \Pi_1^{n+1-} \\ \Pi_2^{n+1-} \\ \vdots \\ \Pi_j^{n+1-} \\ \vdots \\ \Pi_M^{n+1-} \end{pmatrix}, \quad \overrightarrow{R_\Pi} = \begin{pmatrix} h_{0,1}\Pi_1^n - \dfrac{(a_j^n)^2\Delta t}{2\Delta\xi}(u_2^n - u_0^n) \\ h_{0,2}\Pi_2^n - \dfrac{(a_j^n)^2\Delta t}{2\Delta\xi}(u_3^n - u_1^n) \\ \vdots \\ h_{0,j}\Pi_j^n - \dfrac{(a_j^n)^2\Delta t}{2\Delta\xi}(u_{j+1}^n - u_{j-1}^n) \\ \vdots \\ h_{0,M}\Pi_M^n - \dfrac{(a_j^n)^2\Delta t}{2\Delta\xi}(u_{M+1}^n - u_{M-1}^n) \end{pmatrix} \tag{6.3.19}$$

and

$$M_\Pi = \begin{pmatrix} \tilde{c}_1 & \tilde{d}_1 & 0 & \dots & \dots & \dots & 0 \\ \tilde{b}_2 & \tilde{c}_2 & \tilde{d}_2 & 0 & \dots & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & \dots & 0 \\ \dots & 0 & \tilde{b}_j & \tilde{c}_j & \tilde{d}_j & 0 & \dots \\ 0 & \dots & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \ddots & \ddots & \ddots \\ 0 & \dots & \dots & \dots & 0 & \tilde{b}_M & \tilde{c}_M \end{pmatrix} \tag{6.3.20}$$

where

$$\tilde{b}_j = -(a_j^n)^2 \frac{\Delta t^2}{2\Delta\xi^2 h_{0,j-1/2}}, \quad \tilde{c}_j = h_{0,j} + \frac{(a_j^n)^2 \Delta t^2}{2\Delta\xi^2}\Big(\frac{1}{h_{0,j+1/2}} + \frac{1}{h_{0,j-1/2}}\Big)$$

and

$$\tilde{d}_j = -(a_j^n)^2 \frac{\Delta t^2}{2\Delta\xi^2 h_{0,j+1/2}}.$$

Once again, let us observe that matrices M_u and M_Π are diagonally dominant if $h_0 > 0$ for all j, so that the two systems $M_u \overrightarrow{X_u} = \overrightarrow{R_u}$ and $M_\Pi \overrightarrow{X_\Pi} = \overrightarrow{R_\Pi}$ are well-posed. Finally, system (6.3.9) is approximated as follows,

$$\begin{cases} (h_0\tau)_j^{n+1-} = (h_0\tau)_j^n + \dfrac{\Delta t}{\Delta\xi}(u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}) \\ (h_0 u)_j^{n+1-} = (h_0 u)_j^n - \dfrac{\Delta t}{\Delta\xi}(\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}) \\ (h_0\Pi)_j^{n+1-} = (h_0\Pi)_j^n - (a_j^n)^2 \dfrac{\Delta t}{\Delta\xi}(u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}), \end{cases} \tag{6.3.21}$$

while, as far as system (6.3.7) is concerned, we write

$$\begin{cases} L_j^{n+1-} h_j^{n+1-} = L_j^n h_j^n \\ L_j^{n+1-} (hu)_j^{n+1-} = L_j^n (hu)_j^n - \dfrac{\Delta t}{\Delta\xi}(\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}), \end{cases} \tag{6.3.22}$$

where we have set

$$L_j^{n+1-} = L_j^n + \frac{\Delta t}{\Delta\xi}(u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}) \quad \text{with} \quad L_j^n = 1. \tag{6.3.23}$$

6.3.2 Lagrangian step with source term

In this section, we explain how to include the geometric source term in the numerical scheme at second-order of accuracy and, at the same time, how to preserve the zero-velocity stationary solutions of the system. Indeed, we are interested in the well-balanced property of the scheme.

In order to include the source term in the numerical scheme, we observe that now formula (6.3.14) is replaced by

$$\partial_t u = -\frac{1}{h_0}(\partial_\xi \Pi + \frac{g}{\tau}\partial_\xi z)$$

and, thus, we have to change the numerical flux $u^*_{j+\frac{1}{2}}$ accordingly. Considering that

$$\begin{aligned} u^*_{j+\frac{1}{2}} &= u(\xi_{j+\frac{1}{2}}, t^n) + \frac{\Delta t}{2}\partial_t u(\xi_{j+\frac{1}{2}}, t^{n+1-}) = \\ &= u(\xi_{j+\frac{1}{2}}, t^n) - \frac{\Delta t}{2h_0}\Big(\partial_\xi \Pi(\xi_{j+\frac{1}{2}}, t^{n+1-}) + \Big\{\frac{g}{\tau}\partial_\xi z\Big\}(\xi_{j+\frac{1}{2}}, t^n)\Big), \end{aligned} \tag{6.3.24}$$

we simply impose

$$\Big\{\frac{g}{\tau}\partial_\xi z\Big\}^n_{j+1/2} = g\frac{h^n_j + h^n_{j+1}}{2}\frac{z^n_{j+1} - z^n_j}{\Delta\xi} \tag{6.3.25}$$

where we observe that, while the derivative $\partial_\xi \Pi_{j+\frac{1}{2}}$ is computed either at time t^n or at time t^{n+1-} depending on the type of numerical scheme (explicit or implicit), we keep the source term at time t^n in order to simplify the scheme. Thus, we state

$$u^*_{j+\frac{1}{2}} = \frac{u^n_{j+1} + u^n_j}{2} - \frac{\Delta t}{2h_{0,j+1/2}\Delta\xi}(\Pi^{\#}_{j+1} - \Pi^{\#}_j) - \frac{\Delta t}{2h_{0,j+1/2}}\Big\{\frac{g}{\tau}\partial_\xi z\Big\}^n_{j+1/2} \tag{6.3.26}$$

with either $\# = n$ or $\# = n+1-$.

Next, let us see how to approximate the source term in the evolution equation for u. Integrating in the volume $[t^n, t^{n+1}] \times [\xi_{j-1/2}, \xi_{j+1/2}]$, we define the average value

$$-\Big\{\frac{g}{\tau}\partial_\xi z\Big\}^n_j \approx -\frac{1}{\Delta t\Delta\xi}\int_{t^n}^{t^{n+1}}\int_{\xi_{j-1/2}}^{\xi_{j+1/2}}\Big\{\frac{g}{\tau}\partial_\xi z\Big\}(y,\theta)dyd\theta$$

and, using the mid-point rule, we set

$$\Big\{\frac{g}{\tau}\partial_\xi z\Big\}^n_j = \frac{1}{2}\Big(\Big\{\frac{g}{\tau}\partial_\xi z\Big\}^n_{j+1/2} + \Big\{\frac{g}{\tau}\partial_\xi z\Big\}^n_{j-1/2}\Big),$$

with the interface values of the source term given by (6.3.25). Hence, we discretize system (6.3.9) as follows,

$$\begin{cases} (h_0\tau)^{n+1-}_j = (h_0\tau)^n_j + \dfrac{\Delta t}{\Delta\xi}(u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}) \\ (h_0u)^{n+1-}_j = (h_0u)^n_j - \dfrac{\Delta t}{\Delta\xi}(\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}) - \Delta t\Big\{\dfrac{g}{\tau}\partial_\xi z\Big\}^n_j \\ (h_0\Pi)^{n+1-}_j = (h_0\Pi)^n_j - (a^n_j)^2\dfrac{\Delta t}{\Delta\xi}(u^*_{j+\frac{1}{2}} - u^*_{j-\frac{1}{2}}) \end{cases} \tag{6.3.27}$$

with $u^*_{j+\frac{1}{2}}$ and $\Pi^*_{j+\frac{1}{2}}$ defined as in (6.3.26) and (6.3.16) respectively. Finally, the Lagrangian system (6.3.7) with source term reads

$$\begin{cases} \partial_t(L\bar{h}) = 0 \\ \partial_t(L\overline{hu}) + \partial_\xi\bar{p} = -g\bar{h}\partial_\xi\bar{z} \end{cases}$$

and it is approximated as

$$\begin{cases} L^{n+1-}_j h^{n+1-}_j = L^n_j h^n_j \\ L^{n+1-}_j (hu)^{n+1-}_j = L^n_j (hu)^n_j - \frac{\Delta t}{\Delta\xi}(\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}) - \Delta t\{gh\partial_\xi z\}^n_j, \end{cases} \tag{6.3.28}$$

with $u^*_{j+\frac{1}{2}}$ and $\Pi^*_{j+\frac{1}{2}}$ defined as in (6.3.26) and (6.3.16) respectively.

6.3.3 Properties of the Lagrangian system's numerical approximation

This section is devoted to prove both the well-balanced property and the order of accuracy of the proposed numerical approximation.

Theorem 15. *Both the explicit and implicit approximation of the Lagrangian system (6.3.28) are second-order accurate in both space and time under conditions*

$$\begin{cases} \dfrac{(a^n_{j+\frac{1}{2}})^2 - (a^n_{j-\frac{1}{2}})^2}{\Delta x} = \partial_x((h\sqrt{\partial_h p})^2)^n_j + \mathcal{O}(\Delta x) \\ \dfrac{(a^n_{j+\frac{1}{2}})^2 + (a^n_{j-\frac{1}{2}})^2}{2} = ((h\sqrt{\partial_h p})^2)^n_j + \mathcal{O}(\Delta x). \end{cases} \tag{6.3.29}$$

Proof. We directly begin with proof for the implicit scheme as the demonstration for the explicit approximation is part of it. Here, the Taylor expansions are always performed around the point (x_j, t^n) and, for this reason, sometimes we neglect its reference for the sake of conciseness in the notations. Starting from equation $\partial_t(h_0\tau) - \partial_\xi u = 0$ and then proceeding with $\partial_t(h_0 u) + \partial_\xi p = -\frac{g}{\tau}\partial_\xi z$, we prove that the numerical approximation we described is indeed second order accurate. The scheme for the former equation reads

$$\frac{(h_0\tau)_j^{n+1-} - (h_0\tau)_j^n}{\Delta t} - \frac{u^*_{j+1/2} - u^*_{j-1/2}}{\Delta\xi} = 0,$$

where the time increments for $h_0\tau$ give

$$\frac{(h_0\tau)_j^{n+1-} - (h_0\tau)_j^n}{\Delta t} = \partial_t(h_0\tau) + \frac{\Delta t}{2}\partial_t^2(h_0\tau) + \mathcal{O}(\Delta t^2)$$

with

$$\begin{aligned}\partial_t^2(h_0\tau) = \partial_t\partial_\xi u = \partial_\xi\Big(\frac{1}{h_0}\partial_t(h_0 u)\Big) = -\partial_\xi\Big(\frac{1}{h_0}(\partial_\xi p + \frac{g}{\tau}\partial_\xi z)\Big) = \\ = -\partial_\xi\Big(\frac{1}{h_0}\Big)(\partial_\xi p + \frac{g}{\tau}\partial_\xi z) - \frac{1}{h_0}\Big(\partial_\xi^2 p + \frac{g}{\tau}\partial_\xi^2 z + g\partial_\xi\Big(\frac{1}{\tau}\Big)\partial_\xi z\Big)\end{aligned}$$

