UNIVERSIDAD DE MÁLAGA

Programa de Doctorado en Matemáticas

Facultad de Ciencias

Departamento de Análisis Matemático, Estadística e Investigación Operativa y Matemática Aplicada

High order well balanced numerical methods for a multilayer shallow-water model with variable density

Ernesto Guerrero Fernández

PhD Thesis

Advisors:

Dr. Manuel Jesús Castro Díaz Dr. Tomás Morales de Luna

Universidad de Málaga

Marzo 2022

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD DE MÁLAGA

AUTOR: Ernesto Guerrero Fernández

https://orcid.org/0000-0002-8657-3840

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga

Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional:
http://creativecommons.org/licenses/by-nc-nd/4.0/legalcode
Cualquier parte de esta obra se puede reproducir sin autorización
pero con el reconocimiento y atribución de los autores.
No se puede hacer uso comercial de la obra y no se puede alterar, transformar o hacer obras derivadas.

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de Málaga (RIUMA): riuma.uma.es

UNIVERSIDAD DE MÁLAGA

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D. Ernesto Guerrero Fernández
Estudiante del programa de doctorado en Matemáticas de la Universidad de Málaga, autor de la tesis, presentada para la obtención del título de doctor por la Universidad de Málaga, titulada: High order well balanced numerical methods for a multilayer shallow-water model with variable density.

Realizada bajo la tutorización de Manuel Jesús Castro Díaz y dirección de Manuel Jesús Castro Díaz y Tomás Morales de Luna (si tuviera varios directores deberá hacer constar el nombre de todos)

DECLARO QUE:

La tesis presentada es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, conforme al ordenamiento jurídico vigente (Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), modificado por la Ley 2/2019, de 1 de marzo.

Igualmente asumo, ante a la Universidad de Málaga y ante cualquier otra instancia, la responsabilidad que pudiera derivarse en caso de plagio de contenidos en la tesis presentada, conforme al ordenamiento jurídico vigente.

En Málaga, a 21 de Marzo de 2022.

Fdo.: Ernesto Guerrero Fernández Doctorando/a	Fdo.: Manuel Jesús Castro Díaz Tutor/a
Tomás Morales de Luna Director/es de tesis	

D. Manuel Jesús Castro Díaz, Catedrático del Departamento de Análisis Matemático, Estadística e Investigación Operativa, y Matemática Aplicada de la Universidad de Málaga, y D. Tomás Morales de Luna, Profesor del Departamento de Análisis Matemático, Estadística e Investigación Operativa, y Matemática Aplicada de la Universidad de Málaga.

Certifican:

Que D. Ernesto Guerrero Fernández, con grado en Ingeniería en Tecnologías Industriales y máster en Matemática Industrial, ha realizado en el Departamento de Análisis Matemático, Estadística e Investigación Operativa, y Matemática Aplicada de la Universidad de Málaga, bajo nuestra dirección, el trabajo de investigación correspondiente a su Tesis Doctoral, titulado:

High order well balanced numerical methods for a multilayer shallow-water model with variable density.

Revisado el presente trabajo, estimamos que puede ser presentado al Tribunal que ha de juzgarlo. Y para que constate a efectos de lo establecido en el artículo octavo del Real Decreto 99/2011, autorizamos la presentación de este trabajo en la Universidad de Málaga.

Málaga, 21 de Marzo de 2022.

Dr. Manuel Jesús Castro Díaz

Dr. Tomás Morales de Luna

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD
DE MÁLAGA

Esta tesis está dedicada a la memoria de mi madre, Esperanza Fernández Sánchez.

UNIVERSIDAD
DE MÁLAGA

Contents

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD
DE MÁLAGA

List of Figures

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD
DE MÁLAGA

Agradecimientos

Dicen que el demonio está en los detalles.

En una tesis tan larga hay muchos demonios escondidos. Por suerte, he contado con muchas buenas personas que me han ayudado por el camino y sin los cuales me hubiera perdido entre tantos detalles. Me gustaría por tanto agradecer, en primer lugar, a mis tutores Manuel J. Castro Díaz y Tomás Morales de Luna por su dedicación y paciencia. Su implicación académica y, sobre todo, personal han sido una inspiración constante a lo largo de estos años.

Igualmente, en este camino he estado acompañado por todos los integrantes del grupo EDANYA: Carlos Parés, María Luz Muñoz, Jorge Macias, José Manuel González Vida, José María Gallardo, Sergio Ortega, Marc de la Asunción y Carlos Sánchez. Sin olvidar a todos aquellos amigos con los que he tenido el placer de coincidir: Celia Caballero, Hugo Carrillo, Cipriano Escalante, Elena Gaburro, Irene Gómez, Alejandro González, Juan Carlos González, Emmanuel Macca, Ernesto Pimentel, Juan Francisco Rodríguez y Kleiton Schneider. Muchas gracias a todos.

Me gustaría también agradecer a todos aquellos que tan bien me acogieron durante mi estancia en Trento. Empezando por el profesor Michael Dumbser y siguiendo con Saray Busto, Simone Chiocchetti, Ilya Peshkov y Elena Gaburro (¡otra vez!). Gracias.

Por último, quisiera dar las gracias a mi familia. Es gracias a ellos por los que esta tesis ha llegado a buen puerto. Gracias a mi padre, por su entrega abnegada, gracias a mi madre, por su amor incondicional y gracias a mi hermano, por ser una inspiración y referencia. Gracias también a mis tíos, mis tías y mis primos. Hoy no estaría aquí si no fuera por vosotros.

Muchas gracias.

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD
DE MÁLAGA

Introducción y resumen

La derivación de modelos matemáticos y la simulación numérica constituyen un floreciente campo de estudio en la comunidad científica y ha contribuido con avances importantes en una gran variedad de materias, desde la mecánica del sólido hasta la hidrodinámica de fluidos. Los campos más beneficiados son aquellos donde es muy difícil, o imposible, la obtención de datos experimentales del proceso que se quiere modelizar, permitiendo de esta forma profundizar en nuestro conocimiento de la materia. Los flujos geofísicos son un ejemplo notable de este problema. El tamaño de los dominios y el hecho de que los eventos más catastróficos no son necesariamente los más frecuentes en la naturaleza, hace que el uso de modelos predictivos sea de especial relevancia. Además, la correcta simulación numérica de este tipo de fluidos requiere, en primer lugar, una profunda comprensión de la física involucrada y, en segundo lugar, un esquema numérico preciso y robusto para reflejarla correctamente.

Es más, la carga computacional requerida por los modelos unido al tamaño de los dominios computacionales son una dificultad añadida en este tipo de problemas. Existen varios caminos para afrontar esta cuestión. El más intuitivo quizás sea el de la progresiva simplificación del modelo matemático, manteniendo intactos los fenómenos físicos relevantes para el problema. Otra aproximación consiste en modificar el esquema numérico para mejorar su eficiencia pero preservando o mejorando la robustez y precisión del mismo. Finalmente, es posible explorar las opciones disponibles dentro de la optimización de código y la aceleración por *hardware*.

La física que gobierna los flujos geofísicos está regida por las conocidas ecuaciones de Navier-Stokes. Estas ecuaciones pueden simplificarse bajo ciertas hipótesis que permiten reducir la complejidad del modelo y mejorar su eficiencia global. Una forma particular de reducir la complejidad es considerar las ecuaciones de aguas someras (o ecuaciones de Saint-Venant) también conocidas por su denominación en inglés como *shallow-water* [1]. Estas ecuaciones resultan de promediar en la dirección vertical las ecuaciones de Navier-Stokes bajo las siguientes hipótesis:

- La dimensión vertical del dominio H es pequeña comparada con la horizontal L,
$$\frac{H}{L} \ll 1.$$
- Se asume que la presión del fluido tiene un comportamiento hidrostático.

UNIVERSIDAD DE MÁLAGA

- Las variaciones verticales de la componentes horizontales de la velocidad son despreciables y se pueden suponer uniformes en vertical.

El sistema de ecuaciones en derivadas parciales hiperbólico resultante ha sido ampliamente utilizado para resolver toda clase de problemas hidrodinámicos [2–4]. La principal ventaja de las ecuaciones de aguas someras es que permiten reducir en una unidad la dimensión del problema, gracias a la simplificación en la dirección vertical, permitiendo una significativa reducción del esfuerzo computacional.

Sin embargo, esta homogeneización en la dirección vertical también puede suponer una desventaja importante, puesto que buena parte de la información relativa al flujo vertical se pierde. Para sobreponerse a esta limitación se han desarrollado en los últimos años las ecuaciones de aguas someras multicapa. Existen numerosas formas de describir estos modelos, aunque destacaríamos la propuesta por Audusse et. al. in [5]. En este trabajo, el dominio vertical es discretizado en un conjunto de capas, y en cada una de estas capas se usan las hipótesis de aguas someras, admitiendo que la solución puede presentar discontinuidades al cambiar de una capa a la contigua. El intercambio de masa y momento entre capas se logra mediante la incorporación al sistema de ecuaciones de una aproximación del flujo vertical bajo la forma de productos no conservativos. De esta forma se logra obtener un rico perfil vertical del flujo hidrodinámico a coste de incrementar la complejidad del modelo y, por tanto, la carga computacional de forma proporcional al número de capas considerado. De hecho, avances recientes en el desarrollo de sistemas de aguas someras multicapa comprenden un proceso de adaptación vertical en lo referente al número de capas, de acuerdo con la relevancia de los efectos verticales presentes, (ver [6]), lo que permite reducir la carga computacional a la vez que se recupera la información vertical relevante para el problema. En cualquier caso, los sistemas de aguas someras multicapa permiten recuperar efectos que los sistemas de aguas someras tradicionales no pueden (ver [7–10]). Se han aplicado con éxito por distintos autores en años recientes para flujos geofísicos como tsunamis (ver [3, 11–14]), inundaciones (ver [15–19]), o marejadas ciclónicas (*storm-surge* en inglés) (ver [20–22]).

Hemos resaltado cómo los sistemas de aguas someras multicapa permiten recuperar complejos perfiles verticales. Sin embargo, si estos efectos verticales están causados por fenómenos asociados a variaciones en la densidad del fluido, la aproximación estándar de los sistemas de aguas someras multicapa fracasa, puesto que estos efectos no están presentes en las hipótesis de partida. No obstante, las variaciones en la densidad sí pueden jugar un rol significativo en determinados flujos geofísicos. Un ejemplo paradigmático lo podemos encontrar en el estrecho de Gibraltar, donde dos corrientes, una proveniente del Océano Atlántico y la otra del Mar Mediterráneo, se encuentran. Estas dos corrientes son distintas en términos de temperatura y salinidad, por lo que sus respectivas densidades relativas juegan un rol crucial en su interacción. Efectivamente, el Mar Mediterráneo es más cálido y salino que el Océano Atlántico, teniendo por tanto una densidad mayor debido a una mayor evaporación. Por consiguiente, la corriente asociada al Mar Mediterráneo es más densa y tiende a fluir bajo la corriente correspondiente al Océano

UNIVERSIDAD DE MÁLAGA

Atlántico. Capturar este tipo de comportamiento es imposible para sistemas de aguas someras que no tengan en cuenta la densidad de alguna forma.

Para afrontar este problema, se han desarrollado recientemente modelos de aguas someras multicapa que incorporan efectos asociados a las variaciones de densidad. Por ejemplo, en [34, 35] se propone un sistema de aguas someras multicapa que depende de una especie sedimentaria con su propia velocidad de sedimentación. En [36] se propone un sistema de aguas someras de dos capas donde el intercambio entre capas es aproximada por un término fuente heurístico. Podemos encontrar otro ejemplo en [37], donde los autores derivan una aproximación semi-implícita en tiempo para un sistema de aguas someras con una cantidad de capas variable para mejorar la flexibilidad y eficiencia del modelo.

El modelo propuesto para esta tesis hace uso de un trazador dado que es transportado por advección por el fluido. Este trazador está definido en términos de la densidad relativa de tal forma que los términos de presión dependen de él. El modelo difiere de propuestas similares, como [38], en el proceso para derivar el sistema de aguas someras multicapa, lo que resulta en una definición distinta para los términos de transferencia.

A medida que el modelo incorpora más efectos físicos, el coste computacional asociado se incrementa. Además, el sistema de aguas someras multicapa de densidad variable asociado resulta muy sensible a pequeños cambios en la densidad relativa. Por este motivo, cualquier discretización numérica de este modelo debe ser eficiente a la par que robusta. En esta tesis se proponen dos discretizaciones numéricas distintas: una basada en la técnica de volúmenes finitos y la otra en el método de Galerkin discontinuo.

El marco de los volúmenes finitos es especialmente adecuado para aproximar las soluciones de sistemas hiperbólicos no lineales, en las que pueden aparecer discontinuidades en tiempo finito. Una de las principales dificultades que surgen cuando se consideran leyes de equilibrio con términos fuentes singulares, o sistemas hiperbólicos con productos no conservativos es la propia definición de solución débil y su aproximación mediante un método numérico. Este tipo de sistemas aparecen de forma usual en el modelado y simulación de fluidos geofísicos. En particular, el sistema que consideramos en esta tesis se enmarca en esta categoría. Nosotros vamos a considerar el marco propuesto por Dal Maso, LeFloch y Murat en [39] para la definición de las soluciones débiles de sistemas hiperbólicos no conservativos. En este marco, los productos no conservativos se interpretan como una medida de Borel definida en términos de una elección de caminos que permite conectar las discontinuidades de salto finito que aparecen. Desde el punto de vista del Análisis Numérico, nosotros vamos a considerar los esquemas camino-conservativos introducidos por C. Parés en [40]. Estos esquemas son formalmente consistentes con la elección de caminos realizada para definir las soluciones débiles del problema. En este tipo de problemas, la elección de caminos es un problema y debe estar estrechamente ligada a la física del problema. Sin embargo, la correcta elección del camino adecuado para cada problema puede resultar excesivamente complicado. En la práctica, para la gran mayoría de los casos la simple elección de una familia de segmentos es suficiente

UNIVERSIDAD DE MÁLAGA

para producir resultados satisfactorios. Una discusión en profundidad sobre la correcta elección de caminos puede encontrarse en [41].

En realidad, los esquemas numéricos camino-conservativos pueden considerarse como la generalización natural de los esquemas conservativos para los sistemas hiperbólicos con productos no conservativos. En [41] se muestran ejemplos de extensión de diversos esquemas conservativos a problemas no conservativos.

Uno de los resolvedores de Riemann más populares es el método de Roe, que fue extendido a problemas no conservativos por Toumi (véase [42],[45]. Aunque el método de Roe puede proporcionar soluciones satisfactorias (ver por ejemplo [43–45]), exigen el conocimiento explícito de la estructura espectral del sistema, algo que no siempre es alcanzable. Una alternativa consiste en aproximar numéricamente los autovalores del sistema asumiendo el coste computacional extra, lo cual no es siempre posible. Es más, estos esquemas en general no satisfacen una desigualdad de entropía y por tanto se hace necesario incorporar algún tipo de técnica para capturar la solución entrópica en presencia de transiciones (ver [46]). Una alternativa consiste en el uso de Resolvedores de Riemann incompletos que no necesitan la estructura espectral del problema, siendo suficiente una aproximación de las ondas que componen la solución del problema de Riemann. Estos resolvedores suelen ser más eficientes desde el punto de vista computacional pero pueden ser más difusivos, lo cual hace necesario su extensión a alto orden. Los resolvedores de Riemann incompletos de alto orden son la solución más eficiente en general.

En esta tesis vamos a utilizar los llamados métodos *Polinomial Viscosity Matrix* (PVM), introducidos por Castro y Fernández en [48]. Estos métodos permiten la obtención de resolvedores de Riemann incompletos basados en la definición de una matriz de viscosidad que se define mediante una evaluación adecuada de una linealización de Roe. Estos métodos pueden verse como una generalización de los propuestos por Russo y colaboradores en [49]. Además, hay un buen número de resolvedores de Riemann usuales que se pueden interpretar como métodos de tipo PVM. Finalmente, los métodos PVM admiten extensión a alto orden (ver [50]) y aplicaciones en dominios multidimensionales basadas en una reconstrucción de estados (véase [51]).

La otra discretización numérica considerada en esta tesis se basa en la familia de esquemas numéricos de tipo Galerkin Discontinuo (DG, por sus siglas en inglés). Estos métodos datan de los trabajos de Reed y Hill en [52] y fueron subsecuentemente desarrollados por Cockburn y Shu que proveyeron de un marco teórico sólido para sistemas hiperbólicos de leyes de conservación no lineales en una serie de conocidas publicaciones como [53–56] entre otras.

Los métodos numéricos basados en técnicas de tipo DG tienen ventajas e inconvenientes con respecto al método de volúmenes finitos. Ciertamente, mientras que los métodos de volúmenes finitos se basan en considerar una aproximación constante a trozos en la celda y definen un operador de reconstrucción para alcanzar alto orden, los métodos tipo DG aproximan la solución mediante una función polinómica a trozos. De esta forma, alcanzar alto orden en espacio con esquemas de tipo DG es tan fácil como incrementar

UNIVERSIDAD DE MÁLAGA

el orden de dicho polinomio. Asimismo, existen varias formas de alcanzar alto orden en tiempo. Probablemente la más natural consiste en considerar el método de líneas y discretizar el subsiguiente sistema de ecuaciones diferenciales ordinarias con algún método explícito en tiempo tipo Runge-Kutta, lo que lleva a la familia de esquemas Runge-Kutta DG [53–56]. Otra alternativa es usar métodos DG en espacio-tiempo donde se emplean funciones base en espacio tiempo y funciones test. Los primeros trabajos que describen estas técnicas se remontan a [57–67], resultando esquemas numéricos que son condicionalmente estables bajo una adecuada restricción CFL. Otra aproximación para obtener esquemas numéricos de tipo DG en tiempo es mediante el método de Cauchy-Kovaleskaya, desarrollados fundamentalmente en el marco de los sistemas hiperbólicos, por la escuela de Trento: [68–73]. Esta técnica usa una expansión de Taylor de las variables con el objetivo de sustituir derivadas temporales por derivadas espaciales, que pueden ser tratadas por el esquema numérico DG. Sin embargo, el cálculo de los desarrollos de Taylor puede convertirse rápidamente en algo excesivamente complejo. En su lugar, el predictor local en espacio-tiempo ADER DG puede utilizarse para alcanzar orden arbitrario en espacio y tiempo en un solo paso, evitando completamente el procedimiento de Cauchy-Kovaleskaya (ver [74, 75]). La predicción de alto orden proporcionada por el procedimiento ADER se basa en la aproximación del problema localmente en la celda utilizando para ellos polinomios en espacio tiempo. El sistema no lineal resultante se resuelve por medio de un algoritmo de punto fijo local para el elemento y cuya convergencia fue probada en [76]. De esta forma, la combinación de la técnica ADER con el método DG resulta en una nueva familia de esquemas numéricos ADER-DG de orden arbitrariamente alto en espacio y tiempo.

Los esquemas numéricos de tipo DG y ADER-DG han sido aplicados con éxito a flujos geofísicos en el marco de las ecuaciones de aguas someras. Algunos trabajos relacionados con las ecuaciones de aguas someras en una capa pueden encontrarse en [77–80]. Además, aplicaciones de esta técnica para ecuaciones descritas en dos capas inmiscibles pueden encontrarse en [81–83]. En cuanto a esquemas multicapa, el lector puede referirse a [84]. Avances más recientes en reformulaciones hiperbólicas de ecuaciones de aguas someras dispersivas no lineales y que hacen uso de la técnica DG o ADER-DG pueden encontrarse en [85–88].

La técnica DG se ha descrito previamente como una técnica de alto orden y da lugar, en general, a la resolución de problemas no lineales. El hecho de que sea un método de alto orden puede convertirse en un problema en presencia de discontinuidades o en zonas con gradientes muy fuertes de la solución. Ciertamente, el conocido teorema de Godunov establece que esquemas de alto orden perderán monotonía cerca de discontinuidades, y por tanto desarrollarán oscilaciones espurias. La solución por tanto debe corregirse convenientemente. En la literatura podemos encontrar diversas técnicas que permiten controlar la aparición de oscilaciones espurias en presencia de discontinuidades. Por ejemplo, es posible usar un detector de orden óptimo multidimensional (MOOD, ver [89, 90]), que consiste en una aproximación *a posteriori* al problema de la limitación.

UNIVERSIDAD DE MÁLAGA

Con esta técnica, la solución sin limitar del esquema DG o ADER-DG se considera como una solución candidata, y es posteriormente analizada para determinar su idoneidad con respecto a un determinado criterio. Si la solución candidata se considera adecuada, entonces pasa a considerarse como la solución final. Por contra, si la solución es considerada insatisfactoria, entonces la solución candidata se proyecta en la propia celda en una función constante a trozos que se evoluciona en tiempo usando un esquema de volúmenes finitos robusto. Esta solución será finalmente proyectada en un polinomio y pasará a considerarse como la solución final limitada. Por supuesto, el uso de un esquema de volúmenes finitos puede destruir la resolución en cada celda de los esquemas DG. Es por eso, que es necesario considerar una proyección adecuada en una función constante a trozos en cada celda, considerando para ello una submalla local con un paso de espacio adecuado (véase [91]). Una ventaja crucial del limitador MOOD es que permite analizar la solución candidata para cualquier número de propiedades físicas o numéricas, asegurándose de esta forma la coherencia y consistencia de la solución final. Por ejemplo, para detectar oscilaciones espurias asociadas al alto orden es posible considerar un principio del máximo relajado discreto. Además, otras propiedades, como por ejemplo la positividad, pueden imponerse de esta forma. La correcta calibración de estos criterios es muy importante para determinar la admisibilidad de la solución candidata.

Esta no es la única técnica de limitación que existe en el marco DG. Ciertamente, otras estrategias para limitar incluyen el limitador de variación total acotada (TVD, véase [53, 92]), el limitador basado en momento [93] o el limitador basado en momento mejorado [94]. De particular interés resulta la metodología WENO aplicada a la familia DG de esquemas numéricos, principalmente desarrollado por Shu, Qiu and Zhu *et. al* (véase [95–97]). De forma similar a la versión en volúmenes finitos, la técnica WENO para métodos tipo DG aplica una combinación convexa de la solución en una celda computacional y sus vecinos inmediatos. De esta forma las oscilaciones espurias pueden ser eliminadas por el efecto regularizador de considerar una combinación no lineal de las soluciones aproximadas en las celdas vecinas.

Otro problema importante en la simulación numérica en general y la simulación de flujos geofísicos, en particular, es la preservación de soluciones estacionarias. Efectivamente, en la práctica muchas simulaciones numéricas consisten en esencia en la perturbación de un estado de equilibrio. Es más, la convergencia de las soluciones para tiempos largos es a menudo una solución de equilibrio. Por tanto, los esquemas numéricos que preserven de forma exacta (o con una precisión elevada) este tipo de soluciones son de gran utilidad práctica. Un argumento adicional a favor de los esquemas numéricos que preservan soluciones estacionarias surge cuando nos encontramos en el caso de pequeñas perturbaciones de los mencionados estados estacionarios. En estos casos, el ruido numérico introducido por el esquema puede destruir la perturbación que buscábamos simular. Por supuesto, una primera aproximación para resolver este problema puede ser incrementar el orden del método o el número total de elementos. Aunque esta estrategia puede mejorar el comportamiento numérico, no afronta el problema real y puede volverse rápidamente

UNIVERSIDAD DE MÁLAGA

excesivamente costosa computacionalmente.

Los esquemas numéricos que preservan soluciones estacionarias se denominan esquemas exactamente bien equilibrados(o *exactly well-balanced* en inglés). Esta propiedad fue introducida por primera vez por Bermúdez y Vázquez Cendón en [44] para las ecuaciones de aguas someras, donde definieron la *propiedad C*, o la habilidad de un esquema numérico para preservar de forma exacta la soluciones de agua en reposo. Desde entonces, los esquemas numéricos bien equilibrados han sido un campo activo de investigación [10, 15, 98–127]. Recientemente, en [128] los autores propusieron un marco general para construir esquemas de alto orden bien equilibrados para sistemas de leyes de equilibrio en el marco de los volúmenes finitos.

Al igual que en [128], en esta tesis desarrolla una metodología general para definir esquemas numéricos de tipo Galerkin discontinuo de orden alto y bien equilibrados para sistemas de leyes de equilibrio. Asimismo, se realiza un estudio de algunas de las soluciones estacionarias más relevantes para el sistema de aguas someras con densidad variable considerado, en particular las que corresponden a fluidos estratificados. Existen trabajos previos en los que se proponen esquemas de tipo DG que son bien equilibrados para diferentes leyes de equilibrio (véase [129–140]), pero se centran principalmente en un modelo particular, normalmente en las ecuaciones de aguas someras. Nosotros proponemos aquí un procedimiento general que permite abordar con éxito la construcción de esquemas DG bien equilibrados para leyes de equilibrio unidimensionales. Esta técnica también puede extenderse a problemas multidimensionales.

Finalmente, un recurso adicional para mejorar el tiempo computacional del esquema numérico es la optimización de código y la paralelización por *hardware*. Específicamente, las unidades de procesamiento gráfico (GPU, por sus siglas en inglés) modernas ofrecen la alternativa más eficiente en términos de coste-beneficio para obtener bajos tiempos de computación para dominios grandes en espacio y tiempo. Las GPUs ofrecen enormes oportunidades de paralelización gracias a los cientos de unidades de procesamiento optimizadas para ejecutar operaciones en coma flotante y de ejecución multihebra. Esta particular arquitectura permite obtener tiempos de computación que son órdenes de magnitud menores que los pertenecientes a las tradiciones unidades centrales de procesamiento (CPU). La implementación de esquemas numéricos en GPUs ha sido ampliamente usado, también para flujos geofísicos (ver [141]). En esta tesis, se incluye un breve estudio sobre diferentes estrategias para la paralelización en GPU del modelo de aguas someras con densidad variable considerado.

Esta tesis se apoya en las siguientes publicaciones:

- Guerrero Fernández E, Castro Díaz MJ, Morales de Luna T. A Second-Order Well-Balanced Finite Volume Scheme for the Multilayer Shallow Water Model with Variable Density. Mathematics. 2020; 8(5):848. https://doi.org/10.3390/math8050848

- Guerrero Fernández E, Castro Díaz MJ, Dumbser M, Morales de Luna T. An Arbitrary High Order Well-Balanced ADER-DG Numerical Scheme for the

UNIVERSIDAD DE MÁLAGA

Multilayer Shallow-Water Model with Variable Density. J Sci Comput. 2022; 90:52. https://doi.org/10.1007/s10915-021-01734-2

- Guerrero Fernández E, Escalante C, Castro Díaz MJ. Well-Balanced High-Order Discontinuous Galerkin Methods for Systems of Balance Laws. Mathematics. 2022; 10(1):15. https://doi.org/10.3390/math10010015

A continuación, se realiza un resumen de la tesis en español. Nótese que esta síntesis no persigue ser exhaustiva, sino presentar los resultados principales que se han alcanzado durante la tesis. En la versión en inglés se presentan los resultados con mayor detalle.

Capítulo 1: Derivación del modelo y soluciones estacionarias

En este capítulo se detalla la derivación de un modelo de aguas someras multicapa con densidad variable. Estos modelos son adecuados para simular flujos geofísicos donde las variaciones de densidad jueguen un papel importante en la hidrodinámica del fluido. Para derivar el modelo se parte de las ecuaciones de Navier-Stokes con gravedad y se considera además una ecuación para un trazador que se transporta con el fluido a la vez que se difunde en el mismo. Estas ecuaciones son integradas en la dirección vertical para reducir la dimensión del problema.

Posteriormente se considera una ley de presión y una ecuación de estado particular para deducir el modelo particular que acabará discretizándose. Finalmente, también se identifican una familia de soluciones estacionarias que son de especial interés: aquellas con velocidad nula y que presentan un perfil estratificado en densidad.

Tanto la escritura general del modelo como el estudio de las soluciones estacionarias estratificadas constituyen una aportación novedosa de esta tesis, parcialmente recogida en una publicación (véase [100]).

Derivación del modelo

Las ecuaciones de Navier-Stokes compresibles en un espacio de dimensión $d = 2, 3$ son,

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \boldsymbol{v}) = 0, \\ \partial_t (\rho\, \boldsymbol{v}) + \nabla \cdot (\rho \boldsymbol{v} \otimes \boldsymbol{v}) = -g\, \rho \boldsymbol{k} + \nabla \cdot \boldsymbol{\Sigma}, \end{cases} \tag{1}$$

donde $\boldsymbol{v} = (\boldsymbol{u}, w) \in \mathbb{R}^d$ es la función velocidad con $\boldsymbol{u} \in \mathbb{R}^{d-1}$ la velocidad horizontal y $w \in \mathbb{R}$ la vertical. Asimismo, $\boldsymbol{k} \in \mathbb{R}^d$ es el vector unitario, $g \in \mathbb{R}$ es la constante de la gravedad y el tensor de tensiones total viene dado por $\boldsymbol{\Sigma} = -p\boldsymbol{I} + \boldsymbol{\sigma}$, con $\boldsymbol{I}$ el tensor identidad, $p \in \mathbb{R}$ la presión y $\boldsymbol{\sigma}$ el tensor correspondiente a los esfuerzos viscosos.

UNIVERSIDAD DE MÁLAGA

Finalmente, $\rho \in \mathbb{R}$ es la función de densidad no constante. Esta función depende de una ecuación de estado dada R para poder reflejar los cambios en la densidad debidos a la temperatura y/o salinidad. Estos efectos se tienen en cuenta gracias a una función trazadora $T = T(t, \boldsymbol{x}, z)$, con $\boldsymbol{x} = (x_1, \dots, x_{d-1}) \in \mathbb{R}^{d-1}$ las coordenadas horizontales, que es transportada y difundida con el flujo,

$$\partial_t(\rho T) + \nabla \cdot (\rho T \boldsymbol{v}) + \nabla \cdot (\rho \nu_T \nabla T) = 0, \tag{2}$$

y

$$\rho = R(T). \tag{3}$$

Nótese que en el caso en el que la densidad del fluido solo dependa de la temperatura o la salinidad, entonces el trazador T representa la temperatura o salinidad respectivamente.

El volumen fluido se divide en una serie de capas en la dirección vertical, $\alpha = 1, \dots, M$. La presión se asume hidrostática,

$$p_\alpha(t, x, z) = p_{\alpha+\frac{1}{2}} + \rho_\alpha \, g \, (z_{\alpha+\frac{1}{2}} - z), \tag{4}$$

con

$$p_{\alpha+\frac{1}{2}}(t, x) = p_S(t, x) + g \sum_{\beta=\alpha+1}^{M} \rho_\beta h_\beta(t, x), \tag{5}$$

donde $p_{\alpha+\frac{1}{2}}$ es la presión en la interfaz entre dos capas $\Gamma_{\alpha+\frac{1}{2}}(t)$ y p_S denota la presión en la superficie libre, normalmente considerada como cero.

Ahora siguiendo el procedimiento descrito en el capítulo 1 o en [163], si multiplicamos el sistema (1)-(2) por una función test e integramos en la dirección vertical obtenemos el siguiente modelo multicapa de aguas someras,

$$\begin{cases} \partial_t(h_\alpha \rho_\alpha) + \nabla_x(h_\alpha \rho_\alpha \boldsymbol{u}_\alpha) = G_{\alpha+\frac{1}{2}} - G_{\alpha-\frac{1}{2}}, \\ \\ \partial_t(h_\alpha \rho_\alpha T_\alpha) + \nabla_x(h_\alpha \rho_\alpha T_\alpha \boldsymbol{u}_\alpha) - \nabla_x \left(\nu_T h_\alpha \rho_\alpha \nabla_x T_\alpha\right) + \nu_T \left(K_{T,\alpha+\frac{1}{2}} - K_{T,\alpha-\frac{1}{2}}\right) \\ = \left(\dfrac{T_{\alpha+1} + T_\alpha}{2}\right) G_{\alpha+\frac{1}{2}} - \left(\dfrac{T_\alpha + T_{\alpha-1}}{2}\right) G_{\alpha-\frac{1}{2}}, \\ \\ \partial_t(h_\alpha \rho_\alpha \boldsymbol{u}_\alpha) + \nabla_x(h_\alpha \rho_\alpha \boldsymbol{u}_\alpha \otimes \boldsymbol{u}_\alpha) + g h_\alpha \rho_\alpha \nabla_x \eta - \nabla_x(h_\alpha \boldsymbol{\sigma}_H) + K_{\alpha+\frac{1}{2}} - K_{\alpha-\frac{1}{2}} \\ + g h_\alpha \left(\displaystyle\sum_{\beta=\alpha+1}^{M} (\rho_\beta - \rho_\alpha) \nabla_x h_\beta \right) + g h_\alpha \displaystyle\sum_{\beta=\alpha+1}^{M} h_\beta \nabla_x \rho_\beta + \frac{1}{2} g h_\alpha^2 \nabla_x \rho_\alpha \\ = \left(\dfrac{\boldsymbol{u}_{\alpha+1} + \boldsymbol{u}_\alpha}{2}\right) G_{\alpha+\frac{1}{2}} - \left(\dfrac{\boldsymbol{u}_\alpha + \boldsymbol{u}_{\alpha-1}}{2}\right) G_{\alpha-\frac{1}{2}}. \end{cases} \tag{6}$$

con

$$\rho_\alpha = R(T_\alpha), \tag{7}$$

UNIVERSIDAD DE MÁLAGA

donde $R(T)$ es una función dependiente del trazador T que se corresponde con la ecuación de estado.

Pero estas ecuaciones pueden simplificarse si consideramos un fluido incompresible y una ecuación de estado particular donde $R(T) = \rho_0 T$ e identificando el trazador T con θ, la densidad relativa, que se define como

$$\theta = \frac{\rho}{\rho_0} = 1 + \frac{\rho_1}{\rho_0}, \tag{8}$$

con ρ definida como la suma de una densidad de referencia, ρ_0, y una fluctuación respecto a esa referencia, ρ_1.

De esta forma, el modelo final puede resulta:

$$\begin{cases} \partial_t h + \partial_x \left(h \sum_{\beta=1}^{M} l_\beta u_\beta \right) = 0, \\ \partial_t (h\theta_\alpha) + \partial_x (h\theta_\alpha u_\alpha) = \dfrac{1}{l_\alpha} \left(\theta_{\alpha+\frac{1}{2}} G_{\alpha+\frac{1}{2}} - \theta_{\alpha-\frac{1}{2}} G_{\alpha-\frac{1}{2}} \right), \\ \partial_t (h\theta_\alpha u_\alpha) + \partial_x (h\theta_\alpha u_\alpha^2) + gh\theta_\alpha \partial_x \eta + \dfrac{g l_\alpha}{2} \left(h\partial_x (h\theta_\alpha) - h\theta_\alpha \partial_x h \right) \\ + g \sum_{\beta=\alpha+1}^{M} l_\beta \left(h\partial_x (h\theta_\beta) - h\theta_\alpha \partial_x h \right) = \dfrac{1}{l_\alpha} \left(u_{\alpha+\frac{1}{2}} \theta_{\alpha+\frac{1}{2}} G_{\alpha+\frac{1}{2}} - u_{\alpha-\frac{1}{2}} \theta_{\alpha-\frac{1}{2}} G_{\alpha-\frac{1}{2}} \right), \end{cases} \tag{9}$$

donde $\theta_{\alpha+\frac{1}{2}}$ y $u_{\alpha+\frac{1}{2}}$ son la media aritmética en la interfaz $\Gamma_{\alpha+\frac{1}{2}}(t)$, $\alpha = 1, \ldots, M-1$:

$$u_{\alpha+\frac{1}{2}} = \frac{u_{\alpha+1} + u_\alpha}{2}, \qquad \theta_{\alpha+\frac{1}{2}} = \frac{\theta_{\alpha+1} + \theta_\alpha}{2}, \qquad \alpha = 1, \ldots, M-1.$$

Finalmente, los términos de transferencia de masa entre capas, $G_{\alpha\pm\frac{1}{2}}$ se definen mediante la siguiente expresión,

$$G_{\alpha+\frac{1}{2}} = \sum_{\beta=1}^{\alpha} l_\beta \left(\partial_t h + \partial_x (h u_\beta) \right) = \sum_{\beta=1}^{\alpha} l_\beta \left(\partial_x (h u_\beta) - \partial_x \left(h \sum_{\gamma=1}^{M} l_\gamma u_\gamma \right) \right). \tag{10}$$

Soluciones estacionarias

Las soluciones estacionarias con velocidad nula del modelo anterior son soluciones de la siguiente ecuación diferencial ordinaria,

$$\frac{l_\alpha}{2} (\theta_\alpha)' + \sum_{\beta=\alpha+1}^{M} l_\beta (\theta_\beta)' = -\frac{1}{h} \left(\theta_\alpha (h+b)' + \sum_{\beta=\alpha+1}^{M} l_\beta (\theta_\beta - \theta_\alpha) h' \right), \quad \alpha = 1, \ldots, M. \tag{11}$$

UNIVERSIDAD DE MÁLAGA

Esta ecuación tiene un tipo de solución trivial que se corresponde con una densidad relativa constante y una superficie libre constante. Sin embargo, tiene otro tipo de soluciones no triviales donde la densidad relativa no es constante, aunque la superficie libre sí lo sea. Si imponemos estas condiciones sobre la EDO (11) obtenemos la siguiente familia de soluciones estacionarias:

$$\begin{gathered} u_\alpha = 0,\ \eta(x) = z_b(x) + h(x) = \text{cte}, \\ \theta_M(x) = \bar{\theta}_M \geq 1, \\ \theta_\alpha(x) = \bar{\theta}_\alpha\, h^{2(M-\alpha)}(x) + \sum_{\beta=\alpha+1}^{M} S_{2(M-\beta)}(M-\alpha+1)\bar{\theta}_\beta\, h^{2(M-\beta)}(x), \end{gathered} \tag{12}$$

con

$$\begin{aligned} S_\beta(\alpha) &= (\beta+1)\cdot A_{\frac{\beta+2}{2}+1}(\alpha), \\ A_p(k) &= \begin{cases} 1 & \text{si } p \geq k, \\ (p-1)\prod_{\gamma=2}^{k-p}(1+(p-2)C_{\gamma-1}) & \text{si } p < k, \end{cases} \\ C_\gamma &= C_{\gamma-1} - \frac{1}{Q_\gamma}, \\ Q_\gamma &= Q_{\gamma-1} + \gamma + 1, \\ C_0 &= Q_0 = 1. \end{aligned}$$

Preservar este tipo de soluciones será un objetivo crucial de los esquemas numéricos diseñados.