Whereas $u^*_{j+1/2}$ is given by

$$\begin{aligned}u^*_{j+1/2} &= \frac{u^n_{j+1} + u^n_j}{2} - \frac{\Delta t}{2h_{0,j+1/2}\Delta\xi}(\Pi^{n+1-}_{j+1} - \Pi^{n+1-}_j) - \frac{\Delta t}{2}\Big(g\frac{h^n_j + h^n_{j+1}}{2}\frac{z_{j+1} - z_j}{h_{0,j+1/2}\Delta\xi}\Big) \\ &= \frac{u^n_{j+1} + u^n_j}{2} - \frac{\Delta t}{2h_{0,j+1/2}\Delta\xi}(\Pi^n_{j+1} - \Pi^n_j + \Delta t(\partial_t\Pi^n_{j+1} - \partial_t\Pi^n_j) + \\ &+ \frac{\Delta t^2}{2}(\partial_t^2\Pi^n_{j+1} - \partial_t^2\Pi^n_j) + \mathcal{O}(\Delta t^3)) - \frac{\Delta t}{2}\Big(g\frac{h^n_j + h^n_{j+1}}{2}\frac{z_{j+1} - z_j}{h_{0,j+1/2}\Delta\xi}\Big).\end{aligned}$$

Recalling the space increments relations seen in section (6.2.3) and that $h^n_{0,j+1/2} = \frac{h^n_{0,j} + h^n_{0,j+1}}{2}$, we get

$$\begin{aligned}&\frac{u^*_{j+1/2} - u^*_{j-1/2}}{\Delta\xi} = \partial_\xi u - \frac{\Delta t}{2}\Big\{\frac{1}{2}\Big(\frac{1}{h_{0,j+1/2}} + \frac{1}{h_{0,j-1/2}}\Big)\Big(\partial_\xi^2 p + g\partial_\xi h\partial_\xi z + gh\partial_\xi^2 z\Big) + \\ &+ \frac{1}{\Delta\xi}\Big(\frac{1}{h_{0,j+1/2}} - \frac{1}{h_{0,j-1/2}}\Big)\Big(\partial_\xi p + gh\partial_\xi z\Big)\Big\} + \mathcal{O}(\Delta\xi^2) + \mathcal{O}(\Delta t^2) \\ &= \partial_\xi u - \frac{\Delta t}{2}\Big\{\frac{1}{h_0}\Big(\partial_\xi^2 p + g\partial_\xi h\partial_\xi z + gh\partial_\xi^2 z\Big) + \partial_\xi\Big(\frac{1}{h_0}\Big)\Big(\partial_\xi p + gh\partial_\xi z\Big)\Big\} + \mathcal{O}(\Delta\xi^2) + \mathcal{O}(\Delta t^2).\end{aligned}$$

In order to conclude the first part of the proof, we observe that

$$\begin{aligned}&\frac{(h_0\tau)_j^{n+1-} - (h_0\tau)_j^n}{\Delta t} - \frac{u^*_{j+1/2} - u^*_{j-1/2}}{\Delta\xi} = \partial_t(h_0\tau) - \frac{\Delta t}{2}\Big(\partial_\xi\Big(\frac{1}{h_0}\Big)(\partial_\xi p + \frac{g}{\tau}\partial_\xi z) \\ &+ \frac{1}{h_0}\Big(\partial_\xi^2 p + \frac{g}{\tau}\partial_\xi^2 z + g\partial_\xi\Big(\frac{1}{\tau}\Big)\partial_\xi z\Big)\Big) - \partial_\xi u + \frac{\Delta t}{2}\Big\{\frac{1}{h_0}\Big(\partial_\xi^2 p + g\partial_\xi h\partial_\xi z + gh\partial_\xi^2 z\Big) \\ &+ \partial_\xi\Big(\frac{1}{h_0}\Big)\Big(\partial_\xi p + gh\partial_\xi z\Big)\Big\} + \mathcal{O}(\Delta t^2) + \mathcal{O}(\Delta\xi^2) = \\ &= \partial_t(h_0\tau) - \partial_\xi u + \mathcal{O}(\Delta t^2) + \mathcal{O}(\Delta\xi^2)\end{aligned}$$

which proves that the evolution equation for $h_0\tau$ is approximated at second-order of accuracy without particular requirements on the definition of a.

Subsequently, we consider equation $\partial_t(h_0u) + \partial_\xi p = -\frac{g}{\tau}\partial_\xi z$ and recall its numerical approximation,

$$\frac{(h_0u)_j^{n+1-} - (h_0u)_j^n}{\Delta t} + \frac{\Pi^*_{j+1/2} - \Pi^*_{j-1/2}}{\Delta\xi} = -\Big\{\frac{g}{\tau}\partial_\xi z\Big\}_j^n.$$

From the time increments of u, we find

$$\frac{(h_0u)_j^{n+1-} - (h_0u)_j^n}{\Delta t} = \partial_t(h_0u) + \frac{\Delta t}{2}\partial_t^2(h_0u) + \mathcal{O}(\Delta t^2),$$

where in particular

$$\partial_t^2(h_0u) = -\partial_t\Big(\partial_\xi p + \frac{g}{\tau}\partial_\xi z\Big) = -\partial_t\partial_\xi p = -\partial_\xi\Big(\frac{ghh^2}{h_0}\partial_\xi u\Big) = \partial_\xi\Big(\frac{ghh^2}{h_0}\Big)\partial_\xi u + \frac{ghh^2}{h_0}\partial_\xi^2 u$$

as we note that $\partial_t(\frac{g}{\tau}\partial_\xi z) = \partial_t(ghL\partial_x z) = \partial_t(\frac{ghh_0}{h}\partial_x z) = \partial_t(gh_0\partial_x z) = 0$. Then, we consider the numerical flux $\Pi^*_{j+1/2}$,

$$\begin{aligned}\Pi^*_{j+1/2} &= \frac{\Pi^n_{j+1} + \Pi^n_j}{2} - \frac{(a^n_{j+\frac12})^2\Delta t}{2h_{0,j+1/2}\Delta\xi}(u^{n+1-}_{j+1} - u^{n+1-}_j) = \\ &= \frac{\Pi^n_{j+1} + \Pi^n_j}{2} - \frac{(a^n_{j+\frac12})^2\Delta t}{2h_{0,j+1/2}\Delta\xi}(u^n_{j+1} - u^n_j + \Delta t(\partial_t u^n_{j+1} - \partial_t u^n_j) + \\ &+ \frac{\Delta t^2}{2}(\partial_t^2 u^n_{j+1} - \partial_t^2 u^n_j) + \mathcal{O}(\Delta t^3))\end{aligned}$$

which leads to the following space increments

$$\begin{aligned}\frac{\Pi^*_{j+1/2} - \Pi^*_{j-1/2}}{\Delta\xi} &= \partial_\xi\Pi - \frac{\Delta t}{2}\Big(\frac{1}{\Delta\xi}\Big(\frac{(a^n_{j+\frac12})^2}{h_{0,j+1/2}} - \frac{(a^n_{j-\frac12})^2}{h_{0,j-1/2}}\Big)\partial_\xi u \\ &+ \frac12\Big(\frac{(a^n_{j+\frac12})^2}{h_{0,j+1/2}} + \frac{(a^n_{j-\frac12})^2}{h_{0,j+1/2}}\Big)\partial_\xi^2 u\Big) + \mathcal{O}(\Delta t^2) + \mathcal{O}(\Delta\xi^2).\end{aligned}$$

Finally, from the source term we find

$$-\Big\{\frac{g}{\tau}\partial_\xi z\Big\}_j^n = -gh\partial_\xi z + \mathcal{O}(\Delta x^2),$$

which allows us to write

$$\begin{aligned}&\frac{(h_0u)_j^{n+1-} - (h_0u)_j^n}{\Delta t} + \frac{\Pi^*_{j+1/2} - \Pi^*_{j-1/2}}{\Delta\xi} + \Big\{\frac{g}{\tau}\partial_\xi z\Big\}_j^n = \partial_t u + \frac{\Delta t}{2}\Big(\partial_\xi\Big(\frac{ghh^2}{h_0}\Big)\partial_\xi u + \frac{ghh^2}{h_0}\partial_\xi^2 u\Big) + \\ &+ \partial_\xi\Pi - \frac{\Delta t}{2}\Big(\frac{1}{\Delta\xi}\Big(\frac{(a^n_{j+\frac12})^2}{h_{0,j+1/2}} - \frac{(a^n_{j-\frac12})^2}{h_{0,j-1/2}}\Big)\partial_\xi u + \frac12\Big(\frac{(a^n_{j+\frac12})^2}{h_{0,j+1/2}} + \frac{(a^n_{j-\frac12})^2}{h_{0,j+1/2}}\Big)\partial_\xi^2 u\Big) + \\ &+ gh\partial_\xi z + \mathcal{O}(\Delta t^2) + \mathcal{O}(\Delta x^2).\end{aligned}$$

If $a^n_{j+\frac12}$ satisfies condition (6.3.29) $\forall j$, then it follows

$$\begin{aligned}\frac{(h_0u)_j^{n+1-} - (h_0u)_j^n}{\Delta t} &+ \frac{\Pi^*_{j+1/2} - \Pi^*_{j-1/2}}{\Delta\xi} + \Big\{\frac{g}{\tau}\partial_\xi z\Big\}_j^n = \\ &= \partial_t(h_0u) + \partial_\xi\Pi + gh\partial_\xi z + \mathcal{O}(\Delta t^2) + \mathcal{O}(\Delta x^2)\end{aligned}$$

and thus, we have proved that the second-order of accuracy is also reached when considering the evolution equation for u. □

Remark 10. *If we simply take a constant such that $a = \max_j(((h\sqrt{\partial_h p})^2)_j^n)$, both the implicit and explicit schemes are only first-order accurate.*

Theorem 16. *The implicit second-order approximation of the Lagrangian system (6.3.28) preserves the "lake at rest" stationary solution.*

Proof. Assuming to be under the "lake at rest" condition, that is to say $u_j^n = 0$ and $h_j^n + z_j =$ constant for all j, we want to prove that $(Lh)_j^{n+1-} = (Lh)_j^n = h_j^n$, $(Lhu)_j^{n+1-} = (Lhu)_j^n = (hu)_j^n = 0$ and $u^*_{j+\frac{1}{2}} = 0\ \forall j$. In particular, the latter is also needed in order to automatically have a well-balanced transport step. First, we observe that

$$\Pi_{j+1}^n - \Pi_j^n = \frac{g}{2}((h_{j+1}^n)^2 - (h_j^n)^2) = \frac{g}{2}(h_j^n + h_{j+1}^n)(h_{j+1}^n - h_j^n) = -\frac{g}{2}(h_j^n + h_{j+1}^n)(z_{j+1} - z_j)$$

and consequently

$$\begin{aligned} h_{0,j}u_j^{n+1-} &= h_{0,j}u_j^n - \frac{\Delta t}{\Delta\xi}(\Pi^*_{j+\frac{1}{2}} - \Pi^*_{j-\frac{1}{2}}) - \Delta t\{\frac{g\partial_\xi z}{\tau}\}_j^n \\ &= h_{0,j}u_j^n + \frac{\Delta t}{\Delta\xi}\Big(\frac{(a^n_{j+\frac{1}{2}})^2\Delta t}{2h_{0,j+1/2}\Delta\xi}(u_{j+1}^{n+1-} - u_j^{n+1-}) - \frac{(a^n_{j-\frac{1}{2}})^2\Delta t}{2h_{0,j-1/2}\Delta\xi}(u_j^{n+1-} - u_{j-1}^{n+1-})\Big). \end{aligned}$$

Therefore, it is easy to see that $u_j^{n+1-} = u_j^n = 0\ \forall j$. Now, let us consider the numerical flux $u^*_{j+\frac{1}{2}}$. Since we know that $u_j^n = u_{j+1}^n = 0$, we write

$$\begin{aligned} u^*_{j+\frac{1}{2}} &= -\frac{\Delta t}{2h_{0,j+1/2}\Delta\xi}(\Pi_{j+1}^{n+1-} - \Pi_j^{n+1-}) - \frac{\Delta t}{2}g\frac{h_j^n + h_{j+1}^n}{2}\frac{z_{j+1} - z_j}{h_{0,j+1/2}\Delta\xi} \\ &= -\frac{\Delta t}{2h_{0,j+1/2}\Delta\xi}\Big(\Pi_{j+1}^{n+1-} - \Pi_j^{n+1-} - g\frac{h_j^n + h_{j+1}^n}{2}(h_{j+1}^n - h_j^n)\Big) \\ &= -\frac{\Delta t}{2h_{0,j+1/2}\Delta\xi}\Big(\Pi_{j+1}^{n+1-} - \Pi_j^{n+1-} - g\frac{(h_{j+1}^n)^2 - (h_j^n)^2}{2}\Big) \\ &= -\frac{\Delta t}{2h_{0,j+1/2}\Delta\xi}\Big(\Pi_{j+1}^{n+1-} - \Pi_j^{n+1-} - (\Pi_{j+1}^n - \Pi_j^n)\Big) \end{aligned}$$

where we used the relation $h_{j+1}^n + z_{j+1} = h_j^n + z_j$. Then, we easily find

$$\begin{aligned} &h_{0,j}\Pi_j^{n+1-} - (a_j^n)^2\frac{\Delta t}{\Delta\xi}(\frac{\Delta t}{2h_{0,j+1/2}\Delta\xi}(\Pi_{j+1}^{n+1-} - \Pi_j^{n+1-}) - \frac{\Delta t}{2h_{0,j-1/2}\Delta\xi}(\Pi_j^{n+1-} - \Pi_{j-1}^{n+1-})) = \\ &= h_{0,j}\Pi_j^n - (a_j^n)^2\frac{\Delta t}{\Delta\xi}(\frac{\Delta t}{2h_{0,j+1/2}\Delta\xi}(\Pi_{j+1}^n - \Pi_j^n) - \frac{\Delta t}{2h_{0,j-1/2}\Delta\xi}(\Pi_j^n - \Pi_{j-1}^n)), \end{aligned}$$

from which follows that $\Pi_j^{n+1-} = \Pi_j^n\ \forall j$, which concludes our proof.

□

6.3.4 Projection step

Let us now briefly present the numerical approximation of the projection step. Since it has already been explained in [25], here we refer to it and we only give few details. The objective being the projection of the Lagrangian variables Lh, Lhu into the Eulerian ones h, hu, we first introduce the following identity

$$\int_{\xi_1}^{\xi_2} L(\xi,t)X(\xi,t)d\xi = \int_{x(\xi_1,t)}^{x(\xi_2,t)} X(x,t)dx$$

and, defining $\hat{\xi}_{j+1/2}$ such that $x(\hat{\xi}_{j+1/2}, t^{n+1}) = x_{j+1/2}$ and $x(\hat{\xi}_{j+1/2}, t^n) = \hat{\xi}_{j+1/2}$ for all j, we write

$$\begin{aligned} X_j^{n+1} &= \frac{1}{\Delta x}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} X(x, t^{n+1})dx = \frac{1}{\Delta x}\int_{x(\hat{\xi}_{j-\frac{1}{2}}, t^{n+1})}^{x(\hat{\xi}_{j+\frac{1}{2}}, t^{n+1})} X(x, t^{n+1})dx \\ &= \frac{1}{\Delta x}\int_{\hat{\xi}_{j-\frac{1}{2}}}^{\hat{\xi}_{j+\frac{1}{2}}} L(\xi, t^{n+1-})X(\xi, t^{n+1-})d\xi \end{aligned} \tag{6.3.30}$$

where the variable X denotes h and hu. Then, we split the last integral into three parts, namely

$$\begin{aligned} X_j^{n+1} =& \frac{1}{\Delta x}\int_{\hat{\xi}_{j-\frac{1}{2}}}^{\xi_{j-\frac{1}{2}}} L(\xi, t^{n+1-})X(\xi, t^{n+1-})d\xi + \\ &+ \frac{1}{\Delta x}\int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} L(\xi, t^{n+1-})X(\xi, t^{n+1-})d\xi + \frac{1}{\Delta x}\int_{\xi_{j+\frac{1}{2}}}^{\hat{\xi}_{j+\frac{1}{2}}} L(\xi, t^{n+1-})X(\xi, t^{n+1-})d\xi, \end{aligned} \tag{6.3.31}$$

where we assume $\hat{\xi}_{j+1/2}$ given by

$$x_{j+1/2} = x(\hat{\xi}_{j+1/2}, t^{n+1}) \simeq x(\hat{\xi}_{j+1/2}, t^n) + \Delta t \partial_t x(\hat{\xi}_{j+1/2}, t^n) \simeq \hat{\xi}_{j+1/2} + \Delta t u^*_{j+1/2}.$$

In order to be able to approximate these integrals at second-order of accuracy, we also need to reconstruct the Lagrangian variables (LX),

$$(LX)_j^{n+1-}(\xi) = (LX)_j^{n+1-} + \Delta_j^{n+1-}(\xi - \xi_j) \tag{6.3.32}$$

where we use the ENO or MINMOD slope for$\Delta_j^{n+1-} = \Delta_j^{n+1-}((LX)_{j-1}^{n+1-}, (LX)_j^{n+1-}, (LX)_{j+1}^{n+1-})$. Then, exploiting the mid-point rule, from (6.3.31) we get

$$X_j^{n+1-} = (LX)_j^{n+1-} - \frac{\Delta t}{\Delta x}\Big(u^*_{j+\frac{1}{2}}(LX)_{j+\frac{1}{2}}^{n+1-}\Big(\frac{\xi_{j+\frac{1}{2}} + \hat{\xi}_{j+\frac{1}{2}}}{2}\Big) - u^*_{j-\frac{1}{2}}(LX)_{j-\frac{1}{2}}^{n+1-}\Big(\frac{\xi_{j-\frac{1}{2}} + \hat{\xi}_{j-\frac{1}{2}}}{2}\Big)\Big), \tag{6.3.33}$$

with the upwind definition

$$(LX)_{j-\frac{1}{2}}^{n+1-}(\xi) = \begin{cases} (LX)_{j-1}^{n+1-}(\xi) & \text{if} \quad u^*_{j-\frac{1}{2}} > 0 \\ (LX)_j^{n+1-}(\xi) & \text{if} \quad u^*_{j-\frac{1}{2}} \leq 0. \end{cases} \tag{6.3.34}$$

6.4 *A posteriori* limiting approach

The objective of this section is to explain how to avoid the appearance of the well-known spurious oscillations in presence of discontinuities when employing high (in our specific case, second) order schemes. Plenty of strategies have been proposed for this purpose and we simply refer to [27] and the references therein. However, the great majority of these approaches lead to the construction of nonlinear schemes. Since we are working with an implicit scheme for the Lagrangian system, this would imply the rise of the computational cost as systems $M_u\overrightarrow{X_u} = \overrightarrow{R_u}$ and $M_\Pi\overrightarrow{X_\Pi} = \overrightarrow{R_\Pi}$ would not be linear anymore. Moreover, the reason why we would like to employ an implicit-explicit scheme in a subsonic regime is to fasten the numerical computations by using greater time steps. Hence, to increase the computational cost in order to remove the oscillations would indeed be a contradiction. A partial solution to this problem

could be the employ of a so-called *a posteriori* criterion, which would reveal the eventual presence of oscillations after having computed the second-order solution. An example of such a strategy is the MOOD (Multi-dimensional Optimal Order Detection) method, presented for the first time by Clain et al. in [17] and then applied also to the shallow water system in [18]. Afterwards, this strategy has been extended to an implicit-explicit scheme, see the work of Dimarco et al. [20]. In that paper, a posteriori TVD (Total Variation Diminishing) method has been described for a second order implicit-explicit scheme applied to the compressible isentropic Euler system in the case of low-Mach number flows. We highlight that the use of this approach leads to the loss of the second-order of accuracy. Indeed, the idea is composed of the following three steps:

1. Compute the candidate solution using the second-order implicit-explicit numerical method;
2. In order to detect the presence of spurious oscillations, check if the candidate solution satisfies a criterion related to the TVD property;
3. If not, substitute the second-order solution with another one computed using a TVD numerical scheme, which in this case would be given by a convex combination of a first-order TVD method and the second-order scheme.