Capítulo 2: Esquemas numéricos de volúmenes finitos y de Galerkin discontinuo de alto orden bien equilibrados

En este capítulo se presenta una metodología general para el diseño de esquemas numéricos de volúmenes finitos y Galerkin discontinuo de alto orden y bien equilibrados. Estos métodos son ideales para sistemas hiperbólicos no lineales con términos fuentes y productos no conservativos que aparecen en diversos contextos: sistemas de aguas someras, dinámica de gases, flujos multifase, o modelos de aguas someras dispersivos, por ejemplo. Estos sistemas a menudo contienen productos no conservativos, lo que incrementa considerablemente la dificultad tanto desde el punto de vista teórico como continuo. Nosotros usamos el marco teórico desarrollado por Dal Maso, LeFloch y Murat en [39], donde las soluciones débiles se interpretan en términos de medidas de Borel que

UNIVERSIDAD DE MÁLAGA

se definen a partir de la elección de una familia de caminos que conectan los estados que definen una discontinuidad admisible.

Desde el punto de vista del análisis numérico, nosotros consideramos el marco de los esquemas camino-conservativos propuesto por C. Parés [40]. Este marco permite la extensión natural de resolvedores de Riemann convencionales para leyes de conservación a sistemas hiperbólicos con términos fuentes y productos no conservativos. Además, este marco también permite la extensión a alto orden de estos esquemas. Desde su introducción, los esquemas numéricos camino-conservativos han sido ampliamente utilizados en el marco de los volúmenes finitos y de los esquemas de tipo Galerkin discontinuo. En este capítulo hacemos un repaso de algunos de los resolvedores usuales y su extensión a sistemas hiperbólicos no-conservativos. En particular se presta especial atención a los métodos denominados *Polynomial Viscosity Methods* o PVM, y su extensión a alto orden. Además se describe la técnica general para la construcción de esquemas bien equilibrados de alto orden propuesta en [128]

También se abordan en este capítulo métodos numéricos de tipo Galerkin discontinuo. Aunque las herramientas asociadas a los problemas de Riemann son también útiles para esquemas de tipo Galerkin discontinuo, se necesitan nuevas estrategias para lidiar con esquemas numéricos que son fundamentalmente distintos a los métodos de volúmenes finitos. En particular, se necesitarán nuevas técnicas de limitación para evitar las oscilaciones espurias en presencia de discontinuidades, así como modificar adecuadamente los esquemas para que sean bien equilibrados.

Puesto que la mayoría de los resultados expuestos en este capítulo pueden encontrarse en la literatura, no forman parte de este resumen. No obstante, la derivación de esquemas numéricos de tipo Galerkin discontinuo bien equilibrados es una contribución original de esta tesis y procedemos a describirla brevemente. Estos resultados han sido recogidos en una publicación que puede consultarse en [204].

En primer lugar, consideramos el sistema hiperbólico no lineal con término fuente y productos no conservativos,

$$\partial_t w + \partial_x F(w) + B(w)\partial_x w = G(w)\sigma_x. \tag{13}$$

Como es usual, discretizamos el dominio computacional I en un conjunto de celdas I_i, siendo N_s el número total de las mismas. Denotamos por $\boldsymbol{w}_h(x,t^n)$ la aproximación de la solución del sistema (13). $\boldsymbol{w}_h(x,t^n)$ es una función polinómica a trozos de grado N_p que es continua en cada celda I_i pero posiblemente discontinua en las interceldas. Denotamos por $\mathcal{U}_h$ el espacio de las funciones polinómicas a trozos en cada I_i de grado N_p, de forma que $\boldsymbol{w}_h(\cdot,t^n) \in \mathcal{U}_h$. Estos polinomios están definidos en función de una base de Lagrange que interpola los N_p+1 puntos de cuadratura de Gauss-Legendre en el elemento I_i. De esta forma, la solución discreta se puede escribir en esta base $\Phi_l(x)$ siendo $\hat{\boldsymbol{w}}_{i,l}^n$ los diferentes grados de libertad

$$w_h(x,t^n) = \sum_l \hat{w}_{i,l}^n \Phi_l(x) \coloneqq \hat{w}_{i,l}^n \Phi_l(x), \qquad \text{con } x \in I_i, \tag{14}$$

UNIVERSIDAD DE MÁLAGA

donde se utiliza la notación de Einstein sobre dos índices repetidos.

Así, multiplicando (13) por una función test e integrando en la celda I_i resulta la siguiente formulación débil:

$$\int_{I_i} \Phi_k \frac{d}{dt} w_h \, dx + \int_{I_i} \Phi_k \left(\partial_x F(w_h) + B(w_h) \partial_x w_h \right) \, dx = \int_{I_i} \Phi_k G(w_h) \sigma_x \, dx. \tag{15}$$

Utilizando (14) e integrando por partes, esta expresión puede desarrollarse resultando en siguiente esquema numérico semi discreto:

$$\begin{aligned} \int_{I_i} \Phi_k \Phi_l \frac{d}{dt} \hat{w}_{i,l} \, dx - \int_{I_i^\circ} \Phi_k' F(w_h) \, dx \\ + \Phi_{k,i+\frac{1}{2}} D^-_{i+\frac{1}{2}} \left(w^-_{h,i+\frac{1}{2}}, w^+_{h,i+\frac{1}{2}} \right) - \Phi_{k,i-\frac{1}{2}} D^+_{i-\frac{1}{2}} \left(w^-_{h,i-\frac{1}{2}}, w^+_{h,i-\frac{1}{2}} \right) \\ + \int_{I_i^\circ} \Phi_k \left(B(w_h) \partial_x w_h \right) \, dx = \int_{I_i^\circ} \Phi_k G(w_h) \sigma_x \, dx. \end{aligned} \tag{16}$$

Aquí, $f^{\pm}_{i+\frac{1}{2}}$ denota la evaluación de f por la derecha e izquierda de la intercelda $x_{i+\frac{1}{2}}$. Asimismo, $\Phi_{k,i\pm\frac{1}{2}}$ representa la evaluación de la función base en la interfaz del elemento. Puesto que la solución tiene permitido saltar entre celdas, es natural aproximar este flujo numérico mediante un resolvedor de Riemann aproximado, denotado por $D^{\mp}_{i\pm\frac{1}{2}}$.

Este esquema semidiscreto puede ser discretizado en tiempo de varias formas. Es posible considerar el método de líneas e integrar el sistema de EDO (16) mediante métodos clásicos, como puede ser un método de Runge-Kutta. Si por ejemplo escribimos el esquema numérico (16) de la siguiente forma,

$$\partial_t \hat{w}_{i,l}(t) = L(\hat{w}_{i,l}(t)), \tag{17}$$

donde $L(w)$ es un operador espacial, entonces el método de Runge-Kutta TVD de tercer orden se escribiría como:

$$\begin{aligned} \hat{w}^{(1)}_{i,l} &= L\left(\hat{w}^n_{i,l}\right), \\ \hat{w}^{(2)}_{i,l} &= L\left(\hat{w}^n_{i,l} + \frac{\Delta t}{2} \hat{w}^{(1)}_{i,l}\right), \\ \hat{w}^{(3)}_{i,l} &= L\left(\hat{w}^n_{i,l} - \Delta t \hat{w}^{(1)}_{i,l} + 2\Delta t \hat{w}^{(2)}_{i,l}\right), \\ \hat{w}^{n+1}_{i,l} &= \hat{w}^n_{i,l} + \frac{\Delta t}{6}\left(\hat{w}^{(1)}_{i,l} + 4\hat{w}^{(2)}_{i,l} + \hat{w}^{(3)}_{i,l}\right). \end{aligned} \tag{18}$$

Asimismo, es posible utilizar una aproximación unipaso en el marco de los esquemas ADER-DG. La técnica ADER está basada en la aproximación de la solución de un problema de Riemann mediante un algoritmo de punto fijo local. Esta aproximación

UNIVERSIDAD DE MÁLAGA

$q_h(x,t)$, denominada solución predictora, se utiliza entonces para calcular (16) con la precisión deseada. Está basada en una formulación débil del sistema (13) en la celda y sin considerar las posibles iteraciones con las celdas vecinas. Para ello, la solución aproximada se escribe en términos de una base espacio-temporal,

$$q_h(x,t) = \sum_l \theta_l(x,t)\hat{q}_l^i := \theta_l(x,t)\hat{q}_l^i. \tag{19}$$

De esta forma, el sistema general (16) se multiplica por una función test resultando en una formulación débil que es fundamentalmente distinta a (15)-(16), puesto que ahora tanto las funciones test como las funciones base dependen del espacio y del tiempo,

$$\begin{aligned}&\int_{t^n}^{t^{n+1}} \int_{I_i} \theta_k(x,t)\partial_t q_h \, dxdt \\ &+ \int_{t^n}^{t^{n+1}} \int_{I_i} \theta_k(x,t) \left(\partial_x F(q_h) + B(q_h)\partial_x q_h\right) \, dxdt = \int_{t^n}^{t^{n+1}} \int_{I_i} \theta_k(x,t) G(q_h)\sigma_x \, dxdt. \end{aligned} \tag{20}$$

Puesto que estamos interesados en una expresión local, sin interacción con los elementos vecinos, los términos de salto asociados a las discontinuidades en las interfaces de la celda no se tienen en cuenta. Estos se consideran en la etapa correctora del método ADER-DG. Así, integrando por partes obtenemos:

$$\begin{aligned}&\int_{I_i} \theta_k(x,t^{n+1}) q_h(x,t^{n+1}) \, dx - \int_{I_i} \theta_k(x,t^n) q_h^0(x,t^n) \, dx - \int_{t^n}^{t^{n+1}} \int_{I_i} \partial_t \theta_k(x,t) q_h(x,t) \, dxdt \\ &\qquad = -\int_{t^n}^{t^{n+1}} \int_{I_i} \theta_k(x,t) \left(\partial_x F(q_h) + B(q_h)\partial_x q_h - G(q_h)\sigma_x\right) \, dxdt. \end{aligned} \tag{21}$$

La resolución del sistema no lineal (21) es local a cada celda y sus incógnitas son los grados de libertad $\hat{q}_l^i$. Este sistema se puede resolver mediante un algoritmo de punto fijo. Nótese que la elección de una semilla inicial $q_h^0(x,t)$ puede tener un impacto significativo en la velocidad de convergencia del algoritmo. En este trabajo, se utiliza la elección más sencilla, $q_h^0(x,t) = w_h(x,t^n)$.

Usamos ahora la solución que acabamos de obtener para calcular (16), resultando:

$$\begin{aligned}&\int_{I_i} \Phi_k \Phi_l (\hat{w}_{i,l}^{n+1} - \hat{w}_{i,l}^n) \, dxdt - \int_{t^n}^{t^{n+1}} \int_{I_i^\circ} (\Phi_k' F(q_h)) \, dxdt \\ &\quad + \int_{t^n}^{t^{n+1}} \left(\Phi_{k,i+\frac{1}{2}} D^-_{i+\frac{1}{2}} \left(q^-_{h,i+\frac{1}{2}}, q^+_{h,i+\frac{1}{2}} \right) + \Phi_{k,i-\frac{1}{2}} D^+_{i-\frac{1}{2}} \left(q^-_{h,i-\frac{1}{2}}, q^+_{h,i-\frac{1}{2}} \right) \right) dt \\ &\qquad + \int_{t^n}^{t^{n+1}} \int_{I_i^\circ} \Phi_k \left(B(q_h)\partial_x q_h \right) \, dxdt = \int_{t^n}^{t^{n+1}} \int_{I_i} \Phi_k G(q_h)\sigma_x \, dxdt. \end{aligned} \tag{22}$$

UNIVERSIDAD DE MÁLAGA

Esquemas numéricos DG bien equilibrados

A continuación discutimos de forma somera una técnica para preservar soluciones de equilibrio dentro del marco de esquemas DG Runge-Kutta (16)-(18) y ADER-DG (21)-(22). El primer paso consiste en calcular una solución estacionaria local en cada celda computacional y en cada paso de tiempo, $w_i^*(x,t^n),\ x\in I_i$. La solución estacionaria calculada es una solución del siguiente problema de minimización,

$$\begin{cases} \partial_x F(w_i^*)+B(w_i^*)\partial_x w_i^* = G(w_i^*)\sigma_x, \quad x\in I_i, & (23)\\ \dfrac{1}{\Delta x}\displaystyle\int_{I_i} w_i^*(x)\,dx = \dfrac{1}{\Delta x}\int_{I_i} w_h(x,t^n)\,dx, & (24)\\ \text{que minimiza } \displaystyle\int_{I_i}(w_i^*(x)-w_h(x,t^n))^2\,dx. & (25)\end{cases}$$

Este es un problema difícil de resolver en general, aunque el problema anterior es abordable si conocemos la solución general de (23). En general, si la solución estacionaria $w_i^*(x)$ depende de un conjunto de parámetros, el sistema previo se reduce a un sistema de ecuaciones no lineal si las integrales en (24) se aproximan por alguna fórmula de cuadratura. Es más, en la mayor parte de las situaciones, en problemas unidimensionales, las ecuaciones (23) y (24) bastan para determina una solución estacionaria única.

Tal y como hicimos anteriormente, podemos proyectar la solución estacionaria en $\mathcal{U}_h\ x\in I_i$, así definimos

$$w_{i,h}^*(x)=\sum_l \hat{w}_{i,l}^*\Phi_l(x) \coloneqq \hat{w}_{i,l}^*\Phi_l(x), \quad \text{for } x\in I_i. \tag{26}$$

Asimismo, la fluctuación entre la solución y la solución estacionaria se denota de la siguiente forma,

$$\tilde{w}_h(x)=w_h(x)-w_{i,h}^*(x), \quad x\in I_i. \tag{27}$$

Un esquema numérico DG bien equilibrado debe asegurar que, cuando $w_h(x,t^n)=w_{i,h}^*(x)$, la solución $w_h(x,t^{n+1})$ permanece inalterada. Esto se consigue reescribiendo (16) como sigue,

$$\begin{aligned}\int_{I_i}\Phi_k\Phi_l\frac{d}{dt}\hat{w}_{i,l}\,dxdt &- \int_{I_i^\circ}\Phi_k'\Big(F(w_h)-F(w_{i,h}^*)\Big)\,dxdt\\ &+\Phi_{k,i+\frac12}\Big(\tilde{D}^-_{i+\frac12}-F(w_i^*(x_{i+\frac12}))\Big)-\Phi_{k,i-\frac12}\Big(\tilde{D}^+_{i-\frac12}-F(w_i^*(x_{i-\frac12}))\Big)\\ +\int_{I_i^\circ}\Phi_k\Big(B(w_h)\partial_x w_h - B(w_{i,h}^*)\partial_x w_{i,h}^*\Big)\,dxdt &= \int_{I_i}\Phi_k\Big(G(w_h)-G(w_{i,h}^*)\Big)\sigma_x\,dxdt. \end{aligned}\tag{28}$$

donde $\tilde{D}^\pm_{i+\frac12}$ representa el flujo numérico aplicado a los estados reconstruidos,

$$\tilde{D}^\pm_{i+\frac12}=\tilde{D}^\pm_{i+\frac12}(w_i^*(x_{i+\frac12})+\tilde{w}^-_{h,i+\frac12},\sigma_{i+\frac12},w_{i+1}^*(x_{i+\frac12})+\tilde{w}^+_{h,i+\frac12},\sigma_{i+\frac12}). \tag{29}$$

UNIVERSIDAD DE MÁLAGA

Nótese que en (29) se evalúa en los estados reconstruidos

$$w^-_{i+1/2} = w^*_i(x_{i+\frac{1}{2}}) + \tilde{w}^-_{h,i+\frac{1}{2}}, \qquad w^+_{i+1/2} = w^*_{i+1}(x_{i+\frac{1}{2}}) + \tilde{w}^+_{h,i+\frac{1}{2}},$$

donde,

$$\tilde{w}^-_{h,i+\frac{1}{2}} = \tilde{w}_{h|_{I_i}}(x_{i+\frac{1}{2}}), \qquad \tilde{w}^+_{h,i+\frac{1}{2}} = \tilde{w}_{h|_{I_{i+1}}}(x_{i+\frac{1}{2}}).$$

Es claro que el esquema resultante preserva las soluciones estacionarias y, por otro lado, sigue siendo del mismo orden que (16).

Con esto termina la descripción de los esquemas numéricos DG bien equilibrados. Si (28) se discretiza en tiempo usando un esquema de tipo Runge-Kutta, entonces no se necesitan más pasos: si el operador espacial en (17) es bien equilibrado, entonces la combinación convexa del esquema Runge-Kutta mantendrá esta propiedad. Sin embargo, si la aproximación ADER-DG se utiliza, es necesario asegurarse que la solución predictora $q_h(x,t^n)$ no destruye las propiedades de buen equilibrado del método DG. Describimos a continuación como procedemos en dicho caso.

El método ADER bien equilibrado

La aproximación para mantener el buen equilibrado para los métodos de tipo ADER es de hecho muy similar al método para esquemas DG descrito anteriormente. Buscamos ahora que si $w_h(x,t^n)$ es una solución estacionaria, entonces $q_h(x,t^n)$ permanezca inalterada. Para alcanzar esto, modificamos (21) como sigue,

$$\begin{aligned}
&\int_{I_i} \theta_k(x,t^{n+1})\tilde{q}_h(x,t^{n+1})\,dx - \int_{I_i} \theta_k(x,t^n)\tilde{q}^0_h(x,t^n)\,dx - \int_{t^n}^{t^{n+1}}\int_{I_i} \partial_t\theta_k(x,t)\tilde{q}_h(x,t)\,dxdt \\
&= -\int_{t^n}^{t^{n+1}}\int_{I_i} \theta_k(x,t)\left(\left(\partial_x F(q_h) - \partial_x F(w^*_{i,h})\right) + \left(B(q_h)\partial_x q_h - B(w^*_{i,h})\partial_x w^*_{i,h}\right)\right)dxdt \\
&\qquad + \int_{t^n}^{t^{n+1}}\int_{I_i} \theta_k(x,t)\left(G(q_h) - G(w^*_{i,h})\right)\sigma_x\,dxdt, \quad (30)
\end{aligned}$$

donde,

$$\tilde{q}_h(x,t) = q_h(x,t) - w^*_{i,h}(x), \quad x \in I_i. \tag{31}$$

Finalmente, recuperamos $q_h(x,t)$ como:

$$q_h(x,t^{n+1}) = \tilde{q}_h(x,t^{n+1}) + w^*_h(x), \quad x \in I_i. \tag{32}$$

La semilla que utilizamos para resolver el problema anterior no es más que la fluctuación en el instante t^n,

$$\tilde{q}^0_h(x,t^n) = w_h(x,t^n) - w^*_{i,h}(x,t^n), \quad x \in I_i.$$

De esta forma, estamos calculando una aproximación de alto orden de la fluctuación. Si $w_h(x,t)$ es una solución estacionaria, entonces (31) será cero y el procedimiento (30) será

UNIVERSIDAD DE MÁLAGA

cero hasta precisión máquina. La solución estacionaria se recuperaría en tal caso gracias a (32). Finalmente, la aproximación de alto orden $q_h(x, t^n)$ será usada en el esquema DG (28) para obtener un esquema numérico de orden arbitrariamente alto bien equilibrado de tipo ADER-DG.

Capítulo 3: Discretizaciones numéricas

En este capítulo se aplican las técnicas de volúmenes finitos y DG para la discretización del modelo de aguas someras con densidad variable presentado en el Capítulo 1. Hasta donde sabemos, los esquemas que hemos desarrollado son originales y se recogen en las publicaciones [100] y [205]. Este modelo es extremadamente sensible a fluctuaciones en la densidad. Resulta por tanto muy importante que el esquema numérico sea a la vez robusto y preciso y que evite las oscilaciones espurias en densidad, que pueden traducirse en oscilaciones espurias en la superficie libre y en velocidad.

En primer lugar, presentamos brevemente la aproximación numérica mediante volúmenes finitos. Está basada en un resolvedor de Riemann de tipo HLL con una reconstrucción hidrostática para preservar soluciones estacionarias de tipo agua en reposo y mejorar la estabilidad general del esquema. También se considera una aproximación de tipo *upwind* para los términos de transferencia. El segundo orden en espacio se alcanza gracias a un operador de reconstrucción MUSCL: se define un operador de reconstrucción basado en la combinación de las pendientes resultantes de relacionar el estado central y los vecinos. También se considera un limitador de tipo media armónica para limitar la solución cerca de fuertes gradientes o discontinuidades. Finalmente, se considera una estrategia para preservar una familia paramétrica de soluciones estacionarias correspondientes a una estratificación vertical de densidad en agua en reposo.

Finalmente presentamos una discretización alternativa basada en esquemas de tipo Galerkin discontinuo bien equilibrado y de orden arbitrariamente alto en espacio y tiempo. La mayor eficiencia de los esquemas de alto orden y la gran resolución espacial de los esquemas de tipo DG permiten obtener excelentes resultados desde un punto de vista de carga computacional y precisión del esquema. De esta forma, es posible capturar de forma precisa el comportamiento de la solución con mallas muy groseras. Finalmente, como en el caso de volúmenes finitos, también se incluye una estrategia para preservar una familia de soluciones estacionarias no trivial.

Un esquema de volúmenes finitos de segundo orden bien equilibrado

Consideramos el sistema hiperbólico (9) escrito en forma de un sistema hiperbólico con flujos convectivos y productos no conservativos:

$$\partial_t \boldsymbol{w} + \partial_x \boldsymbol{F}_C(\boldsymbol{w}) + \boldsymbol{P}(\boldsymbol{w}, \eta, \partial_x \boldsymbol{w}, \partial_x \eta) - \boldsymbol{T}(\boldsymbol{w}, \partial_x \boldsymbol{w}) = \boldsymbol{0}, \tag{33}$$

UNIVERSIDAD DE MÁLAGA

donde w son las variables conservadas,

$$\boldsymbol{w} = (h \,|\, h\theta_\alpha \,|\, h\theta_\alpha u_\alpha)^T \in \mathbb{R}^{2M+1}. \tag{34}$$

$\boldsymbol{F}_C(\boldsymbol{w})$ es el flujo convectivo,

$$\boldsymbol{F}_C(\boldsymbol{w}) = \left(h\sum_{\beta=1}^{M} l_\beta u_\beta \,\middle|\, h\theta_\alpha u_\alpha \,\middle|\, h\theta_\alpha u_\alpha^2 \right)^T \in \mathbb{R}^{2M+1}. \tag{35}$$

Los términos de presión, que dependen de la densidad relativa, la batimetría y la profundidad del agua están definidos por,

$$\boldsymbol{P}(\boldsymbol{w}, \eta, \partial_x \boldsymbol{w}, \partial_x \eta) = (0 \,|\, \boldsymbol{0} \,|\, P_\alpha) \in \mathbb{R}^{2M+1}, \tag{36}$$

con

$$P_\alpha = gh\theta_\alpha \partial_x \eta + \frac{gl_\alpha}{2}(h\partial_x h\theta_\alpha - h\theta_\alpha \partial_x h) + g\sum_{\beta=\alpha+1}^{M} l_\beta (h\partial_x h\theta_\beta - h\theta_\alpha \partial_x h). \tag{37}$$

Finalmente, los términos de transferencia $\boldsymbol{T}(\boldsymbol{w}, \partial_x \boldsymbol{w})$ correspondientes al intercambio de masa, densidad y momento entre capas son:

$$\boldsymbol{T}(\boldsymbol{w}, \partial_x \boldsymbol{w}) = \left(0 \,\middle|\, \frac{1}{l_\alpha}(\theta_{\alpha+\frac{1}{2}} G_{\alpha+\frac{1}{2}} - \theta_{\alpha-\frac{1}{2}} G_{\alpha-\frac{1}{2}}) \,\middle|\, \frac{1}{l_\alpha}(u_{\alpha+\frac{1}{2}}\theta_{\alpha+\frac{1}{2}} G_{\alpha+\frac{1}{2}} - u_{\alpha-\frac{1}{2}}\theta_{\alpha-\frac{1}{2}} G_{\alpha-\frac{1}{2}}) \right)^T \in \mathbb{R}^{2M+1}. \tag{38}$$

Como es habitual en el marco de los volúmenes finitos, el dominio computacional I es discretizado en una malla con celdas de igual tamaño $I_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ $i = 1, \ldots, N_s$, donde N_s es el número total de celdas con longitud constante $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$. En cada celda I_i se considera la aproximación de la solución en el instante de tiempo t^n:

$$\boldsymbol{w}_i^n \approx \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \boldsymbol{w}(x, t^n)\, dx.$$

Esquema de tipo HLL de primer orden

Como se discute en profundidad en la versión en inglés de esta tesis y siguiendo [48], consideramos el esquema numérico de tipo HLL para (33) dado por:

$$\boldsymbol{w}_i^{n+1} = \boldsymbol{w}_i^n - \frac{\Delta t}{\Delta x}\left(\boldsymbol{D}^+_{i-\frac{1}{2}}(\boldsymbol{w}_{i-1}^n, \boldsymbol{w}_i^n, z_{B,i-1}, z_{B,i}) + \boldsymbol{D}^-_{i+\frac{1}{2}}(\boldsymbol{w}_i^n, \boldsymbol{w}_{i+1}^n, z_{B,i}, z_{B,i+1}) \right), \tag{39}$$

UNIVERSIDAD DE MÁLAGA

donde

$$\boldsymbol{D}^{-}_{i+\frac{1}{2}} = \frac{1}{2}\left((1-\alpha_{1,i+\frac{1}{2}})\boldsymbol{E}_{i+\frac{1}{2}} - \alpha_{0,i+\frac{1}{2}}(\boldsymbol{w}_{i+1}-\boldsymbol{w}_i)\right) + \boldsymbol{F}_C(\boldsymbol{w}_i), \tag{40}$$

$$\boldsymbol{D}^{+}_{i+\frac{1}{2}} = \frac{1}{2}\left((1+\alpha_{1,i+\frac{1}{2}})\boldsymbol{E}_{i+\frac{1}{2}} + \alpha_{0,i+\frac{1}{2}}(\boldsymbol{w}_{i+1}-\boldsymbol{w}_i)\right) - \boldsymbol{F}_C(\boldsymbol{w}_{i+1}), \tag{41}$$

y

$$\boldsymbol{E}_{i+\frac{1}{2}} = \boldsymbol{F}_C(\boldsymbol{w}_{i+1}) - \boldsymbol{F}_C(\boldsymbol{w}_i) + \boldsymbol{P}_{i+\frac{1}{2}} - \boldsymbol{T}_{i+\frac{1}{2}}.$$

Aquí, $\boldsymbol{P}_{i+\frac{1}{2}}$ y $\boldsymbol{T}_{i+\frac{1}{2}}$ son discretizaciones de los términos de presión y transferencia respectivamente en la intercelda $x_{i+\frac{1}{2}}$ y cuya expresión se detalla en la Sección 3.2.

El esquema anterior no es bien equilibrado para las soluciones correspondientes al agua en reposo. Para construir un esquema de primer orden que lo sea usamos la técnica de reconstrucción conocida como reconstrucción hidrostática y que describiremos brevemente a continuación.

Reconstrucción hidrostática

La reconstrucción hidrostática consiste esencialmente en una técnica de reconstrucción que puede identificarse con una elección muy particular de caminos que conectan los estados $\boldsymbol{w}^n_i$, $z_{B,i}$ con $\boldsymbol{w}^n_{i+1}$ y $z_{B,i+1}$. Los estados reconstruimos que consideramos en esta memoria son los siguientes:

$$z_{B,i+\frac{1}{2}} = \max\left(z_{B,i}, z_{B,i+1}\right), \tag{42}$$

y los estados reconstruidos para la profundidad,

$$h^{HR,-}_{i+\frac{1}{2}} = \left(h_i + z_{B,i} - z_{B,i+\frac{1}{2}}\right)_+, \qquad h^{HR,+}_{i+\frac{1}{2}} = \left(h_{i+1} + z_{B,i+1} - z_{B,i+\frac{1}{2}}\right)_+. \tag{43}$$

Finalmente, usando esta expresión podemos definir los estados reconstruidos que dejan invariantes las densidades y las velocidades en cada una de las interceldas:

$$\boldsymbol{w}^{HR,\pm}_{i+\frac{1}{2}} = \left(h^{HR,\pm}_{i+\frac{1}{2}} \,\middle|\, h^{HR,\pm}_{i+\frac{1}{2}}\theta_{\alpha,i} \,\middle|\, h^{HR,\pm}_{i+\frac{1}{2}}\theta_{\alpha,i}u_{\alpha,i}\right)^T \in \mathbb{R}^{2M+1}. \tag{44}$$

Nos encontramos finalmente en disposición de redefinir el esquema numérico original (39) como,

$$\begin{aligned}\boldsymbol{w}^{n+1}_i = \boldsymbol{w}^n_i - \frac{\Delta t}{\Delta x}\Big(&\boldsymbol{D}^{+}_{i-\frac{1}{2}}(\boldsymbol{w}^{HR-}_{i-\frac{1}{2}}, \boldsymbol{w}^{HR+}_{i-\frac{1}{2}}, z_{B,i-\frac{1}{2}}, z_{B,i-\frac{1}{2}}) \\ &+\boldsymbol{D}^{-}_{i+\frac{1}{2}}(\boldsymbol{w}^{HR-}_{i+\frac{1}{2}}, \boldsymbol{w}^{HR+}_{i+\frac{1}{2}}, z_{B,i+\frac{1}{2}}, z_{B,i+\frac{1}{2}}) + \boldsymbol{S}^{+}_{i-\frac{1}{2}} + \boldsymbol{S}^{-}_{i+\frac{1}{2}}\Big),\end{aligned} \tag{45}$$

donde el término $\boldsymbol{S}^{\pm}_{i+\frac{1}{2}}$ corresponde a la corrección asociada con la reconstrucción hidrostática y garantiza tanto la consistencia del esquema, como el carácter bien

UNIVERSIDAD DE MÁLAGA

equilibrado del esquema numérico para las soluciones estacionarias correspondientes al agua en reposo y densidad constante,

$$\boldsymbol{S}^{\pm}_{i+\frac{1}{2}} = \boldsymbol{P}^{\pm}_{i+\frac{1}{2}} - \boldsymbol{T}^{\pm}_{i+\frac{1}{2}}.$$

Las expresiones de $\boldsymbol{P}^{\pm}_{i+\frac{1}{2}}$ y $\boldsymbol{T}^{\pm}_{i+\frac{1}{2}}$ corresponden a discretizaciones de la presión y los términos de transferencia respectivamente usando los valores de la reconstrucción hidrostática (44) y se definen en la Sección 3.2.2.

El esquema numérico resultante es de primer orden en espacio y tiempo, y es capaz de preservar soluciones estacionarias correspondientes al agua en reposo con densidad constante. También es capaz de preservar la positividad para un $CFL \leq 0.5$, siempre y cuando la batimetría sea suave.

Aproximación de segundo orden

Para alcanzar segundo orden en espacio, combinamos el esquema de primer orden (45) con un operador de reconstrucción espacial de segundo orden $\boldsymbol{R}_i(x)$ $x \in I_i$.

En primer lugar, siguiendo [41], la extensión de alto orden del esquema numérico de primer orden (45) es:

$$\begin{aligned} \boldsymbol{w}_i'(t) &= -\frac{1}{\Delta x}\left(\boldsymbol{D}^{+}_{i-\frac{1}{2}}(t) + \boldsymbol{D}^{-}_{i+\frac{1}{2}}(t)\right) \\ &\quad -\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \left(\boldsymbol{P}(\boldsymbol{R}_i, R_i^{\eta}, \partial_x \boldsymbol{R}_i, \partial_x R_i^{\eta}) - \boldsymbol{T}(\boldsymbol{R}_i, \partial_x \boldsymbol{R}_i)\right) dx, \end{aligned} \tag{46}$$

donde

$$\begin{aligned} \boldsymbol{D}^{+}_{i-\frac{1}{2}}(t) &= \boldsymbol{D}^{+}_{i-\frac{1}{2}}(\boldsymbol{w}^{HR-}_{i-\frac{1}{2}}(t), \boldsymbol{w}^{HR+}_{i-\frac{1}{2}}(t), z_{B,i-\frac{1}{2}}, z_{B,i-\frac{1}{2}}) + \boldsymbol{S}^{+}_{i-\frac{1}{2}}, \\ \boldsymbol{D}^{-}_{i+\frac{1}{2}}(t) &= \boldsymbol{D}^{-}_{i+\frac{1}{2}}(\boldsymbol{w}^{HR-}_{i+\frac{1}{2}}(t), \boldsymbol{w}^{HR+}_{i+\frac{1}{2}}(t), z_{B,i+\frac{1}{2}}, z_{B,i+\frac{1}{2}}) + \boldsymbol{S}^{-}_{i+\frac{1}{2}}. \end{aligned}$$

En general, dada una variable f, denotamos por $f^{\pm}_{i+\frac{1}{2}}$ la evaluación del operador de reconstrucción de dicha variable, R^f, en la intercelda $x_{i+\frac{1}{2}}$. De esta forma, $f^{-}_{i+\frac{1}{2}} = R^f_i(x_{i+\frac{1}{2}})$ y $f^{+}_{i+\frac{1}{2}} = R^f_{i+1}(x_{i+\frac{1}{2}})$. Además, $\boldsymbol{w}^{HR\pm}_{i+\frac{1}{2}}$ es el resultado de aplicar el procedimiento de reconstrucción hidrostática a los estados reconstruidos $\boldsymbol{w}^{\pm}_{i+\frac{1}{2}}$. Esto se explica con más detalle en la Sección 3.2.4. Nótese que el término integral en (46) debe aproximarse por una fórmula de cuadratura de al menos segundo orden. Nosotros usamos aquí la fórmula del punto medio.

Para el esquema de volúmenes finitos se considera un operador de reconstrucción MUSCL. En cada celda se define un operador lineal a trozos de la forma,

$$\boldsymbol{R}_i(x) = \boldsymbol{w}_i + \boldsymbol{\sigma}_i(x - x_i), \tag{47}$$

UNIVERSIDAD DE MÁLAGA

donde $\boldsymbol{\sigma}_i$ proporciona la pendiente de la reconstrucción para cada variable en (44). Esta pendiente debe limitarse en presencia de fuertes gradientes o discontinuidades, de tal forma que se preserve el segundo orden en regiones suaves. Nosotros usamos un limitador basado en la media geométrica *avg* que se define como

$$[\boldsymbol{\sigma}_i]_k = \text{avg}\left(\frac{[\boldsymbol{w}_{i+1} - \boldsymbol{w}_i]}{\Delta x}, \frac{[\boldsymbol{w}_i - \boldsymbol{w}_{i-1}]}{\Delta x}\right), \tag{48}$$

donde

$$\text{avg}(a,b) = \begin{cases} \dfrac{|a|\, b + a\, |b|}{|a| + |b|} & \text{if } |a| + |b| > 0, \\ 0 & \text{en otro caso.} \end{cases} \tag{49}$$

Buen equilibrado para una familia de soluciones estacionarias

Las propiedades de buen equilibrado del esquema numérico descrito hasta ahora pueden mejorarse para preservar un mayor número de soluciones estacionarias, correspondientes a soluciones estacionarias para perfiles de densidad estratificados.

El primer paso consiste en calcular la solución estacionaria $\boldsymbol{w}_i^*(x)$ en cada celda para cada tiempo como las descritas en (12). Nótese que $\boldsymbol{w}_i^*(x)$ queda determinada conociendo un conjunto de parámetros que denotaremos por $h_{e,i}$ y $\theta_{1,e,i}, \ldots \theta_{M,e,i}$ y que pueden determinarse a partir de los promedios en la celda. Recordamos que estas soluciones estacionarias vienen caracterizadas por velocidad nula en cada capa y superficie libre constante.

Una vez que la solución estacionaria $\boldsymbol{w}_i^*(x)$ está calculada en cada celda, el operador de reconstrucción se escribe en términos de la fluctuación de la solución con respecto a la solución estacionaria,

$$\tilde{\boldsymbol{w}}_i^n(t) = \boldsymbol{w}_i^n - \boldsymbol{w}_i^*(x_i).$$

El próximo paso consiste en aplicar el operador de reconstrucción de segundo orden definido en el paso anterior a las fluctuaciones. El operador de reconstrucción aplicado a estos operadores se denota por $\tilde{\boldsymbol{R}}_i(x)$. De esta forma, la reconstrucción de las variables de estado (44) se define como,

$$\boldsymbol{R}_i(x) = \boldsymbol{w}_i^*(x) + \tilde{\boldsymbol{R}}_i(x).$$

Un esquema Galerkin discontinuo de orden arbitrariamente alto

Describimos ahora el esquema ADER-DG de alto orden de un solo paso para el modelo de aguas someras con densidad variable, que hemos desarrollado en esta tesis. Al igual que

UNIVERSIDAD DE MÁLAGA

el esquema de volúmenes finitos considerado, se trata de un esquema numérico original para este problema.

Recordamos que la solución $\boldsymbol{w}_h(x,t^n)$ de los esquemas DG, definida en (14), es una función polinómica a trozos de grado N_p en cada celda I_i y no necesariamente continuo. Aplicando la forma general de un esquema DG al problema que estamos estudiando resulta:

$$\int_{t^n}^{t^{n+1}} \int_{I_i} \Phi_k \partial_t \boldsymbol{w}_h \, dxdt + \int_{t^n}^{t^{n+1}} \int_{I_i} \Phi_k \left(\partial_x \boldsymbol{F}_C(\boldsymbol{w}_h) \, dxdt + \boldsymbol{P}(\boldsymbol{w}_h, \partial_x \boldsymbol{w}_h, \partial_x \eta) - \boldsymbol{T}(\boldsymbol{w}_h, \partial_x \boldsymbol{w}_h) \right) dxdt = \boldsymbol{0}. \quad (50)$$

Utilizando (14), esta fórmula puede reescribirse en función del predictor local espacio-tiempo $\boldsymbol{q}_h$, tal que

$$\begin{aligned} \left(\int_{I_i} \Phi_k \Phi_l \, dx \right) \left(\hat{\boldsymbol{w}}_{i,l}^{n+1} - \hat{\boldsymbol{w}}_{i,l}^{n} \right) - \int_{t^n}^{t^{n+1}} \int_{I_i^\circ} \Phi_k' \boldsymbol{F}_C(\boldsymbol{q}_h) \, dxdt \\ + \int_{t^n}^{t^{n+1}} \Phi_{k,i+\frac{1}{2}} \mathcal{D}_{i+\frac{1}{2}}^{-} \left(\boldsymbol{q}_{h,i+\frac{1}{2}}^{-}, \boldsymbol{q}_{h,i+\frac{1}{2}}^{+}, z_{b_{h,i+\frac{1}{2}}}^{-}, z_{b_{h,i+\frac{1}{2}}}^{+} \right) dt \\ + \int_{t^n}^{t^{n+1}} \Phi_{k,i-\frac{1}{2}} \mathcal{D}_{i-\frac{1}{2}}^{+} \left(\boldsymbol{q}_{h,i-\frac{1}{2}}^{-}, \boldsymbol{q}_{h,i-\frac{1}{2}}^{+}, z_{b_{h,i-\frac{1}{2}}}^{-}, z_{b_{h,i-\frac{1}{2}}}^{+} \right) dt \\ + \int_{t^n}^{t^{n+1}} \int_{I_i^\circ} \Phi_k \left(\boldsymbol{P}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h, \partial_x \eta_h) - \boldsymbol{T}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h) \right) dxdt = \boldsymbol{0}. \quad (51) \end{aligned}$$

Aquí, $\mathcal{D}_{i\mp\frac{1}{2}}^{\pm}$ representa el flujo numérico en las interfaces de las celdas mientras que $z_{b_{h,i\pm\frac{1}{2}}}^{\pm}$ son los valores extrapolados de la batimetría en las interceldas. Nótese que el primer término integral de (51) se corresponde con la matriz de masa, que es diagonal puesto que la base utilizada es ortogonal.

Puesto que la solución puede ser discontinua en las interceldas, usaremos un resolvedor de Riemann aproximado para aproximar el flujo entre las mismas, que en nuestro caso es un resolvedor de tipo HLL.

Predictor en espacio tiempo ADER-DG

Para evolucionar el esquema numérico DG en tiempo, se usa una discretización ADER. Recordamos que el predictor $\boldsymbol{q}_h(x,t)$ es una aproximación de alto orden en el intervalo $[t^n,\ t^{n+1}]$. Esta aproximación se calcula resolviendo un problema de Cauchy local, sin interacción con los estados vecinos. El primer paso consiste en expandir la solución del predictor $\boldsymbol{q}_h$ en el elemento I_i en términos de una base en espacio-tiempo local,

$$\boldsymbol{q}_h(x,t) = \sum_l \theta_l(x,t) \hat{\mathbf{q}}_l^i \coloneqq \theta_l(x,t) \hat{\mathbf{q}}_l^i \,, \quad (52)$$

UNIVERSIDAD DE MÁLAGA

con l un multi-índice y donde la función base en espacio tiempo $\theta_l(x,t) = \varphi_{l_0}(\tau)\varphi_{l_1}(\xi)$ se define mediante el producto tensorial en espacio-tiempo de los polinomios de Lagrange de grado N_p que pasan por los N_p+1 puntos de los nodos de cuadratura de Gauss-Legendre. Finalmente, aplicamos el procedimiento ADER para el sistema multicapa de aguas someras con densidad variable multiplicando el sistema de EDP (33) por una función test θ_k e integrada en el volumen de control $I_i \times [t^n, t^{n+1}]$:

$$\int_{t^n}^{t^{n+1}} \int_{I_i} \theta_k(x,t)\partial_t \boldsymbol{q}_h \, dxdt + \int_{t^n}^{t^{n+1}} \int_{I_i} \theta_k(x,t) \left(\partial_x \boldsymbol{F}_C(\boldsymbol{q}_h)\, dxdt + \boldsymbol{P}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h, \partial_x \eta_h) - \boldsymbol{T}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h)\right) dxdt = \mathbf{0}. \quad (53)$$

Esta ecuación puede ser reescrita como:

$$\begin{aligned} &\int_{I_i} \theta_k(x,t^{n+1})\boldsymbol{q}_h(x,t^{n+1})\, dx \\ &\qquad - \int_{I_i} \theta_k(x,t^n)\boldsymbol{q}_h^0(x,t^n)\, dx - \int_{t^n}^{t^{n+1}} \int_{I_i} \partial_t \theta_k(x,t)\boldsymbol{q}_h(x,t)\, dxdt \\ &= -\int_{t^n}^{t^{n+1}} \int_{I_i} \theta_k(x,t) \left(\partial_x \boldsymbol{F}_C(\boldsymbol{q}_h) + \boldsymbol{P}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h, \partial_x \eta_h) - \boldsymbol{T}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h)\right) dxdt. \quad (54) \end{aligned}$$

Los grados de libertad $\hat{\mathbf{q}}_l^i$ del polinomio en espacio tiempo $\boldsymbol{q}_h$ se determinan resolviendo el problema no lineal resultante mediante un algoritmo de tipo punto fijo. Nosotros usamos como semilla inicial la solución en el instante precedente: $\boldsymbol{q}_h^0(x,t) = \boldsymbol{w}_h(x,t^n)$.

A continuación describimos el procedimiento de limitación que se ha utilizado en esta tesis para evitar la aparición de oscilaciones espurias en presencia de discontinuidades de la solución o en zonas donde el gradiente de la solución es grande.

Un limitador subcelda *a posteriori* basado en volúmenes finitos

Hasta ahora, la aproximación ADER-DG propuesta no está limitada, en el sentido de que no existe ningún mecanismo para prevenir oscilaciones espurias cerca de fuertes gradientes o discontinuidades asociadas al alto orden. En esta tesis se discuten dos aproximaciones para limitar una solución en el marco de los métodos de tipo DG: la aproximación WENO, basada en una reescritura de la solución en la celda ponderada con la solución en las celdas vecinas (véase [95–97]) y la aproximación MOOD, basada en combinar adecuadamente un resolvedor DG y otro en volúmenes finitos en determinadas zonas espaciales del dominio (véase [89, 90]). Es esta última estrategia la que se describe a continuación.

En la estrategia MOOD consideramos, en primer lugar, la solución sin limitar del esquema DG como una solución candidata. Ésta se analiza localmente y se considera

UNIVERSIDAD DE MÁLAGA

como solución válida o final si verifica un conjunto de criterios. Si la solución se rechaza en una determinada celda, se considera una submalla en dicha celda y se proyecta la solución en el instante t^n en la nueva malla usando funciones constante a trozos en cada nueva subcelda de forme que se garantice la conservación. A continuación se evoluciona la solución en dichas subceldas con un esquema robusto de volúmenes finitos. Por último, usando la evolución de estos estados constantes en cada subcelda, se reconstruye un polinomio de grado N_p en la celda original y se corrigen las celdas vecinas para garantizar la conservación.

Una de las ventajas de la técnica MOOD es que permite verificar un amplio número de propiedades físicas y numéricas. Particularmente, para el esquema de aguas someras multicapa con densidad variable se comprueba la positividad de la columna de agua h y que la densidad relativa sea siempre mayor o igual que la unidad. Para la detección de discontinuidades, se utiliza un principio del máximo discreto relajado aplicado *a posteriori*. Véase la Sección 3.3.2 para más detalles.

Preservando soluciones estacionarias en el marco de los esquema ADER-DG

Al igual que hicimos en volúmenes finitos, vamos a modificar el esquema DG para que sea exactamente bien equilibrado para las soluciones estacionarias estratificadas con velocidad nula y superficie libre constante definidas en (12). Como en el caso del esquema de volúmenes finitos, el primer paso consiste en calcular, para cada celda y en cada instante de tiempo, la solución estacionaria $\boldsymbol{w}^*_{i,h}(x),\ x \in I_i$. Para ello, las constantes de las que depende la expresión (12) deben ser adecuadamente calculadas. Nótese que, por definición, las soluciones estacionarias satisfacen los términos de presión (11) en cada celda,

$$\boldsymbol{P}(\boldsymbol{w}^*_{i,h}, \partial_x \boldsymbol{w}^*_{i,h}, \partial_x \bar{\eta}_i) = 0. \tag{55}$$

Teniendo en cuenta dicha igualdad podemos reescribir el esquema numérico como sigue

$$\begin{aligned}\left(\int_{I_i} \Phi_k \Phi_l \, dx\right) \left(\hat{\boldsymbol{w}}^{n+1}_{i,l} - \hat{\boldsymbol{w}}^{n}_{i,l}\right) - \int_{t^n}^{t^{n+1}} \int_{I_i^\circ} \left(\Phi_k' \boldsymbol{F}_C(\boldsymbol{q}_h) - \Phi_k \boldsymbol{T}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h)\right) \, dxdt \\ + \int_{t^n}^{t^{n+1}} \Phi_{k,i+\frac{1}{2}} \mathcal{D}^-_{i+\frac{1}{2}} \left(\boldsymbol{q}^-_{h,i+\frac{1}{2}}, \boldsymbol{q}^+_{h,i+\frac{1}{2}}, z_{b^-_{h,i+\frac{1}{2}}}, z_{b^+_{h,i+\frac{1}{2}}}\right) \, dt \\ + \int_{t^n}^{t^{n+1}} \Phi_{k,i-\frac{1}{2}} \mathcal{D}^+_{i-\frac{1}{2}} \left(\boldsymbol{q}^-_{h,i-\frac{1}{2}}, \boldsymbol{q}^+_{h,i-\frac{1}{2}}, z_{b^-_{h,i-\frac{1}{2}}}, z_{b^+_{h,i-\frac{1}{2}}}\right) \, dt \\ + \int_{t^n}^{t^{n+1}} \int_{I_i^\circ} \Phi_k \left(\boldsymbol{P}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h, \partial_x \eta_h) - \boldsymbol{P}(\boldsymbol{w}^*_{i,h} \partial_x \boldsymbol{w}^*_{i,h}, \partial_x \bar{\eta}_i)\right) \, dxdt = \boldsymbol{0}.\end{aligned} \tag{56}$$

El último paso para asegurarse que el esquema anterior es bien equilibrado es calcular los valores extrapolados $\boldsymbol{q}^{\pm}_{h,i\pm\frac{1}{2}}$ adecuadamente. Esto se hace extrapolando la fluctuación

UNIVERSIDAD DE MÁLAGA

con respecto a la solución estacionaria,

$$\tilde{\boldsymbol{q}}^{-}_{h,i+\frac{1}{2}} = (\boldsymbol{q}_{h,i} - \boldsymbol{w}^{*}_{i,h})(x_{i+\frac{1}{2}}).$$

El valor final extrapolado en la intercelda se recupera mediante la siguiente expresión,

$$\boldsymbol{q}^{-}_{h,i+\frac{1}{2}} = \boldsymbol{w}^{*}_{i}(x_{i+\frac{1}{2}}) + \tilde{\boldsymbol{q}}^{-}_{h,i+\frac{1}{2}}. \tag{57}$$

Por supuesto, este proceso debe realizarse para todas las interceldas.

De igual forma, la etapa predictora de alto orden ADER también debe ser modificada para que preserve soluciones estacionarias. El algoritmo es adaptado para que calcule fluctuaciones con respecto a la solución estacionaria y queda como sigue:

$$\begin{aligned}\int_{I_i} \theta_k(x,t^{n+1})\tilde{\boldsymbol{q}}_h(x,t^{n+1})\,dx - \int_{I_i} \theta_k(x,t^{n})\tilde{\boldsymbol{q}}^{0}_h(x,t^{n})\,dx - \int_{t^n}^{t^{n+1}}\int_{I_i} \partial_t\theta_k(x,t)\tilde{\boldsymbol{q}}_h(x,t)\,dxdt \\ = -\int_{t^n}^{t^{n+1}}\int_{I_i} \theta_k(x,t)\left(\partial_x \boldsymbol{F}_C(\boldsymbol{q}_h) - \boldsymbol{T}(\boldsymbol{q}_h,\partial_x\boldsymbol{q}_h)\right)\,dxdt \\ -\int_{t^n}^{t^{n+1}}\int_{I_i} \theta_k(x,t)\left(\boldsymbol{P}(\boldsymbol{q}_h,\partial_x\boldsymbol{q}_h,\partial_x\eta_h) - \boldsymbol{P}(\boldsymbol{w}^{*}_{i,h},\partial_x\boldsymbol{w}^{*}_{i,h}\partial_x\bar{\eta}_i)\right)\,dxdt,\end{aligned} \tag{58}$$

donde $\tilde{\boldsymbol{q}}_h(x,t)$ es la proyección en $\mathcal{U}_h$ de la fluctuación respecto de la solución estacionaria,

$$\tilde{\boldsymbol{q}}_h(x,t) = \boldsymbol{q}_h(x,t) - \boldsymbol{w}^{*}_{i,h}(x).$$

Restricción al paso de tiempo

Los esquemas de tipo DG sufren de una número Courant-Friedrichs-Lewy (CFL) que se reduce conforme aumenta el orden del polinomio de aproximación N. En particular, la expresión viene dada por,

$$\Delta t \leq \frac{1}{d}\frac{\mathrm{CFL}}{2N+1}\frac{\Delta x}{|\lambda_{max}|}. \tag{59}$$

Esta severa restricción debe ser vista con perspectiva. Aunque ciertamente el paso de tiempo que impone es pequeño, la mayor resolución de los esquemas DG permiten obtener buenos resultados con mallas groseras o incluso muy groseras.

Capítulo 4: Experimentos numéricos

En este capítulo se abordan diversos experimentos numéricos que persiguen mostrar la robustez y eficacia del modelo y su discretizaciones numéricas, que hemos descrito en los capítulos precedentes, así como su validación con datos de laboratorio. Por motivos

UNIVERSIDAD DE MÁLAGA

de espacio no se reproducen en este resumen los resultados, que pueden consultarse en la versión completa en inglés. Por tanto, nos limitaremos a comentar las principales simulaciones realizadas y las conclusiones que de ellas pueden extraerse.

Con el fin de comparar ambas métodos de discretización se han realizado una serie de test numéricos comunes. Los experimentos están diseñados para emular situaciones que pueden darse potencialmente en flujos geofísicos. En particular, se incluyen una perturbación de una solución estacionaria, una rotura de presa de densidad relativa sobre un obstáculo o la evolución de un perfil suave de densidad, entre otros. En particular, estos test están diseñados para comprobar las buenas propiedades de los esquemas propuestos, así como el orden de aproximación o el carácter bien equilibrado de los esquemas.

De un interés particular resulta la simulación de una rotura de presa en densidad relativa en un canal plano, para la que están disponibles datos de laboratorio con los que es posible comparar los resultados obtenidos. Se estudia además la convergencia a la solución correcta conforme se aumenta el número de capas. Los resultados para ambas discretizaciones son excelentes y permiten estimar las condiciones necesarias que debe cumplir una simulación en condiciones reales.

Es también interesante resaltar la excelente resolución de los esquemas de tipo DG frente a los esquemas de volúmenes finitos. Como se muestran en varias simulaciones, los esquemas DG de alto orden permiten obtener soluciones numéricas de gran calidad con mallas relativamente groseras, mientras si usamos el método de volúmenes finitos serían necesarias mallas con una mayor resolución (hasta 10 veces más finas).

Capítulo 5: Conclusiones y trabajo futuro

Esta tesis incorpora avances en modelado matemático y análisis numérico. En particular se propone un nuevo modelo para el modelado y simulación de fluidos estratificados y dos discretizaciones diferentes del mismo. También se describe una estrategia general para la construcción de esquemas de tipo DG que sean bien equilibrados.

En el Capítulo 1, se presenta un modelo general de aguas someras multicapa donde los efectos asociados a la densidad se tienen en cuenta a través de una función trazador que se transporta y difunde con el flujo. Su objetivo es resolver el problema presente en modelos de aguas someras tradicionales que ignoran como la densidad puede afectar a la hidrodinámica general del fluido cuando estos efectos son relevantes. El modelo se define usando las hipótesis de los modelos de aguas someras multicapa, donde la velocidad y densidad son constantes en cada capa. Se supone además que la presión es hidrostática y depende de la densidad relativa. Como consecuencia, el modelo final (9) incluye términos no conservativos directamente relacionados con la expresión de la presión y de los términos de transferencia. Por tanto, es necesario un tratamiento numérico cuidadoso para poder discretizar con éxito el modelo. Además, se han analizado las soluciones estacionarias correspondientes a velocidad nula, estratificaciones en densidad

UNIVERSIDAD DE MÁLAGA

y superficie libre constante.

En el Capítulo 2, se presenta una metodología general para diseñar métodos de volúmenes finitos y de Galerkin discontinuo de alto orden y bien equilibrados. Puesto que el modelo desarrollado posee productos no conservativos, se usa el marco de los esquemas camino-conservativos. Asimismo, se presentan varias estrategias para preservar soluciones estacionarias o mejorar la eficiencia general del resolvedor de Riemann. Estos resultados se usan para dar soporte a los métodos de tipo Galerkin discontinuo que se presentan más tarde. En este contexto, se incluyen aportaciones novedosas en la forma de una estrategia general para diseñar esquemas numéricos DG bien equilibrados. Además, se realiza una nueva propuesta en el sempiterno problema de la limitación en el marco de resolvedores de tipo DG. La principal ventaja de esta propuesta es que no rompe la propiedad de buen equilibrado del método numérico. Finalmente se proponen diferentes test numéricos para verificar la eficiencia de los esquemas DG bien equilibrados y el limitador propuesto.

El Capítulo 3 está dedicado a una discretización *ad hoc* del modelo mutlicapa de aguas someras con densidad variable mediante las técnicas generales previamente revisadas. Se deriva un esquema numérico de volúmenes finitos bien equilibrado y de segundo orden que hace uso de una reconstrucción hidrostática. El esquema numérico es capaz de preservar soluciones estacionaras o recuperar soluciones estacionarias tras una perturbación. Además, se propone un esquema de tipo ADER-DG para el mismo modelo. Este método presenta algunas ventajas evidentes: permite obtener un orden arbitrariamente alto tanto en espacio, permitiendo esquemas de alto orden en un solo paso; la mayor resolución en la celda inherente a los métodos DG permiten considerar mallas groseras o muy groseras a la vez que se mantiene una gran resolución; las técnicas de buen equilibrado propuestas son aplicadas para preservar soluciones estacionarias en el marco de los métodos DG. Además, el esquema DG resultante satisface un principio del máximo discreto para la densidad relativa, gracias a su limitador *a posteriori* de tipo MOOD basado en el resolvedor de volúmenes finitos. Este limitador comprobará la solución numérica para asegurar que todas las propiedades físicas y numéricas deseadas se cumplen. En caso de que no sea así, permite intercambiar el resolvedor por uno más robusto basado en volúmenes finitos.

Finalmente, el Capítulo 4 incluye varios experimentos numéricos. Estos experimentos han sido diseñados para resaltar las ventajas y propiedades de los esquemas numéricos propuestos. También se incluye una comparación entre los dos esquemas numéricos desarrollados, el de volúmenes finitos y el método DG. En particular, se han realizado test numéricos donde se estudia el orden de los métodos, se verifica empíricamente la propiedad de buen equilibrado o un experimento donde están disponibles datos de laboratorio para la validación del modelo y los esquemas numéricos, entre otros. Estas simulaciones muestran la eficiencia del modelo para simular flujos geofísicos donde la densidad juega un papel importante. Los experimentos también subrayan la importancia de los esquemas de alto orden como una característica clave para mejorar el rendimiento. Esto es especialmente cierto en el caso de los métodos de tipo DG, donde pueden obtenerse

UNIVERSIDAD DE MÁLAGA

excelentes resultados incluso con mallas muy groseras. Es de singular importancia la habilidad de ambas discretizaciones numéricas para correlacionarse con datos empíricos. Los excelentes resultados pueden explicarse por el elevado número de capas considerado y porque los efectos de la aceleración vertical en la simulación son despreciables para una pluma de densidad que avanza a velocidades pequeñas. En general, este capítulo permite entender los requisitos en términos de discretización y números de capas verticales que son necesarios para aplicar el modelo en escenarios reales.

Trabajo futuro

A continuación presentamos algunos apuntes sobre posibles trabajos futuros que pueden derivarse de la presente tesis. Sería interesante rescribir el presente modelo de aguas someras multicapa en términos de σ-coordenadas. De esta forma, sería posible una mejor definición de los flujos verticales. En particular estos se describirían en términos de un flujo direccional en las interceldas que separan las celdas en vertical, en lugar de como flujos no conservativos con una discretización *upwind*, que es lo que hemos considerado en este trabajo. Además, el modelo actual en coordenadas cartesianas podría ser escrito en coordenadas esféricas, más adecuadas para simular problemas de gran tamaño, como la evolución de corrientes en el Estrecho de Gibraltar, por ejemplo. Además sería necesario la incorporación de las fuerzas de Coriolis.

Otras oportunidades de investigación residen en mejorar la discretización numérica. Especialmente el esquema numérico ADER-DG puede ser expandido en dimensión por dimensión y así poder abordar la discretización de problemas en dominios bidimensionales.

Dentro del campo de los modelos de aguas someras, los modelos no hidrostáticos se han vuelto muy populares por su habilidad para capturar mejor algunos efectos físicos relacionados con la aceleración vertical del fluido. Sería interesante extender el presente modelo de aguas someras con presión variable a una versión no hidrostática para estudiar la influencia de estos efectos en flujos estratificados. Este estudio podría estar relacionado con un análisis más general sobre la hiperbolicidad del modelo, un problema que no ha sido abordado en esta tesis.

Finalmente, esta tesis ha presentado una serie de resultados en el ámbito de los esquemas numéricos bien equilibrados que son un decidido paso adelante en el campo, especialmente para la familia de resolvedores de tipo DG. El uso de estas técnicas DG pude extenderse para abordar esquemas bien equilibrados en dominios bidimensionales.

UNIVERSIDAD DE MÁLAGA

Abstract

This thesis presents some novel contributions within the framework of multilayer shallow-water models and well-balanced high order numerical schemes. In the first place, a multilayer shallow-water model with variable density is derived. A particular state equation is chosen such that the model can be written in terms of its relative density. The model is derived under the hydrostatic assumption for the pressure and consist on a system of partial differential equations with non-conservative products and source terms. As it is usual in multilayer systems, a closure relation is needed for the vertical transference terms, which are then expressed in terms of horizontal velocities. Finally, the model presents some stationary solutions that go beyond the usual lake-at-rest type stationary solutions. Since the model admits density fluctuations, stationary solutions corresponding to a vertical stratification in relative density are also possible. However, due to the multilayer shallow-water approach, the vertical profile depends on the given topography and the number of layers chosen. This may be seen as a discrete version for the multilayer approach for such vertical stratification profile.

A complete review of the existing techniques regarding finite volume methods based on path-conservative Riemann solvers and discontinuous Galerkin (DG) methods is also included. Of particular interest is the novel contribution of this thesis where a general strategy to design well-balanced DG methods for general systems of balanced laws is given. The strategy includes a limiting technique that preserves the well-balanced property of the numerical scheme. Moreover, it is compatible with several time discretization methods, such as ADER or Runge-Kutta. It has been successfully tested with the Burgers' equation, the Euler equations with gravity and the one layer shallow-water equations.

The next step consists on providing numerical discretizations for the multilayer shallow-water model with variable density. The discretization proposed here is based on the finite volume method and the discontinuous Galerkin method, and both of them are novel contributions of this thesis. The finite volume method provides a second-order accurate approximation of the PDE system by means of a hydrostatic reconstruction that is able to preserve the water at rest solution with constant density and also, non-trivial stationary solutions corresponding to the a vertical stratification of relative density. Likewise, the DG based numerical discretization allows for arbitrary high order in space and time in one step thanks to the use of the ADER-DG technique. This approach is also well-balanced for trivial and non-trivial stationary solutions. The limiting strategy

UNIVERSIDAD DE MÁLAGA

is based on the MOOD paradigm, with a proper switching between the DG solver and a robust finite volume solver.

Finally, several numerical experiments are proposed that seek to prove the accuracy and robustness of the proposed numerical discretizations. They include a test studying the order of accuracy, where it can be seen that the desired order is achieved for both numerical solvers. Several well-balanced tests are also discussed. Non-trivial stationary solutions are not only preserved, but even recovered after a small perturbation. Next, several experiments are considered, including a lock exchange in relative density and a comparison with experimental data, where the numerical results yield excellent data agreement. The last experiment corresponds to a lock exchange in a two-dimensional channel.

As closure and in order to go beyond the results already presented, two appendices are included. The first one addresses the numerical implementation in two dimensional domains of the finite volume method proposed in this thesis for the multilayer shallow-water model with variable density. The second appendix explains the parallelization performed in CPU and in GPU, including a comparison of different parallelization strategies running in several hardware architectures.

All these novel contributions have been published in a total of three scientific publications:

- Guerrero Fernández E, Castro Díaz MJ, Morales de Luna T. A Second-Order Well-Balanced Finite Volume Scheme for the Multilayer Shallow Water Model with Variable Density. Mathematics. 2020; 8(5):848. https://doi.org/10.3390/math8050848

- Guerrero Fernández E, Castro Díaz MJ, Dumbser M, Morales de Luna T. An Arbitrary High Order Well-Balanced ADER-DG Numerical Scheme for the Multilayer Shallow-Water Model with Variable Density. J Sci Comput. 2022; 90:52. https://doi.org/10.1007/s10915-021-01734-2

- Guerrero Fernández E, Escalante C, Castro Díaz MJ. Well-Balanced High-Order Discontinuous Galerkin Methods for Systems of Balance Laws. Mathematics. 2022; 10(1):15. https://doi.org/10.3390/math10010015

UNIVERSIDAD DE MÁLAGA

Introduction

Model derivation and numerical simulation constitutes a thriving research topic in the scientific community and has contributed to some key achievements in a wide range of fields, from solid mechanics to fluids hydrodynamics. Particularly, those fields where experimental data is very difficult or impossible to obtain is often the greatest beneficiary of the numerical simulation, allowing to deepen our understanding of the underlying processes in a wide range of situations. Specifically, geophysical flows are notably suited to this kind of approximation. However, the numerical simulation of fluid hydrodynamics requires, on the one hand, a deep understanding of the physics involved and, on the other hand, robust and accurate numerical schemes that reflect the physical behavior of the model.

Moreover, another important problem that often arises in numerical simulations is the computational load associated with complex models and large computational domains. There are several approaches to address this issue. For instance, efforts may focus on simplifying the model while keeping the relevant physical phenomena intact. Another approach, however, focuses on modifying the numerical scheme to improve the efficiency while maintaining the accuracy and robustness. Lastly, it is also possible to explore the various options available in the framework of code optimization and hardware acceleration.

The underlying physics in geophysical flows are governed by the well-known Navier-Stokes equations. These equations can be simplified under some hypotheses, which allows to improve the overall efficiency so that the resulting systems is less complex. One way of reducing the complexity of the model correspond to the well-known shallow-water (or Saint-Venant) equations [1]. The shallow-water equations results from depth-averaging in the vertical direction the Navier-Stokes equations under the following hypotheses:

- The average vertical dimension of the domain H is small compared with the horizontal one L,
$$\frac{H}{L} \ll 1.$$
- The pressure of the fluid is assumed to be hydrostatic.
- The horizontal velocity does not depend on the vertical direction.

UNIVERSIDAD DE MÁLAGA

The resulting hyperbolic system of partial differential equations has been widely used for solving all kind of hydrodynamic problems [2–4]. The main advantage of the shallow-water system consists on reducing the dimension of the problem by one, thanks to its simplification in the vertical direction, allowing a significant reduction of the computational effort.

However, homogenization in the vertical dimension can also be an important drawback, since the vertical flux information is lost. In order to overcome this limitation, multilayer shallow-water models have been derived in recent years. One possible approach to introduce multilayer shallow water models is the one described by Audusse et al. in [5]. In this work, the vertical domain is discretized into a set of vertical layers and the shallow-water hypothesis are considered in each layer, allowing the solution to be discontinuous across layer interfaces. Mass and momentum exchange between layers is accomplished by incorporating into the equations an approximation of the vertical flux in the form of non-conservative terms. In this way, a rich vertical profile of the fluid hydrodynamics can be is obtained at the cost of increasing computational effort, proportional to the total number of layers. In fact, recent development in multilayer shallow-water model concern the local change in the number of layers according to the presence of relevant vertical effects (see [6]), which allows to reduce computational load while keeping the vertical information. In any case, the traditional multilayer approach is able to capture vertical effects that single layer shallow-water equations fail to describe (see [7–10]). It has been successfully applied by different authors over the years to geophysical flows such as tsunamis ([3, 11–14]), flooding events ([15–19]), or storm-surges ([20–22]).

Recent developments in multilayer shallow-water models have been done, which integrate dispersive effects by considering non-hydrostatic pressure functions (see [23, 24] or [25]). In this way, the dispersive properties of the scheme are enhanced and the resulting model yields more accurate physical solutions at the cost of greater computational effort.

Alternative approaches to capture vertical effects while avoiding the direct numerical simulation of the full three dimensional Navier-Stokes equations is the so-called σ-coordinate models [26]. While these models are also employed for the solution of the full three-dimensional Navier-Stokes equations in oceanography and weather prediction, they can also be considered within the framework of shallow-water type models. In these models, the free surface is defined as a function of time and the horizontal coordinate, allowing to compute the free surface directly like in the shallow-water models. In this way, the total water height is reparametrized in the unit interval, placing the free surface in the upper boundary. This approach provides a full new set of models that allow turbulence, solute transport and short wave propagation (see [27–29]). For an overview about the history of efficient semi-implicit discretizations of hydrostatic and non-hydrostatic shallow water flows, including flows with multi-valued free surface, the reader is referred to the work of Casulli, see e.g. [27, 30–33].

As it has been discussed, multilayer shallow-water models are able to capture complex vertical profiles. However, if the vertical stratification is caused by density effects, the

UNIVERSIDAD DE MÁLAGA

standard approximation of the multilayer shallow-water model fails, since density effects are not present in the underlying hypotheses. Nevertheless, density variations can play a significant role in geophysical currents. One classical example occurs at the Strait of Gibraltar, where two currents, one originated at the Mediterranean Sea and the other from the Atlantic Ocean, meet. These two currents are different in terms of temperature and salinity, and therefore its respective densities take an essential role in its interaction. Indeed, the Mediterranean is warmer and has a higher salinity concentration due to greater evaporation. Therefore, the Mediterranean current is denser and tends to flow under the current of the Atlantic Ocean. Capturing this kind of behavior is impossible for multilayer shallow-water models that do not incorporate these buoyancy effects.

In order to address this issue, multilayer shallow-water models with density effects have been developed. For instance, in [34, 35] a multilayer shallow-water model with density depending on sediment species, with their own settling velocity, is proposed. In [36], a two layer shallow-water model is considered where the water entrainment is approximated by an heuristic-dependent source term. Another example is [37], where the authors derive a semi-implicit time discretization approach for a variable density shallow water model with a variable number of layers to improve flexibility and efficiency.

The model considered in this thesis makes use of a given tracer that is advected with the flow. This tracer is defined in terms of the relative density and the pressure terms depends on it. In this way, the relative density is taken into account. The model differs from other proposals like [38] in the procedure to obtain the multilayer system. This derives mainly in a different definition of the transfer terms.

As the model incorporates more physical effects, the associated computational cost increases. In addition, the considered system of partial differential equations is very sensitive to small changes in relative density. For this reason, any numerical discretization of the aforementioned model must be efficient and robust. In this thesis, two different numerical discretizations are proposed. One is based in a finite volume discretization while the second one is based on a Discontinuous Galerkin method.

The finite volume framework is especially well-suited to deal with the typical discontinuities that are developed by non-linear system of hyperbolic balance laws. In the multilayer framework, additional non-conservative terms appear due to the transfer terms. The main difficulty that arise, from both a mathematical and numerical analysis perspective, is the numerical treatment of non-conservative products, which adds a considerable complexity to the proper definition of weak solutions. In addition, many geophysical problems and relevant systems of balanced laws fall within this category, including the ones used in this thesis. However, this non-conservative products can be interpreted as a Borel measure in the sense introduced by Dal Maso, LeFloch and Murat in [39]. In this way, the numerical flux of the Riemann problem with non-conservative products can be expressed in terms of a free-chosen path linking two states. This constitutes a cornerstone of the path-conservative methods introduced by C. Parés in [40]. In fact, the proper choice of the path linking two states should be closely related

UNIVERSIDAD DE MÁLAGA

with the physics of the problem. However, the computation of the adequate path for each problem can quickly become a cumbersome task. In some practical problems, the simple choice of the straight segment linking two states will give sufficiently accurate results. More insight on the election of paths can be found in [41].

As in conservation laws, Roe Path-conservative methods (see [42–45]) produce accurate results, but they demand the explicit knowledge of the eigenstructure of the system, something that is not always feasible. Alternatively, one could approximate the eigenvalues by means of some approximation technique, obtaining numerically the spectral information of the system at the cost of further increasing the computational effort. Moreover, Roe schemes do not satisfy, in general, an entropy inequality and therefore an entropy-fix technique has to be added to properly capture the entropy solution in the presence of non smooth transitions (see [46]). Of course, incomplete Riemann solver such as Rusanov or HLL, which uses less information relative to the characteristic field, are simpler and present some advantages regarding computational load (see [47]). This improvement in computational effort comes at the cost of a subpar resolution at discontinuities compared with complete Riemann solvers.

For this thesis, a class of computationally fast first order finite volume solvers, the so-called Polynomial Viscosity Matrix (PVM) methods, introduced by Castro et al. in [48], are used. The PVM methods provide incomplete Riemann solver techniques based on the definition of viscosity matrices computed by a suitable polynomial evaluation of a Roe linearisation for general conservative and non-conservative hyperbolic systems. In fact, PVM methods can be interpreted as a natural generalization for non-conservative systems of the methods proposed in [49] for balanced laws. Finally, PVM methods admit a high order extension based on a polynomial reconstruction of states (see [50]) and multi-dimensional application (see [51]).

Another numerical discretization considered in this thesis is the family of Discontinuous Galerkin (DG) methods. The DG method itself dates back to the early work by Reed and Hill in [52]. It was later developed by Cockburn and Shu, who set a sound theoretical foundation for these numerical schemes in a series of well-known publications ([53–56] among others), where the DG methods were extended to non-linear hyperbolic systems of conservation laws.

The DG approach to numerical discretization has some advantages and disadvantages with respect to the finite volume approach. Indeed, in opposition to the finite volume method that consider the solution to be constant within the discretization cell and define some reconstruction operator to achieve high order, in the DG approach the solution within a cell is already a polynomial. In this way, high order in space in the DG framework is simply achieved by increasing the order of the aforementioned polynomial. Moreover, there are several strategies to reach high order in time with DG methods. Probably, the most natural way is to use the method of lines and consider an explicit high order time discretization via Runge-Kutta schemes, leading to the family of Runge-Kutta discontinuous Galerkin schemes [53–56]. Space-time DG methods where space-time

UNIVERSIDAD DE MÁLAGA

basis and test functions are employed were first developed in [57–59]. Other interesting references are [60–67]. The resulting schemes are conditionally stable under a CFL condition. Yet another approach to build high order space time DG approximation is to consider the Cauchy-Kovaleskaya procedure, first developed by Toro and Titarev in the finite volume framework (see [68–73]). This technique uses a Taylor expansion of the variables in order to substitute time derivatives by spatial derivatives that can be then treated by the DG numerical scheme. However, the computation of the Taylor expansion can quickly become too complex and cumbersome. Instead, the ADER (Arbitrary high order DErivative Riemann problem) local space time DG predictor procedure can be used to reach arbitrary high order in time, while completely avoiding the Cauchy-Kovaleskaya procedure (see [74, 75]). The ADER high order approximation of the solution is based on the solution of a generalized Riemann problem by means of an element-local fixed point algorithm, the convergence of which was proven in [76]. In this way, the combination of the ADER technique with the DG method result in a new family of arbitrary high order in space and time one-step ADER-DG numerical schemes.

The DG and the ADER-DG numerical schemes have been successfully applied to geophysical flows within the shallow-water framework. Some works regarding a one layer shallow-water model can be found in [77–80]. Additionally, application of this technique to immiscible two-layer models can be found in [81–83]. Regarding multilayer models, the reader is referred to [84]. More recently, advances on hyperbolic reformulations of nonlinear dispersive shallow water systems that makes use of the DG or ADER-DG techniques are found on [85–88].

The DG approach was previously described as inherently high-order, thanks to its definition in terms of polynomials within the cell. This high-order nature can become a problem near strong gradients or discontinuities. Indeed, the well-known Godunov's theorem states that high order numerical schemes will lose monotony near discontinuities, thus developing spurious oscillations. Several approaches exist to address this problem of limiting. For instance, it is possible to use a multi-dimensional optimal order detection (MOOD, see [89, 90]), which is *a posteriori* approach to the problem of limiting. In this technique, the ADER-DG unlimited solution is considered a candidate solution, and subsequently analyzed to determine its suitability against a number of criteria. If the candidate is deemed as adequate, then it becomes the final solution. However, if the solution is considered unsatisfactory, then the candidate solution is projected within the cell into piecewise constant states, that are then solved using a robust finite volume numerical scheme. The projection of the finite volume solution on a polynomial will then become the limited final solution in the cell. Note that the use of a finite volume solver could destroy the subcell resolution of the DG scheme. In order to preserve it, an optimal subcell discretization can be used (see [91]). A key advantage of the MOOD limiter is that it allows to analyze the candidate solution for any number of physical and numerical properties, ensuring in this way the coherence and consistency of the final solution. For instance, to detect spurious oscillation associated with high order near discontinuities it is

UNIVERSIDAD DE MÁLAGA

possible to consider a relaxed discrete maximum principle. Additionally, other numerical properties, like positivity, can be enforced in this way. The correct calibration of these criteria is very important to determine the admissibility of the candidate solution.