Here we follow the same approach and, as such, we have to understand what is the criterion the numerical solution should satisfy in order to be sure that there are no oscillations in the outputs, and, in the negative case, which kind of numerical scheme we should employ to compute a robust solution with no spurious oscillations.

For the latter issue, if oscillations are detected in the numerical outputs at the second step, we substitute the second-order Lagrangian solution $\mathbf{Q}^{n+1-} = (\tau^{n+1-}, u^{n+1-}, \Pi^{n+1-})^t$ with another one, given by a convex combination of $\mathbf{Q}^{n+1-}$ and a first-order TVD solution which we denote by $\mathbf{Q}^{n+1-}_{ARS}$. $\mathbf{Q}^{n+1-}_{ARS}$ is found using an implicit first-order Godunov-type scheme applied to the acoustic system, for which we refer to [14]. The numerical fluxes of the scheme are found using an Approximate Riemann Solver (ARS), defined by solving the Riemann problem associated with the relaxation system 6.3.9. We highlight that this first-order method satisfies the well-balanced property as well. Finally, the new Lagrangian solution would be given by

$$\mathbf{Q}^{n+1-}_{Lim} = \theta \mathbf{Q}^{n+1-} + (1-\theta)\mathbf{Q}^{n+1-}_{ARS} \tag{6.4.1}$$

with $\theta \in [0,1]$ a constant parameter which has to be chosen. Note that, if $\theta = 1$, we recover the second-order method, while, if $\theta = 0$, $\mathbf{Q}^{n+1-}_{Lim}$ reduces itself to $\mathbf{Q}^{n+1-}_{ARS}$. Thus, it is clear that the closer θ is to 0, the more diffusive the solution is but the less the oscillations associated to the discontinuities are. Following [20], we take $\theta = \sqrt{2} - 1$. Then, the transport step is solved as for the second-order scheme, see section 6.3.4. Indeed, we already use a ENO or Minmod slope when defining the reconstructed states thus, nothing more has to be done. Let us highlight that the new implicit-explicit method will still be well-balanced.

Then, in the following, we describe which kind of criterion is employed in the second step of the algorithm.

6.4.1 Spurious oscillations' detection criteria

Here we detail the different criteria we could use to detect a solution with spurious oscillations.

1. As a first option and following [20], we simply ask for the Riemann invariants associated to the shallow water systems to satisfy a Discrete Maximum Principle (DMP).

When working with a scalar equation, we know that the unknown satisfies the maximum discrete principle; however this is not true when moving to a system of equations. On the other hand, we are aware that each Riemann invariant of the system satisfies, in a certain sense, a kind of scalar equation. In particular, the eigenvalues of the shallow water system are given by $\lambda_\pm = u \pm c$, while the Riemann invariants read $\psi_\pm = u \pm 2c$. Hence, in the homogeneous case, few algebraic computations show that the evolution equations associated with the Riemann invariants $\psi_\pm$ can be written as follows,

$$\partial_t \psi_+ + \lambda_+ \partial_x \psi_+ = 0$$
$$\partial_t \psi_- + \lambda_- \partial_x \psi_- = 0.$$

Thus, after having computed the candidate numerical solution $\mathbf{U}^{n+1} = (h^{n+1}, (hu)^{n+1})^t$, we check if the following criterion is satisfied:

$$\max_{j=1}^{M} |\psi_{+,j}^{n+1}| \leq \max_{j=1}^{M} |\psi_{+,j}^{n}| \quad \text{and} \quad \max_{j=1}^{M} |\psi_{-,j}^{n+1}| \leq \max_{j=1}^{M} |\psi_{-,j}^{n}|. \tag{6.4.2}$$

2. As second option, we refer to [18] and we consider the usual discrete maximum principle but only applied to the free surface $H = h + z$. Namely, we check if

$$\min(H_{i-1}^n, H_i^n, H_{i+1}^n) \leq H_i^{n+1} \leq \max(H_{i-1}^n, H_i^n, H_{i+1}^n) \quad \forall i = 1, \dots, M$$

or alternatively, we use its relaxation version

$$\min(H_{i-1}^n, H_i^n, H_{i+1}^n) - \varepsilon \leq H_i^{n+1} \leq \max(H_{i-1}^n, H_i^n, H_{i+1}^n) + \varepsilon \quad \forall i = 1, \dots, M$$

where we take $\varepsilon = 10^{-12}$. Such a criterion succeeds in removing the oscillations but it is actually activated even when it would not be necessary, as it considers as oscillations the smooth extrema of the solution as well. Thus, additional criteria are required in order to obtain a more accurate solution. As a consequence, the following conditions on the second-order derivative of the free surface H are checked as well.

First of all, it is useful to define $\partial^2_{\max} H_i = \max(\partial_x^2 H_{i-1}^n, \partial_x^2 H_i^n, \partial_x^2 H_{i+1}^n)$ and $\partial^2_{\min} H_i = \min(\partial_x^2 H_{i-1}^n, \partial_x^2 H_i^n, \partial_x^2 H_{i+1}^n)$ where we approximate the second-order derivative in space as $\partial_x^2 H_i^n = \dfrac{H_{i+1}^n - 2H_i^n + H_{i-1}^n}{\Delta x^2}$. Then, let us enumerate the additional conditions.

- Oscillations' detection: $\partial^2_{\max} H_i \partial^2_{\min} H_i \geq -\varepsilon$,
- Plateau's detection: $\max(|\partial^2_{\max} H_i|, |\partial^2_{\min} H_i|) \leq \varepsilon$,
- Smoothness' detection:

$$\frac{\min(|\partial^2_{\max} H_i|, |\partial^2_{\min} H_i|)}{\max(|\partial^2_{\max} H_i|, |\partial^2_{\min} H_i|)} \geq 1/2 - \varepsilon.$$

If none of this criteria is fulfilled, we use the solution given by the convex combination of the second and first order scheme. Let us observe that, since our numerical scheme is implicit-explicit, it is sufficient to find a troubled cell to modify the solution in all the domain. If we would have used an explicit scheme, it could have been possible to modify the solution only in the critical cell, as it is done for instance in [18]. Thus, we are including additional diffusion to the numerical scheme.

3. Lastly, another idea could be to only accept the second-order solutions which belong to the so-called invariant region. Referring to [23], we define the invariant region as

$$\Sigma = \{(h, hu)^t \quad | \quad \psi_+^{n+1} \leq \psi_+^n \quad \text{and} \quad \psi_-^{n+1} \geq \psi_-^n\}$$

However this last option reveals itself to be the most diffusive among the criteria we saw before.

For the numerical simulations (section 6.5), we will always use the second criterion, as the first one is the less restrictive and does not always remove the spurious oscillations, while the third criterion makes the solution even more diffusive.

Remark 11. *In order to have a less diffusive numerical solution, another option is to choose a non constant in space θ in formula (6.4.1). Namely, when considering the second criterion, if it is not satisfied in the cell j, we impose $\theta_j = \sqrt{2} - 1$ otherwise we use $\theta_j = 0.75$. Observe that, if the criterion is satisfied, we do not use $\theta_j = 1$ but a smaller value in order to be sure to not have oscillations in the outputs. Consequently, when we have to define the new solution $\boldsymbol{Q}_{Lim}^{n+1-}$, instead of directly use formula (6.4.1), we first define the new numerical flux given by*

$$\boldsymbol{F}_{j+1/2}^{Lim} = \theta_{j+1/2}\boldsymbol{F}_{j+1/2}(\boldsymbol{Q}) + (1-\theta_{j+1/2})\boldsymbol{F}_{j+1/2}^{ARS}(\boldsymbol{Q})$$

and similarly for the source term

$$\boldsymbol{S}_{j+1/2}^{Lim} = \theta_{j+1/2}\boldsymbol{S}_{j+1/2}(\boldsymbol{Q}) + (1-\theta_{j+1/2})\boldsymbol{S}_{j+1/2}^{ARS}(\boldsymbol{Q}),$$

with $\theta_{j+1/2} = \min(\theta_j, \theta_{j+1})$. Note that $\boldsymbol{F}_{j+1/2}$, $\boldsymbol{S}_{j+1/2}$ are second-order accurate while $\boldsymbol{F}_{j+1/2}^{ARS}$, $\boldsymbol{S}_{j+1/2}^{ARS}$ are only first-order accurate. Then, $\boldsymbol{Q}_{Lim}^{n+1-}$ is found by exploiting the usual finite volume formula, namely

$$\boldsymbol{Q}_{j,Lim}^{n+1-} = \boldsymbol{Q}_j^{n+1-} - \frac{\Delta t}{\Delta \xi}(\boldsymbol{F}_{j+1/2}^{Lim} - \boldsymbol{F}_{j-1/2}^{Lim}) + \Delta t \boldsymbol{S}_j^{Lim}.$$

Unless otherwise specified, the numerical results presented in section 6.5 are obtained using θ constant in space.

6.5 Numerical simulations

In this section we propose some numerical tests and outputs to validate the numerical schemes. We mainly compare four numerical methods:

- "LP-HO": the second-order accurate implicit-explicit Lagrange-Projection scheme which is not TVD;
- "LP-ARS": the first-order accurate implicit-explicit Lagrange-Projection scheme obtained with the ARS and whose transport step is first-order accurate as well;
- "LP-DMP": the implicit-explicit Lagrange-projection scheme described in section 6.4. Here, the Lagrangian step approximation is obtained as a convex combination of the first and second-order implicit approximations (6.4.1), when the second-order candidate solution does not satisfy the criterion which ensures that there are no oscillations. Here we use the second criterion, namely the discrete maximum principle for the free surface with the additional checks on its second-order derivative;
- "LP-RI": as for "LP-DMP" but using the first criterion, that is to say the DMP for the Riemann invariants.

As far as the time step is concerned, we define it both for the Lagrangian and projection steps. Then, the minimum between the two is taken. The two time steps are defined as

$$\Delta t \leq \mathrm{CFL}_l \frac{\Delta x}{\max_j\{\max(\tau_j^n, \tau_{j+1}^n) a_{j+\frac{1}{2}}\}}, \tag{6.5.1}$$

for the Lagrangian system and

$$\Delta t \leq \text{CFL}_t \frac{\Delta x}{\max_j \{u^+_{j-\frac{1}{2}} - u^-_{j+\frac{1}{2}}\}},$$

with

$$u^+_{j-\frac{1}{2}} = \max(u^*_{j-\frac{1}{2}}, 0) \quad \text{and} \quad u^-_{j+\frac{1}{2}} = \min(u^*_{j+\frac{1}{2}}, 0)$$

for the transport system. Usually, for explicit first-order schemes we should take $\text{CFL}_t = 0.5$ and $\text{CFL}_l = 0.5$. Since we are considering implicit-explicit method, we could neglect the acoustic condition (6.5.1). Thus, we usually take $\text{CFL}_l = 1$ unless otherwise specified. Moreover, let us note that, in order to determine the transport time step, we compute the star values for the velocity u^* with the explicit scheme. Then, we verify that the obtained time step actually satisfies the above condition. Note that we usually compute u^* at second order of accuracy, with the exception of the one for the LP-ARS method.