This is not, by any means, the only limiting technique in the DG framework. Indeed, other limiter strategies include the minmod type total variation bounded (TVB) limiter [53, 92], the moment-based limiter [93] or the improved moment limiter [94]. Of particular interest is also the WENO methodology applied to the DG family of numerical schemes, mainly developed by Shu, Qiu and Zhu *et al.* (see [95–97]). Similar to the finite volume version, the WENO technique for DG methods applies a convex combination of the solution within a computational cell and its immediate neighbors. In this way, spurious oscillation can be purged by the regularization effect of the weighted combination of the smooth solutions at the neighbors cells.

Another major issue in numerical simulations in general and the simulation of geophysical flows in particular is the preservation of stationary solutions. Indeed, in practice it is common to find relevant simulations that consist essentially in a perturbation of an *equilibrium* state. Furthermore, some times we are interested in long time integration to achieve the convergence towards an equilibrium. Therefore, numerical schemes must be able to preserve these steady states exactly or, at least, with an enhanced accuracy. An additional argument in favor of numerical schemes that preserve stationary solutions arise when we are in the case of a small perturbation of such steady states. In those cases, the numerical noise introduced by the scheme can effectively destroy the perturbation. Of course, a first approximation to solve this issue can be increasing the number of discretization points or the order of the numerical scheme. While this strategy can improve the numerical behavior, it does not address the underlying problem and can quickly become computationally expensive. Moreover, failing to preserve stationary solutions has a direct impact on the long time stability of the numerical scheme.

Numerical schemes that can exactly preserve stationary solutions are called well-balanced. This property was first introduced by Bermúdez and Vázquez-Cendón in [44] for the sallow water equations, where they defined the *C-property*, or the ability of a numerical scheme to solve a set of stationary solutions exactly or, at least, with enhanced accuracy. Since then, well-balanced numerical schemes have been an active research topic, like [10, 15, 98–127]. Recently, in [128] the authors propose well-balanced high order schemes for systems of balance laws in the framework of the finite volume method for both continuous and discontinuous source term.

Following that work, in this thesis we also develop a general framework for high-order well balanced Discontinuous Galerkin methods for general systems of balance laws. To our knowledge, this is the first time that a general procedure for designing well-balanced methods in the DG framework is developed. Nevertheless, there also exits a wide literature about the derivation of well-balanced method in the DG framework for different problems (see for example [129–140]). Moreover, as we are interested in the simulation of perturbations around a stratified fluid, a study of some relevant stationary

UNIVERSIDAD DE MÁLAGA

solutions for the shallow-water model with variable pressure is also discussed in this thesis.

Finally, an additional resource to improve the overall computation time is code optimization and hardware parallelization. Specifically, modern Graphics Processor Units (GPUs) offer the most cost-efficient approximation to obtain reasonable computational times for big domains in space and time. GPUs offer massive parallel opportunities, thanks to the hundreds of processing units optimized for performing floating point operations and multithreaded execution. This special architecture allows to obtain computational times that are order of magnitude lower than the comparable CPU. The implementation of numerical schemes on GPUs has been widely used, also in the framework of geophysical flows simulations (see [141]). In this thesis, a brief study of different GPU parallelization strategies is included as an appendix in order to explore different approximations for the shallow water model with variable density.

This thesis is based in the following publications:

- Guerrero Fernández E, Castro Díaz MJ, Morales de Luna T. A Second-Order Well-Balanced Finite Volume Scheme for the Multilayer Shallow Water Model with Variable Density. Mathematics. 2020; 8(5):848. https://doi.org/10.3390/math8050848

- Guerrero Fernández E, Castro Díaz MJ, Dumbser M, Morales de Luna T. An Arbitrary High Order Well-Balanced ADER-DG Numerical Scheme for the Multilayer Shallow-Water Model with Variable Density. J Sci Comput. 2022; 90:52. https://doi.org/10.1007/s10915-021-01734-2

- Guerrero Fernández E, Escalante C, Castro Díaz MJ. Well-Balanced High-Order Discontinuous Galerkin Methods for Systems of Balance Laws. Mathematics. 2022; 10(1):15. https://doi.org/10.3390/math10010015

Outline of the thesis

The outline of this thesis is as follows:

- In chapter 1 a derivation of the multilayer shallow-water model with variable density considered in this thesis is included, alongside a discussion on some relevant stationary solutions.

- In chapter 2, a broad introduction on designing high order well-balanced finite volume and DG methods is described. In this chapter, we also describe a general strategy to design high-order well-balanced DG numerical schemes.

- Chapter 3 is dedicated to particular discretizations of the multilayer shallow-water model with variable pressure. Includes a finite volume and ADER-DG discretization.

UNIVERSIDAD DE MÁLAGA

- Chapter 4 is devoted to check the properties and efficiency of the numerical schemes developed in this thesis. Several experiments are considered, including well-balanced test and comparison with laboratory data.
- Finally, chapter 5 offers some conclusions and speculate with some possible future work related with the results obtained in this thesis.

UNIVERSIDAD DE MÁLAGA

Chapter 1

Model derivation and stationary solutions

The shallow water equations are widely used when simulating geophysical flows. They are an excellent balance between the complexity of fluid hydrodynamics and the computational load associated with the huge computational domains characteristic of geophysical flows. However, the standard shallow water hypothesis assumes a homogeneous vertical behavior that does not allow to recover vertical information for complex fluid hydrodynamics that are essential in some flows. To address this problem, multilayer shallow water models were developed in the recent years (see [7–10]).

The multilayer approach divides the fluid column in a number of computational layers, applying in each one the same hypotheses of the shallow water: vertically averaged horizontal velocity and hydrostatic pressure (see [5]). Mass and momentum transfer between layer is accounted for by an approximation of the vertical fluxes in the form of non-conservative products (see [7, 142] or [9]). In this way, a rich vertical profile for the velocity can be recovered at the cost of greater computational effort.

Under the multilayer approach, complex vertical profiles of geophysical flows can be accurately simulated. However, standard multilayer shallow-water systems do not take into account the stratification of the fluid due to density changes. Here, we propose a multilayer shallow-water system for a varying density fluid. The procedure carried out here corresponds to an extension of the one presented in [7] for the variable density case.

The model described here is quite different from other approximations that assume immiscible layers (see [43, 143]) or consider the water entrainment as an heuristic-dependent source term (see [36]). Here, the density variations freely circulate through the layers. Consequentially, it is also very different with respect to multilayer shallow-water model where the pressure terms are taken into account as external forces that affect the layers (see [144]). Therefore, the model is similar to [34, 35], where the density differences are due to suspended sediment with its own settling velocity. It is also similar to the one in [38], where the density variations are taken into account through a given

UNIVERSIDAD DE MÁLAGA

tracer.

The main differences with the aforementioned models and this work involves the model derivation and, in particular, the transfer terms that will be discussed later on in this Chapter. Additionally, a study of some relevant stationary solutions of the model is also included. The correct definition of the *equilibrium* states are of paramount importance to design numerical schemes that are well-balanced, in other words, that preserve the aforementioned stationary solutions exactly. The well-balanced property is crucial for the long term numerical stability of the schemes and very important in geophysical flow simulations. Not only the trivial stationary solutions are studied but also non-trivial ones, corresponding to the fluid stratification in density, a common stationary solution. The exact preservation of stationary solutions, especially for the shallow water equations, is a relevant research topic (see [44, 145, 146], or [41], and the references therein).

1.1 Model derivation

We begin by deriving a multilayer shallow-water model with density dependent pressure terms that depend on a general convection-diffusion equation. This model can be derived from the free surface compressible Navier-Stokes equations in a d-dimensional space ($d = 2, 3$),

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \boldsymbol{v}) = 0, \\ \partial_t (\rho\, \boldsymbol{v}) + \nabla \cdot (\rho \boldsymbol{v} \otimes \boldsymbol{v}) = -g\, \rho \boldsymbol{k} + \nabla \cdot \boldsymbol{\Sigma}, \end{cases} \tag{1.1.1}$$

where $\boldsymbol{v} = (\boldsymbol{u}, w) \in \mathbb{R}^d$ is the velocity function with $\boldsymbol{u} \in \mathbb{R}^{d-1}$ the horizontal velocity and $w \in \mathbb{R}$ the vertical one. Additionally, $\boldsymbol{k} \in \mathbb{R}^d$ is the unit vector pointing downwards, $g \in \mathbb{R}$ is the gravity constant and the total stress tensor is given by $\boldsymbol{\Sigma} = -p\boldsymbol{I} + \boldsymbol{\sigma}$, with $\boldsymbol{I}$ the identity tensor, $p \in \mathbb{R}$ the pressure term, and $\boldsymbol{\sigma}$ the tensor corresponding to the viscous terms, given by

$$\boldsymbol{\sigma} = \frac{1}{2}\mu \left((\nabla \boldsymbol{v}) + (\nabla \boldsymbol{v})^T \right) = \left(\begin{array}{c|c} \boldsymbol{\sigma}_H & \sigma_{xz} \\ \hline \sigma'_{xz} & \sigma_{zz} \end{array} \right).$$

Finally, $\mu \in \mathbb{R}$ is the dynamic viscosity and $\rho \in \mathbb{R}$ is the non-constant density function. This density function will depend on a given equation of state R in order to reflect changes in density due to temperature and/or salinity. These effects are taken into account thanks to a tracer function $T = T(t, \boldsymbol{x}, z)$, with $\boldsymbol{x} = (x_1, \ldots, x_{d-1}) \in \mathbb{R}^{d-1}$ the horizontal coordinates, that is advected and diffused with a diffusivity constant $\nu_T \in \mathbb{R}$ within the flow,

$$\partial_t (\rho T) + \nabla \cdot (\rho T \boldsymbol{v}) + \nabla \cdot (\rho \nu_T \nabla T) = 0, \tag{1.1.2}$$

and

$$\rho = R(T). \tag{1.1.3}$$

UNIVERSIDAD DE MÁLAGA

Note that in the case that the density of the fluid is only temperature dependent or salinity dependent, then the tracer T represent temperature or salinity respectively.

We now proceed to derive a multilayer shallow-water approach to the system (1.1.1)-(1.1.2). As usual in the multilayer approach, the fluid domain $\Omega_F(t)$ at a given time $t \in [0, T_f]$ $T_f \in \mathbb{R}^+$, is divided along the vertical dimension z into a set of $M \in \mathbb{N}^*$ layers of thickness $h_\alpha(t,x)$ (see Figure 1.1). The $M+1$ interfaces between layers, $\Gamma_{\alpha+\frac{1}{2}}(t)$, are described as,

$$\Gamma_{\alpha+\frac{1}{2}}(t) = \left\{(\boldsymbol{x}, z) \in \Omega_F(t), \text{ such that } z = z_{\alpha+\frac{1}{2}}(t, \boldsymbol{x})\right\}, \quad \alpha = 0, 1, \ldots, M, \tag{1.1.4}$$

while the layer is defined as

$$\Omega_\alpha(t) = \left\{(\boldsymbol{x}, z) \in \Omega_F(t), \text{ such that } z_{\alpha-\frac{1}{2}}(t, \boldsymbol{x}) < z < z_{\alpha+\frac{1}{2}}(t, \boldsymbol{x})\right\}, \quad \alpha = 1, \ldots, M. \tag{1.1.5}$$

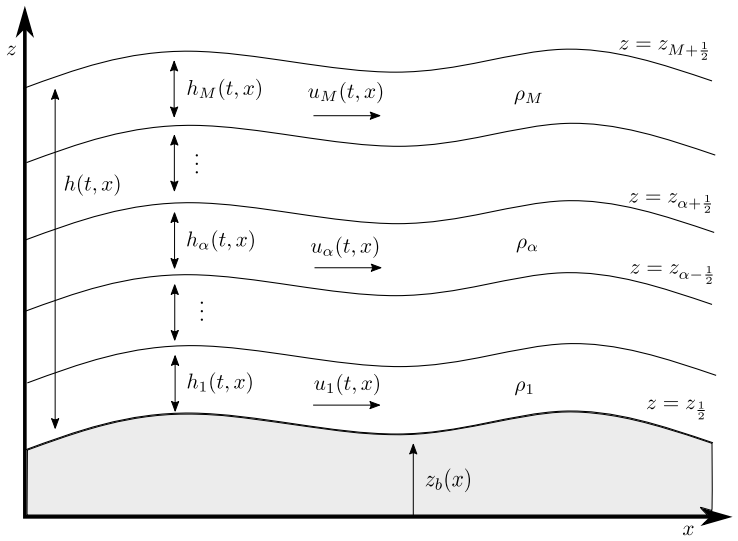


Figure 1.1: Sketch of the multilayer approach in one dimension.

The projection of $\Omega_F(t)$ onto the horizontal plane is denoted by $I_F(t)$. Additionally, the bottom and free surface functions at $\Gamma_B(t)$ and $\Gamma_S(t)$ are denoted by z_B and η respectively. Notice that the total height of the fluid column is given by $h = \eta - z_B = \sum_{\beta=1}^{M} h_\beta$. Moreover, we have $\partial\Omega_F(t) = \Gamma_B(t) \cup \Gamma_S(t) \cup \Theta(t)$, where $\Theta(t)$ is the inflow/outflow boundary, which we assume here to be vertical.

We shall also set the following notation for a generic function f and $\alpha = 0, 1, \ldots, M$,

$$f^-_{\alpha+\frac{1}{2}} \coloneqq (f_{|\Omega_\alpha(t)})_{|\Gamma_{\alpha+\frac{1}{2}}(t)} \quad \text{and} \quad f^+_{\alpha+\frac{1}{2}} \coloneqq (f_{|\Omega_{\alpha+1}(t)})_{|\Gamma_{\alpha+\frac{1}{2}}(t)}.$$

UNIVERSIDAD DE MÁLAGA

Meanwhile, if the function is continuous across $\Gamma_{\alpha+\frac{1}{2}}$ we shall simply denote

$$f_{\alpha+\frac{1}{2}} \coloneqq f^-_{\alpha+\frac{1}{2}} = f^+_{\alpha+\frac{1}{2}}.$$

Additionally, the ∇_x operator stands for $\nabla_x = (\partial_{x_1}, \ldots, \partial_{x_{d-1}})$. Also, the vector $\boldsymbol{n}_{\alpha+\frac{1}{2}}$ is the space unit normal vector to the interface $\Gamma_{\alpha+\frac{1}{2}}$ outwards to the layer $\Omega_\alpha(t)$ for any given time t and for $\alpha = 0, 1, \ldots, N$. This vector is defined by

$$\boldsymbol{n}_{\alpha+\frac{1}{2}} = \frac{\left(\nabla_x z_{\alpha+\frac{1}{2}}, -1\right)'}{\sqrt{1 + \left|\nabla_x z_{\alpha+\frac{1}{2}}\right|^2}}. \tag{1.1.6}$$

Meanwhile the space-time unit normal vector to $\Gamma_{\alpha+\frac{1}{2}}$ is,

$$\boldsymbol{n}_{T,\alpha+\frac{1}{2}} = \frac{\left(\partial_t z_{\alpha+\frac{1}{2}}, \nabla_x z_{\alpha+\frac{1}{2}}, -1\right)'}{\sqrt{1 + \left|\nabla_x z_{\alpha+\frac{1}{2}}\right|^2 + \left|\partial_t z_{\alpha+\frac{1}{2}}\right|^2}}. \tag{1.1.7}$$

The jump of the concatenated pair $(a; b)$ through $\Gamma_{\alpha+\frac{1}{2}}$ is denoted as

$$[(a;b)]_{\Gamma_{\alpha+\frac{1}{2}}} = \Big((a;b)_{|\Omega_{\alpha+1}(t)} - (a;b)_{|\Omega_\alpha(t)}\Big)_{|\Gamma_{\alpha+\frac{1}{2}}}.$$

Finally, we denote a quantity within a layer with the following simplified notation,

$$\boldsymbol{v}_{|\Omega_\alpha(t)} \coloneqq \boldsymbol{v}_\alpha \coloneqq (\boldsymbol{u}_\alpha, w_\alpha)', \qquad \rho_{|\Omega_\alpha} \coloneqq \rho_\alpha, \qquad T_{|\Omega_\alpha} \coloneqq T_\alpha, \qquad p_{|\Omega_\alpha} \coloneqq p_\alpha.$$

1.1.1 Weak solutions with discontinuities

We consider that the piece-wise weak solution $(\boldsymbol{v}, \rho, p)$, which is smooth within each layer $\Omega_\alpha(t)$ but possibly discontinuous across layer interfaces $\Gamma_{\alpha+\frac{1}{2}}(t)$ for $\alpha = 1, \ldots, M-1$. More explicitly, for $(\boldsymbol{v}, \rho, p)$ to be a solution of (1.1.1), the following conditions must be fulfilled:

- In each layer $\Omega_\alpha(t)$, $(\boldsymbol{v}, \rho, p)$ is a standard weak solution of the system (1.1.1).
- The normal flux conditions are satisfied at each layer interface $\Gamma_{\alpha+\frac{1}{2}}(t)$ for $\alpha = 0, 1, \ldots, M$:
  - For the mass equation,

$$[(\rho; \rho\,\boldsymbol{v})]_{|\Gamma_{\alpha+\frac{1}{2}}(t)} \cdot \boldsymbol{n}_{T,\alpha+\frac{1}{2}} = 0. \tag{1.1.8}$$

UNIVERSIDAD DE MÁLAGA

- For the momentum conservation law,

$$[\rho\,\boldsymbol{v};\rho\,\boldsymbol{v}\otimes\boldsymbol{v}-\boldsymbol{\Sigma}]_{|\Gamma_{\alpha+\frac{1}{2}}(t)}\cdot\boldsymbol{n}_{T,\alpha+\frac{1}{2}}=0. \tag{1.1.9}$$

Additionally, the solution T of the tracer equation (1.1.2) must also be a standard weak solution of the system (1.1.1)-(1.1.2) and fulfill the following normal flux condition at each layer interface $\Gamma_{\alpha+\frac{1}{2}}(t)$ for $\alpha=0,1,\ldots,M$:

$$[\rho\,T;\rho\,T\boldsymbol{v}-\rho\nu_T\nabla T]_{|\Gamma_{\alpha+\frac{1}{2}}(t)}\cdot\boldsymbol{n}_{T,\alpha+\frac{1}{2}}=0.$$

Following the usual approach in multilayer shallow water models, the following hypotheses are assumed:

- The horizontal velocity $\boldsymbol{u}_\alpha$, the density function ρ_α and the tracer T_α are constant in the vertical direction and thus they do not depend on z inside each layer.
- Both the vertical velocity w_α and the pressure terms p_α are linear in z inside each layer.

Of course, there is no such a particular set $\big(\boldsymbol{v}_\alpha,\rho_\alpha,T_\alpha,p_\alpha\big)$ that is a solution of the complete equations in each layer $\Omega_\alpha(t)$, $\alpha=1,\ldots,M$. Instead, a reduced weak formulation with particular test functions will be derived in the following. However, first we focus on given a thoroughly description of the jump conditions for the mass and momentum equations.

1.1.1.1 Mass conservation jump conditions

The mass conservation jump conditions are satisfied provided that the transfer terms at each layer are balanced,

$$G_{\alpha+\frac{1}{2}}\coloneqq G^-_{\alpha+\frac{1}{2}}=G^+_{\alpha+\frac{1}{2}}, \tag{1.1.10}$$

where

$$\begin{cases} G^+_{\alpha+\frac{1}{2}}=\rho_{\alpha+1}\big(\partial_t z_{\alpha+\frac{1}{2}}+\boldsymbol{u}_{\alpha+1}\cdot\nabla_x z_{\alpha+\frac{1}{2}}-w^+_{\alpha+\frac{1}{2}}\big),\\ G^-_{\alpha+\frac{1}{2}}=\rho_{\alpha}\big(\partial_t z_{\alpha+\frac{1}{2}}+\boldsymbol{u}_{\alpha}\cdot\nabla_x z_{\alpha+\frac{1}{2}}-w^-_{\alpha+\frac{1}{2}}\big), \end{cases} \tag{1.1.11}$$

that has been obtained from (1.1.8) and (1.1.10).

1.1.1.2 Momentum conservation jump conditions

The momentum conservation jump condition (1.1.9) is

$$\left[\Big(\rho\,\boldsymbol{v};\rho\,\boldsymbol{v}\otimes\boldsymbol{v}-\boldsymbol{\Sigma}\Big)\right]_{|\Gamma_{\alpha+\frac{1}{2}}(t)}\cdot\boldsymbol{n}_{T,\alpha+\frac{1}{2}}=0.$$

UNIVERSIDAD DE MÁLAGA

which can be rewritten as

$$\big[\boldsymbol{\Sigma}\big]_{|\Gamma_{\alpha+\frac{1}{2}}(t)} \cdot \big(\nabla_x z_{\alpha+\frac{1}{2}}, -1\big)' = \left[\left(\rho\,\boldsymbol{v}; \rho\,\boldsymbol{v}\otimes\boldsymbol{v}\right)\right]_{|\Gamma_{\alpha+\frac{1}{2}}(t)} \cdot \big(\partial_t z_{\alpha+\frac{1}{2}}, \nabla_x z_{\alpha+\frac{1}{2}}, -1\big)'.$$

Using (1.1.10), we have that

$$\left[\left(\rho\,\boldsymbol{v}; \rho\,\boldsymbol{v}\otimes\boldsymbol{v}\right)\right]_{|\Gamma_{\alpha+\frac{1}{2}}(t)} \cdot \big(\partial_t z_{\alpha+\frac{1}{2}}, \nabla_x z_{\alpha+\frac{1}{2}}, -1\big) = G_{\alpha+\frac{1}{2}}[\boldsymbol{v}]_{|\Gamma_{\alpha+\frac{1}{2}}(t)},$$

and

$$\big[\boldsymbol{\Sigma}\big]_{|\Gamma_{\alpha+\frac{1}{2}}(t)} \cdot \big(\nabla_x z_{\alpha+\frac{1}{2}}, -1\big) = G_{\alpha+\frac{1}{2}}[\boldsymbol{v}]_{|\Gamma_{\alpha+\frac{1}{2}}(t)}.$$

Therefore, condition (1.1.9) may be rewritten as,

$$\big[\boldsymbol{\Sigma}\big]_{|\Gamma_{\alpha+\frac{1}{2}}(t)} \cdot \boldsymbol{n}_{\alpha+\frac{1}{2}} = \frac{1}{\sqrt{1+\left|\nabla_x z_{\alpha+\frac{1}{2}}\right|^2}} G_{\alpha+\frac{1}{2}}[\boldsymbol{v}]_{|\Gamma_{\alpha+\frac{1}{2}}(t)}. \tag{1.1.12}$$

The total stress tensor across $\Gamma_{\alpha+\frac{1}{2}}$ is

$$\boldsymbol{\Sigma}^{\pm}_{\alpha+\frac{1}{2}} = -p_{\alpha+\frac{1}{2}}\boldsymbol{I} + \boldsymbol{\sigma}^{\pm}_{\alpha+\frac{1}{2}}, \tag{1.1.13}$$

where $p_{\alpha+\frac{1}{2}}$ is the kinematic pressure, assumed as continuous across the layer interface $\Gamma_{\alpha+\frac{1}{2}}$ and $\boldsymbol{\sigma}^{\pm}_{\alpha+\frac{1}{2}}$ is the limit approximation of $\boldsymbol{\sigma}(\boldsymbol{v})$ at $\Gamma_{\alpha+\frac{1}{2}}$, defined as $\boldsymbol{\sigma}(\boldsymbol{v}) = \frac{1}{2}\mu\left(\nabla(\boldsymbol{u}) + \nabla(\boldsymbol{u})^T\right)$. Combining (1.1.12) with (1.1.13), we obtain,

$$\left(\boldsymbol{\sigma}^{+}_{\alpha+\frac{1}{2}} - \boldsymbol{\sigma}^{-}_{\alpha+\frac{1}{2}}\right) \cdot \boldsymbol{n}_{\alpha+\frac{1}{2}} = \frac{1}{\sqrt{1+\left|\nabla_x z_{\alpha+\frac{1}{2}}\right|^2}} G_{\alpha+\frac{1}{2}}[\boldsymbol{v}]_{|\Gamma_{\alpha+\frac{1}{2}}(t)}. \tag{1.1.14}$$

Moreover, $\boldsymbol{\sigma}^{\pm}_{\alpha+\frac{1}{2}}$ should also satisfy a consistency condition,

$$\frac{1}{2}\left(\boldsymbol{\sigma}^{+}_{\alpha+\frac{1}{2}} + \boldsymbol{\sigma}^{-}_{\alpha+\frac{1}{2}}\right) = \widetilde{\boldsymbol{\sigma}}_{\alpha+\frac{1}{2}} = \left(\begin{array}{c|c} \widetilde{\boldsymbol{\sigma}}_{H,\alpha+\frac{1}{2}} & \widetilde{\sigma}_{xz,\alpha+\frac{1}{2}} \\ \hline \widetilde{\sigma}'_{xz,\alpha+\frac{1}{2}} & \widetilde{\sigma}_{zz,\alpha+\frac{1}{2}} \end{array}\right), \tag{1.1.15}$$

where $\widetilde{\boldsymbol{\sigma}}_{\alpha+\frac{1}{2}}$ is an approximation of $\boldsymbol{\sigma}(\boldsymbol{v})_{|\Gamma_{\alpha+\frac{1}{2}}}$ that we define as follows:

$$\widetilde{\boldsymbol{\sigma}}_{\alpha+\frac{1}{2}} = \frac{1}{2}\mu\frac{(\rho_{\alpha+1}+\rho_\alpha)}{2}\widetilde{D}_{\alpha+\frac{1}{2}}, \tag{1.1.16}$$

UNIVERSIDAD DE MÁLAGA

with

$$\widetilde{D}_{\alpha+\frac{1}{2}} = \begin{pmatrix} D_H\left(\frac{\boldsymbol{u}_{\alpha+1}+\boldsymbol{u}_\alpha}{2}\right) & \left(\nabla_x\left(\frac{w^+_{\alpha+\frac{1}{2}}+w^-_{\alpha+\frac{1}{2}}}{2}\right)\right)' + \boldsymbol{Q}_{H,\alpha+\frac{1}{2}} \\ \nabla_x\left(\frac{w^+_{\alpha+\frac{1}{2}}+w^-_{\alpha+\frac{1}{2}}}{2}\right) + (\boldsymbol{Q}_{H,\alpha+\frac{1}{2}})' & 2Q_{v,\alpha+\frac{1}{2}} \end{pmatrix} \tag{1.1.17}$$

where $\boldsymbol{Q}_{\alpha+\frac{1}{2}} = \boldsymbol{Q}(\widetilde{\boldsymbol{u}})$ at $\Gamma_{\alpha+\frac{1}{2}}$ and $\boldsymbol{Q}$ satisfies the following equation

$$\boldsymbol{Q} - \partial_z \boldsymbol{v} = 0, \text{with } \boldsymbol{Q} = (\boldsymbol{Q_H}, Q_v). \tag{1.1.18}$$

To approximate $\boldsymbol{Q}$, a solution of (1.1.18), we approximate $\boldsymbol{v}$ by $\widetilde{\boldsymbol{u}}$ that is a linear interpolation in z, such that $\widetilde{\boldsymbol{u}}_{|z=\frac{1}{2}(z_{\alpha-\frac{1}{2}}+z_{\alpha+\frac{1}{2}})} = \boldsymbol{u}_\alpha$. Finally, we can solve the system defined by (1.1.14) and the equation resulting from multiplying scalarly (1.1.15) by $\boldsymbol{n}_{\alpha+\frac{1}{2}}$. In this way, we can obtain the expressions of $\boldsymbol{\sigma}^\pm_{\alpha+\frac{1}{2}}$ that satisfy the jump condition and the consistency condition on the interface. We can solve it and we obtain,

$$\boldsymbol{\sigma}^\pm_{\alpha+\frac{1}{2}} \cdot \boldsymbol{n}_{\alpha+\frac{1}{2}} = \widetilde{\boldsymbol{\sigma}}_{\alpha+\frac{1}{2}} \cdot \boldsymbol{n}_{\alpha+\frac{1}{2}} \pm \frac{1}{\sqrt{1+\left|\nabla_x z_{\alpha+\frac{1}{2}}\right|^2}} G_{\alpha+\frac{1}{2}}[\boldsymbol{v}]_{|\Gamma_{\alpha+\frac{1}{2}}(t)}. \tag{1.1.19}$$

1.1.1.3 Diffusion gradient decomposition

The same arguments can be applied to approximate the limits of the gradient of the tracer T at both sides of the layer interface $\Gamma_{\alpha+\frac{1}{2}}$, denoted by $\nabla^\pm_{\alpha+\frac{1}{2}}(T)$. In this way, the decomposition of the gradient is,

$$\nu_T \rho^\pm_{\alpha+\frac{1}{2}} \nabla^\pm_{\alpha+\frac{1}{2}}(T) \cdot \boldsymbol{n}^+_{\alpha+\frac{1}{2}} = \nu_T \widetilde{\nabla(T)}_{\alpha+\frac{1}{2}} \cdot \boldsymbol{n}_{\alpha+\frac{1}{2}} \pm \frac{1}{\sqrt{1+\left|\nabla_x z_{\alpha+\frac{1}{2}}\right|^2}} G_{\alpha+\frac{1}{2}}[T]_{|\Gamma_{\alpha+\frac{1}{2}}(t)}. \tag{1.1.20}$$

where $\widetilde{\nabla(T)}_{\alpha+\frac{1}{2}}$ is a consistent approximation of the tensor at the interface $\Gamma_{\alpha+\frac{1}{2}}$,

$$\widetilde{\nabla(T)}_{\alpha+\frac{1}{2}} = \frac{1}{2}\left(\rho^+_{\alpha+\frac{1}{2}} \nabla T^+_{\alpha+\frac{1}{2}} + \rho^-_{\alpha+\frac{1}{2}} \nabla T^-_{\alpha+\frac{1}{2}}\right). \tag{1.1.21}$$

1.1.2 Vertical velocity

When setting the horizontal velocities as independent of the vertical direction z, we are in fact assuming a linear profile for the vertical velocity w. Since $\boldsymbol{v}_\alpha$ is a solution of the mass equation in (1.1.1) in Ω_α, the vertical integration of the mass equation in (1.1.1) yields,

$$w_\alpha(t,\boldsymbol{x},z) = w^+_{\alpha-\frac{1}{2}}(t,\boldsymbol{x}) - (z - z_{\alpha-\frac{1}{2}})\frac{1}{\rho_\alpha}\left(\partial_t \rho_\alpha + \nabla_x \cdot (\rho_\alpha \boldsymbol{u}_\alpha)\right), \quad \text{for } \alpha = 1, \ldots, M.$$

UNIVERSIDAD DE MÁLAGA

From conditions (1.1.10), we can also deduce,

$$w^+_{\alpha+\frac{1}{2}} = \frac{1}{\rho_{\alpha+1}}\big((\rho_{\alpha+1}-\rho_\alpha)\partial_t z_{\alpha+\frac{1}{2}} + (\rho_{\alpha+1}\boldsymbol{u}_{\alpha+1}-\rho_\alpha\boldsymbol{u}_\alpha)\cdot\nabla_x z_{\alpha+\frac{1}{2}} + \rho_\alpha w^-_{\alpha+\frac{1}{2}}\big).$$

Therefore, using the horizontal velocities obtained from the model, we can compute the linear vertical velocities in the layers by following these steps:

- First, the quantity $w^+_{\frac{1}{2}}$ is determined from the given mass exchange through the bottom, $G_{\frac{1}{2}}$ and using (1.1.10) by

$$w^+_{\frac{1}{2}} = \boldsymbol{u}_1\cdot\nabla_x z_B + \partial_t z_B - \frac{1}{\rho_1}G_{\frac{1}{2}}.$$

- Then, for $\alpha = 1,\dots,N$ and $z\in[z_{\alpha+\frac{1}{2}}, z_{\alpha-\frac{1}{2}}]$, we set

$$w_\alpha(t,\boldsymbol{x},z) = w^+_{\alpha-\frac{1}{2}}(t,\boldsymbol{x}) - (z-z_{\alpha-\frac{1}{2}})\frac{1}{\rho_\alpha}\left(\partial_t\rho_\alpha + \nabla_x\cdot(\rho_\alpha\boldsymbol{u}_\alpha)\right),$$

and

$$w^+_{\alpha+\frac{1}{2}} = \frac{1}{\rho_{\alpha+1}}\left((\rho_{\alpha+1}-\rho_\alpha)\partial_t z_{\alpha+\frac{1}{2}} + (\rho_{\alpha+1}\boldsymbol{u}_{\alpha+1}-\rho_\alpha\boldsymbol{u}_\alpha)\cdot\nabla_x z_{\alpha+\frac{1}{2}} + \rho_\alpha w^-_{\alpha+\frac{1}{2}}\right).$$

Note that $w^-_{\alpha+\frac{1}{2}}$ is computed from evaluating $w_\alpha(t,\boldsymbol{x},z_{\alpha+\frac{1}{2}})$.

In this way, an enhanced velocity profile is recuperated and the vertical velocity can be recovered from the horizontal one.

1.1.3 A particular weak solution with hydrostatic pressure

As already stated, the pressure terms are assumed hydrostatic,

$$p_\alpha(t,x,z) = p_{\alpha+\frac{1}{2}} + \rho_\alpha\, g\,(z_{\alpha+\frac{1}{2}} - z), \tag{1.1.22}$$

with

$$p_{\alpha+\frac{1}{2}}(t,x) = p_S(t,x) + g\sum_{\beta=\alpha+1}^{M}\rho_\beta h_\beta(t,x). \tag{1.1.23}$$

Here, $p_{\alpha+\frac{1}{2}}$ is the kinematic pressure at the layer interface $\Gamma_{\alpha+\frac{1}{2}}(t)$ and p_S denotes the pressure at the free surface, usually set to zero.

UNIVERSIDAD DE MÁLAGA

Let us consider the weak formulation of system (1.1.1)-(1.1.2) in $\Omega_\alpha(t)$. A weak formulation of the original equations in $\Omega_\alpha(t)$ for $\alpha = 1, \dots, M$ should satisfy

$$\begin{cases} \displaystyle\int_{\Omega_\alpha(t)} (\partial_t \rho_\alpha + \nabla \cdot (\rho_\alpha \boldsymbol{v}_\alpha))\varphi \, d\Omega = 0, \\ \displaystyle\int_{\Omega_\alpha(t)} (\partial_t (\rho_\alpha T_\alpha) + \nabla \cdot (\rho_\alpha T \boldsymbol{v}_\alpha) + \nabla \cdot (\nu_T \rho_\alpha \nabla T))\, \varphi \, d\Omega = 0, \\ \displaystyle\int_{\Omega_\alpha(t)} \partial_t (\rho_\alpha \boldsymbol{v}) \boldsymbol{\vartheta} \, d\Omega + \int_{\Omega_\alpha(t)} \nabla \cdot (\rho_\alpha \, \boldsymbol{v} \otimes \boldsymbol{v}) \boldsymbol{\vartheta} \, d\Omega = - \int_{\Omega_\alpha(t)} g \, \rho_\alpha \boldsymbol{k} \boldsymbol{\vartheta} \, d\Omega + \int_{\Omega_\alpha(t)} (\nabla \cdot \boldsymbol{\Sigma}) \boldsymbol{\vartheta} \, d\Omega, \end{cases} \tag{1.1.24}$$

for all $\varphi \in L^2(\Omega_\alpha(t))$ and for all $\boldsymbol{\vartheta} \in H^1(\Omega_\alpha(t))$ with $\boldsymbol{\vartheta}_{|\partial I_F} = 0$. We consider the following particular vertical structure of the test function,

$$\partial_z \varphi = 0, \tag{1.1.25}$$

$$\boldsymbol{\vartheta}(t, x, z) = \Big(\vartheta_H(t, x), (z - z_B) V(t, x) \Big), \tag{1.1.26}$$

where ϑ_H and $V(t, x)$ are smooth functions that do not depend on z. In this way, the following computations are considered.

1.1.3.1 Mass conservation equation

We choose a scalar test function $\varphi = \phi(t, x)$ independent of the vertical direction z. Then,

$$\begin{aligned} 0 &= \int_{\Omega_\alpha(t)} (\partial_t \rho_\alpha + \nabla \cdot (\rho_\alpha \boldsymbol{v}_\alpha))\varphi \, d\Omega \\ &= \int_{I_F(t)} \phi(t, x) \Big\{ \int_{z_{\alpha-\frac{1}{2}}}^{z_{\alpha+\frac{1}{2}}} (\partial_t \rho_\alpha + \nabla_x \cdot (\rho_\alpha \boldsymbol{u}_\alpha) + \partial_z (\rho_\alpha w_\alpha)) \, dz \Big\} dx \\ &= \int_{I_F(t)} \phi(t, x) \Big\{ \partial_t \int_{z_{\alpha-\frac{1}{2}}}^{z_{\alpha+\frac{1}{2}}} \rho_\alpha \, dz + \nabla_x \cdot \int_{z_{\alpha-\frac{1}{2}}}^{z_{\alpha+\frac{1}{2}}} \rho_\alpha \boldsymbol{u}_\alpha \, dz \\ &\quad - \rho_\alpha \partial_z z_{\alpha+\frac{1}{2}} - \rho_\alpha \boldsymbol{u}_\alpha \cdot \nabla_x z_{\alpha+\frac{1}{2}} + \rho_\alpha w^-_{\alpha+\frac{1}{2}} + \rho_\alpha \partial_z z_{\alpha-\frac{1}{2}} + \rho_\alpha \boldsymbol{u}_\alpha \cdot \nabla_x z_{\alpha-\frac{1}{2}} - \rho_\alpha w^+_{\alpha-\frac{1}{2}} \Big\} dx, \end{aligned}$$

for all $\phi(t, \cdot) \in L^2(I_F(t))$. Finally, taking into account that $h_\alpha = z_{\alpha+\frac{1}{2}} - z_{\alpha-\frac{1}{2}}$ and the expression for the transfer terms between layers (1.1.11), we obtain the final expression for the mass equation:

$$\partial_t (\rho_\alpha h_\alpha) + \nabla_x \cdot (\rho_\alpha h_\alpha \boldsymbol{u}_\alpha) = G_{\alpha+\frac{1}{2}} - G_{\alpha-\frac{1}{2}}, \quad \alpha = 1, \dots, M. \tag{1.1.27}$$

Note that the transfer terms G_{N+1} and $G_{\frac{1}{2}}$ correspond with a net mass contribution in the free surface and bottom respectively. They correspond to raining or filtration effects and they may be set to zero if these effects are not present.

UNIVERSIDAD DE MÁLAGA

1.1.3.2 Momentum conservation equation

After some calculations very similar to those in [34] we arrive to the following expression of the momentum equation,

$$\begin{aligned}
&\partial_t(h_\alpha\rho_\alpha\boldsymbol{u}_\alpha)+\nabla_x(h_\alpha\rho_\alpha\boldsymbol{u}_\alpha\otimes\boldsymbol{u}_\alpha)+gh_\alpha\rho_\alpha\nabla_x\eta-\nabla_x(h_\alpha\,\boldsymbol{\sigma}_H)\\
&+gh_\alpha\left(\sum_{\beta=\alpha+1}^{M}(\rho_\beta-\rho_\alpha)\nabla_x h_\beta\right)+gh_\alpha\sum_{\beta=\alpha+1}^{M}h_\beta\nabla_x\rho_\beta+\frac{1}{2}gh_\alpha^2\nabla_x\rho_\alpha\\
&=\boldsymbol{u}^-_{\alpha+\frac{1}{2}}G_{\alpha+\frac{1}{2}}-\boldsymbol{u}^+_{\alpha-\frac{1}{2}}G_{\alpha-\frac{1}{2}}-\boldsymbol{\sigma}^-_{\alpha+\frac{1}{2}}\cdot\boldsymbol{n}_{\alpha+\frac{1}{2}}\sqrt{1+|\nabla_x z_{\alpha+\frac{1}{2}}|^2}+\boldsymbol{\sigma}^+_{\alpha-\frac{1}{2}}\cdot\boldsymbol{n}_{\alpha-\frac{1}{2}}\sqrt{1+|\nabla_x z_{\alpha-\frac{1}{2}}|^2},
\end{aligned}\tag{1.1.28}$$

for $\alpha=1,\dots,M$. Note that the transfer terms can be written in the following way,

$$\left(\frac{\boldsymbol{u}^+_{\alpha+\frac{1}{2}}+\boldsymbol{u}^-_{\alpha+\frac{1}{2}}}{2}\right)G_{\alpha+\frac{1}{2}}-\frac{1}{2}[\boldsymbol{u}]_{|\Gamma_{\alpha+\frac{1}{2}}}G_{\alpha+\frac{1}{2}}-\left(\frac{\boldsymbol{u}^+_{\alpha-\frac{1}{2}}+\boldsymbol{u}^-_{\alpha-\frac{1}{2}}}{2}\right)G_{\alpha-\frac{1}{2}}-\frac{1}{2}[\boldsymbol{u}]_{|\Gamma_{\alpha-\frac{1}{2}}}G_{\alpha-\frac{1}{2}}.\tag{1.1.29}$$

Likewise, the stress terms at the interface can be written using (1.1.19) as,

$$-\widetilde{\boldsymbol{\sigma}}_{\alpha+\frac{1}{2}}\cdot\boldsymbol{n}_{\alpha+\frac{1}{2}}\sqrt{1+|\nabla_x z_{\alpha+\frac{1}{2}}|^2}+\frac{1}{2}[\boldsymbol{u}]_{|\Gamma_{\alpha+\frac{1}{2}}}G_{\alpha+\frac{1}{2}}+\widetilde{\boldsymbol{\sigma}}_{\alpha-\frac{1}{2}}\cdot\boldsymbol{n}_{\alpha-\frac{1}{2}}\sqrt{1+|\nabla_x z_{\alpha-\frac{1}{2}}|^2}+\frac{1}{2}[\boldsymbol{u}]_{|\Gamma_{\alpha-\frac{1}{2}}}G_{\alpha-\frac{1}{2}}.\tag{1.1.30}$$

Finally, by taking into account (1.1.29) and (1.1.30) we arrive to the following expression for the momentum equation,

$$\begin{aligned}
&\partial_t(h_\alpha\rho_\alpha\boldsymbol{u}_\alpha)+\nabla_x(h_\alpha\rho_\alpha\boldsymbol{u}_\alpha\otimes\boldsymbol{u}_\alpha)+gh_\alpha\rho_\alpha\nabla_x\eta-\nabla_x(h_\alpha\,\boldsymbol{\sigma}_H)\\
&+gh_\alpha\left(\sum_{\beta=\alpha+1}^{M}(\rho_\beta-\rho_\alpha)\nabla_x h_\beta\right)+gh_\alpha\sum_{\beta=\alpha+1}^{M}h_\beta\nabla_x\rho_\beta+\frac{1}{2}gh_\alpha^2\nabla_x\rho_\alpha\\
&=\left(\frac{\boldsymbol{u}^+_{\alpha+\frac{1}{2}}+\boldsymbol{u}^-_{\alpha+\frac{1}{2}}}{2}\right)G_{\alpha+\frac{1}{2}}-\left(\frac{\boldsymbol{u}^+_{\alpha-\frac{1}{2}}+\boldsymbol{u}^-_{\alpha-\frac{1}{2}}}{2}\right)G_{\alpha-\frac{1}{2}}\\
&-\widetilde{\boldsymbol{\sigma}}_{\alpha+\frac{1}{2}}\cdot\boldsymbol{n}_{\alpha+\frac{1}{2}}\sqrt{1+|\nabla_x z_{\alpha+\frac{1}{2}}|^2}+\widetilde{\boldsymbol{\sigma}}_{\alpha-\frac{1}{2}}\cdot\boldsymbol{n}_{\alpha-\frac{1}{2}}\sqrt{1+|\nabla_x z_{\alpha-\frac{1}{2}}|^2}.
\end{aligned}\tag{1.1.31}$$

This equation can be written in the more compact form,

$$\begin{aligned}
&\partial_t(h_\alpha\rho_\alpha\boldsymbol{u}_\alpha)+\nabla_x(h_\alpha\rho_\alpha\boldsymbol{u}_\alpha\otimes\boldsymbol{u}_\alpha)+gh_\alpha\rho_\alpha\nabla_x\eta-\nabla_x(h_\alpha\,\boldsymbol{\sigma}_H)\\
&+gh_\alpha\left(\sum_{\beta=\alpha+1}^{M}(\rho_\beta-\rho_\alpha)\nabla_x h_\beta\right)+gh_\alpha\sum_{\beta=\alpha+1}^{M}h_\beta\nabla_x\rho_\beta+\frac{1}{2}gh_\alpha^2\nabla_x\rho_\alpha\\
&=\left(\frac{\boldsymbol{u}^+_{\alpha+\frac{1}{2}}+\boldsymbol{u}^-_{\alpha+\frac{1}{2}}}{2}\right)G_{\alpha+\frac{1}{2}}-\left(\frac{\boldsymbol{u}^+_{\alpha-\frac{1}{2}}+\boldsymbol{u}^-_{\alpha-\frac{1}{2}}}{2}\right)G_{\alpha-\frac{1}{2}}-\left(K_{\alpha+\frac{1}{2}}-K_{\alpha-\frac{1}{2}}\right),
\end{aligned}\tag{1.1.32}$$

UNIVERSIDAD DE MÁLAGA

with

$$K_{\alpha\pm\frac{1}{2}} = \widetilde{\boldsymbol{\sigma}}_{\alpha\pm\frac{1}{2}} \cdot \boldsymbol{n}_{\alpha\pm\frac{1}{2}} \sqrt{1 + |\nabla_x z_{\alpha\pm\frac{1}{2}}|^2}. \tag{1.1.33}$$

1.1.3.3 Convection/Diffusion equation

By following the same steps as in the previous equations we arrive to the following expression for the tracer equation,

$$\begin{aligned} &\partial_t(h_\alpha \rho_\alpha T_\alpha) + \nabla_x(h_\alpha \rho_\alpha T_\alpha \boldsymbol{u}_\alpha) - \nabla_x\left(\nu_T h_\alpha \rho_\alpha \nabla_x T_\alpha\right)) = T^-_{\alpha+\frac{1}{2}} G_{\alpha+\frac{1}{2}} - T^+_{\alpha-\frac{1}{2}} G_{\alpha-\frac{1}{2}} \\ &- \nu_T (\rho \nabla T)^-_{\alpha+\frac{1}{2}} \cdot \boldsymbol{n}_{\alpha+\frac{1}{2}} \sqrt{1 + |\nabla_x z_{\alpha+\frac{1}{2}}|^2} + \nu_T (\rho \nabla T)^+_{\alpha-\frac{1}{2}} \cdot \boldsymbol{n}_{\alpha-\frac{1}{2}} \sqrt{1 + |\nabla_x z_{\alpha-\frac{1}{2}}|^2}, \end{aligned} \tag{1.1.34}$$

for $\alpha = 1, \ldots, M$. Using (1.1.20) and (1.1.21) the tracer equation can be rewritten as,

$$\begin{aligned} &\partial_t(h_\alpha \rho_\alpha T_\alpha) + \nabla_x(h_\alpha \rho_\alpha T_\alpha \boldsymbol{u}_\alpha) - \nabla_x\left(\nu_T h_\alpha \rho_\alpha \nabla_x T_\alpha\right)) = \left(\frac{T^+_{\alpha+\frac{1}{2}} + T^-_{\alpha+\frac{1}{2}}}{2}\right) G_{\alpha+\frac{1}{2}} - \left(\frac{T^+_{\alpha-\frac{1}{2}} + T^-_{\alpha-\frac{1}{2}}}{2}\right) G_{\alpha-\frac{1}{2}} \\ &- \nu_T \widetilde{\nabla T}_{\alpha+\frac{1}{2}} \cdot \boldsymbol{n}_{\alpha+\frac{1}{2}} \sqrt{1 + |\nabla_x z_{\alpha+\frac{1}{2}}|^2} + \nu_T \widetilde{\nabla T}_{\alpha-\frac{1}{2}} \boldsymbol{n}_{\alpha-\frac{1}{2}} \sqrt{1 + |\nabla_x z_{\alpha-\frac{1}{2}}|^2}. \end{aligned} \tag{1.1.35}$$

Again, a more compact form may be considered,

$$\begin{aligned} &\partial_t(h_\alpha \rho_\alpha T_\alpha) + \nabla_x(h_\alpha \rho_\alpha T_\alpha \boldsymbol{u}_\alpha) - \nabla_x\left(\nu_T h_\alpha \rho_\alpha \nabla_x T_\alpha\right)) \\ &= \left(\frac{T^+_{\alpha+\frac{1}{2}} + T^-_{\alpha+\frac{1}{2}}}{2}\right) G_{\alpha+\frac{1}{2}} - \left(\frac{T^+_{\alpha-\frac{1}{2}} + T^-_{\alpha-\frac{1}{2}}}{2}\right) G_{\alpha-\frac{1}{2}} - \nu_T \left(K_{T,\alpha+\frac{1}{2}} - K_{T,\alpha-\frac{1}{2}}\right), \end{aligned} \tag{1.1.36}$$

with

$$K_{T,\alpha\pm\frac{1}{2}} = \widetilde{\nabla T}_{T,\alpha\pm\frac{1}{2}} \cdot \boldsymbol{n}_{T,\alpha\pm\frac{1}{2}} \sqrt{1 + |\nabla_x z_{T,\alpha\pm\frac{1}{2}}|^2}. \tag{1.1.37}$$

UNIVERSIDAD DE MÁLAGA

1.1.3.4 Final system of equations

The final system of equations for $\alpha = 1, \dots, M$ is:

$$\begin{cases} \partial_t(h_\alpha \rho_\alpha) + \nabla_x(h_\alpha \rho_\alpha \boldsymbol{u}_\alpha) = G_{\alpha+\frac{1}{2}} - G_{\alpha-\frac{1}{2}}, \\ \\ \partial_t(h_\alpha \rho_\alpha T_\alpha) + \nabla_x(h_\alpha \rho_\alpha T_\alpha \boldsymbol{u}_\alpha) - \nabla_x\left(\nu_T h_\alpha \rho_\alpha \nabla_x T_\alpha\right) + \nu_T \left(K_{T,\alpha+\frac{1}{2}} - K_{T,\alpha-\frac{1}{2}}\right) \\ = \left(\dfrac{T_{\alpha+1} + T_\alpha}{2}\right) G_{\alpha+\frac{1}{2}} - \left(\dfrac{T_\alpha + T_{\alpha-1}}{2}\right) G_{\alpha-\frac{1}{2}}, \\ \\ \partial_t(h_\alpha \rho_\alpha \boldsymbol{u}_\alpha) + \nabla_x(h_\alpha \rho_\alpha \boldsymbol{u}_\alpha \otimes \boldsymbol{u}_\alpha) + g h_\alpha \rho_\alpha \nabla_x \eta - \nabla_x(h_\alpha \boldsymbol{\sigma}_H) + K_{\alpha+\frac{1}{2}} - K_{\alpha-\frac{1}{2}} \\ + g h_\alpha \left(\displaystyle\sum_{\beta=\alpha+1}^{M} (\rho_\beta - \rho_\alpha) \nabla_x h_\beta \right) + g h_\alpha \displaystyle\sum_{\beta=\alpha+1}^{M} h_\beta \nabla_x \rho_\beta + \frac{1}{2} g h_\alpha^2 \nabla_x \rho_\alpha \\ = \left(\dfrac{\boldsymbol{u}_{\alpha+1} + \boldsymbol{u}_\alpha}{2}\right) G_{\alpha+\frac{1}{2}} - \left(\dfrac{\boldsymbol{u}_\alpha + \boldsymbol{u}_{\alpha-1}}{2}\right) G_{\alpha-\frac{1}{2}}. \end{cases} \tag{1.1.38}$$

with

$$\rho_\alpha = R(T_\alpha), \tag{1.1.39}$$

where $R(T)$ is a function that depends on the tracer T that correspond with the state equation. Note that in the system (1.1.38), the fact that the tracer and the velocity are constant within a layer has been used.

1.1.4 Closure of the model

In order to give an explicit expression for the transfer terms, the diffusivity term at the tracer equation is neglected. This is in fact equivalent to assume incompressibility of the fluid. Indeed, the mass conservation equation and the tracer evolution equation are,

$$\partial_t \rho + \nabla \cdot (\rho \boldsymbol{v}) = 0, \tag{1.1.40}$$

$$\partial_t(\rho T) + \nabla(\rho T \boldsymbol{v}) = 0. \tag{1.1.41}$$

Combining these two equations we get,

$$\partial_t T + \boldsymbol{v} \cdot \nabla(T) = 0. \tag{1.1.42}$$

Now, considering the state equation (1.1.3) and multiplying the prior expression by $\partial R/\partial T$ we obtain,

$$\partial_t \rho + \boldsymbol{v} \cdot \nabla(\rho) = 0. \tag{1.1.43}$$

By combining (1.1.40) and (1.1.43) the incompressibility equation can be deduced,

$$\nabla \cdot \boldsymbol{v} = 0. \tag{1.1.44}$$

UNIVERSIDAD DE MÁLAGA

The tracer equation in the final multilayer shallow-water system (1.1.38) can be rewritten as,

$$\partial_t(h_\alpha\rho_\alpha T_\alpha) + \nabla_x(h_\alpha\rho_\alpha T_\alpha \boldsymbol{u}_\alpha) = T_\alpha\left(G_{\alpha+\frac{1}{2}} - G_{\alpha-\frac{1}{2}}\right) + \nu_\alpha, \tag{1.1.45}$$

with

$$\nu_\alpha = \frac{1}{2}G_{\alpha+\frac{1}{2}}(T_{\alpha+1} - T_\alpha) + \frac{1}{2}G_{\alpha-\frac{1}{2}}(T_\alpha - T_{\alpha-1}) + \nu_T\left(K_{T,\alpha+\frac{1}{2}} - K_{T,\alpha-\frac{1}{2}}\right). \tag{1.1.46}$$

By combining equation (1.1.45) with the mass equation in (1.1.38) we get,

$$\rho_\alpha\left(\partial_t h_\alpha + \nabla_x(h_\alpha \boldsymbol{u})\right) + \frac{\nu_\alpha}{\rho_\alpha}\frac{\partial R}{\partial T}(T_\alpha) = G_{\alpha+\frac{1}{2}} - G_{\alpha-\frac{1}{2}}. \tag{1.1.47}$$

Additional simplifications can be performed by considering the layer thickness proportional to the total height of the water, $h_\alpha = l_\alpha h$ for $\alpha = 1, \ldots, M$, and with l_α positive constants such that

$$\sum_{\alpha=1}^{M} l_\alpha = 1. \tag{1.1.48}$$

Then, summing the equation (1.1.47) from $\beta = 1, \ldots, \alpha$, results,

$$G_{\alpha+\frac{1}{2}} - G_{\frac{1}{2}} = \left(\sum_{\beta=1}^{\alpha} l_\beta\rho_\beta\right)\partial_t h + \sum_{\beta=1}^{\alpha} l_\beta\rho_\beta\partial_x(hu_\beta) + \sum_{\beta=1}^{\alpha}\frac{\nu_\beta}{\rho_\beta}\frac{\partial R}{\partial T}(T_\beta). \tag{1.1.49}$$

The particular case of $\alpha = M$ yields,

$$G_{M+\frac{1}{2}} - G_{\frac{1}{2}} = \left(\sum_{\gamma=1}^{M} l_\gamma\rho_\gamma\right)\partial_t h + \sum_{\gamma=1}^{M} l_\gamma\rho_\gamma\partial_x(hu_\gamma) + \sum_{\gamma=1}^{M}\frac{\nu_\gamma}{\rho_\gamma}\frac{\partial R}{\partial T}(T_\gamma). \tag{1.1.50}$$

Combining equations (1.1.49) and (1.1.50) an explicit expression for the transfer terms can be obtained,

$$G_{\alpha+\frac{1}{2}} = G_{\frac{1}{2}} + L_\alpha\left(G_{M+\frac{1}{2}} - G_{\frac{1}{2}} - \sum_{\gamma=1}^{M} l_\gamma\rho_\gamma\partial_x(hu_\gamma)\right) + \sum_{\beta=1}^{\alpha} l_\beta\rho_\beta\partial_x(hu_\beta) + \Psi_\alpha, \tag{1.1.51}$$

where

$$L_\alpha \coloneqq \frac{\sum_{\beta=1}^{\alpha}\rho_\beta l_\beta}{\sum_{\gamma=1}^{M}\rho_\gamma l_\gamma} \tag{1.1.52}$$

and

$$\Psi_\alpha \coloneqq \sum_{\beta=1}^{\alpha}\frac{\nu_\beta}{\rho_\beta}\frac{\partial R}{\partial T}(T_\beta) - L_\alpha\sum_{\gamma=1}^{M}\frac{\nu_\gamma}{\rho_\gamma}\frac{\partial R}{\partial T}(T_\gamma). \tag{1.1.53}$$

UNIVERSIDAD DE MÁLAGA

1.2 A particular equation of state

The system of equations (1.1.38) can be further simplified by considering some preliminary hypotheses in the Navier-Stokes system of equations (1.1.1)-(1.1.2). If the incompressibility hypothesis is assumed and we consider that $R(T) = \rho_0 T$, then

$$\begin{aligned} \nabla \cdot \boldsymbol{v} &= 0, \\ \partial_t(\rho_0 T) + \nabla \cdot (\rho_0 T\, \boldsymbol{v}) &= 0, \\ \partial_t(\rho_0 T\, \boldsymbol{v}) + \nabla \cdot (\rho_0 T \boldsymbol{v} \otimes \boldsymbol{v}) &= -g\, \rho_0 T \boldsymbol{k} + \nabla \cdot \boldsymbol{\Sigma}. \end{aligned} \tag{1.2.1}$$

The total density ρ is defined as the sum of a reference density ρ_0 and the fluctuation from that reference ρ_1,

$$\rho = \rho_0 + \rho_1. \tag{1.2.2}$$

The Navier–Stokes equations (1.2.1) can be expressed in terms of relative quantities by dividing by ρ_0. Additionally, the tracer function T is renamed as θ to identify it with the transport of relative density as follows,

$$\begin{aligned} \nabla \cdot \boldsymbol{v} &= 0, \\ \partial_t(\theta) + \nabla \cdot (\theta\, \boldsymbol{v}) &= 0, \\ \partial_t(\theta\, \boldsymbol{v}) + \nabla \cdot (\theta \boldsymbol{v} \otimes \boldsymbol{v}) &= -g\, \theta \boldsymbol{k} + \frac{1}{\rho_0} \nabla \cdot \boldsymbol{\Sigma}, \end{aligned} \tag{1.2.3}$$

with

$$\theta = \frac{\rho}{\rho_0} = 1 + \frac{\rho_1}{\rho_0}. \tag{1.2.4}$$

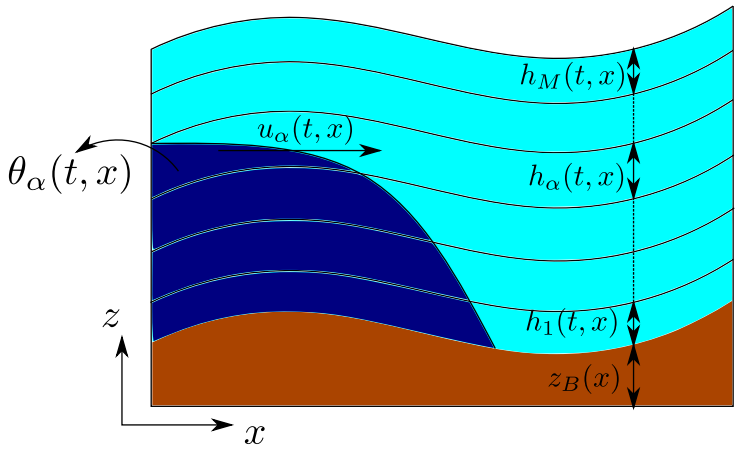


Figure 1.2: Sketch of the multilayer approach in one dimension with relative density.

UNIVERSIDAD DE MÁLAGA

The same multilayer approach can be performed in the alternative system (1.2.3). The hydrostatic pressure depends now on the relative density:

$$p_\alpha(t,x,z) = p_{\alpha+\frac{1}{2}} + \theta_\alpha\, g\,(z_{\alpha+\frac{1}{2}} - z), \tag{1.2.5}$$

with

$$p_{\alpha+\frac{1}{2}}(t,x) = p_S(t,x) + g\sum_{\beta=\alpha+1}^{M} \theta_\beta h_\beta(t,x). \tag{1.2.6}$$

Under these hypotheses, the weak formulation (1.1.24) yields the following system of equations. For simplicity, we write them in one dimensional form,

$$\begin{cases} \partial_t h_\alpha + \partial_x(h_\alpha u_\alpha) = G_{\alpha+\frac{1}{2}} - G_{\alpha-\frac{1}{2}}, \\ \partial_t(h_\alpha\theta_\alpha) + \partial_x(h_\alpha\theta_\alpha u_\alpha) = \left(\dfrac{\theta_{\alpha+1}+\theta_\alpha}{2}\right)G_{\alpha+\frac{1}{2}} - \left(\dfrac{\theta_\alpha+\theta_{\alpha-1}}{2}\right)G_{\alpha-\frac{1}{2}}, \\ \partial_t(h_\alpha\theta_\alpha u_\alpha) + \partial_x(h_\alpha\theta_\alpha u_\alpha^2) + g h_\alpha\theta_\alpha\partial_x\eta - \partial_x(h_\alpha\,\sigma_H) + \nu\left(K_{\alpha+\frac{1}{2}} - K_{\alpha-\frac{1}{2}}\right) \\ + g h_\alpha\left(\displaystyle\sum_{\beta=\alpha+1}^{M}(\theta_\beta-\theta_\alpha)\partial_x h_\beta\right) + g h_\alpha\sum_{\beta=\alpha+1}^{M} h_\beta\partial_x\theta_\beta + \frac{1}{2}g h_\alpha^2\partial_x\theta_\alpha \\ = \left(\dfrac{u_{\alpha+1}+u_\alpha}{2}\right)\left(\dfrac{\theta_{\alpha+1}+\theta_\alpha}{2}\right)G_{\alpha+\frac{1}{2}} - \left(\dfrac{u_\alpha+u_{\alpha-1}}{2}\right)\left(\dfrac{\theta_\alpha+\theta_{\alpha-1}}{2}\right)G_{\alpha-\frac{1}{2}}. \end{cases}$$

Under the same hypothesis as in Subsection 1.1.4, both the model and the mass transfer terms can be further simplified. The final model, neglecting the horizontal viscosity terms and rewriting the pressure terms, is:

$$\begin{cases} \partial_t h + \partial_x\left(h\displaystyle\sum_{\beta=1}^{M} l_\beta u_\beta\right) = 0, \\ \partial_t(h\theta_\alpha) + \partial_x(h\theta_\alpha u_\alpha) = \dfrac{1}{l_\alpha}\left(\theta_{\alpha+\frac{1}{2}}G_{\alpha+\frac{1}{2}} - \theta_{\alpha-\frac{1}{2}}G_{\alpha-\frac{1}{2}}\right), \\ \partial_t(h\theta_\alpha u_\alpha) + \partial_x(h\theta_\alpha u_\alpha^2) + g h\theta_\alpha\partial_x\eta + \dfrac{g l_\alpha}{2}\left(h\partial_x(h\theta_\alpha) - h\theta_\alpha\partial_x h\right) \\ + g\displaystyle\sum_{\beta=\alpha+1}^{M} l_\beta\left(h\partial_x(h\theta_\beta) - h\theta_\alpha\partial_x h\right) = \dfrac{1}{l_\alpha}\left(u_{\alpha+\frac{1}{2}}\theta_{\alpha+\frac{1}{2}}G_{\alpha+\frac{1}{2}} - u_{\alpha-\frac{1}{2}}\theta_{\alpha-\frac{1}{2}}G_{\alpha-\frac{1}{2}}\right), \end{cases} \tag{1.2.7}$$

where $\theta_{\alpha+\frac{1}{2}}$ and $u_{\alpha+\frac{1}{2}}$ are the arithmetic mean at the interfaces $\Gamma_{\alpha+\frac{1}{2}}(t)$, $\alpha = 1,\dots,M-1$:

$$u_{\alpha+\frac{1}{2}} = \frac{u_{\alpha+1}+u_\alpha}{2}, \qquad \theta_{\alpha+\frac{1}{2}} = \frac{\theta_{\alpha+1}+\theta_\alpha}{2}, \qquad \alpha = 1,\dots,M-1,$$

UNIVERSIDAD DE MÁLAGA

with $u_{\frac{1}{2}} = u_1$, $u_{M+\frac{1}{2}} = u_M$ and $\theta_{\frac{1}{2}} = \theta_1$, $\theta_{M+\frac{1}{2}} = \theta_M$. Additionally, no transfer terms through the bottom and free surface is assumed, that is, $G_{\frac{1}{2}} = G_{M+\frac{1}{2}} = 0$. Note that the incompressibility hypotheses considered in this Section allows to simplify the density in the transfer terms (1.1.11). Hence, we have written the mean of the relative density in the momentum equation in system (1.2.7).

The simplified explicit expression for the transfer terms (1.1.51) with constant density are,

$$G_{\alpha+\frac{1}{2}} = \sum_{\beta=1}^{\alpha} l_\beta \left(\partial_t h + \partial_x(hu_\beta)\right) = \sum_{\beta=1}^{\alpha} l_\beta \left(\partial_x(hu_\beta) - \partial_x\left(h\sum_{\gamma=1}^{M} l_\gamma u_\gamma\right)\right). \tag{1.2.8}$$

In this way, the mass exchange terms between layers can be defined exclusively in terms of the horizontal velocities in the layers.

Remark 1.2.1. *Note that by removing the stress terms in* (1.2.7) *the simplification* (1.1.29)-(1.1.30) *is no longer valid. Then, the jump terms at* (1.1.29) *are not simplified and should be considered in system* (1.2.7)*. However, since the literature agree on not reflecting these jump terms, we have chosen to not write them in the model. Nonetheless, these terms will play an important role in the numerical discretization, where they have to be properly taken into account.*

While the full spectral information of the model (1.2.7) is unknown, empirically we observe that the model remains hyperbolic for all the numerical simulations performed. A further study on the hyperbolicity of multilayer models with variable density is found in [37]. In this reference, a linearization of the equations is performed and the eigenstructure of the corresponding system is studied. The authors conclude that the model is hyperbolic as long as the velocity profile remains moderate, within normal values under the hydrostatic regime. Additionally, in [38] the authors study the hyperbolicity of a two layer version of a shallow-water model with variable density and they conclude that such model is strictly hyperbolic. Nevertheless, note that in this thesis a HLL type Riemann solver has been used, so only a bound for the maximum and minimum eigenvalues is needed. Therefore, even in presence of an eventual loss of hyperbolicity, this will not result in a catastrophic failure of the numerical scheme.

Furthermore, some previous work at [35] allows to give a bound to the maximum and minimum wave speeds using the results at [147]. In that work, it was proved that, if all the roots of the characteristic polynomial are real, then they are bounded by,

$$\frac{a_{n-1}}{na_n} \pm (n-1)\left(\frac{a_{n-1}^2}{n^2 a_n^2} - \frac{2a_{n-2}}{n(n-1)a_n}\right)^{\frac{1}{2}},$$

where a_i are the coefficient of the characteristic polynomial, being a_n the one corresponding to the leading term.

UNIVERSIDAD DE MÁLAGA

Using this result, we can find a bound for the maximum and minimum wave propagation speed $\lambda_{\max}$, $\lambda_{\min}$,

$$\lambda_{\min} \geq \bar{u} - \Psi, \quad \lambda_{\max} \leq \bar{u} + \Psi, \tag{1.2.9}$$

where

$$\Psi = \sqrt{\frac{2M-1}{2M}\left(2\sum_{\alpha=1}^{M}(\bar{u}-u_\alpha)^2 + gh\left(1+\frac{1}{M}\sum_{\beta=1}^{M}(2\beta-1)\theta_\beta\right)\right)}, \tag{1.2.10}$$

and

$$\bar{u} = \frac{1}{M}\sum_{\alpha=1}^{M} u_\alpha. \tag{1.2.11}$$

1.3 Stationary solutions

As was stated previously, stationary solutions of the system (1.2.7) are of particular interest to preserve. Let us discuss the stationary solutions associated with the full PDE system (1.2.7). In particular, we are interested in stationary solutions with $u_\alpha = 0$ for $\alpha = 1, \ldots, M$. Stationary solutions with zero velocity should satisfy,

$$\begin{aligned} P_\alpha &:= g\frac{l_\alpha}{2}\partial_x(h^2\theta_\alpha) + h\partial_x p_{\alpha+\frac{1}{2}} + gh\theta_\alpha\partial_x z_{\alpha-\frac{1}{2}} \\ &= gh\theta_\alpha\partial_x\eta + \frac{gl_\alpha}{2}\left(h\partial_x(h\theta_\alpha) - h\theta_\alpha\partial_x h\right) + g\sum_{\beta=\alpha+1}^{M} l_\beta\left(h\partial_x(h\theta_\beta) - h\theta_\alpha\partial_x h\right). \end{aligned} \tag{1.3.1}$$

Considering that stationary solutions do not depend on time, we get from previous expression the following ODE:

$$gh\theta_\alpha(h+b)' + \frac{gl_\alpha}{2}\left(h^2(\theta_\alpha)' + h\theta_\alpha h' - h\theta_\alpha h'\right) + g\sum_{\beta=\alpha+1}^{M} l_\beta\left(h^2(\theta_\beta)' + h\theta_\beta h' - h\theta_\alpha h'\right) = 0, \tag{1.3.2}$$

where the prime mark denotes the derivative with respect to the spatial variable and $\eta = h + b$ corresponds to the free surface. By further developing the prior expression, we obtain,

$$\theta_\alpha(h+b)' + \frac{l_\alpha}{2}h(\theta_\alpha)' + \sum_{\beta=\alpha+1}^{M} l_\beta h(\theta_\beta)' + l_\beta(\theta_\beta - \theta_\alpha)h' = 0. \tag{1.3.3}$$

UNIVERSIDAD DE MÁLAGA

In other words, this pressure terms should be so that the following system of ODEs are satisfied:

$$\frac{l_\alpha}{2}(\theta_\alpha)' + \sum_{\beta=\alpha+1}^{M} l_\beta(\theta_\beta)' = -\frac{1}{h}\left(\theta_\alpha(h+b)' + \sum_{\beta=\alpha+1}^{M} l_\beta(\theta_\beta - \theta_\alpha)h'\right), \quad \alpha = 1, \dots, M. \tag{1.3.4}$$

System (1.3.4) has trivial stationary solution with a constant relative density function $\theta_\alpha = K_1$, $\alpha = 1, \dots, M$, in which case the stationary solutions are the well known lake-at-rest stationary solution with constant free surface,

$$u_\alpha = 0,\ \theta_\alpha = K_1, \eta = h + b = K_2,$$

where $K_1 \geq 1$ and K_2 are two given constant. However, there are also non-trivial stationary solutions that are a solution of the ODE system (1.3.4) with a non-constant relative density profile. Of course, there are an infinite number of solutions for any given function $b(x)$, but we are interested exclusively in stationary solutions that are also stable. Therefore, we consider stationary solutions that correspond to a constant free surface and a vertical density profile given by a relative density stratification:

$$u = 0, \quad \eta(x) = b(x) + h(x) = K, \quad \theta(z) = \theta_{surface} + \alpha(\eta - z). \tag{1.3.5}$$

Unfortunately, the stratification profile (1.3.5) is not a solution of (1.3.4), except for the particular case of $b(x)$ is constant, in which case the density profile is trivial. Indeed, this is a direct consequence of the vertical discretization in layers of the multilayer approach, where the variables are vertically averaged and the layers are not allowed to vanish. This results in discrete stratified solutions for the multilayer system that is not space independent for the relative density θ, unless the free surface and bathymetry functions are constant.

Nevertheless, the ODE system (1.3.4) may be solved recursively if we assume a vertical stratification profile like (1.3.5). We begin by solving the upper layer $\alpha = M$ for θ_M,

$$(\theta_M)' = \frac{2}{h l_M}(h+b)'\theta_M,$$

and subsequentially going downwards, $\alpha = M-1, \dots, 1$, computing θ_α. Following this procedure, we can derive the set of stationary solutions:

$$\begin{gathered} u_\alpha = 0,\ \eta(x) = z_b(x) + h(x) = K, \\ \theta_M(x) = \bar{\theta}_M \geq 1, \\ \theta_\alpha(x) = \bar{\theta}_\alpha\, h^{2(M-\alpha)}(x) + \sum_{\beta=\alpha+1}^{M} S_{2(M-\beta)}(M-\alpha+1)\bar{\theta}_\beta\, h^{2(M-\beta)}(x), \end{gathered} \tag{1.3.6}$$

UNIVERSIDAD DE MÁLAGA

with

$$\begin{aligned} S_\beta(\alpha) &= (\beta+1)\cdot A_{\frac{\beta+2}{2}+1}(\alpha), \\ A_p(k) &= \begin{cases} 1 & \text{if } p \geq k, \\ (p-1)\prod_{\gamma=2}^{k-p}(1+(p-2)C_{\gamma-1}) & \text{if } p < k, \end{cases} \\ C_\gamma &= C_{\gamma-1} - \frac{1}{Q_\gamma}, \\ Q_\gamma &= Q_{\gamma-1} + \gamma + 1, \\ C_0 &= Q_0 = 1. \end{aligned}$$

This profile, clearly non-linear, can be seen in figures 1.3 and 1.4 for a non-constant smooth bathymetry function. In particular, the relative density distribution $\theta_\alpha(x)$, $\alpha = 1, \ldots, M$ is depicted through a heat map in a channel with $M = 5$, $M = 10$, $M = 20$, and $M = 25$. Observe that stationary solutions (1.3.6) could be seen as approximations of (1.3.5).

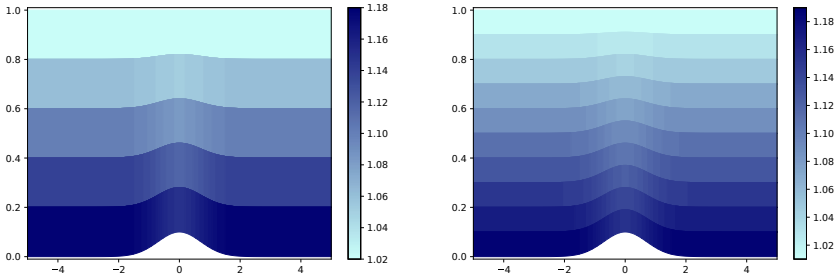


Figure 1.3: Solution of the ODE (1.3.4) for a stratified fluid with $M = 5$ (left) and $M = 10$ (right).

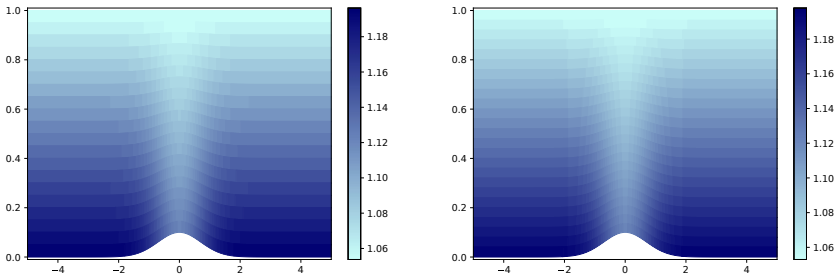


Figure 1.4: Solution of the ODE (1.3.4) for a stratified fluid with $M = 20$ (left) and $M = 25$ (right).

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD
DE MÁLAGA

Chapter 2

Well-Balanced finite volume and discontinuous Galerkin methods

2.1 Introduction

In this chapter, a general methodology for designing high order well-balanced finite volume and discontinuous Galerkin numerical schemes is presented. These methods are suited for hyperbolic systems with non-conservative products and/or source terms that appear in many different context: shallow-water systems, gas dynamic, multiphase flows or dispersive shallow-water models, for instance. These systems often involve the computation of non-conservative terms that increase considerably the difficulty of a thoroughly theoretical description. Indeed, these terms may not make sense in the distributional framework when the solution of the system develops discontinuities.