The reference solution is usually computed with $M = 1000$ cells and the explicit second-order version of the LP-ARS scheme. Note that the space step is always computed as $\Delta x = \frac{L}{M}$, with L the length of the channel. Lastly, if not otherwise specified, in the numerical simulations we use transmissive boundary conditions.

6.5.1 Riemann problem with null topography

We start by showing an academic test case, a Riemann Problem (RP) with a left rarefaction and a right shock. The length of the channel is $L = 50m$ and the dam is placed in the middle. The ending time is $t_{end} = 5s$. The initial condition is given by null velocity, flat topography and water height $h_L = 1m$ if $x < L/2$, $h_R = 0.5m$ if $x > L/2$. Figure 6.5.1 shows the numerical outputs. We observe that the LP-HO scheme is less diffusive than the other methods but it presents many oscillations in correspondence of the shock. Then, we note that the LP-DMP scheme is more accurate than the LP-ARS method and removes the oscillations present in the LP-HO solution. Only a small pike remains in correspondence with the shock.

6.5.2 Riemann problem with non-zero topography

Imposing the length of the channel $L = 1500$m, let us see another Riemann problem. This time, we define a non-flat topography,

$$z(x, t=0) = \begin{cases} 4e^{2-\frac{150}{x-487.5}} & \text{if} \quad 487.5 < x \leq 562.5 \\ 8 - 4e^{2-\frac{150}{637.5-x}} & \text{if} \quad 562.5 < x \leq 637.5 \\ 8 & \text{if} \quad 637.5 < x \leq 862.5 \\ 8 - 4e^{2-\frac{150}{x-862.5}} & \text{if} \quad 862.5 < x \leq 937.5 \\ 4e^{2-\frac{150}{1012.5-x}} & \text{if} \quad 937.5 < x \leq 1012.5 \\ 0 & \text{otherwise,} \end{cases}$$

while the free surface $H = h + z$ is given by $H = 20$m if $x \leq 750$m and $H = 15$m if $x > 750$m. Finally, we take zero velocity at time $t = 0$, while for the ending time we use $t_{end} = 50$s, see also [15]. In figure 6.5.2, the results are reported. We can draw similar conclusions to the previous test case: the LP-HO outputs reveal themselves to be accurate with the exception of the shock zone, where many oscillations are present. The LP-DMP method eliminates these oscillations with the drawback to make the numerical solution more diffusive in all the domain, even if it always remains more accurate than the first-order scheme.

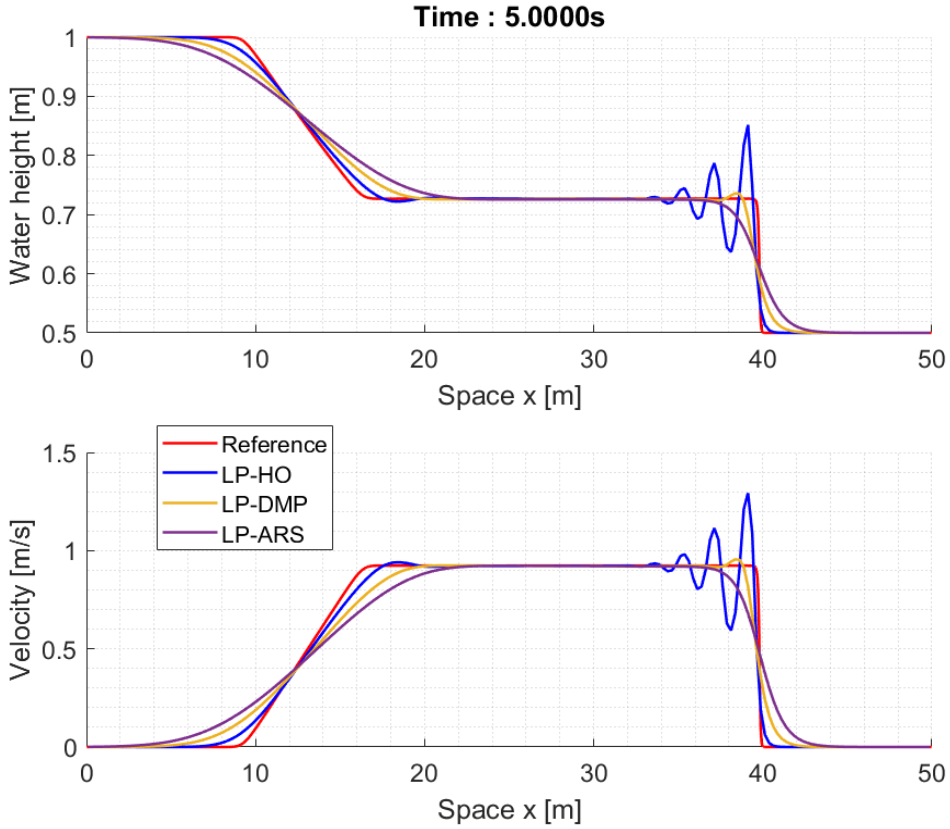


Figure 6.5.1: RP of section 6.5.1, water height (up) and velocity (bottom), $M = 200$ cells. LP-HO (blue), LP-DMP (yellow), LP-ARS (violet) and reference (red) solution are reported.

Finally, in figure 6.5.3, we compare the LP-DMP solution obtained using θ constant and non-constant in space. As expected, both schemes remove almost all the second-order oscillations but to use θ non-constant in space gives a less diffusive solution.

6.5.3 Stationary solution and perturbation

This numerical test is useful to verify if the scheme is indeed able to preserve the "lake at rest" stationary solution, see [14]. Thus, let us consider as initial condition null velocity, $h(x, t = 0) + z(x, t = 0) = 3m$ and

$$z(x, t = 0) = \begin{cases} 2 + 0.25(cos(10\pi(x - 0.5)) + 1) & \text{if} \quad 1.4 < x < 1.6 \\ 2 & \text{otherwise.} \end{cases}$$

The length of the channel is $L = 2.0m$. The numerical method proved to be able to maintain the steady state up to an error of order 10^{-15}.

As a second step, let us introduce small perturbations, namely we impose

$$h(x, t = 0) = \begin{cases} 3 - z(x, t = 0) + 0.001 & \text{if} \quad 1.1 < x < 1.2 \\ 3 - z(x, t = 0) & \text{otherwise.} \end{cases}$$

In figure 6.5.4 we insert the numerical results at time $t_{end} = 0.2s$ and we compare the three numerical methods. As for the previous numerical test, we observe that the second-order accurate method is the least diffusive but with some oscillations that are removed when employing the LP-DMP scheme. Again, the LP-DMP outputs are considered better than the first-order

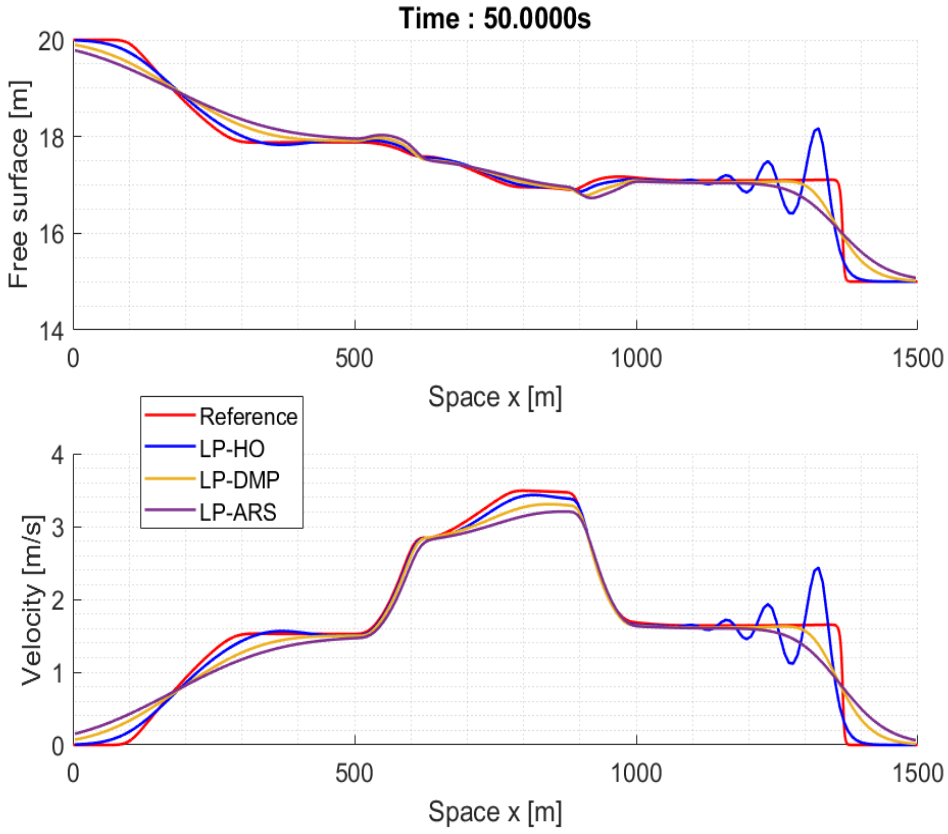


Figure 6.5.2: RP of section 6.5.2, free surface (up) and velocity (bottom), $M = 200$ cells. LP-HO (blue), LP-DMP (yellow), LP-ARS (violet) and reference (red) solution are reported.

ones. However, we also observe that, for this test case, the LP-RI solution is more accurate than the LP-DMP one, thus for this test case it would be preferable to exploit the first criterion, namely the discrete maximum principle applied to the Riemann invariants. Finally, let us remark that, for this test, we could have used a larger time step value for the implicit scheme, even up to 20 times greater. Of course, the outputs would be too diffusive if we would really use such a large time step.