To address this issue, path-conservative methods were introduced by Parés in [40]. This theory is actually supported by the notion of weak solution for a non-conservative system given by Dal Maso, LeFloch and Murat in [39], where the definition of weak solutions is based on a choice of a family of paths connecting two arbitrary states. In this way, a strategy to develop numerical schemes for many well-known systems of conservation laws with non-conservative terms is provided. Furthermore, its extension to high order can be done as well. Since their introduction, path-conservative methods has been widely used in the finite volume framework (see [148–150]) and the DG and ADER DG framework (see [151–153]) among others. In fact, the design of numerical methods for systems of conservation laws with or without non-conservative products constitutes an active research topic for the scientific community. See for instance [10, 38, 42, 102, 104, 105, 108, 109, 113–116, 121–123, 125, 126, 133, 149, 151, 154–160] among others.

As it has been previously discussed, the well-balanced property of numerical schemes is of paramount importance in many practical applications. In this chapter, it will be shown that, in fact, this property is closely related to the definition of the Riemann solver.

Additionally, a class of efficient first order path-conservative schemes, the so-called

UNIVERSIDAD DE MÁLAGA

Polynomial Viscosity Methods, is described. Furthermore, the extension to high order using reconstruction operators in the framework of finite volume is also contemplated.

Discontinuous Galerkin methods are also addressed in this chapter. While the tools associated with the Riemann problem are also useful for the discontinuous Galerkin methods, new strategies are required when dealing with numerical schemes that are fundamentally different from the finite volume methods, especially in its ability to achieve high order. Discontinuous Galerkin methods possess some unique advantages and disadvantages with respect to finite volume methods. They allow to reach very high order easily, resulting in an enhanced cell resolution, which make them extremely efficient. However, there are some important caveats that must also be considered. As it is discussed in this chapter, DG methods suffer from a very restrictive CFL condition. Moreover, there is no obvious way to easily limit them in the presence of strong gradient traversing the system. In this chapter, two different limiting approaches are considered. The overall efficiency of DG methods depends then on the correct balance between the total number of cell, the order of the method (and associated CFL restriction) and the total amount of cell that need limiting, as well as the particular limiting technique chosen. Additionally, in this chapter some novel contribution regarding DG methods are described that consist on a new approach for the limiting problem and a general strategy to achieve well-balanced schemes.

2.2 Finite volume path-conservative numerical schemes

We consider a general one-dimensional non-conservative hyperbolic system of the form

$$\partial_t w + \partial_x F(w) + B(w)\partial_x w = G(w)\sigma_x, \quad x \in I,\ t > 0, \tag{2.2.1}$$

with initial condition

$$w(x,0) = w_0(x), \quad x \in I. \tag{2.2.2}$$

Here, I is a open set of $\mathbb{R}$, such that $I \subset \mathbb{R}$, while the unknown $w(x,t) = (w_1(x,t), \ldots, w_N(x,t))^T$ takes values in Ω, being an open convex set of $\mathbb{R}^N$ called set of states. F is a regular function from Ω to $\mathbb{R}^N$ called flux-function. B is a regular matrix function from Ω to $\mathcal{M}_N(\mathbb{R})$, S is a function from Ω to $\mathbb{R}^N$, and $\sigma(x)$ is a known bounded function from I to $\mathbb{R}$.

System (2.2.1) could be rewritten as a general quasi-linear hyperbolic system of the form

$$\partial_t W + \mathcal{A}(W)\partial_x W = 0, \quad x \in I \subset \mathbb{R},\ t > 0, \tag{2.2.3}$$

with initial condition

$$W(x,0) = W_0(x), \quad x \in I \subset \mathbb{R}, \tag{2.2.4}$$

UNIVERSIDAD DE MÁLAGA

setting $W = (w, \sigma)^T \in O = \Omega \times \mathbb{R} \subset \mathbb{R}^M$, $M = N + 1$, and the matrix $\mathcal{A}(W)$ defined with the following block structure,

$$\mathcal{A}(W) = \left(\begin{array}{c|c} A(W) & -G(W) \\ \hline 0 & 0 \end{array} \right), \tag{2.2.5}$$

where $A(W) = J_F(W) + B(W)$, $J_F(W)$ is the Jacobian matrix of the flux function F,

$$J_F(W) = \frac{\partial F}{\partial W}(W).$$

The system (2.2.3) is assumed to be strictly hyperbolic, i.e, the matrix $\mathcal{A}(W)$ has M real and distinct eigenvalues

$$\lambda_1(W) < \ldots < \lambda_M(W),$$

and eigenvectors,

$$\mathcal{R}_1(W) < \ldots < \mathcal{R}_M(W),$$

for all $W \in O$. Therefore, the matrix $\mathcal{A}(W)$ is diagonalizable. Additionally, the characteristic field $\mathcal{R}_i$ is assumed to be either non linear,

$$\nabla \lambda_i(W) \cdot \mathcal{R}_i(W) \neq 0, \quad \forall W \in O,$$

or linearly degenerate,

$$\nabla \lambda_i(W) \cdot \mathcal{R}_i(W) = 0, \quad \forall W \in O.$$

In order to define the notion of weak solution for system (2.2.3), the equations are integrated in an arbitrary space-time control domain $[a, b] \times [t_0, t_1]$,

$$\int_a^b W(x, t_1)\, dx = \int_a^b W(x, t_0)\, dx - \int_{t_0}^{t_1} \int_a^b \mathcal{A}(W(x, t)) \partial_x W(x, t)\, dx\, dt. \tag{2.2.6}$$

The difficulty arises from the fact that $W(x, t)$ may be discontinuous and the last integral has no sense in the distributional framework. The theory developed by Dal Maso, LeFloch, and Murat in [39] provides a framework to define the product $\mathcal{A}(W)\partial_x W$ for a function W with bounded variation, provided that a family of Lipschitz continuous paths $\Phi\colon [0, 1] \times O \times O \to O$ is prescribed, which must satisfy certain regularity and compatibility conditions. Particularly,

$$\Phi(0; W_L, W_R) = W_L, \qquad \Phi(1; W_L, W_R) = W_R, \tag{2.2.7}$$

and

$$\Phi(s; W, W) = W. \tag{2.2.8}$$

In this theory, if $W(x)$ is a bounded variation function, then the product $\mathcal{A}(W)\partial_x W$ is defined as a locally bounded measure that coincides with the distributional derivative in

UNIVERSIDAD DE MÁLAGA

the special case for which $\mathcal{A}(W)$ is the Jacobian matrix for some function $F(W)$. More details regarding this theory can be found in [39].

To apply this theory to (2.2.3), the families of paths linking two states W_L and W_R can be interpreted as a way of giving sense to the last integral term at (2.2.6) for piecewise smooth functions W. More explicitly, given a bounded variation function $W : [a, b] \to \mathbb{R}^M$, we define:

$$\int\!\!\!\!\!\!-_a^b \mathcal{A}(W(x))\partial_x W(x)\,dx = \int_a^b \mathcal{A}(W(x))\partial_x W(x)\,dx + \sum_\ell \int_0^1 \mathcal{A}(\Phi(s; W_\ell^-, W_\ell^+))\partial_s \Phi(s; W_\ell^-, W_\ell^+)\,ds, \quad (2.2.9)$$

where W_ℓ^- and W_ℓ^+ stands for the limit of W to the left and right side of the ℓ-th discontinuity respectively. Note that the set of discontinuities of a bounded variation function is countable. Additionally, in (2.2.9) the family of paths has been used to determine the weight of the Dirac measure at the ℓ-discontinuity.

Defining the integral in this way, a weak solution of the PDE system (2.2.3) can be seen as a function satisfying,

$$\int_a^b W(x, t_1)\,dx = \int_a^b W(x, t_0)\,dx - \int_{t_0}^{t_1} \int\!\!\!\!\!\!-_a^b \mathcal{A}(W(x,t))\partial_x W(x,t)\,dx\,dt, \quad (2.2.10)$$

for every space-time rectangle $[a, b] \times [t_0, t_1]$.

As in the fully conservative case, weak solutions still must satisfy the generalized Rankine-Hugoniot condition across a discontinuity,

$$\xi(W^+ - W^-) = \int_0^1 \mathcal{A}(\Phi(s; W^-, W^+))\partial_s \Phi(s; W^-, W^+)\,ds, \quad (2.2.11)$$

where ξ is the speed of propagation of the discontinuity, and W^- and W^+ are the left and right limits of the solution at the discontinuity. Note that, if $\mathcal{A}(W)$ is the Jacobian matrix for some function $F(W)$, then the equation (2.2.11) reduces to the standard Rankine-Hugoniot condition:

$$\xi(W^+ - W^-) = F(W^+) - F(W^-), \quad (2.2.12)$$

independently of the path Φ.

Analogously to the conservative case, the uniqueness of solution requires adding an entropy condition to the definition of weak solution. Let us consider an entropy-entropy flux pair of (2.2.3) $(\mathcal{H}, Q)$, with $\mathcal{H}$ a convex function $\mathcal{H} \colon O \to \mathbb{R}$, called entropy, and Q also function $Q \colon O \to \mathbb{R}$, called entropy flux, such that

$$\nabla Q(W) = \nabla \mathcal{H}(W) \cdot \mathcal{A}(W). \quad (2.2.13)$$

UNIVERSIDAD DE MÁLAGA

Then, a weak solution W is an entropy solution if it satisfies the following inequality,

$$\partial_t \mathcal{H}(W) + \partial_x Q(W) \leq 0, \tag{2.2.14}$$

in the sense of distributions.

As it has been discussed, the notion of weak solution depends strongly on the family of paths chosen. In fact, the proper choosing of the adequate path becomes an important issue since an oversimplified understanding of the theory may lead to the wrong premise that path-conservative methods provide wrong solutions. On the contrary, path-conservative methods converge with the expected order of accuracy under a suitable CFL condition, with the same stability properties as their conservative counterparts, and any convergence failure can be explained by the intrinsic nature of non-conservative systems. The main difficulty comes from the fact that the limits of numerical solutions may differ from the correct ones. Likewise, as stated before, weak solutions of non-conservative systems may be defined in an infinite number of ways, hence the importance of a proper understanding of the meaning of weak solutions for a PDE system of the form (2.2.3).

In order to study this issue, let us consider the equivalent equations of the system (2.2.3), where higher order terms are present, corresponding to vanishing diffusion and/or dispersion limit terms. Let us consider an example with a vanishing viscosity term

$$\partial_t W + \mathcal{A}(W)\partial_x W = \varepsilon R(W)\partial_{xx} W, \tag{2.2.15}$$

where $R(W)$ is a positive definite matrix. The adequate notion of weak solution, and therefore the correct choice of the family of paths, should be consistent with the traveling waves of the regularized system (see [161]), defined as the solution of (2.2.15),

$$W_\epsilon(x,t) = V\left(\frac{x-\xi t}{\epsilon}\right), \tag{2.2.16}$$

that satisfies

$$\lim_{s\to-\infty} V(s) = W^-, \quad \lim_{s\to+\infty} V(s) = W^+, \quad \lim_{s\to\pm\infty} V'(s) = 0.$$

If there exists a traveling wave of speed ξ linking the states W^-, W^+, the limit when ϵ tends to 0 of W_ϵ,

$$W(x,t) = \begin{cases} W^- & \text{if } x < \xi t, \\ W^+ & \text{if } x > \xi t, \end{cases}$$

should be an admissible weak solution of the non-conservative hyperbolic system. By setting V into (2.2.15) we get,

$$-\xi V' + \mathcal{A}(V)\, V' = RV''. \tag{2.2.17}$$

And by integrating this expression from $-\infty$ to ∞ and taking into account the boundary conditions, the following jump condition is obtained,

$$\int_{-\infty}^{\infty} \mathcal{A}(V(s))\, V'(s)\, ds = \xi(W^+ - W^-).$$

UNIVERSIDAD DE MÁLAGA

Comparing this jump condition with the generalized Rankine-Hugoniot condition (2.2.11), it becomes obvious that the good choice for the path connecting the states W^- and W^+ would be, after a reparametrization, the viscous profile.

Note that for systems of conservation laws the jump condition reduces to the standard Rakine-Hugoniot condition, independent of the form of the diffusion term $R(W)$. On the contrary, for non-conservative systems the jump condition depends explicitly on the particular choice of the matrix R. In this specific choice resides the issue in which many numerical methods fail in converging to the correct weak solutions: the limits of the numerical solutions satisfy a jump condition which is related to the numerical viscosity of the method rather that to the physically relevant one. Of course, this problem affects all numerical methods where the small scale effects corresponding to the higher order terms are not taken into account, regardless if the numerical schemes is designed under the path-conservative framework or not.

2.2.1 Path-conservative numerical schemes

In this section, we present a brief introduction of path-conservative schemes to discretize system (2.2.3). As usual, the domain I is discretized into a set of conforming computational cells $I_i = [x_{i-1/2}, x_{i+1/2}]$. For simplicity, a uniform cell size Δx is assumed. We denote by x_i the center of the cell I_i, such that $x_i = (i-1/2)\Delta x$, and $x_{i+\frac{1}{2}}$ the cell interface, $x_{i+1/2} = i\Delta x$. Likewise, Δt denotes the time step size, such that $t^n = n\Delta t$. The cell-averaged piecewise approximation of the solution $W(x,t)$ within a cell I_i at time t^n is denoted as W_i^n,

$$W_i^n = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} W(x,t^n)\,dx. \tag{2.2.18}$$

Now, from the discrete version of (2.2.10),

$$\frac{1}{\Delta x}\int_{I_i} W(x,t^{n+1})\,dx = \frac{1}{\Delta x}\int_{I_i} W(x,t^n)\,dx - \frac{\Delta t}{\Delta x}\frac{1}{\Delta t}\int_{t^n}^{t^{n+1}}\oint_{I_i}\mathcal{A}(W)\partial_x W\,dx\,dt, \tag{2.2.19}$$

the following definition of a path-conservative numerical method is given by Parés in [40],

Definition 1. *Given a family of paths Φ, a numerical scheme is said to be Φ-conservative if it can be written under the form:*

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x}\big(\mathcal{D}^+_{i-1/2} + \mathcal{D}^-_{i+1/2}\big), \tag{2.2.20}$$

where

$$\mathcal{D}^\pm_{i+1/2} = \mathcal{D}^\pm\big(W_i^n, W_{i+1}^n\big),$$

$\mathcal{D}^-$ and $\mathcal{D}^+$ being two continuous functions from $O\times O$ to $\mathbb{R}^N$ satisfying

$$\mathcal{D}^\pm(W,W) = 0 \quad \forall W \in O, \tag{2.2.21}$$

UNIVERSIDAD DE MÁLAGA

and

$$\mathcal{D}^-(W_L, W_R) + \mathcal{D}^+(W_L, W_R) = \int_0^1 \mathcal{A}\big(\Phi(s; W_L, W_R)\big)\partial_s \Phi(s; W_L, W_R)\, ds, \tag{2.2.22}$$

for every pair $(W_L, W_R) \in O \times O$.

This definition of path-conservative methods corresponds to a generalization of classical conservative methods for systems of conservation laws. Indeed, if there are no non-conservative products, then $\mathcal{A}(W)$ is the Jacobian of a flux function $\widetilde{F} = (F, 0)$ and (2.2.22) is reduced to,

$$\mathcal{D}^-(W_L, W_R) + \mathcal{D}^+(W_L, W_R) = \widetilde{F}(W_R) - \widetilde{F}(W_L), \tag{2.2.23}$$

and we can write,

$$\mathcal{F}(W_L, W_R) = \mathcal{D}^-(W_L, W_R) + \widetilde{F}(W_L), \tag{2.2.24}$$

or, equivalently,

$$\mathcal{F}(W_L, W_R) = \widetilde{F}(W_R) - \mathcal{D}^+(W_L, W_R). \tag{2.2.25}$$

Then, (2.2.21) leads to

$$\mathcal{F}(W, W) = \widetilde{F}(W),$$

and therefore $\mathcal{F}$ is a numerical flux consistent with F. Furthermore, combining (2.2.24)-(2.2.25) and (2.2.20), the path-conservative method can be written in the conservative form,

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x}\big(\mathcal{F}_{i+1/2} - \mathcal{F}_{i-1/2}\big), \tag{2.2.26}$$

with

$$\mathcal{F}_{i+1/2} = \mathcal{F}\big(W_i^n, W_{i+1}^n\big).$$

Consequentially, if the PDE system (2.2.3) involves conservation laws, like the mass equation in the shallow-water model, a path-conservative method will be conservative for these equations.

Path-conservative methods have been successfully applied in a wide range of problem, well beyond the shallow-water equations. Some examples of this includes the Saint Venant-Exner [162], turbidity currents [163], Ripa model [148], two-modes shallow-water system [164], Baer-Nunziato model [165], Pitman-Le model [149], Savage-Hutter models [166], Bingham shallow-water system [167], blood flow [150] and two-phase flows [168], among others.

2.2.2 Path-conservative numerical schemes: examples

In this subsection, we present the extension of some well-known Riemann solvers that are commonly used for the discretization of conservation laws to non-conservative systems.

UNIVERSIDAD DE MÁLAGA

2.2.2.1 Roe method

Path-conservative Roe method is based on the following extension of a Roe linearization of non-conservative systems introduced in [169]:

Definition 2. *Given a family of paths Φ, a function $\mathcal{A}_\Phi \colon O \times O \mapsto M_{M\times M}(\mathbb{R})$ is called a Roe linearization if it verifies the following properties:*

- *for each $W_L, W_R \in O$, $\mathcal{A}_\Phi(W_L, W_R)$ has M distinct real eigenvalues,*
- *$\mathcal{A}_\Phi(W, W) = \mathcal{A}(W)$ for every $W \in O$,*
- *for any $W_L, W_R \in O$,*

$$\mathcal{A}_\Phi(W_L, W_R) \cdot (W_R - W_L) = \int_0^1 \mathcal{A}(\Phi(s; W_L, W_R))\partial_s \Phi(s; W_L, W_R)\, ds. \tag{2.2.27}$$

Note that, if the system is conservative and $\mathcal{A}$ is the Jacobian of some flux function $\mathcal{F}$, that is $\mathcal{A} = \frac{d\mathcal{F}}{dW}$, then (2.2.27) reduces to the usual Roe property,

$$\mathcal{A}_\Phi(W_L, W_R) \cdot (W_R - W_L) = \mathcal{F}(W_R) - \mathcal{F}(W_L),$$

and therefore the usual notion of Roe matrix is recovered.

Once the linearization has been chosen, the corresponding Roe scheme can be written as a path-conservative scheme in the form (2.2.20) with

$$\mathcal{D}^\pm(W_L, W_R) \;=\; \mathcal{A}^\pm_\Phi(W_L, W_R) \cdot (W_R - W_L), \tag{2.2.28}$$

where, as usual,

$$\mathcal{A}^\pm_\Phi(W_L, W_R) = \frac{1}{2}\big(\mathcal{A}_\Phi(W_L, W_R) \pm |\mathcal{A}_\Phi(W_L, W_R)|\big), \tag{2.2.29}$$

where

$$|\mathcal{A}_\Phi(W_L, W_R)| = K_\Phi(W_L, W_R) \cdot |L_\Phi(W_L, W_R)| \cdot K_\Phi(W_L, W_R)^{-1}. \tag{2.2.30}$$

Here, $|L_\Phi(W_L, W_R)|$ corresponds to the diagonal matrix whose coefficients are the absolute value of the eigenvalues of $\mathcal{A}_\Phi(W_L, W_R)$ and $K_\Phi(W_L, W_R)$ is a $M \times M$ matrix whose columns are the associated eigenvectors.

Using (2.2.28) and (2.2.29), an explicit expression for the numerical fluxes can be given,

$$\mathcal{D}^\pm(W_L, W_R) = \frac{1}{2}\mathcal{A}_\Phi(W_L, W_R) \pm \frac{1}{2}\,|\mathcal{A}_\Phi(W_L, W_R)|\,, \tag{2.2.31}$$

thus, completing the definition of the Roe numerical scheme (2.2.20).

UNIVERSIDAD DE MÁLAGA

An alternative writing can be given for systems like (2.2.1). Firstly, we recall that the solution W can be expressed as $W = (w, \sigma)^T$, and we consider the following paths,

$$\Phi = \begin{pmatrix} \Phi_w \\ \\ \Phi_\sigma \end{pmatrix}. \tag{2.2.32}$$

Additionally, the following linearization of (2.2.5) is considered, as in [170],

$$\mathcal{A}_\Phi(W_L, W_R) = \left(\begin{array}{c|c} A_\Phi(W_L, W_R) & -G_\Phi(W_L, W_R) \\ \hline 0 & 0 \end{array} \right),$$

where

$$A_\Phi(W_L, W_R) = J_F(w_L, w_R) + B_\Phi(W_L, W_R). \tag{2.2.33}$$

Here, $J_F(w_L, w_R)$ is a Roe linearization of the Jacobian of the flux function F in the usual sense,

$$J_F(w_L, w_R) \cdot (w_R - w_L) = F(w_R) - F(w_L); \tag{2.2.34}$$

and $B_\Phi(W_L, W_R)$ is a matrix satisfying,

$$B_\Phi(W_L, W_R) \cdot (w_R - w_L) = \int_0^1 B(\Phi_w(s; W_L, W_R)) \partial_s \Phi_w(s; W_L, W_R)\, ds, \tag{2.2.35}$$

while $G_\Phi(W_L, W_R)$ is a vector satisfying,

$$G_\Phi(W_L, W_R) \cdot (\sigma_R - \sigma_L) = \int_0^1 G(\Phi_w(s; W_L, W_R)) \partial_s \Phi_\sigma(s; W_L, W_R)\, ds. \tag{2.2.36}$$

Where the path expression (2.2.32) has been used. It is easy to check that, if (2.2.34)-(2.2.36) are satisfied, then (2.2.33) is a Roe linearization, provided that $A_\Phi(W_L, W_R)$ has N real distinct and not vanishing eigenvalues:

$$\lambda_1(W_L, W_R) < \cdots < \lambda_N(W_L, W_R).$$

The numerical scheme (2.2.20) can finally be written in terms of the solution W by taking into account the structure of the matrix $|\mathcal{A}_\Phi|$ and the Roe property (2.2.34),

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x} \left(D^+_{i-1/2} + D^-_{i+1/2} \right),$$

and

$$\begin{aligned} D^\pm_{i+1/2}(W_i, W_{i+1}) = \frac{1}{2} \big(& F(w_{i+1}) - F(w_i) + B_{i+1/2}(w_{i+1} - w_i) - G_{i+1/2}(\sigma_{i+1} - \sigma_i) \big) \\ & \pm \frac{1}{2} \left| A_{i+1/2} \right| \left(w_{i+1} - w_i - A^{-1}_{i+1/2} G_{i+1/2}(\sigma_{i+1} - \sigma_i) \right) \big), \end{aligned} \tag{2.2.37}$$

with $A_{i+1/2} = A_\Phi(W_i, W_{i+1})$, $B_{i+1/2} = B_\Phi(W_i, W_{i+1})$, $G_{i+1/2} = G_\Phi(W_i, W_{i+1})$.

UNIVERSIDAD DE MÁLAGA

2.2.2.2 Polynomial Viscosity Matrix (PVM) methods

While Roe methods are an effective way to approximate the numerical solution of general hyperbolic systems, the required computation of the full spectral information of the system can become a cumbersome task. Furthermore, the eigenvalues and eigenvectors of some systems, like the multilayer shallow-water models, are in fact unknown and while they could be numerically approximated at each cell interface for all simulation times, this strategy may require too much computational effort. Here, we present some ideas, mainly developed in a series of publications in [171], [172] and [48], to address this issue based on the decomposition of Roe matrices (2.2.29) that overcomes this difficulty. The idea is to replace (2.2.29) by:

$$\widehat{\mathcal{A}}^{\pm}_{\Phi}(W_L, W_R) = \frac{1}{2}\left(\mathcal{A}_{\Phi}(W_L, W_R) \pm \mathcal{Q}_{\Phi}(W_L, W_R)\right), \tag{2.2.38}$$

where

$$\mathcal{Q}_{\Phi}(W_L, W_R) = \left(\begin{array}{c|c} Q_{\Phi}(W_L, W_R) & -Q_{\Phi}(W_L, W_R) A^{-1}_{\Phi}(W_L, W_R) G_{\Phi}(W_L, W_R) \\ \hline 0 & 0 \end{array}\right).$$

In the previous formulas, $Q_{\Phi}(W_L, W_R)$ plays the role of a numerical viscosity matrix, that could be seen as an approximation of $|A_{\Phi}(W_L, W_R)|$. In this way, the numerical scheme defined by (2.2.20)-(2.2.31) is redefined by replacing the absolute value of the intermediate matrix by an approximation $\mathcal{Q}_{\Phi}$, which is easier to compute and that plays the role of viscosity matrix. Consequently, by choosing different viscosity matrix $\mathcal{Q}_{\Phi}$, different numerical methods will be obtained. For instance, it is obvious that the following choice,

$$Q_{\Phi}(W_L, W_R) = |A_{\Phi}(W_L, W_R)|,$$

corresponds to the Roe method just reviewed. Another example consists on the Rusanov method, which only needs an estimation of the largest wave speed in absolute value, and therefore it corresponds to the following viscosity matrix choice,

$$Q_{\Phi}(W_L, W_R) = \max(|\lambda_i(W_L, W_R)|)\mathcal{I}, \quad i = 1, \ldots, N, \tag{2.2.39}$$

where $\mathcal{I}$ is the identity matrix. Of course, this rough estimation is highly diffusive, especially for the waves corresponding to the eigenvalues with smaller absolute value.

Thus, the strategy of the *Polynomial Viscosity Methods*, PVM hereafter, introduced in [48] is to consider viscosity matrices of the following form,

$$Q_{\Phi}(W_L, W_R) = f(A_{\Phi}(W_L, W_R)), \tag{2.2.40}$$

where, $f : \mathbb{R} \mapsto \mathbb{R}$ satisfies the following properties:

1. $f(x) \geq 0, \forall x \in \mathbb{R}$,

UNIVERSIDAD DE MÁLAGA

2. $f(x)$ is *easy* to evaluate,

3. the graph of $f(x)$ is as close as possible to the graph of $|x|$.

Moreover, if $f(0) > 0$, then no entropy-fix techniques are required to avoid the appearance of non-entropy discontinuities at the numerical solutions.

The definition of the function f plays a major role in the stability of the numerical scheme. In [48] it is proven that the numerical scheme is L^∞-stable if the graph of the function $f(x)$ is above the one corresponding to the absolute value function in the interval containing the eigenvalues,

$$f(x) \geq |x|, \quad \forall\, x \in [\lambda_{1,i+1/2}, \lambda_{N,i+1/2}],\ \forall\, i \in \mathbb{Z}, \tag{2.2.41}$$

where $\lambda_{j,i+1/2} \equiv \lambda_j(W_i, W_{i+1}),\ j = 1, \ldots, N$, are the eigenvalues of the linearized matrix $A_\Phi(W_i, W_{i+1})$ and the usual CFL condition is assumed,

$$\frac{\Delta t}{\Delta x} \max_{i,j} |\lambda_{j,i+1/2}| \leq 1. \tag{2.2.42}$$

The next step consists on approximating the function $f(x)$ by an easy to construct and evaluate polynomial $P_\ell^{i+\frac{1}{2}}(x)$ of degree ℓ, that is evaluated on the linearized Roe matrix:

$$Q_\Phi(W_L, W_R) = P_\ell^{i+\frac{1}{2}}(A_\Phi(W_L, W_R)) \tag{2.2.43}$$

and

$$P_\ell^{i+\frac{1}{2}}(x) = \sum_{j=0}^{\ell} \alpha_j^{i+\frac{1}{2}} x^j. \tag{2.2.44}$$

This idea dates back to the work of Harten, Lax and van Leer [173], who proposed to choose f as a linear polynomial that interpolates the absolute value function $|x|$ at the smallest and larges eigenvalue, resulting in a considerable improvement with respect to the local Lax-Friedrichs method. This approximation, which results in the HLL method, has been further developed by several authors (the interested reader can refer to [174] for a complete review).

Additional work related to the approximation of $|A_\Phi|$ by means of a polynomial that interpolates $|x|$ can be found in Degond *et. al.* in [49]. In this case, the polynomial does not interpolate the absolute value function at the exact eigenvalues. This approach is the one extended in [48] to provide a general framework for the Polynomial Viscosity Matrix (PVM). In fact, many well-know numerical schemes can be written as a PVM method, such as the Roe method, the Lax-Friedrichs, Rusanov, HLL, FORCE, MUSTA, etc.

UNIVERSIDAD DE MÁLAGA

The numerical scheme (2.2.37) can be written in terms of the PVM polynomial (2.2.43),

$$\begin{aligned} D^{\pm}_{i+1/2}(W_i, W_{i+1}) \\ &= \frac{1}{2}\big(F(w_{i+1}) - F(w_i) + B_{i+1/2}(w_{i+1} - w_i) - G_{i+1/2}(\sigma_{i+1} - \sigma_i)\big) \\ &\quad \pm \frac{1}{2} Q_{i+1/2}\big(w_{i+1} - w_i - A^{-1}_{i+1/2} G_{i+1/2}(\sigma_{i+1} - \sigma_i))\big), \end{aligned} \tag{2.2.45}$$

where $Q_{i+1/2} = P_\ell^{i+1/2}(A_{i+1/2})$. Note that the term,

$$C = Q_{i+1/2} A^{-1}_{i+1/2} G_{i+1/2}, \tag{2.2.46}$$

that can be interpreted as the up-winding discretization of the source term, can in fact lose meaning if any eigenvalue of $A_{i+1/2}$ vanishes. In that case, the problem is categorized as resonant, and they present additional difficulties. Here we propose to use the strategy developed in [175] to formally avoid it.

Definition 3. *A numerical method PVM is said to be upwind if*

$$P_\ell^{i+1/2}(A_{i+1/2}) = \begin{cases} A_{i+1/2}, & \text{if } \lambda_{1,i+1/2} > 0, \\ -A_{i+1/2}, & \text{if } \lambda_{N,i+1/2} < 0, \end{cases}$$

and we denote as PVM-U. Thus, if

$$P_\ell^{i+1/2}(x) = \begin{cases} x, & \text{if } \lambda_{1,i+1/2} > 0, \\ -x, & \text{if } \lambda_{N,i+1/2} < 0, \end{cases}$$

then the resulting PVM method is upwind.

We will now review some examples of PVM methods, particularly those classic methods that can be rewritten as PVM type methods. The following notation will be considered: PVM-ℓ $(S_0, \dots, S_k)$ denotes a numerical scheme which use a polynomial of degree ℓ to define the viscosity matrix $Q_{i+\frac{1}{2}}$ whose coefficients depend on the parameters $S_0, \dots, S_k$. These parameters will be related to the approximation of the wave speed considered. Additionally, if the method is upwind, we will write PVM-U. In order to simplify the notation, the superindex $i + \frac{1}{2}$ will be dropped from the interpolation polynomial and its coefficients. Thus, $P_\ell(x)$ will be preferred over $P_\ell^{i+\frac{1}{2}}(x)$.

2.2.2.3 PVM-(N-1)U $(\lambda_1, \dots, \lambda_N)$ or Roe method

As stated before, the Roe method correspond to the following choice,

$$Q_{i+\frac{1}{2}} = |A_{i+\frac{1}{2}}|. \tag{2.2.47}$$

UNIVERSIDAD DE MÁLAGA

Then, to rewrite the Roe method as a PVM method it is enough to consider,

$$|A_{i+\frac{1}{2}}| = \sum_{j=0}^{N-1} \alpha_j A^j_{i+\frac{1}{2}}, \tag{2.2.48}$$

where α_j, $j = 0, \dots, N$ are the solution to the following linear system:

$$\begin{pmatrix} 1 & \lambda_{1,i+\frac{1}{2}} & \dots & \lambda^{N-1}_{1,i+\frac{1}{2}} \\ 1 & \lambda_{2,i+\frac{1}{2}} & \dots & \lambda^{N-1}_{2,i+\frac{1}{2}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_{N,i+\frac{1}{2}} & \dots & \lambda^{N-1}_{N,i+\frac{1}{2}} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} = \begin{pmatrix} |\lambda_{1,i+\frac{1}{2}}| \\ |\lambda_{2,i+\frac{1}{2}}| \\ \vdots \\ |\lambda_{N,i+\frac{1}{2}}| \end{pmatrix}. \tag{2.2.49}$$

Note that this is a Vandermonde matrix, and thus it always has a unique solution if and only if $\lambda_{l,i+\frac{1}{2}} \neq \lambda_{j,i+\frac{1}{2}}$, $l \neq j$ with $l, j = 1, \dots, N$.

2.2.2.4 PVM-0 (S_0) methods: Rusanov, Lax-Friedrichs and Lax-Friedrichs modified methods

The simplest choice for a PVM methods correspond to,

$$P_0(x) = S_0. \tag{2.2.50}$$

Therefore, $P_0(x)$ correspond to the constant horizontal line (see Figure 2.1). Additionally, for stability demands we have that,

$$\max_j |\lambda_{j,i+\frac{1}{2}}| \leq S_0 \leq \frac{\Delta x}{\Delta t}.$$

Therefore, some choices for S_0 correspond to the Rusanov, Lax-Friedrichs and Lax-Friedrichs modified,

$$S_0 \in \{S_{Rus},\, S_{LF},\, S^{mod}_{LF}\}, \tag{2.2.51}$$

where

$$S_{Rus} = \max_j |\lambda_{j,i+\frac{1}{2}}|, \qquad S_{LF} = \frac{\Delta x}{\Delta t}, \qquad S^{mod}_{LF} = \alpha \frac{\Delta x}{\Delta t}. \tag{2.2.52}$$

2.2.2.5 PVM-1U (S_L, S_R) or HLL method

We consider now the PVM method corresponding to a first order polynomial (see Figure 2.2),

$$P_1(x) = \alpha_0 + \alpha_1 x. \tag{2.2.53}$$

UNIVERSIDAD DE MÁLAGA

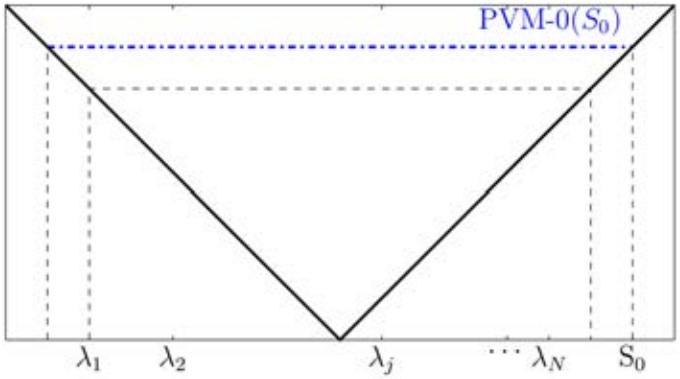


Figure 2.1: PVM-0 (S_0) polynomial.

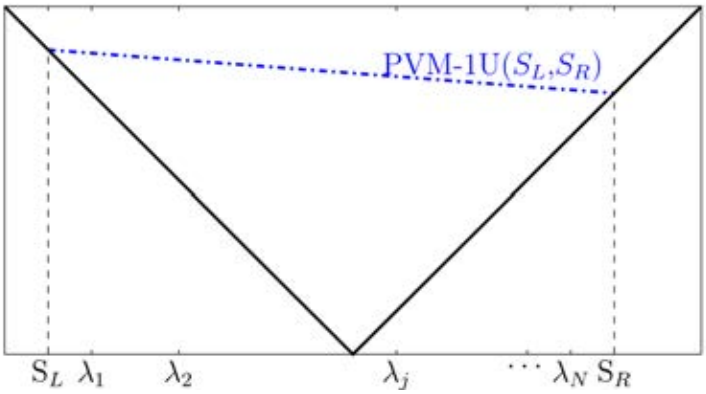


Figure 2.2: PVM-1U (S_L, S_R) polynomial.

In order to define the coefficients α_0 y α_1, we impose that,

$$P_1(S_L) = |S_L|, \quad P_1(S_R) = |S_R|,$$

where S_L and S_R are, respectively, an approximation of the minimum and maximum wave speed propagation. A possibility is to assume $S_L = \lambda_{1,i+\frac{1}{2}}$, $S_R = \lambda_{N,i+\frac{1}{2}}$, although it is also possible to use the choice by Davis in [176],

$$S_L = \min(\lambda_{1,i+\frac{1}{2}}, \lambda_{1,i}), \quad S_R = \max(\lambda_{N,i+\frac{1}{2}}\lambda_{N,i}),$$

where $\lambda_{i,1} < \ldots < \lambda_{i,N}$ are the eigenvalues of the matrix $A_{i+\frac{1}{2}}$. After some computations, we arrive to,

$$\alpha_0 = \frac{S_R|S_L| - S_L|S_R|}{S_R - S_L}, \quad \alpha_1 = \frac{|S_R| - |S_L|}{S_R - S_L}. \tag{2.2.54}$$

Note that if $S_L > 0 \Longrightarrow P_1(x) = x$ and if $S_R < 0 \Longrightarrow P_1(x) = -x$. Then the resulting method is *upwind*. Additionally, if the system (2.2.3) is conservative, then the flux is also

UNIVERSIDAD DE MÁLAGA

conservative and the method coincides with the classic HLL method as described in [173]. Thus, the described PVM-1U(S_L, S_R) scheme gives a natural generalization of the HLL method for non-conservative problems.

2.2.2.6 PVM-2(S_0) or FORCE type methods

The PVM methods with a second order polynomial (see Figure 2.3) are now considered:

$$P_2(x) = \alpha_0 + \alpha_2 x^2, \tag{2.2.55}$$

such that

$$P_2(S_0) = S_0, \quad P_2'(S_0) = 1,$$

where S_0 is given by (2.2.51). After some simple computations we obtain,

$$\alpha_0 = \frac{S_0}{2}, \quad \alpha_2 = \frac{1}{2S_0}. \tag{2.2.56}$$

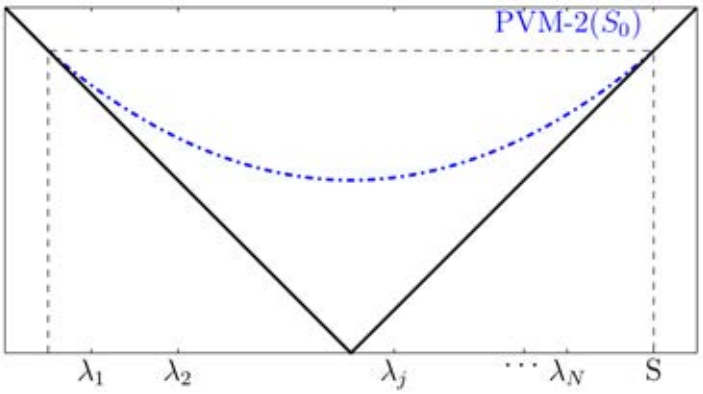


Figure 2.3: PVM-2 (S_0) polynomial.

Note that, since $\alpha_0 = 0$, the PVM-2 (S_0) is not upwind in the sense given by definition 3. Additionally, if $S_0 = S_{LF}$, then the PVM-2(S_{LF}) method coincides with the FORCE method (see [171], [177]). In this case, the PVM-2(S_{LF}) depends on $\frac{\Delta x}{\Delta t}$ and the method can be seen as a combination of the Lax-Friedrichs and Lax-Wendroff methods.

Finally, the GFORCE scheme can be obtained by imposing,

$$P_2(S_{LF}^{mod}) = S_{LF}^{mod}, \quad P_2'(S_{LF}^{mod}) = \frac{2\alpha}{1+\alpha},$$

where the following expression is derived,

$$\alpha_0 = \frac{S_{LF}^{mod}}{1+\alpha}, \quad \alpha_2 = \frac{1}{S_{LF}^{mod}} \frac{\alpha}{1+\alpha}. \tag{2.2.57}$$

which is linearly L^∞ stable under the usual CFL condition.

UNIVERSIDAD DE MÁLAGA

2.2.3 High order extension

In this section we review the general framework to define high order finite volume methods for PDE system (2.2.3) based on the first order path-conservative numerical schemes and the use of a reconstruction operator. The cell average of the solution W of (2.2.3) in the cell I_i at time t is denoted as,

$$W_i(t) = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} W(x,t)\,dx.$$

From (2.2.3), we are able to obtain the following equation:

$$W_i' = -\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \mathcal{A}(W(x,t))\partial_x W(x,t)\,dx. \tag{2.2.58}$$

Let us recall the definition of a reconstruction operator.

Definition 4. *A reconstruction operator of order s is an operator that provides a smooth function at every cell $x \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ from the cell average value W_i and its neighbors W_j, that constitutes the stencil $\mathcal{S}_i$,*

$$R_i(x) = R_i(x; \{W_j\}_{j\in\mathcal{S}_i}). \tag{2.2.59}$$

Then,

$$W^+_{i-\frac{1}{2}} = W(x_{i-\frac{1}{2}}) + \mathcal{O}(\Delta x^s), \tag{2.2.60}$$
$$W^-_{i+\frac{1}{2}} = W(x_{i+\frac{1}{2}}) + \mathcal{O}(\Delta x^s), \tag{2.2.61}$$

with

$$W^+_{i-\frac{1}{2}} = R_i(x_{i-\frac{1}{2}}), \tag{2.2.62}$$
$$W^-_{i+\frac{1}{2}} = R_i(x_{i+\frac{1}{2}}). \tag{2.2.63}$$

The states $W^+_{i-\frac{1}{2}}$ and $W^-_{i+\frac{1}{2}}$ are called the reconstructed states and are computed by taking the limit of the reconstruction operator,

$$\lim_{x\to x^+_{i-\frac{1}{2}}} R_i(x) = W^+_{i-\frac{1}{2}}, \qquad \lim_{x\to x^-_{i+\frac{1}{2}}} R_i(x) = W^-_{i+\frac{1}{2}}.$$

Following [40], we consider the semi-discrete method:

$$W_i' = -\frac{1}{\Delta x}\left(\mathcal{D}^+_{i-\frac{1}{2}} + \mathcal{D}^-_{i+\frac{1}{2}} + \int_{I_i}\mathcal{A}(R_i(x))\partial_x R_i(x)\,dx\right), \tag{2.2.64}$$

UNIVERSIDAD DE MÁLAGA

where $\mathcal{D}^{\pm}_{i+\frac{1}{2}}$ are evaluated at $W^{\pm}_{i+\frac{1}{2}}(t)$,

$$\mathcal{D}^{\pm}_{i+\frac{1}{2}} = \mathcal{D}^{\pm}(W^{-}_{i+\frac{1}{2}}(t), W^{+}_{i+\frac{1}{2}}(t)). \tag{2.2.65}$$

Here, $\{W^{\pm}_{i+\frac{1}{2}}(t)\}$ stands for the reconstructed states associated with $\{W_i(t)\}$, and

$$R_i(x) \equiv R_i(x; W_{i-l}(t), \ldots, W_{i+r}(t)). \tag{2.2.66}$$

Note that in (2.2.64) the approximation functions are used to approximate the regular part of the weak integral in (2.2.58) while the terms $\mathcal{D}^{\pm}_{i-1/2}$ are used to split the Dirac measures corresponding to the discontinuities at the intercells.

The semidiscrete method (2.2.64) is a high order in space system of ordinary differential equations and must be discretized in time with a suitable high order in time solver. It is common to use TVD Runge-Kutta numerical schemes (see [178, 179]). But other fully implicit alternatives can also be considered. For instance, the Cauchy-Kovaleskaya procedure [68] that substitutes time derivatives with space derivatives via successive differentiation of the governing PDE with respect to space and time. Alternatively, the ADER method [74] allows to reach arbitrary high order in time in one step. This method will be discussed in detail later.

Note that the scheme (2.2.64) can also be represented by,

$$\begin{aligned} w_i' = &-\frac{1}{\Delta x}\left(D^{+}_{i-\frac{1}{2}}(w^{-}_{i-\frac{1}{2}}, w^{+}_{i-\frac{1}{2}}, \sigma^{-}_{i-\frac{1}{2}}, \sigma^{+}_{i-\frac{1}{2}}) + D^{-}_{i+\frac{1}{2}}(w^{-}_{i+\frac{1}{2}}, w^{+}_{i+\frac{1}{2}}, \sigma^{-}_{i+\frac{1}{2}}, \sigma^{+}_{i+\frac{1}{2}}) \right) \\ &- \frac{1}{\Delta x}\left(\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} B(R^{w}_i(x)) \frac{dR^{w}_i(x)}{dx}\, dx - \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} G(R_i(x)) \frac{dR^{\sigma}_i(x)}{dx} \right). \end{aligned} \tag{2.2.67}$$

Where $R^{w}_i(x)$ stands for the high order reconstruction operator applied to $w_i(x)$, while $R^{\sigma}_i(x)$ is similarly applied to $\sigma(x)$. Note that in the expression (2.2.67) we can use the explicit definition of $\sigma(x)$ or, alternatively, apply the high order reconstruction operator to $\sigma(x)$.

In high order methods, there are significant differences between the conservative and the non-conservative cases. Whereas in the conservative case the high order numerical scheme only depends on the order of approximation of the reconstructed values at the intercells, this is not the case for non-conservative systems. In this case, the order of reconstruction depends on the approximation properties of the reconstruction operator on the complete cell. Thus, the order of approximation for non-conservative systems is $\alpha = \min(s, s_1, s_2)$, where s_1 and s_2 are the order of accuracy of the reconstruction operator and its derivative inside the cell (see [50]):

$$\begin{aligned} R_i(x) &= W(x) + O\left(\Delta x^{s_1}\right) \quad \forall\, x \in I_i, \\ \frac{d}{dx} R_i(x) &= \frac{d}{dx} W'(x) + O\left(\Delta x^{s_2}\right) \quad \forall\, x \in I_i. \end{aligned}$$

UNIVERSIDAD DE MÁLAGA

For the most common reconstruction techniques, it is often the case that $s_2 \leq s_1 \leq s$ and therefore the order of accuracy of (2.2.64) is s_2 for non-conservative systems, assuming a small loss of accuracy for these cases. This effect has been observed numerically for Roe-WENO methods in [50]. Nevertheless, this error estimation is rather pessimistic: in practice, the order of the observed error is usually s_1: see [51] or [50]. Additionally, in [120] a technique to avoid the explicit computation of the derivation of the reconstruction operator so that the order of accuracy is now $\min(s, s_1)$ is introduced. It is based on the use of the trapezoidal rule and Romberg extrapolation for the numerical approximation of the integrals in (2.2.64).

Moreover, in [180] a discussion on high order finite volume central schemes for conservatives and non-conservatives hyperbolic systems on staggered grids is included. The method is developed under the path-conservative paradigm.

2.2.3.1 MUSCL reconstruction operator

Several strategies exist to define high order reconstruction operators, usually computed by means of interpolation or approximation techniques. For instance, the ENO, WENO or hyperbolic reconstruction techniques falls into this category (see [181], [182], [183], [179], [74], [184], [185]).

We limit to the MUSCL (monotone upwind scheme for conservation laws) reconstruction operator (see [186]), the one used in this thesis, for the case of regular meshes. In this way, the MUSCL reconstruction operator defines, in each cell I_i at each time,

$$W_i(x,t) = W_i(t) + (x - x_i)\,\delta_i. \tag{2.2.68}$$

Here, δ_i is an approximation of the derivative with respect to x of the solution at the cell I_i. This approximation must be coupled with some mechanism to limit the slope of (2.2.68) in the presence of strong gradients or discontinuities, while preserving the second order accuracy of the operator in regular regions. Particularly, the average (avg) slope limiter is considered for the definition of δ_i,

$$\delta_i = \mathrm{avg}\left(\frac{W_{i+1} - W_i}{\Delta x}, \frac{W_i - W_{i-1}}{\Delta x}\right), \tag{2.2.69}$$

with

$$\mathrm{avg}(a,b) = \begin{cases} \dfrac{|a|\,b + a\,|b|}{|a| + |b|} & \text{if } |a| + |b| > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.2.70}$$

Note that the expression in (2.2.70) is computed componentwise.

UNIVERSIDAD DE MÁLAGA

2.2.3.2 MUSCL-Hancock reconstruction operator

The previous high order reconstruction operator can be combined with a linear reconstruction in time that avoids the use of an ODE solver for the semidiscrete method, allowing for implicit second order finite volume methods. This approach is the so called MUSCL-Hancock reconstruction operator, as described in [186]. In that work, the following space time function is defined for each cell I_i and for all time step,

$$W_i(x,t) = W_i(t) + (x - x_i)\,\delta_i + (t - t^n)\partial_t W_i(t). \tag{2.2.71}$$

The time derivative that appears in equation (2.2.71) can be approximated using the PDE (2.2.1),

$$\partial_t W_i = -\partial_x F(W_i) - B(W_i)\partial_x W_i + G(W_i)\sigma_x, \tag{2.2.72}$$

and approximating the spatial derivatives at time t^n. Note that in (2.2.71), the same definition of δ_i (2.2.69) is considered, alongside the slope limiter (2.2.70).

2.2.4 Well-balanced path conservative finite volume schemes

Stationary solutions of general PDE systems like (2.2.3), successively denoted as W^*, are solutions that do not depend on time, that is, $\partial_t W$ is zero,

$$\mathcal{A}(W^*)\partial_x W^* = 0 \qquad \forall x \in I,\ t \geq 0. \tag{2.2.73}$$

Note that (2.2.73) implies that 0 is an associated eigenvalue of $\mathcal{A}$ and $\partial_x W$ is an associated eigenvector for every x. Therefore, $x \mapsto W(x)$ can be seen as a parametrization of an integral curve of a linearly degenerate characteristic field whose corresponding eigenvalue takes the value 0 through the curve. We will denote as Γ the set of all the integral curves γ of a linearly degenerate field of $\mathcal{A}$ such that the corresponding eigenvalue vanish on Γ. With the definition of this set we can give a general definition of first order well-balanced methods:

Definition 5. *A first order path-conservative finite volume scheme* (2.2.20) *is said to be exactly well-balanced if, given any pair of states W_L and W_R belonging to $\gamma \in \Gamma$, one has that*

$$\mathcal{D}^{\pm}(W_L, W_R) = 0. \tag{2.2.74}$$

As a consequence, when we have two states W_L and W_R, both belonging to the same integral curve $\gamma \in \Gamma$, such that the stationary contact discontinuity at x^* is,

$$W(x,t) = \begin{cases} W_L & x < x^*, \\ W_R & x > x^*. \end{cases} \tag{2.2.75}$$

then we want $W(x,t)$ to be a weak solution. In the conservative case this is evident, since W_L and W_R both belong to the same integral curve and this is equivalent to preserve the

UNIVERSIDAD DE MÁLAGA

corresponding Riemann invariant and therefore it will be an admissible weak solution. But, as it has been stated before, in the non-conservative case the notion of weak solution where (2.2.75) is an admissible solution must be chosen. Therefore, the family of paths must be taken so that (2.2.75) satisfies the Rankine-Hugoniot condition (2.2.11). To achieve this, it is enough to ensure that the chosen path linking two states W_L and W_R belonging to the same integral curve of a linearly degenerated field is a parametrization of the arc of this integral curve linking the states.

To check this for hyperbolic systems (2.2.1), let us recall that the matrix $\mathcal{A}(W)$, with $W = (w, \sigma)$, has the following structure,

$$\mathcal{A}(W) = \left(\begin{array}{c|c} J_F(w) + B(w) & -G(w) \\ \hline 0 & 0 \end{array} \right),$$

and the following eigenvalues and eigenvectors,

$$\lambda_1(w) < \ldots < \lambda_{M-1}(w), 0,$$

$$\mathcal{R}_1(w) < \ldots < \mathcal{R}_{M-1}(w)$$

and

$$\mathcal{R}_M(W) = \begin{pmatrix} (J_F(w) + B(w))^{-1} G(w) \\ 1 \end{pmatrix}.$$

Note that $\lambda_1(U), \ \ldots \lambda_{M-1}(w)$ are the eigenvalues of matrix $A(w) = J_F(w) + B(w)$ and $\mathcal{R}_i(w)$, their corresponding eigenvectors. In the definition of $\mathcal{R}_M(W)$, we assume that $A(w)$ is regular.

Observe that the set Γ is defined by the integral curves of the linearly degenerated field associated with the eigenvector $\mathcal{R}_M(W)$, *i.e*, the following ODE system,

$$\frac{dW}{ds} = \mathcal{R}_M(W), \tag{2.2.76}$$

which can be written as,

$$\begin{cases} \dfrac{dw}{ds} = (J_F(w) + B(w))^{-1} G(w), \\ \dfrac{d\sigma}{ds} = 1. \end{cases} \tag{2.2.77}$$

If we choose σ as the parameter, the previous system can be written as,

$$\frac{dw}{d\sigma} = (J_F(w) + B(w))^{-1} G(w), \tag{2.2.78}$$

or alternatively,

$$(J_F(w) + B(w)) \frac{dw}{d\sigma} = G(w), \tag{2.2.79}$$

UNIVERSIDAD DE MÁLAGA

which is clearly a reparametrization of (2.2.73). Therefore, a pair of states W_L and W_R belong to the same integral curve $\gamma \in \Gamma$ if and only if there exist a solution of the ODE system,

$$(J_F(V) + B(V))\frac{dV}{d\sigma} = G(V), \tag{2.2.80}$$

such that,

$$\begin{cases} V(\sigma_L) = w_L, \\ V(\sigma_R) = w_R. \end{cases} \tag{2.2.81}$$

It is clear then that a solution $V(\sigma)$ of (2.2.80), is a stationary solution $w(x) = V(\sigma(x))$ since it satisfies (2.2.73) in the smooth regions and (2.2.80)-(2.2.81) at the discontinuities.

Remark 2.2.1. *In the case of a resonant problem, when some of the eigenvalues of $\mathcal{A}(W)$ are zero, then the Cauchy problem* (2.2.78) *may not have a solution or have more than one. In the case of multiple solution, a criteria must be set to decide what are the admissible discontinuities that the numerical method must preserve (see [175]).*

A possible choice for a family of paths that connects two states belonging to the same integral curve over admissible stationary solutions at a contact discontinuity is,

$$\Phi(\xi; W_L, W_R) = \begin{pmatrix} \Phi_w(\xi; W_L, W_R) \\ \Phi_\sigma(\xi; W_L, W_R) \end{pmatrix} = \begin{pmatrix} V(\sigma_L + \xi(\sigma_R - \sigma_L)) \\ \sigma_L + \xi(\sigma_R - \sigma_L) \end{pmatrix}, \quad \xi \in [0, 1]. \tag{2.2.82}$$

Effectively, it is easy to check that this family of paths satisfy the Rankine-Hugoniot condition (2.2.11) with zero speed.

Finally, we can conclude that if a numerical method is designed such that the family of paths linking two states that are the limits of an admissible jump at the discontinuity points of σ reduces to (2.2.82), then the first order path-conservative numerical method (2.2.20) is exactly well-balanced.

2.2.4.1 Generalized hydrostatic reconstruction

The generalized hydrostatic reconstruction is a technique that allows to define first order path-conservative numerical schemes that are exactly well-balanced. This strategy was first proposed in [98] in the framework of shallow water systems, and was later generalized in [187] and [188]. This technique is later used in this thesis to define a well-balanced method that preserve stationary solution corresponding to the lake-at-rest steady states for the multilayer shallow water system. The generalized hydrostatic reconstruction can be seen as a particular choice of paths that guarantee (2.2.82). The method is defined as follows: let us consider two states $W_L = (w_L, \sigma_L)^T$ and $W_R = (w_R, \sigma_R)^T$ that must be connected by a given path. First, an intermediate value σ_0 between σ_L and σ_R is chosen. In particular, in this thesis the following intermediate value has been chosen,

$$\sigma_0 = \min(\sigma_L, \sigma_R).$$

UNIVERSIDAD DE MÁLAGA

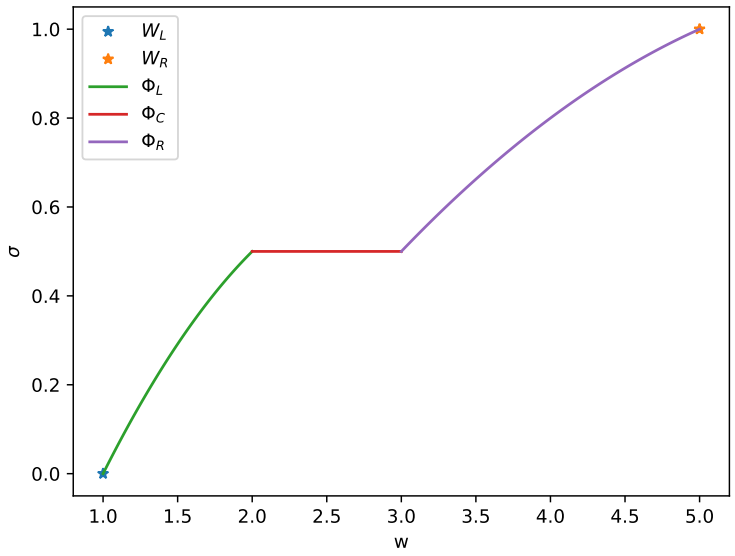


Figure 2.4: Sketch of the hydrostatic reconstruction technique.

Next, equation (2.2.80) is solved (if possible) with initial condition $V(\sigma_L) = w_L$ and $V(\sigma_R) = w_R$ respectively and the solution of these Cauchy problems are denoted by $V_L(\sigma)$ and $V_R(\sigma)$. Now, we propose to use a path to link W_L and W_R that is composed of three pieces in the following way:

- The path Φ_L connecting $W_L = (w_L, \sigma_L)^T$ to $(V_L(\sigma_0), \sigma_0)^T$, given by $(V_L(\sigma), \sigma)^T$.
- The straight segment Φ_C linking $(V_L(\sigma_0), \sigma_0)^T$ and $(V_R(\sigma_0), \sigma_0)^T$.
- And finally, the path Φ_R connecting $(V_R(\sigma_0), \sigma_0)^T$ to $W_R = (w_R, \sigma_R)^T$, given by $(V_R(\sigma), \sigma)^T$.

Observe that the proposed path is a composition of three paths, some of them could be reduced to a single point, that can be parametrized over $[0, 1]$. Note that if W_L and W_R belongs to the same integral curve $\gamma \in \Gamma$ of a linearly degenerated field, then $V_L = V_R = V$ and the path is reduced to (2.2.82). Figure 2.4 offers a sketch of the composition of path of the generalized hydrostatic reconstruction.

2.2.5 High order well-balanced reconstruction operators

The design of well-balanced first order numerical methods has been previously discussed in this thesis. However, achieving high order in space through reconstruction operators

UNIVERSIDAD DE MÁLAGA

like (2.2.59) may destroy the well-balanced character of the numerical scheme. Here, we follow [128] to define high order finite volume schemes that are exactly well-balanced. As pointed out in [128], it is important to consider exactly well-balanced reconstruction operators that allow to reach high order in space while maintaining the ability to preserve stationary solutions. A high order reconstruction operator $R_i(x)$ is said to be exactly well-balanced for a stationary solution $W^*(x)$ if

$$R_i(x; \{W_j\}_{j\in\mathcal{S}_i}) = W^*(x), \qquad \forall x \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}],\ \forall i. \tag{2.2.83}$$

In this way, the high order numerical scheme (2.2.64) is exactly well-balanced provided that both the first order finite volume method and the reconstruction operator are exactly well-balanced. Note that, for smooth stationary solutions, this strategy does not require a first order well-balanced finite volume method since a exactly well-balanced operator is enough.

In general, reconstruction operators are not well-balanced, since they are based on non-linear approximation of the solution using the average at a given cell and its neighbors. Nonetheless, in [128] a procedure to build well-balanced reconstruction operators for a given family of cell values $\{W_i\}$ is presented:

1. First, find the stationary solution $W^*(x)$ defined in the stencil of I_i, $\mathcal{S}_i$, such that,

$$\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} W_i^*(x)\,dx = W_i. \tag{2.2.84}$$

   Note that it may not be possible to solve this equation. In this case, one may take $W_i^* = 0$.

2. Then the fluctuations within the stencil $\mathcal{S}_i$ of the solution and the computed stationary solution are considered,

$$V_j = W_j - \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} W_i^*(x)\,dx, \quad j \in \mathcal{S}_i. \tag{2.2.85}$$

3. Next, the standard reconstruction operator is applied to these fluctuations $\{V_j\}_{j\in\mathcal{S}_i}$,

$$RV_i(x) = RV_i(x; \{V_j\}_{j\in\mathcal{S}_i}). \tag{2.2.86}$$

4. Finally, the well-balanced high order reconstruction operator is defined as,

$$R_i(x) = W_i^*(x) + RV_i(x). \tag{2.2.87}$$

UNIVERSIDAD DE MÁLAGA

It is easy to see that the reconstruction operator $R_i(x)$ is exactly well-balanced provided that the reconstruction operator $RV_i(x)$ is exact for the null operator. Moreover, this reconstruction operator is conservative,

$$\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} R_i(x)\,dx = W_i, \quad \forall i,$$

provided that $RV_i(x)$ is also conservative.

The main difficulty of this method comes at the first step, where a solution to the equation (2.2.84) must be found, which may not be feasible for some systems. Note that, if there are no stationary solutions and W_i^* is taken as zero, then this is equivalent to a standard reconstruction procedure. This does not mean that the well-balanced property of the operator is lost. Rather, in this case the average solution in the stencil cannot be a stationary solution and therefore there is no local solution to preserve. Conversely, if there are multiple stationary solutions satisfying (2.2.84), then a criteria is needed to choose between them depending on the problem (see [189]).

In practice, the integral terms of the well-balanced procedure are approximated by a quadrature formula,

$$W_i = \sum_{k=0}^{M} \alpha_k^i W_i(x_k^i), \quad \forall i, \tag{2.2.88}$$

where α_k^i and x_k^i with $k = 0, \ldots, M$ are the weights and nodes of the quadrature formula respectively. In this case, the first two steps of the well-balanced procedure is modified as follows:

1. First, find the stationary solution $W^*(x)$ defined in the stencil of I_i, $\mathcal{S}_i$, such that,

$$\sum_{k=0}^{M} \alpha_k^i W_i^*(x_k^i) = W_i, \tag{2.2.89}$$

   if possible. Otherwise, take $W_i^* = 0$.

2. Then, the fluctuations within the stencil $\mathcal{S}_i$ of the solution and the computed stationary solution are considered,

$$V_j = W_j - \sum_{k=0}^{M} \alpha_k^i W_i^*(x_k^i). \tag{2.2.90}$$

Note that the first step can be significantly simplified if we restrict ourselves to preserve only a family of relevant stationary solutions instead of all stationary solutions. For instance, if W_i^* is of the form

$$W_i^* = c_i + s(x), \quad x \in I_i, \tag{2.2.91}$$

UNIVERSIDAD DE MÁLAGA

where $s(x)$ is a know function and $c_i \in \mathbb{R}$. The integration of (2.2.84) using a quadrature formula leads to,

$$c_i = \sum_{k=0}^{M} \alpha_k^i (W_i - s(x_k^i)), \tag{2.2.92}$$

which completely characterizes W_i^*. Another example are stationary solutions of the form,

$$W_i^* = c_i s_i(x), \quad x \in I_i. \tag{2.2.93}$$

The constant c_i can be computed in a similar way and yields,

$$c_i = \frac{\sum_{k=0}^{M} \alpha_k^i W_i}{\sum_{k=0}^{M} \alpha_k^i s(x_k^i)}. \tag{2.2.94}$$

Note that the procedure can be generalized for k-parameter family of stationary solutions, as described in [128],

$$W_i^*(x; c_{i,1}, \ldots, c_{i,k}), \qquad c_{i,r} \in \mathbb{R}. \tag{2.2.95}$$

Note that if a quadrature formula is used in (2.2.64) to approximate the integral term, then the well-balanced character of the scheme could be affected. To overcome this difficulty, we propose to subtract the term $\mathcal{A}(W_i^*(x))\partial_x W_i^*(x)$ within the cell I_i:

$$W_i' = -\frac{1}{\Delta x}\left(\mathcal{D}_{i-\frac{1}{2}}^{+} + \mathcal{D}_{i+\frac{1}{2}}^{-} + \int_{I_i} (\mathcal{A}(R_i(x))\partial_x R_i(x) - \mathcal{A}(W_i^*(x))\partial_x W_i^*(x))\, dx\right). \tag{2.2.96}$$

Note that the added term is zero. Next, the integral term is approximated by a quadrature formula

$$W_i' = -\frac{1}{\Delta x}\left(\mathcal{D}_{i-\frac{1}{2}}^{+} + \mathcal{D}_{i+\frac{1}{2}}^{-} + \sum_{k=0}^{M} \alpha_k^i \left(\mathcal{A}(R_i(x_k^i))\partial_x R_i(x_k^i) - \mathcal{A}(W_i^*(x_k^i))\partial_x W_i^*(x_k^i)\right)\right). \tag{2.2.97}$$

Approximating the integral term in this way we ensure that it is zero if $R_i(x) = W_i^*(x)$, while keeping the high order properties of the model intact.

Finally, by taking into account the structure of $\mathcal{A}$ in (2.2.5), expressions (2.2.96) and (2.2.97) can be written as,

$$\begin{aligned} w_i' = &-\frac{1}{\Delta x}\left(D_{i-\frac{1}{2}}^{+} + D_{i+\frac{1}{2}}^{-} + F(w_{i-\frac{1}{2}}^{+}) - F(w_{i-\frac{1}{2}}^{*}) - F(w_{i+\frac{1}{2}}^{-}) + F(w_{i+\frac{1}{2}}^{*})\right) \\ &- \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \left(B(R_i^w(x))\frac{dR_i^w(x)}{dx} - B(w_i^*(x))\frac{dw_i^*(x)}{dx}\right) dx \\ &- \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \left(G(R_i^w(x)) - G(w_i^*(x))\right)\frac{d\sigma(x)}{dx}\, dx, \end{aligned} \tag{2.2.98}$$

UNIVERSIDAD DE MÁLAGA

and

$$w_i' = -\frac{1}{\Delta x}\left(D^+_{i-\frac{1}{2}} + D^-_{i+\frac{1}{2}} + F(w^+_{i-\frac{1}{2}}) - F(w^*_{i-\frac{1}{2}}) - F(w^-_{i+\frac{1}{2}}) + F(w^*_{i+\frac{1}{2}})\right) - \frac{1}{\Delta x}\sum_{k=0}^{M}\alpha_k^i\left(B(R_i^w(x_k^i))\frac{dR_i^w(x_k^i)}{dx} - B(w_i^*(x_k^i))\frac{dw_i^*(x_k^i)}{dx}\right)dx - \frac{1}{\Delta x}\sum_{k=0}^{M}\alpha_k^i\left(G(R_i^w(x_k^i)) - G(w_i^*(x_k^i))\right)\frac{d\sigma(x)}{dx}\,dx. \quad (2.2.99)$$

Note that in both expression (2.2.98) and (2.2.99) the explicit knowledge of $\sigma(x)$ has been used.

2.3 Discontinuous Galerkin numerical schemes

In this section, we recall the DG method to discretize non-conservative hyperbolic systems of the form (2.2.1)

$$\partial_t w + \partial_x F(w) + B(w)\partial_x w = G(w)\sigma_x. \quad (2.3.1)$$

As before, the computational domain I is covered with a set of conforming cells $I_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$, $i = 1, \ldots, N_s$, where N_s is the total number of cells with a constant length $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$. As usual, the computational grid is the union of all the elements I_i,

$$I = \bigcup_{i=1}^{N_s} I_i. \quad (2.3.2)$$

The discrete solution of the PDE system (2.3.1) at time t^n in each subset I_i for the discontinuous Galerkin is denoted by $w_h(x, t^n)$ and is described in terms of piecewise polynomials of degree N_p in the spatial dimensions. We shall denote by $\mathcal{U}_h$ the space of piecewise polynomials up to degree N_p so that $w_h(x, t^n) \in \mathcal{U}_h$. In this thesis, a nodal basis defined by the Lagrange interpolation polynomials over the (N_p+1) Gauss-Legendre quadrature nodes on the element I_i is considered. We stress that the piecewise solution w_h may be discontinuous across elements, allowing jump discontinuities at cell interfaces. Within the element I_i, the discrete solution w_h is written in terms of the nodal spatial basis functions $\Phi_l(x)$ and some unknown degrees of freedom $\hat{w}^n_{i,l}$,

$$w_h(x, t^n) = \sum_l \hat{w}^n_{i,l}\Phi_l(x) \coloneqq \hat{w}^n_{i,l}\Phi_l(x), \qquad \text{for } x \in I_i, \quad (2.3.3)$$

where the Einstein summation convention over two repeated indices has been considered. The spatial basis functions are defined on the reference interval $[0, 1]$ and the transformation from physical coordinates $x \in I_i$ to reference coordinates $\xi \in [0, 1]$ is given by the

UNIVERSIDAD DE MÁLAGA

linear mapping $x = x(\xi) = x_i - \dfrac{\Delta x}{2} + \xi \Delta x$. With this choice, the spatial basis functions is written in terms of the nodal basis function $\varphi_k(\xi)$, which satisfy the interpolation property $\varphi_k(\xi_j) = \delta_{kj}$, where δ_{kj} is the usual Kronecker symbol, ξ_j are the nodal quadrature points, and the resulting basis is by construction orthogonal. Therefore, we can write,

$$\Phi_k(x) = \varphi_k(\xi).$$

Furthermore, due to this particular choice of a nodal basis, all integral operators can be decomposed into a sequence of one-dimensional operators acting only on the $N_p + 1$ degrees of freedom in each dimension.

The DG method results then from multiplying the system of PDE (2.3.1) with a test function $\Phi_k \in \mathcal{U}_h$ and integrating over the space control volume I_i. This leads to the following semi-discrete weak formulation where we seek to find $w_h(x, t^{n+1}) \in \mathcal{U}_h$ such that, for all I_i,

$$\int_{I_i} \Phi_k \frac{d}{dt} w_h \, dx + \int_{I_i} \Phi_k \left(\partial_x F(w_h) + B(w_h) \partial_x w_h \right) \, dx = \int_{I_i} \Phi_k G(w_h) \sigma_x \, dx, \tag{2.3.4}$$

holds for all test functions $\Phi_k(x) \in \mathcal{U}_h$. Expression (2.3.4) can be integrated by parts to obtain a numerical scheme where the numerical flux at the intercells appears explicitly,

$$\begin{aligned} \int_{I_i} \Phi_k \Phi_l \frac{d}{dt} \hat{w}_{i,l} \, dx - \int_{I_i^\circ} \Phi_k' F(w_h) \, dx \\ + \Phi_{k,i+\frac{1}{2}} D^-_{i+\frac{1}{2}} \left(w^-_{h,i+\frac{1}{2}}, w^+_{h,i+\frac{1}{2}} \right) - \Phi_{k,i-\frac{1}{2}} D^+_{i-\frac{1}{2}} \left(w^-_{h,i-\frac{1}{2}}, w^+_{h,i-\frac{1}{2}} \right) \\ + \int_{I_i^\circ} \Phi_k \left(B(w_h) \partial_x w_h \right) \, dx = \int_{I_i^\circ} \Phi_k G(w_h) \sigma_x \, dx. \end{aligned} \tag{2.3.5}$$

Here, $f^\pm$ denotes the evaluation of f in the cell interfaces $i \pm \frac{1}{2}$. Note that the spatial dependency of Φ_k has been dropped in order to make the notation less cumbersome and that $\Phi_{k,i\pm\frac{1}{2}}$ denotes the evaluation of the spatial basis at the element interface. Likewise, I_i° denotes the interior of the cell I_i. In the DG framework, the discrete solution at each cell I_i, w_h, is allowed to jump across element interfaces, naturally leading to the appearance of a numerical flux associated with this discontinuity across the cell interfaces. As usual in the DG framework, this is achieved via numerical flux functions in the form of approximate Riemann solvers. These numerical fluxes are denoted by $D^\pm_{i\pm\frac{1}{2}}$ and in this work they will be defined using the path-conservative approach explained in Section 2.2 within the framework of finite volume methods. The extension of these techniques to the discontinuous Galerkin finite element framework dates back to the work in [151–153].

The semi discrete numerical scheme (2.3.5) admits several ways of time discretizations. The methods of lines may be considered, thus integrating the system in space and

UNIVERSIDAD DE MÁLAGA

obtaining a system of ODEs that can be subsequently integrated using classical Runge-Kutta methods, for instance. This is the case for the classical Runge-Kutta DG schemes in [190], where only a weak form in space of the governing PDE system is obtained, while time is kept continuous. Another approach consist on one-step, fully discrete version of (2.3.5) which can be derived under the ADER-DG framework. We proceed now to give a brief description on both methods. The reader interested in further information can refer to [53–56, 92, 179] for DG Runge-Kutta methods or [78–82, 85] for ADER-DG methods.

2.3.1 TVD Runge-Kutta time discretization

The semi-discrete numerical scheme (2.3.5) can be written in the following form,

$$\partial_t \hat{w}_{i,l}(t) = L(\hat{w}_{i,l}(t)), \tag{2.3.6}$$

where $L(w)$ is a spatial discretization operator. Expression (2.3.6) is a ODE system that can be evolved explicitly in time by any ODE solver. For instance, an explicit total variation diminishing (TVD) [178, 179] scheme can be considered, up to the desired order. An example of this is given by the third order Runge-Kutta scheme, that reads,

$$\begin{aligned}
\hat{w}_{i,l}^{(1)} &= L\left(\hat{w}_{i,l}^{n}\right), \\
\hat{w}_{i,l}^{(2)} &= L\left(\hat{w}_{i,l}^{n} + \frac{\Delta t}{2}\hat{w}_{i,l}^{(1)}\right), \\
\hat{w}_{i,l}^{(3)} &= L\left(\hat{w}_{i,l}^{n} - \Delta t\hat{w}_{i,l}^{(1)} + 2\Delta t\hat{w}_{i,l}^{(2)}\right), \\
\hat{w}_{i,l}^{n+1} &= \hat{w}_{i,l}^{n} + \frac{\Delta t}{6}\left(\hat{w}_{i,l}^{(1)} + 4\hat{w}_{i,l}^{(2)} + \hat{w}_{i,l}^{(3)}\right).
\end{aligned} \tag{2.3.7}$$

2.3.2 ADER time discretization

It is possible to achieve high order in both time and space in a singe step thanks to an unlimited high order accurate (ADER) technique. The ADER method was first introduced by Toro and Titarev in the finite volume method framework in a series of papers [69–72] and later extended to DG methods (see [62, 74, 75]). It has been successfully applied to solve hyperbolic systems of conservation laws in [78–82, 85], for instance. The ADER approach is based on the approximated solution of Riemann problems by means of a fixed point algorithm in each element locally. This solution $q_h(x,t)$, called the predictor solution, is then used to compute (2.3.5) with the desired accuracy and it is based on a weak formulation of the governing system (2.3.1) *in the small*, i.e. without considering the interaction with the neighbour elements. For an element I_i, the predictor solution q_h is now expanded in terms of a local space-time basis,

$$q_h(x,t) = \sum_l \theta_l(x,t)\hat{q}_l^i \coloneqq \theta_l(x,t)\hat{q}_l^i, \tag{2.3.8}$$

UNIVERSIDAD DE MÁLAGA

with the multi-index $l = (l_0, l_1)$ and where again the Einstein notation has been used. The space-time basis functions $\theta_l(x,t) = \varphi_{l_0}(\tau)\varphi_{l_1}(\xi)$ are again generated from the same one-dimensional nodal basis functions $\varphi_k(\xi)$ as before, i.e. the Lagrange interpolation polynomials of degree N_p passing through N_p+1 Gauss-Legendre quadrature nodes. The spatial mapping $x = x(\xi)$ is also the same as before and the physical time is mapped to the reference time $\tau \in [0,1]$ via $t = t^n + \tau\Delta t$. The multiplication of the PDE system (2.3.1) with a test function θ_k and integration over the space-time control volume $I_i \times [t^n, t^{n+1}]$ yields the following weak form of the governing PDE, which we remark that it is different from (2.3.5), since now the test and basis functions are both space and time dependent:

$$\int_{t^n}^{t^{n+1}} \int_{I_i} \theta_k(x,t)\partial_t q_h \, dxdt + \int_{t^n}^{t^{n+1}} \int_{I_i} \theta_k(x,t) \left(\partial_x F(q_h) + B(q_h)\partial_x q_h\right) \, dxdt = \int_{t^n}^{t^{n+1}} \int_{I_i} \theta_k(x,t) G(q_h)\sigma_x \, dxdt. \tag{2.3.9}$$

Since we are interested in a local expression, without any interactions with the neighbor elements, the jump terms associated with the discontinuities at the cell interfaces are not taken into account at this stage. Instead, they will be accounted for in the final corrector step of the ADER-DG method. In this way, we can compute the first integral term of the prior expression to obtain:

$$\int_{I_i} \theta_k(x,t^{n+1}) q_h(x,t^{n+1}) \, dx - \int_{I_i} \theta_k(x,t^n) q_h^0(x,t^n) \, dx - \int_{t^n}^{t^{n+1}} \int_{I_i} \partial_t \theta_k(x,t) q_h(x,t) \, dxdt = -\int_{t^n}^{t^{n+1}} \int_{I_i} \theta_k(x,t) \left(\partial_x F(q_h) + B(q_h)\partial_x q_h - G(q_h)\sigma_x\right) \, dxdt. \tag{2.3.10}$$

Using the local space-time ansatz (2.3.8), equation (2.3.10) becomes a local nonlinear system for the unknown degrees of freedom $\hat{q}_l^i$ of (2.3.8) that can be solved via a simple and fast converging fixed point iteration algorithm detailed in [74, 191]. The convergence of such an algorithm was proved in [76]. Particularly, for linear homogeneous systems, the iteration converges in a finite number of, at most, $N_p + 1$ steps since the associated iteration matrix is nilpotent [192].