6.5.4 Steady flow over a bump: fluvial regime

Here we test if the numerical scheme is able to recover a non-zero velocity steady state by imposing steady boundary conditions. In particular, considering a channel of length $L = 4$ m, we consider the following topography

$$z(x) = \begin{cases} \dfrac{\cos(10\pi(x-1))+1}{4} & \text{if} \quad 1.9 \leq x \leq 2.1 \\ 0 & \text{otherwise} \end{cases}$$

and we assume the steady state to be given by

$$\begin{cases} (hu)_{eq} = K_1 = 1 \\ \dfrac{u_{eq}^2}{2} + g(h_{eq} + z) = K_2 = 25. \end{cases} \tag{6.5.2}$$

Thus, on the left boundary we impose $q(x = 0, t) = q_{eq}$, while on the right we ask for $h(x = L, t) = h_{eq}$. Regarding the initial condition, the latter does not coincide with the steady state,

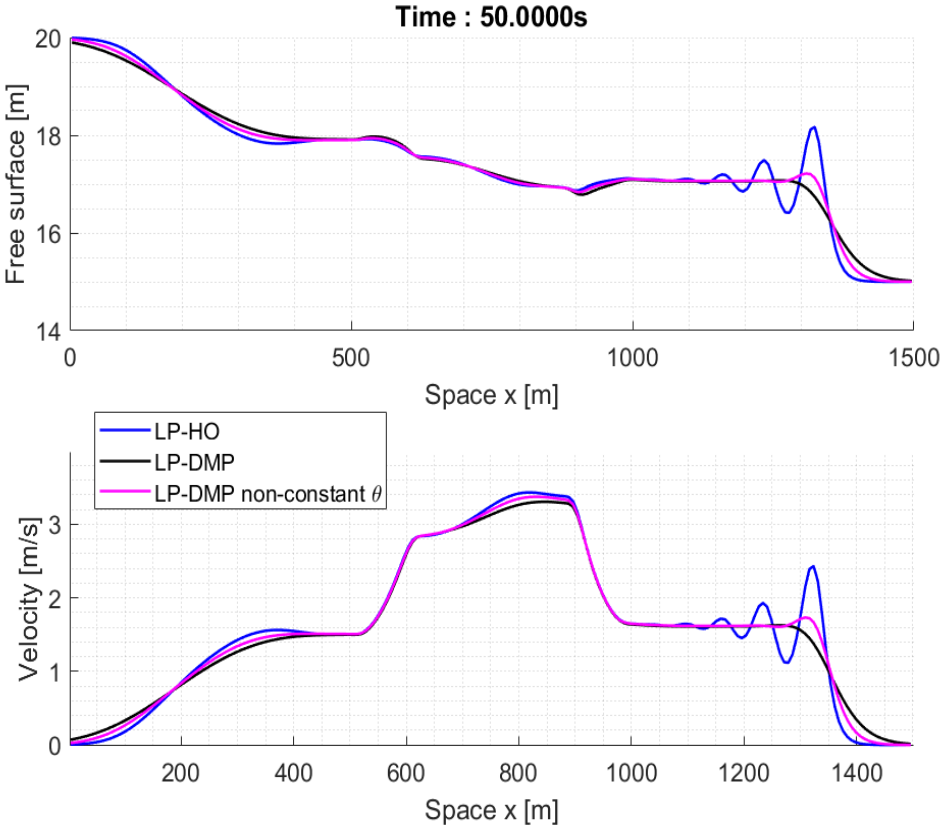


Figure 6.5.3: RP of section 6.5.2, free surface (up) and velocity (bottom), $M = 200$ cells. LP-HO (blue), LP-DMP with constant θ (black) and LP-DMP with non-constant θ (magenta) are reported.

and in particular we take

$$h(x, t = 0) = h_{eq} \quad \text{and} \quad q(x, t = 0) = 0.$$

Finally, we use $M = 1000$ cells, refer to [14] for more details about this test. Then, in figure 6.5.5 we insert the results considering the "LP-HO" and the "LP-ARS" schemes. We observe that both schemes are able to recover the steady state with the exception of the middle region, which is in agreement with what shown in [14]. Moreover, observe that for this test case we neglected the acoustic time step condition (6.5.1), in order to have faster simulations as we take $t_{End} = 200$ s. Indeed, in this way we obtain a time step which is at least four times larger, depeding on the flow regime.

6.5.5 Subsonic regime and order of accuracy

Finally, we consider a numerical test in the subsonic regime, indeed the Froude number is of order 10^{-2}. We take a channel of length $L = 20m$. The initial condition is given by null velocity and

$$\begin{cases} z_{IC} = 0.1 - 0.01e^{-(x-10)^2} \\ h_{IC} = 2 - 0.1e^{-(x-10)^2}. \end{cases}$$

We take $t_{end} = 0.5s$ and we refer to paper [10] for this test case. No oscillations are present in the LP-HO results (figure 6.5.6), however the criterion evidently results to be too strict and the LP-DMP solution reveals itself to be more diffusive than the LP-HO one. Once again, considering only the transport CFL condition, here we could have used a larger time step, namely about 18 times larger than the acoustic one.

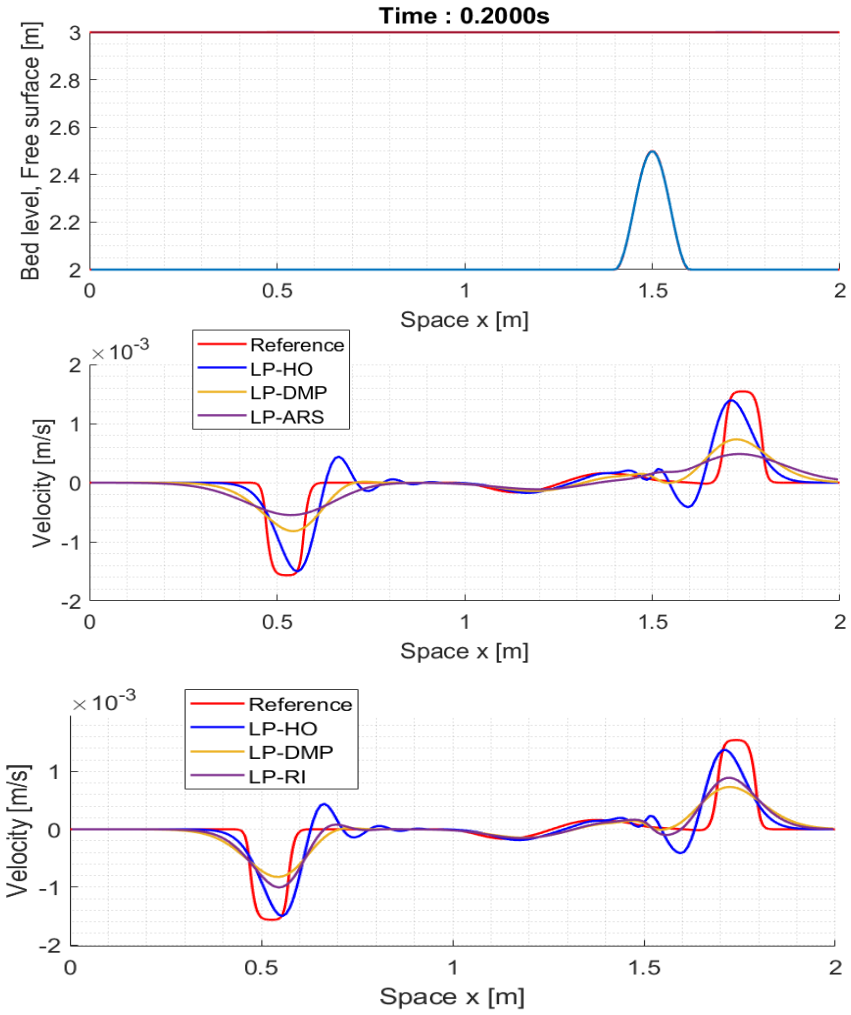


Figure 6.5.4: Test case of section 6.5.3 (perturbation of a "lake at rest" solution), free surface and velocity, $M = 200$ cells. In the middle graphic LP-HO (blue), LP-DMP (yellow), LP-ARS (violet) and reference (red) solution are reported. In the bottom graphic we insert the LP-HO (blue), LP-DMP (yellow), LP-RI (violet) and reference (red) solution

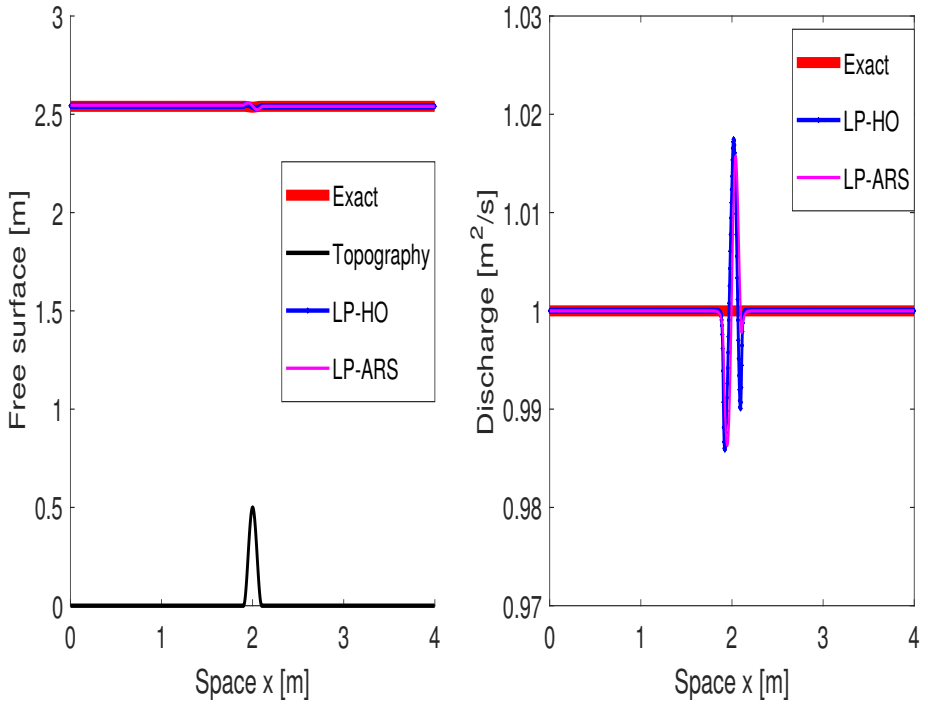


Figure 6.5.5: Steady flow over a bump: fluvial regime, section 6.5.4. Free surface and topography (left) and discharge (right), $M = 1000$ cells. LP-HO (blue), LP-ARS (magenta) and exact (red) solution are reported.

Lastly, since the solution is smooth, we use this test case to check the order of accuracy of the schemes. For this purpose, the reference solution is computed using $M = 2048$ cells. In table 6.5.1 we show that the LP-HO scheme described in section 6.3 seems to reach the second-order of accuracy.

Method	Mesh M	Variable	err $\mathbf{L}^1$	$O(\mathbf{L}^1)$	Variable	err $\mathbf{L}^1$	$O(\mathbf{L}^1)$
2nd-order	64	h	0.0549	–	hu	0.2215	–
	128		0.0199	1.4629		0.0814	1.4439
	256		0.0056	1.8206		0.0237	1.7794
	512		0.0015	1.8753		0.0065	1.8578
	1024		0.0004	1.7871		0.0019	1.7685

Table 6.5.1: Errors and empirical convergence rates for norm $\mathbf{L}^1$. Mesh of size $M = (64, 128, 256, 512, 1024)$, LP-HO numerical scheme.

6.6 Concluding remarks

In this work we have presented a second-order well-balanced implicit-explicit Lagrange-projection scheme for the shallow-water equations. This scheme results to be particularly interesting in subsonic regimes, where the CFL condition in the acoustic step is much more restrictive than the one in the transport step. Indeed, in this kind of situation, the implicit-explicit Lagrange-projection method allows to use large time step values and thus, to obtain faster numerical simulations.

In order to detect the well-known spurious oscillations in presence of discontinuities when employing a second-order scheme, we have exploited an *a posteriori* limiting approach. In particular, the Lagrangian second-order solution is substituted with a convex combination of the

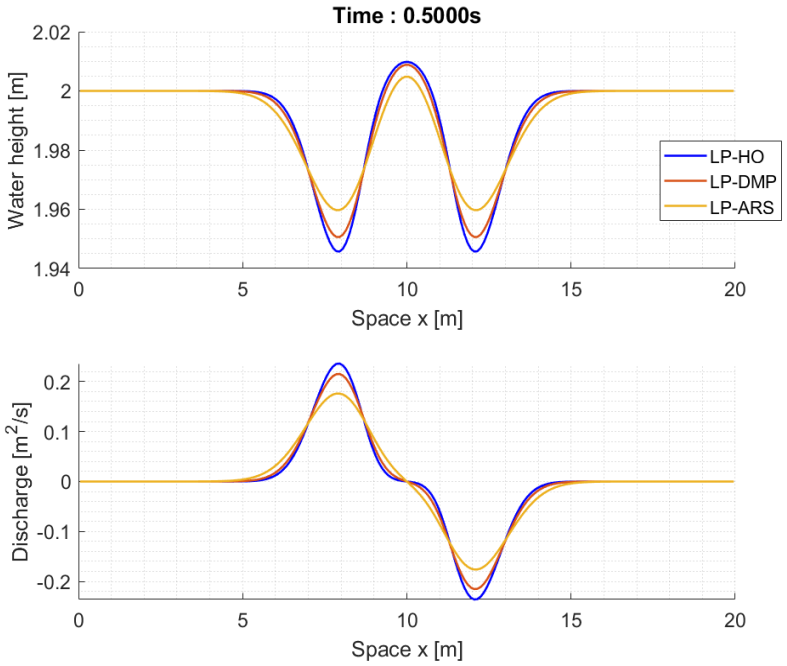


Figure 6.5.6: Subsonic test case of section 6.5.5, water height (up) and discharge (bottom), $M = 200$ cells. LP-HO (blue), LP-DMP (red) and LP-ARS (yellow) solution are reported.

first and second-order implicit approximations when the second-order candidate solution does not satisfy a criterion which ensures that the solution does not have any spurious oscillations. However, on one hand, it is clear that using such a strategy leads to the loss of the second-order of accuracy. On the other hand, we consider the results to be satisfying as the LP-DMP numerical scheme is less diffusive than the implicit first-order LP-ARS scheme.

Further works are expected to deal with the 2D extension of the numerical scheme.

6.7 Bibliography

[1] 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: 2050-2065, 2004. 10.1137/S1064827503431090.

[2] A. Bermudez, M. E. Vazquez, *Upwind methods for hyperbolic conservation laws with source terms.* Computers & Fluids, 23(8): 1049-1071, 1994. 10.1016/0045-7930(94)90004-3.

[3] C. Berthon, A. Duran, F. Foucher, K. Saleh, Khaled and J. Zabsonré. *Improvement of the Hydrostatic Reconstruction Scheme to Get Fully Discrete Entropy Inequalities.* Journal of Scientific Computing. 80, 2019. 10.1007/s10915-019-00961-y.

[4] C. Berthon, F. Foucher. *Hydrostatic Upwind Schemes for Shallow-Water Equations.* Springer Proceedings in Mathematics, 4: 97-105, 2011. 10.1007/978-3-642-20671-9_11

[5] C. Berthon, F. Foucher. *Efficient well–balanced hydrostatic upwind schemes for shallow–water equations.* Journal of Computational Physics. 231(15): 4993–5015, 2012. 10.1016/j.jcp.2012.02.031

[6] C. Berthon, R. Loubère and V. Michel-Dansac. *Second-order well-balanced scheme for the shallow-water equations with topography.* XVI International Conference on Hyperbolic Problems Theory, Numerics, Applications, Aachen (Germany), Springer Proceed, 2016. 10.1007/978-3-319-91545-6_13

[7] M. Billaud Friess, B. Boutin, F. Caetano, G. Faccanoni, S. Kokh, F. Lagoutière, L. Navoret. *A second order anti-diffusive Lagrange-remap scheme for two-component flows.* ESAIM: Proceedings, EDP Sciences, 32: 149-162, 2011. 10.1051/proc/2011018.

[8] F. Bouchut, C. Chalons, S. Guisset. *An entropy satisfying two-speed relaxation system for the barotropic Euler equations. Application to the numerical approximation of low Mach number flows.* Numerische Mathematik, Springer Verlag, 145: 35-76, 2020. 10.1007/s00211-020-01111-5.

[9] M. J. Castro Díaz, C. Chalons and T. Morales De Luna. *A fully well-balanced Lagrange-Projection type scheme for the Shallow-water equations.* SIAM J. Numer. Anal., 56(5): 3071–3098, 2018. 10.1137/17M1156101.

[10] M. J. Castro Díaz, E. D. Fernández-Nieto and A.M. Ferreiro. *Sediment transport models in Shallow Water equations and numerical approach by high order finite volume methods.* Computers & Fluids, 37(3): 299-316, 2008. 10.1016/j.compfluid.2007.07.017

[11] C. Chalons and A. Del Grosso. *Exploring different possibilities for second-order well-balanced Lagrange-projection numerical schemes applied to shallow water Exner equations.* International Journal for Numerical Methods in Fluids. 1- 31, 2022. 10.1002/fld.5064

[12] C. Chalons, M. Girardin, S. Kokh. *An All-Regime Lagrange-Projection like scheme for the gas dynamics equations on unstructured meshes.* Communications in Computational Physics. 20(1): 188-233 2016. 10.4208/cicp.260614.061115a.

[13] C. Chalons, M. Girardin, S. Kokh. *An all-regime Lagrange-Projection like scheme for 2D homogeneous models for two-phase flows on unstructured meshes.* Journal of Computational Physics, Elsevier, 335: 885-904, 2017. 10.1016/j.jcp.2017.01.017.

[14] C. Chalons, P. Kestener, S. Kokh, and M. Stauffert. *A large time-step and well-balanced Lagrange-Projection type scheme for the Shallow-water equations.* Communications in Mathematical Sciences. 15(3): 765–788, 2017. 10.4310/CMS.2017.v15.n3.a9.

[15] C. Chalons, M. Stauffert. *A well-balanced Discontinuous-Galerkin Lagrange-Projection scheme for the Shallow Water Equations.* 2017. hal-01612292

[16] A. Chinnayya, A.-Y. LeRoux, N. Seguin. *A well-balanced numerical scheme for the approximation of the shallow-water equations with topography: the resonance phenomenon.* International Journal on Finite Volume (electronic), 1(1), 2004. hal-00017378

[17] S. Clain, S. Diot, R. Loubère. *A high-order finite volume method for hyperbolic systems: Multi-dimensional Optimal Order Detection (MOOD).* Journal of Computational Physics, Elsevier, 230, 2011. ffhal-00518478v2f

[18] S. Clain, J. Figueiredo. *The MOOD method for the non-conservative shallow-water system.* 145: 99-128, 2014. 10.1016/j.compfluid.2016.11.013.

[19] A. Del Grosso and C. Chalons. *Second-order well balanced Lagrange-Projection schemes for Blood Flow Equations.* Calcolo 58, 43, 2021. 10.1007/s10092-021-00434-5

[20] G. Dimarco, R. Loubère, V. Michel-Dansac, M.-H. Vignal. *Second order Implicit-Explicit Total Variation Diminishing schemes for the Euler system in the low Mach regime.* Journal of Computational Physics, Elsevier, 372: 178-201, 2018. 10.1016/j.jcp.2018.06.022.

[21] F. Duboc, C. Enaux, S. Jaouen, H. Jourdren, M. Wolff. *High-order dimensionally split Lagrange-remap schemes for compressible hydrodynamics.* Comptes Rendus Mathematique, 348(1–2): 105-110, 2010. 10.1016/j.crma.2009.12.008.

[22] T. Gallouët, J.-M. Hérard, N. Seguin. *Some approximate Godunov schemes to compute shallow-water equations with topography*. Computers and Fluids, 32(4) : 479-513, 2003. 10.1016/S0045-7930(02)00011-7

[23] Y. Jiang, H. Liu. *An invariant-region-preserving limiter for DG schemes to isentropic Euler equations*. Numer. Methods Partial Differential Eq.. 35, 2018. 10.1002/num.22274.

[24] S. Jin and Z. Xin. *The relaxation schemes for systems of conservation laws in arbitrary space dimensions*. Communications on pure and applied mathematics, 48(3): 235-276, 1995. 10.1002/cpa.3160480303.

[25] T. Morales De Luna, M. J. Castro Díaz and C. Chalons. *High order fully well-balanced Lagrange-Projection scheme for Shallow-water*. Commun. Math. Sci., 18(3): 781–807, 2020. 10.4310/CMS.2020.v18.n3.a9

[26] I. Suliciu. *On the thermodynamics of fluids with relaxation and phase transitions. Fluids with relaxation.* Int. J. Engag. Sci. 36: 921-947, 1998.

[27] E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics*, Third Edition. Springer–Verlag, 2009.

[28] C. B. Vreugdenhil. *Numerical Methods for Shallow Water Flow.* Dordrecht; Boston: Kluwer Academic Publishers, 1994. Part of the Water Science and Technology Library book series (WSTL, volume 13).

[29] G. Whitham. Linear and nonlinear waves, volume 226. Wiley, New York, NY, USA, 1974.

Chapter 7

Conclusions et perspectives

Dans cette thèse, nous avons développé et implémenté de nombreux schémas de type Lagrange-Projection (LP) et nous les avons adaptés à différents modèles mathématiques. Toutes les méthodes numériques étaient équilibres dans le sens qu'elles préservent les solutions stationnaires à vitesse nulle du système considéré. En fonction du modèle mathématique, nous avons soit étendu la méthode numérique au second ordre de précision, soit développé sa version implicite-explicite. Seulement dans le chapitre 6, nous avons conçu une méthode LP implicite-explicite du second ordre de précision pour les équations de Saint-Venant.

Nous avons donc couvert de nombreux aspects, tant en termes de modèles que de méthodes numériques, mais il est clair qu'il serait très intéressant de mener des études plus approfondies étant donné les nombreuses pistes de recherche possibles.

Le premier modèle mathématique que nous avons considéré est le système du flux sanguin. Ainsi, dans le chapitre 2, nous avons développé deux schémas LP du second ordre capables de préserver la solution stationnaire "homme au repos éternel". Cependant, nous soulignons qu'il pourrait également être intéressant de considérer les états stationnaires à faible nombre de Shapiro, qui pourraient être plus facilement préservés que les états stationnaires classiques avec vitesse non nulle. En effet, les artères sont connues pour avoir un nombre de Shapiro moyen d'ordre 10^{-2}, ce qui pourrait aussi impliquer une condition CFL restrictive sur le pas de temps en raison des ondes acoustiques rapides. Par conséquent, des travaux futurs pourraient également traiter une formulation implicite de l'étape Lagrangienne afin d'accélérer les simulations numériques.

Par ailleurs, le modèle mathématique pourrait être étendu à un modèle plus complexe, en considérant des autres paramètres non constants en espace, comme la pression externe, ou en incluant d'autres forces, comme la friction. Il est donc clair que la propriété équilibre changerait en conséquence et que d'autres études seraient nécessaires.

Ensuite, nous avons tourné notre intérêt vers le modèle de Saint-Venant. Un tel système est connu pour être simple, mais efficace lorsqu'il s'agit de la simulation numérique d'écoulements de fluides pour lesquels l'échelle verticale est beaucoup plus petite que l'échelle horizontale. Ainsi, en fonction des situations particulières que l'on cherche à simuler, différents modèles liés existent.

En considérant par exemple des écoulements multicouche, on sait qu'il n'est généralement pas simple d'étendre la stratégie de type Lagrange-projection et, donc, de formuler le système résultant en coordonnées lagrangiennes. Ainsi, dans le chapitre 3, nous avons développé un schéma LP implicite-explicite pour les équations de Saint-Venant à deux couches, en obtenant en effet des résultats encourageants. Cependant, comme il ne s'agit que d'une première étude, il est clair que d'autres améliorations sont nécessaires. En effet, nous avons considéré une

version simplifiée du système lagrangien, mais nous visons également à analyser l'ensemble complet d'équations et à dériver un solveur de Riemann approché pour celui-ci. De plus, on pourrait essayer de tester et d'adapter la stratégie numérique à des simulations plus réalistes et complexes ou considérer également un modèle mathématique multicouche plus complet.

En dehors du modèle à deux couches, nous avons également examiné le système de Saint-Venant-Exner afin de simuler le transport de sédiments au long du lit. Un tel modèle est généralement connu pour être complexe à simuler numériquement. En effet, si une approche découplée est utilisée, des instabilités pourraient être facilement trouvées dans les tests numériques, principalement en raison des différences des structures propres du système de Saint-Venant avec et sans l'équation d'Exner. Pour cette raison, nous avons développé trois stratégies numériques différentes pour inclure l'équation d'Exner dans l'approche LP. Le second ordre de précision a également été atteint, ouvrant la voie à des schémas d'ordre de précision supérieur. Cependant, nous pouvons nous attendre à une augmentation des oscillations parasites lorsque nous étendons les méthodes numériques à un ordre de précision supérieur. Des études complémentaires devraient donc être menées dans ce sens, par exemple en utilisant une approche de type MOOD. En outre, étant donné que nous avons principalement considéré la version unidimensionnelle du modèle, des améliorations supplémentaires pourraient être liées à l'extension de ces schémas numériques à deux dimensions. En effet, seule la stratégie qui met à jour l'élévation du lit dans l'étape de transport a été étendue à deux dimensions.

Une fois que le système Saint-Venant-Exner a été considéré, l'étape suivante naturelle a été d'inclure les processus d'érosion et de dépôt des sédiments dans le modèle, ce qui a donné lieu au chapitre 5. Dans ce chapitre, une stratégie numérique analogue a été exploitée. De plus, on sait que le transport de sédiments est généralement un processus lent où le temps caractéristique associé aux sédiments est beaucoup plus grand que celui correspondant aux fluides. Pour cette raison, une version implicite-explicite de la méthode numérique a été développée, même si seulement pour une des trois stratégies et au premier ordre ou précision. Ainsi, d'autres améliorations pourraient également être apportées à cet égard. En particulier, il pourrait être intéressant d'étendre les stratégies implicites-explicites à deux dimensions également.

Comme dernier sujet, nous avons considéré le développement de méthodes implicites-explicites du second ordre pour les systèmes hyperboliques de lois de conservation, qui est connu pour être un problème très épineux. Nous avons donc essayé d'obtenir le second ordre de précision en utilisant une technique différente de celle utilisée dans les autres chapitres. Il serait évidemment intéressant d'essayer d'atteindre des ordres de précision encore plus élevés. Cependant, afin de préserver la linéarité du système acoustique, nous n'avons pas inclus de limiteurs de flux dans l'approximation numérique. En conséquence, des oscillations pourraient apparaître en présence de discontinuités, car nous utilisons une méthode du second ordre. Une première tentative pour éliminer ces instabilités a été faite, en utilisant une approche de limitation a posteriori. Cependant, la conséquence est la perte de la précision du second ordre. Ainsi, des efforts pourraient être faits pour améliorer cette approche, y compris en sélectionnant de meilleurs critères pour détecter les instabilités.

Conclusions and perspectives

In this PhD thesis we have designed and implemented numerous Lagrange-Projection (LP) schemes and adapted them to different mathematical models. All the numerical methods were well-balanced in the sense that they preserve the zero-velocity stationary solutions of the model. Depending on the mathematical model, we either extended the numerical method to second-order of accuracy or developed its implicit-explicit version. Only in chapter 6 we built an implicit-explicit LP method of second order of accuracy for the shallow water equations.

Hence, we have covered many aspects, both in terms of models and numerical methods, but it is clear that it would be very interesting to carry out more extensive studies given the many possible paths of research.

The first mathematical model we considered is the blood flow system. Hence, in chapter 2, we have developed two second-order LP schemes able to preserve the so-called "man at eternal rest" stationary solution. However, we highlight that it could also be interesting to consider the low-Shapiro number steady states, which could be more easily preserved than the classical steady states with non-zero velocity. As a matter of fact, arteries are known to have an average Shapiro number of order 10^{-2}, which could also imply a restrictive CFL condition on the time step due to the fast acoustic waves. Hence, future works are equally expected to deal with an implicit formulation of the Lagrangian step in order to speed up the numerical simulations.

Regardless, the mathematical model could be extended to a more complex one, by considering additional non-constant parameters in space, as the external pressure, or by including other forces, as the friction one. Clearly, the well-balanced property would change accordingly and further studies would be needed.

Then, we turned our interest towards the shallow water model. Such a system is known to be simple, yet effective when it comes to the numerical simulation of fluid flows for which the vertical scale is much smaller than the horizontal one. Hence, depending on the particular situations one aims to simulate, different related models exist.

Considering for instance multi-layers flows, it is known that it is generally not straightforward to extend the Lagrange-projection strategy and, in particular, to formulate the resulting system in Lagrangian coordinates. Hence, in chapter 3, we have developed an implicit-explicit LP scheme for the two-layer shallow water equations with promising results. However, being only a first study, it is clear that further improvements are needed. Indeed, we have considered a simplified version of the Lagrangian system but, certainly, we aim to analyze the complete set of equations and to derive an approximate Riemann solver for it. Furthermore, one could try to test and adapt the numerical strategy to more realistic and complex simulations or to consider a more comprehensive multi-layer mathematical model.

Other than the two-layer model, we have also examined the shallow water Exner system in

order to simulate bedload sediment transport. Such a model is generally known to be complex to approximate. Indeed, if a decoupled approach is used, instabilities could be easily found in the numerical simulations, mainly due to differences in the eigenstructures of the shallow water system with and without the Exner equation. For this reason we have developed three different numerical strategies to include the Exner equation in the LP approach. The second-order of accuracy was attained as well, paving the way for schemes of higher order of accuracy. However, we could expect spurious oscillations to increase when extending the numerical methods to higher order of accuracy. Hence, further studies should be done in this sense, for instance by using a MOOD-type approach to combine first and high-order approximations. Apart from that, since we mainly considered the one-dimensional version of the model, additional improvements could be related to the extension of these numerical schemes to two dimensions. Indeed, only the strategy that updates the bed elevation in the transport step has been extended to two dimensions.

Once the shallow water Exner system has been considered, a natural next step was to include the erosion and deposition processes of sediments in the model, giving arise to chapter 5. There, an analogous numerical strategy was exploited. Moreover, sediment transport is generally a slow process where the characteristic time associated with sediments is much larger than the one corresponding to fluids. For this reason, an implicit-explicit version of the numerical method was developed, even if only for one of the three strategies and at first-order or accuracy. Thus, further improvements could be also done in this regard. In particular, it could be interesting to extend the implicit-explicit strategies to two dimensions as well.

As a last topic, we considered the development of second-order implicit-explicit methods for hyperbolic systems of balance laws, which is actually known to be a very demanding problem. Hence, we tried to obtain the second-order of accuracy by using a different technique with respect to the one used in the other chapters. Clearly, it would be interesting to try to reach even higher orders of accuracy. However, in order to preserve the linearity of the acoustic system, we did not included flux limiters in the numerical approximation. As a consequence, oscillations could appear in presence of discontinuities as we are using a second order method. A first attempt to remove these instabilities has been made, by using a posteriori limiting approach. However, a consequence is the loss of the second-order of accuracy. Thus, efforts could be done to improve this approach, also by selecting better criteria to detect the instabilities.