As in many iterative algorithm, the choice of the initial guess $q_h^0(x,t)$ for the predictor algorithm has a huge impact in the convergence rate and therefore the global computational efficiency of the scheme. In order to speed up the algorithm, several strategies exist. For instance, in [191] the authors propose a second-order accurate MUSCL-Hancock-type approach based on discrete derivatives computed at time t^n to compute the initial guess. Another strategy is detailed in [193], where using an extrapolation of q_h from the previous time interval $[t^{n-1}, t^n]$ is suggested. Alternatively, it is possible to better approximate the initial guess by means of a Taylor series expansion of

UNIVERSIDAD DE MÁLAGA

the solution $w_h(x,t^n)$, and then use a continuous extension Runge-Kutta scheme (CERK), as it is discussed at [194]. For more details, one may refer to [195, 196].

In fact, in [75] was proved that if a polynomial of degree N_p-1 is chosen, it is sufficient to use a single Picard iteration to solve (2.3.10) to the desired accuracy. For an efficient task-based formalism of ADER-DG schemes, see [197]. In the simulations of this thesis, we consider the initial guess to be $q_h^0(x,t) = w_h(x,t^n)$.

Finally, the predictor solution of the ADER problem $q_h(x,t)$ is used to achieve high order in space and time. The discrete version of equation (2.3.5) would then yield,

$$\begin{aligned}\int_{I_i} \Phi_k \Phi_l (\hat{w}_{i,l}^{n+1} - \hat{w}_{i,l}^{n})\, dxdt - \int_{t^n}^{t^{n+1}} \int_{I_i^\circ} \Phi_k' F(q_h)\, dxdt \\ + \int_{t^n}^{t^{n+1}} \left(\Phi_{k,i+\frac{1}{2}} D^-_{i+\frac{1}{2}} \left(q^-_{h,i+\frac{1}{2}}, q^+_{h,i+\frac{1}{2}} \right) + \Phi_{k,i-\frac{1}{2}} D^+_{i-\frac{1}{2}} \left(q^-_{h,i-\frac{1}{2}}, q^+_{h,i-\frac{1}{2}} \right) \right) dt \\ + \int_{t^n}^{t^{n+1}} \int_{I_i^\circ} \Phi_k B(q_h) \partial_x q_h \, dxdt = \int_{t^n}^{t^{n+1}} \int_{I_i} \Phi_k G(q_h) \sigma_x \, dxdt. \end{aligned} \quad (2.3.11)$$

2.3.2.1 Time step restriction

The DG method suffers from a severe Courant-Friedrichs-Lewy (CFL) number that decreases with the order of the approximation polynomial, N_p. The expression for the time step restriction for the DG method in multiple dimensions d is,

$$\Delta t \leq \frac{1}{d} \frac{1}{2N_p+1} \frac{\Delta x}{|\lambda_{max}|}, \quad (2.3.12)$$

where $|\lambda_{max}|$ is an approximation of the maximum wave speed and it will depend on the target hyperbolic system (2.3.1). However, this severe restriction has to be seen with perspective. While it is true that it imposes a small time step, the enhanced subcell resolution of the DG scheme allows for very coarse meshes, compensating the strict $\frac{1}{2N_p+1}$ restriction.

2.3.3 Limiting procedure

In this section we present two limiting procedures for the DG method used in this thesis that follow completely different strategies in order to avoid spurious oscillations associated with systems of conservation laws near strong gradients or discontinuities. Indeed, the DG schemes discussed so far are unlimited, in the sense that there is no mechanism to prevent the appearance of Gibbs oscillations near discontinuities. The two limiting strategies that will be discussed act in two steps: first they detect which cells need limiting, and then they introduce some kind of numerical viscosity into the solution in these regions. In any case, the two methods relies on different strategies to limit the solution. Particularly, the

UNIVERSIDAD DE MÁLAGA

MOOD strategies relies on an adequate switching to a robust finite volume method, while the WENO strategies focus on introducing some weights to purge spurious oscillations.

2.3.3.1 MOOD

The MOOD (Multi-dimension Optimal Order Detection) [91] strategy acts after the unlimited solution of (2.3.5) (or (2.3.11)) is computed. This solutions is now considered as a candidate solution $w_h^c(x, t^{n+1})$ and it will remain unchanged if it is deemed acceptable, while it will be overridden if it fails to fulfill some suitability criteria. Cells not considered suitable are denominated troubled cells. In these cells, the solution will be computed again using an explicit second order accurate MUSCL-Hancock finite volume method. To achieve this, the unlimited solution of the numerical scheme $w_h(x, t^n)$ is projected into a subgrid of K_s elements in I_i, denoted by $\mathcal{S}_{i,j}$, and verifying $I_i = \bigcup_j \mathcal{S}_{i,j}$ for $j = 1, \ldots, K_s$. In fact, the projection consists of a set of piecewise constant subcell averages that can be seen as an alternative data representation, denoted by $v_h(x, t^n)$. These averages are a L_2 projection that preserves the mean of $w_h(x, t^n)$ in $\mathcal{S}_{i,j}$,

$$v_h(x, t^n) = \frac{1}{|\mathcal{S}_{i,j}|} \int_{\mathcal{S}_{i,j}} w_h(x, t^n)\, dx, \quad \forall x \in \mathcal{S}_{i,j} \subset I_i. \tag{2.3.13}$$

These subcell averages $v_h(x, t^n)$ are evolved by means of an explicit second order finite volume solver. This solver is expected to be robust and able to preserve the desired physical and numerical properties. Additionally, the limiting tools available for finite volume solvers, like the one presented in this thesis, can be used. Once the subcell averages are evolved in time, the solution $v_h(x, t^{n+1})$ has to be reconstructed back into a polynomial for the DG solver. This is accomplished through a classical least square reconstruction operator that preserves the average of the projected solutions,

$$\int_{\mathcal{S}_{i,j}} w_h(x, t^{n+1})\, dx = \int_{\mathcal{S}_{i,j}} v_h(x, t^{n+1})\, dx, \qquad \mathcal{S}_{i,j} \subset I_i. \tag{2.3.14}$$

Since a subcell resolution $K_s > N_p + 1$ is admitted, this problem may be overdetermined. Hence, the following constraint must also be considered,

$$\int_{I_i} w_h(x, t^{n+1})\, dx = \int_{I_i} v_h(x, t^{n+1})\, dx. \tag{2.3.15}$$

Finally, the final limited solution $w_h(x, t^{n+1})$ is assembled from the candidate solution in regions where the solutions is considered adequate and the reconstructed solution in regions where it was deemed necessary. In any case, the finite volume solution $v_h(x, t^{n+1})$ is kept in thesis in case that its cell is deemed as troubled in the next time step. If this is the case, then the finite volume solver will use $v_h(x, t^{n+1})$ as initial data instead of the projected polynomial from (2.3.14).

UNIVERSIDAD DE MÁLAGA

The main advantage of this approach to limitation is that it allows for any number of physical and numerical properties to be verified. Indeed, the nature of the limiter, which evaluates the suitability of the candidate solution $w_h^c(x, t^{n+1})$ *a posteriori*, allows to consider the cell as troubled not only in the presence of discontinuities, but also if the solution does not fulfill some prescribed properties. For instance, it is common to ensure the positivity of the column of water in the context of shallow-water equations, or check that there are not any NaN present at the solution. Additionally, a relaxed discrete maximum principle can be used to detect discontinuities *a posteriori*, according to [91]. In this way, at each time step the following expression is considered,

$$\min_{y\in\mathcal{V}_i}(v_h(y,t^n)) - \delta \leq v_h(x,t^{n+1}) \leq \max_{y\in\mathcal{V}_i}(v_h(y,t^n)) + \delta, \quad \forall x \in I_i, \tag{2.3.16}$$

where the projection $v_h(x, t^{n+1})$ is used as a discrete form of the polynomial $w_h(x, t^{n+1})$, $\mathcal{V}_i$ is a set containing I_i and its neighbor cells and δ is a small value that relaxes the criteria to allow some very small overshoot or undershoot and avoid roundoff errors that would arise if (2.3.16) is applied strictly.

Furthermore, the number of subcells that constitute the subgrid K_s has to be adequately chosen. According to [91], the optimal choice is the subgrid satisfying $K_s = 2N_p + 1$. This choice is optimal in the sense that it allows to keep the same time step calculated for the DG polynomial and also to have a CFL number close to the theoretical maximum for the finite volume numerical scheme. Finally, a carefully implementation of the numerical flux between a troubled and non-troubled cell must also be considered. Indeed, the non-troubled cell has been computed with a numerical flux that is no longer valid since the numerical scheme applied in the troubled cell has effectively changed. Thus, it is important to update the flux in the non-troubled cell to be consistent with the flux calculated in the troubled cell, and keep intact the conservation properties of the numerical scheme.

2.3.3.2 Weighted essentially non-oscillatory (WENO) limiter

Another alternative to limiting for discontinuous Galerkin methods is an extension of the weighted essentially non-oscillatory limiters, developed mainly in the finite volume framework, to the Discontinuous Galerkin numerical schemes. This procedure is based on the work by Zhong and Shu in [95], where the authors describe this type of limiters for Runge–Kutta DG methods. As before, the limiter can be viewed as a two-step procedure: in a first step the troubled cells are identified and then the proper limiter is applied by introducing some numerical diffusion to the polynomial solution.

Detector of troubled cells There are several ways of detecting that a cell requires limiting (see for instance [95] and references therein). In this thesis, the TVD minmod

UNIVERSIDAD DE MÁLAGA

limiter [56] is used, and it is defined as follows. We first define,

$$\bar{w}_i = \frac{1}{\Delta x}\int_{I_i} w_h(x,t^n)\,dx \tag{2.3.17}$$

and,

$$\widehat{w}_R = w^-_{i+\frac{1}{2}} - \bar{w}_i, \qquad \widehat{w}_L = \bar{w}_i - w^+_{i-\frac{1}{2}}, \tag{2.3.18}$$

where $w^{\pm}_{i\mp\frac{1}{2}}$ is the evaluation of the solution $w_h(x,t^n)$ at the cell interfaces. Then, the minmod is applied as follows:

$$\sigma_R = \text{minmod}\left(\widehat{w}_R, \Delta^+\bar{w}_i, \Delta^-\bar{w}_i\right), \qquad \sigma_L = \text{minmod}\left(\widehat{w}_L, \Delta^+\bar{w}_i, \Delta^-\bar{w}_i\right), \tag{2.3.19}$$

with,

$$\Delta^+\bar{w}_i = \bar{w}_{i+1} - \bar{w}_i, \qquad \Delta^-\bar{w}_i = \bar{w}_i - \bar{w}_{i-1}, \tag{2.3.20}$$

and the following minmod functions,

$$\text{minmod}(a,b,c) = \begin{cases} a & \text{if } |a| \le M\Delta x, \\ \text{m}(a,b,c) & \text{otherwise,} \end{cases} \tag{2.3.21}$$

and,

$$\text{m}(a,b,c) = \begin{cases} \text{Sign}(a)\min(|a|,|b|,|c|) & \text{if } \text{sign}(a) = \text{sign}(b) = \text{sign}(c), \\ 0 & \text{otherwise.} \end{cases} \tag{2.3.22}$$

Here, M is a TVD parameter to be chosen properly for each target PDE system. A cell is marked as troubled if $\sigma_R \neq \widehat{w}_R$ or $\sigma_L \neq \widehat{w}_L$.

WENO limiter Once the troubled cells have been detected in the previous step, the objective consists now on reconstructing a new polynomial solution on the troubled cell that is a convex combination of the polynomial of the troubled cell and its neighbors. In this way, spurious oscillations are supposed to be prevented. Unlike the WENO strategy at [95], here a preliminary step is taken where, if the cell is deemed as troubled, the polynomial of order $N_p > 1$ in $I_j,\ j \in \{i-i, i, i+1\}$ will be projected into a first order polynomial with the following form:

$$w_{p,j}(x) = \frac{1}{\Delta x}\int_{I_j} w_{h,j}(x)\,dx + (x - x_{j,c})\frac{1}{\Delta x}\int_{I_j}\frac{\partial}{\partial_x} w_{h,j}(x)\,dx. \tag{2.3.23}$$

Note that this polynomial preserves the cell average of $w_{i,h}(x)$. In (2.3.23), the subindex p denotes the projected polynomial, while $x_{j,c}$ stands for the barycenter of the cell I_j and the time notation has been dropped in order to simplify the notation. It is important to

UNIVERSIDAD DE MÁLAGA

note that the global order of $w_{p,i}(x)$ is still N_p, it has not been reduced to a first order polynomial.

The projected polynomial is used to rewrite the solution $w_{i,h}(x,t^n)$ on the cell I_i as a convex combination as follows:

$$w_{i,h}^{(\text{new})}(x) = \omega_{i-1}\overline{w}_{p,i-1}(x) + \omega_i w_{p,i}(x) + \omega_{i+1}\overline{w}_{p,i+1}(x), \quad x \in I_i, \tag{2.3.24}$$

with,

$$\begin{aligned}
\overline{w}_{p,i-1}(x) &= w_{p,i-1}(x) - \frac{1}{\Delta x}\int_{I_i} w_{p,i-1}(x)\,dx + \frac{1}{\Delta x}\int_{I_i} w_{p,i}(x)\,dx,\\
\overline{w}_{p,i+1}(x) &= w_{p,i+1}(x) - \frac{1}{\Delta x}\int_{I_i} w_{p,i+1}(x)\,dx + \frac{1}{\Delta x}\int_{I_i} w_{p,i}(x)\,dx.
\end{aligned} \tag{2.3.25}$$

In this way, $w_{i,h}^{(\text{new})}(x)$ has the same cell average as $w_{i,h}(x)$.

The normalized weights for the WENO convex combination in (2.3.24) follows the classical procedure for WENO limiters and are defined as follows:

$$\omega_l = \frac{\bar{\omega}_l}{\sum_s \omega_s}, \tag{2.3.26}$$

with,

$$\omega_l = \frac{\gamma_l}{(\epsilon + \beta_l)^r}, \tag{2.3.27}$$

where γ_l are some user defined weights. Likewise, r and ϵ are user defined parameters. Typically, $\gamma_i = 0.998$ for the element I_i and $\gamma_l = 0.001$ for the neighbors elements, while some reasonable parameters for r and ϵ are $\epsilon = 10^{-6}$ and $r = 2$. Finally, the smooth indicators β_l are defined as,

$$\beta_l = \sum_{s=1}^{2}\int_{I_i} \Delta x^{2s-1}\left(\frac{\partial^s}{\partial x^s} w_{p,l}(x)\right)^2 dx. \tag{2.3.28}$$

The general limiting procedure (2.3.23)-(2.3.28) is applied once per time step for either Runge-Kutta or ADER time discretization. However, it can yield unsatisfactory results at very high order. A possibly strategy to address this issue consists on applying the limiter two consecutive times. The first time, the procedure (2.3.23)-(2.3.28) is used to produce a candidate solution $w_h^c(x)$. This candidate solution is evaluated once more by the WENO detector described in Subsection 2.3.3.2 and it remains unchanged if the cell is not deemed as troubled. However, if the element is considered troubled, then the limiting strategy (2.3.23)-(2.3.28) is repeated with the following projection operator:

$$w_{p,i}(x) = \frac{1}{\Delta x}\int_{I_i} w_{i,h}(x)\,dx. \tag{2.3.29}$$

Additionally, a reduction of the user defined parameter r can be considered. This will ensure a smoother discrete solution for highly oscillating regions near strong discontinuities.

UNIVERSIDAD DE MÁLAGA

2.3.4 Well-balanced discontinuous Galerkin methods

We proceed now to detail a well-balanced technique for the Runge-Kutta DG methods (2.3.5)-(2.3.7) and the ADER-DG method (2.3.10)-(2.3.11). Likewise, a strategy to preserve stationary solutions even when the solution needs limiting will also be discussed. The first step is, in fact, very similar to the procedure for finite volume methods in Section 2.2.5. First, a stationary solution $w_i^*(x,t^n),\ x\in I_i$, is computed locally for each cell. As before, the stationary solution is computed at each time step t^n, hence the time dependence. However, time notation will be subsequently dropped in order to simplify the notation. The stationary solution is a solution of the following minimization problem,

$$\begin{cases} \partial_x F(w_i^*) + B(w_i^*)\partial_x w_i^* = G(w_i^*)\sigma_x, \quad x\in I_i, & (2.3.30)\\ \dfrac{1}{\Delta x}\displaystyle\int_{I_i} w_i^*(x)\,dx = \dfrac{1}{\Delta x}\int_{I_i} w_h(x,t^n)\,dx, & (2.3.31)\\ \text{that minimizes } \displaystyle\int_{I_i}(w_i^*(x)-w_h(x,t^n))^2\,dx. & (2.3.32) \end{cases}$$

As in the finite volume case in Section 2.2.5, this problem is in general difficult to solve exactly. However, if the stationary solution $w_i^*(x)$ depends on a set of parameters, the previous systems is generally reduced to a nonlinear system of equations if the integrals in (2.3.31) are approximated by some quadrature formula defined in terms of the Gaussian nodal points. Moreover, in some situations, like the examples (2.2.91) or (2.2.93), equations (2.3.30) and (2.3.31) are enough to determine uniquely the local stationary solution.

Furthermore, we introduce the notation $w_{i,h}^*(x)$ for the projection of $w_h^*(x)$ onto $\mathcal{U}_h\ x\in I_i$,

$$w_{i,h}^*(x) = \sum_l \hat{w}_{i,l}^*\Phi_l(x) \coloneqq \hat{w}_{i,l}^*\Phi_l(x), \quad \text{for } x\in I_i. \tag{2.3.33}$$

As it was the case for the finite volume method, the fluctuation of the solution and the stationary solution is now considered with the following polynomial,

$$\tilde{w}_h(x) = w_h(x) - w_{i,h}^*(x), \quad x\in I_i, \tag{2.3.34}$$

where the time notation t^n has been dropped for simplicity.

A well-balanced DG numerical method must then ensure that, when $w_h(x,t^n) = w_{i,h}^*(x)$, the solution $w_h(x,t^{n+1})$ remains unchanged. This is achieved by considering

UNIVERSIDAD DE MÁLAGA

the following version of (2.3.5):

$$\int_{I_i} \Phi_k \Phi_l \frac{d}{dt}\hat{w}_{i,l}\, dxdt - \int_{I_i^\circ} \Phi_k' \Big(F(w_h) - F(w_{i,h}^*) \Big)\, dxdt + \Phi_{k,i+\frac{1}{2}} \Big(\tilde{D}^-_{i+\frac{1}{2}} - F(w_i^*(x_{i+\frac{1}{2}})) \Big) - \Phi_{k,i-\frac{1}{2}} \Big(\tilde{D}^+_{i-\frac{1}{2}} - F(w_i^*(x_{i-\frac{1}{2}})) \Big) + \int_{I_i^\circ} \Phi_k \Big(B(w_h)\partial_x w_h - B(w_{i,h}^*)\partial_x w_{i,h}^* \Big)\, dxdt = \int_{I_i} \Phi_k \Big(G(w_h) - G(w_{i,h}^*) \Big) \sigma_x\, dxdt, \tag{2.3.35}$$

where $\tilde{D}^{\pm}_{i+\frac{1}{2}}$ stand for the numerical flux applied to the reconstructed values

$$\tilde{D}^{\pm}_{i+\frac{1}{2}} = \tilde{D}^{\pm}_{i+\frac{1}{2}}(w_i^*(x_{i+\frac{1}{2}}) + \tilde{w}^-_{h,i+\frac{1}{2}}, \sigma_{i+\frac{1}{2}}, w_i^*(x_{i+\frac{1}{2}}) + \tilde{w}^+_{h,i+\frac{1}{2}}, \sigma_{i+\frac{1}{2}}). \tag{2.3.36}$$

Note that in (2.3.36) the solution is not directly evaluated at the cell interface. Instead, the fluctuation $\tilde{w}_h(x)$ is evaluated and the final state for the Riemann solver is obtained as the sum of this value and the evaluation of the stationary solution at the cell interface,

$$w_i^*(x_{i+\frac{1}{2}}) + \tilde{w}^-_{h,i+\frac{1}{2}}, \qquad w_i^*(x_{i+\frac{1}{2}}) + \tilde{w}^+_{h,i+\frac{1}{2}},$$

where,

$$\tilde{w}^-_{h,i+\frac{1}{2}} = \tilde{w}_{h|_{I_i}}(x_{i+\frac{1}{2}}), \qquad \tilde{w}^+_{h,i+\frac{1}{2}} = \tilde{w}_{h|_{I_{i+1}}}(x_{i+\frac{1}{2}}).$$

Additionally, we stress that the solution of system (2.3.35) is the same as (2.3.5) up to the order of the method since the added terms add zero by definition. However, it is easy to see that when $w_h(x,t^n) = w_{i,h}^*(x)$, expression (2.3.35) will be zero up to machine precision. We remark that $w_h(x,t^n) = w_{i,h}^*(x)$ is not enough to make the DG method (2.3.5) well-balanced since all the integrals have to be approximated by quadrature formulas which may destroy the well-balanced character of the scheme. However, the numerical method (2.3.35) does not suffer from this problem.

This finishes the description of the well-balanced technique for the Discontinuous Galerkin method. If (2.3.35) is discretized in time using a Runge-Kutta method, then no special steps must be taken: if the spatial discretization operator L in (2.3.6) is already exactly well-balanced, then the convex combination of the Runge-Kutta scheme will maintain this property. However, if an ADER-DG approach is used, it is necessary to ensure that the predictor solution $q_h(x,t^n)$ does not destroy the well-balanced properties of the DG method. We proceed now to describe a solution to this problem.

2.3.4.1 Well-balanced ADER method

The approach to make the ADER method well-balanced is in fact very similar to the described method for the DG scheme. We seek now that, if $w_h(x,t^n)$ is a stationary

UNIVERSIDAD DE MÁLAGA

solution, then $q_h(x,t^n)$ remains unchanged. Therefore, in order to accomplish this, the predictor algorithm (2.3.10) is modified as follows,

$$\int_{I_i}\theta_k(x,t^{n+1})\tilde{q}_h(x,t^{n+1})\,dx-\int_{I_i}\theta_k(x,t^n)\tilde{q}_h^0(x,t^n)\,dx-\int_{t^n}^{t^{n+1}}\int_{I_i}\partial_t\theta_k(x,t)\tilde{q}_h(x,t)\,dxdt$$
$$=-\int_{t^n}^{t^{n+1}}\int_{I_i}\theta_k(x,t)\left(\left(\partial_xF(q_h)-\partial_xF(w_{i,h}^*)\right)+\left(B(q_h)\partial_xq_h-B(w_{i,h}^*)\partial_xw_{i,h}^*\right)\right)dxdt$$
$$+\int_{t^n}^{t^{n+1}}\int_{I_i}\theta_k(x,t)\left(G(q_h)-G(w_{i,h}^*)\right)\sigma_x\,dxdt,\quad(2.3.37)$$

where,

$$\tilde{q}_h(x,t)=q_h(x,t)-w_{i,h}^*(x),\quad x\in I_i.\quad(2.3.38)$$

Therefore, the final step to recuperate a high order approximation of the solution $w_h(x,t^{n+1})$ is,

$$q_h(x,t)=\tilde{q}_h(x,t)+w_h^*(x),\quad x\in I_i.\quad(2.3.39)$$

Note that the initial condition $\tilde{q}_h^0(x,t^n)$ is also a fluctuation,

$$\tilde{q}_h^0(x,t^n)=w_h(x,t^n)-w_{i,h}^*(x,t^n),\quad x\in I_i.$$

In this way, we are computing a high order approximation of the fluctuation of the solution with respect to the stationary solution, rather than the solution itself. If $w_h(x,t)$ is a stationary solution, then (2.3.38) will be zero and the ADER procedure (2.3.37) will result exactly zero. The stationary solution would be then recuperated thanks to (2.3.39). Finally, this high order approximation $q_h(x,t^n)$ will be used in the DG method (2.3.35) to obtain an exactly well-balanced arbitrary high order in time and space fully discrete ADER-DG numerical scheme.

Finally, we stress that, although the techniques here have been described for the particular case of one dimensional problems, all methods can be extended in a dimension by dimension way for the multidimensional case.

2.3.5 Well-balanced limiting procedure

All limiting strategies described in this thesis can be adapted to preserve stationary solutions. Observe that if we are only interested on smooth stationary solutions, the limiter should not be active in their presence. However, roundoff errors or local *extrema* can accidentally activate the limiter and, in these cases, it is desirable that the limiter does not destroy the well-balanced property of the numerical scheme.

In the case of the MOOD limiter, the adaption to a well-balanced limiter is straightforward: since it is based in a switching from a DG numerical scheme to a finite volume one, then it is a matter of applying all the well-balanced tools in Section 2.2 to

UNIVERSIDAD DE MÁLAGA

the chosen finite volume solver. Special care should be taken with the projection and reconstruction operator so that they do not destroy the well-balanced property. In the case of the WENO limiter, considering the fluctuation with respect to the stationary solution can ensure the well-balanced property of the limiter. In this way, the projection of the polynomial (2.3.23) in I_j $j \in \{i-1, i, i+1\}$ is now with respect to the fluctuation,

$$\tilde{w}_{p,j}(x) = \frac{1}{\Delta x}\int_{I_j} \tilde{w}_{h,j}(x)\,dx + (x - x_{j,c})\frac{1}{\Delta x}\int_{I_j} \frac{\partial}{\partial_x}\tilde{w}_{h,j}(x)\,dx, \tag{2.3.40}$$

denoted by $\tilde{w}_h$.

Likewise, the limited fluctuation $\tilde{w}_{i,h}(x, t^n)$ on the element I_i is written as,

$$\tilde{w}_{i,h}^{(\text{new})}(x) = \omega_{i-1}\overline{\tilde{w}}_{p,i-1}(x) + \omega_i\tilde{w}_{p,i}(x) + \omega_{i+1}\overline{\tilde{w}}_{p,i+1}(x), \tag{2.3.41}$$

with,

$$\begin{aligned}
\overline{\tilde{w}}_{p,i-1}(x) &= \tilde{w}_{p,i-1}(x) - \frac{1}{\Delta x}\int_{I_i} \tilde{w}_{p,i-1}(x)\,dx + \frac{1}{\Delta x}\int_{I_i} \tilde{w}_{p,i}(x)\,dx,\\
\overline{\tilde{w}}_{p,i+1}(x) &= \tilde{w}_{p,i+1}(x) - \frac{1}{\Delta x}\int_{I_i} \tilde{w}_{p,i+1}(x)\,dx + \frac{1}{\Delta x}\int_{I_i} \tilde{w}_{p,i}(x)\,dx.
\end{aligned} \tag{2.3.42}$$

Since we have made use of the fluctuation to calculate the new limited solution, the final limited solution can be recuperated with,

$$w_{i,h}^{(\text{new})}(x) = w_{i,h}^{*}(x) + \tilde{w}_{i,h}^{(\text{new})}(x). \tag{2.3.43}$$

The normalized weights for the WENO convex combination in (2.3.41) are actually the same that the ones proposed at (2.3.26) and (2.3.27), save for the smooth indicator β_l which is now defined in terms of the fluctuation as,

$$\beta_l = \sum_{s=1}^{2}\int_{I_i} \Delta x^{2s-1}\left(\frac{\partial^s}{\partial x^s}\tilde{w}_{p,l}(x)\right)^2 dx. \tag{2.3.44}$$

Note that the use of the fluctuation $\tilde{w}_h(x)$ will ensure that the stationary solution is preserved, since the procedure (2.3.40)-(2.3.43) is clearly exact for the null operator.

2.3.6 Well-balanced discontinuous Galerkin methods: examples

Since the well-balanced methods for the discontinuous Galerkin numerical scheme is a novel proposal of this thesis, it is pertinent to include some numerical tests where the well-balanced properties and the well-balanced limiter are shown. In particular, simulations with the shallow-water model, with the Burgers' equation and the Euler system are included. Additionally, a simulation far from the *equilibria* is also considered to show

UNIVERSIDAD DE MÁLAGA

the capabilities of the limiter. For the sake of completeness, both the Runge-Kutta and ADER time discretization are considered.

Unless stated otherwise, for all simulations the CFL condition is set to 0.9 and the limiter parameters are set to $M = 1$, $r = 2$ for the candidate solution and $M = 1$ and $r = 0.25$ if the WENO limiter is computed twice. Finally, $\gamma_c = 0.998$ for the central cell and $\gamma = 0.001$ for the neighbors cells.

2.3.6.1 Burgers' equation

The one dimensional scalar law Burgers' equation with source terms is,

$$\partial_t u + \partial_x \left(\frac{u^2}{2} \right) = u^2 \sigma_x. \tag{2.3.45}$$

The stationary solutions of this equation are given by,

$$u(x) = C e^{\sigma(x)}, \quad C \in \mathbb{R}, \tag{2.3.46}$$

where $\sigma(x)$ is a known continuous function.

Preservation of a stationary solution We want to ensure that a stationary solution of the form (2.3.46) is adequately preserved. To achieve this, the following initial condition is considered,

$$u(x) = e^{\sigma(x)}, \quad \sigma(x) = x. \tag{2.3.47}$$

The computational domain is $I = [-1, 1]$ and Dirichlet boundary condition are imposed at both boundaries. Figure 2.5 and Table 2.1 show the results of the experiment at final time $t = 10$ s. As it can be seen, only the well-balanced schemes are able to preserve the stationary solution.

Perturbed stationary solutions We consider now a perturbation of the stationary solution (2.3.47) given by

$$u = e^{\sigma(x)} + 0.3\, e^{-200(x+0.5)^2}, \quad \sigma(x) = x. \tag{2.3.48}$$

The computational domain $I = [-1, 1]$ is discretized with a mesh of $N_s = 100$ elements. Dirichlet boundary conditions are set on the left boundary while free-flow boundary conditions are set on the right one. The final time $t = 10$ s can be seen in Figure 2.6 for the fourth order ADER and Runge-Kutta numerical scheme. As it can be seen, once the perturbation has left the domain, only the well-balanced schemes are able to recuperate the stationary solution in contrast with the non well-balanced schemes.

UNIVERSIDAD DE MÁLAGA

Table 2.1: Numerical validation of the well-balanced and non well-balanced methods for the stationary problem (2.3.47) (Burges's' equation). Table contains L_1 errors for the variable u at final time $t = 10$ s for both Runge-Kutta and ADER DG schemes of order $N_p + 1 = 1,\ 2,\ 3,\ 4$. Uniform Cartesian meshes of N_s elements has been used.

		Non Well-balanced		**Well-balanced**	
$\boldsymbol{N}$	$\boldsymbol{N_s}$	**Runge-Kutta**	**ADER**	**Runge-Kutta**	**ADER**
0	25	1.39×10^{-01}	1.39×10^{-01}	0	0
	50	6.52×10^{-02}	6.52×10^{-02}	2.66×10^{-17}	0
	100	3.17×10^{-02}	3.17×10^{-02}	1.55×10^{-17}	0
	200	1.56×10^{-02}	1.56×10^{-02}	0	0
1	25	6.47×10^{-04}	5.37×10^{-04}	8.88×10^{-18}	0
	50	1.60×10^{-04}	1.34×10^{-04}	0	0
	100	3.96×10^{-05}	3.33×10^{-05}	2.22×10^{-18}	1.11×10^{-18}
	200	9.85×10^{-06}	8.29×10^{-06}	2.22×10^{-18}	5.55×10^{-19}
2	25	4.49×10^{-06}	3.80×10^{-06}	0	2.47×10^{-18}
	50	5.91×10^{-07}	5.25×10^{-07}	7.40×10^{-18}	0
	100	7.57×10^{-08}	6.85×10^{-08}	2.47×10^{-18}	0
	200	9.57×10^{-09}	8.75×10^{-09}	1.23×10^{-18}	9.25×10^{-19}
3	25	2.56×10^{-08}	1.10×10^{-08}	4.63×10^{-18}	3.09×10^{-18}
	50	1.70×10^{-09}	7.85×10^{-10}	1.54×10^{-18}	7.53×10^{-18}
	100	1.09×10^{-10}	5.61×10^{-11}	2.32×10^{-18}	4.15×10^{-18}
	200	6.91×10^{-12}	4.00×10^{-12}	1.54×10^{-18}	2.27×10^{-18}

Limited simulation Finally, we seek to check the robustness of the proposed well-balanced WENO limiter. The following initial condition is considered,

$$u = 1 + \sin(x), \quad \sigma = 0. \tag{2.3.49}$$

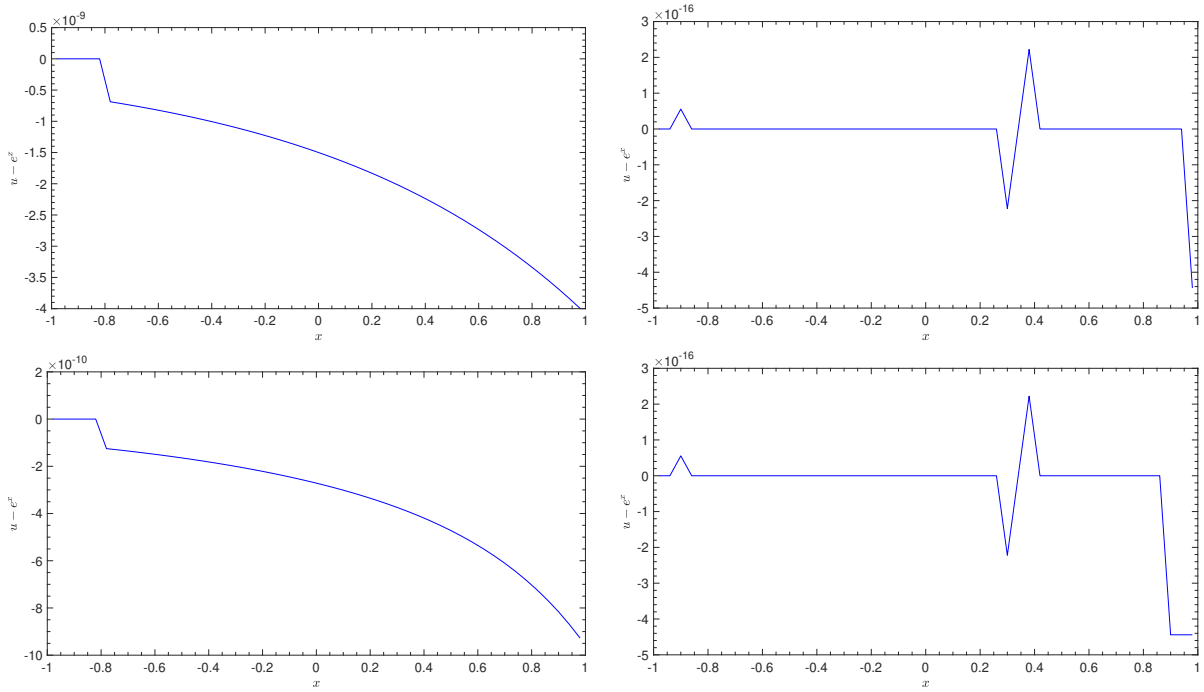
On this occasion, the computational domain is $I = [0, 2\pi]$ and it is discretized with a coarse mesh of $N_s = 80$ elements. Periodic boundary conditions are set for both boundaries. Figure 2.7 depicts the solution for a fourth order ADER and Runge-Kutta numerical schemes at final time $t = 1.5$ s. As it can be seen, both models provide satisfactory results without spurious oscillations, even for coarse meshes.

2.3.6.2 Well-balanced and non well-balanced schemes comparison

This experiment consist on a comparison between the well-balanced and non-well-balanced schemes for a solution which is not a perturbation of an equilibrium state. The following

UNIVERSIDAD DE MÁLAGA

Figure 2.5: Fluctuation $(u - e^{\sigma})$ for the stationary problem (2.3.47) for the Burgers' equation at time $t = 10$ s with the fourth order non well-balanced (left) and well-balanced (right) numerical schemes. An uniform Cartesian mesh of $N_s = 50$ elements has been used. Upper and lower panels stands for Runge-Kutta and ADER time-marching discretizations respectively.



initial condition is set:

$$u = 0.1e^{-200(x+0.5)^2}, \quad \sigma = x. \tag{2.3.50}$$

The domain $I = [-1, 1]$ is discretized with 100 elements and periodic boundary conditions are considered. The results are depicted in Figure 2.8 for a Runge-Kutta scheme of second and third order. The reference solution has been computed with a first order method with 20000 discretization points. The results conclude that the well-balanced and non well-balanced methods are in fact very similar regardless of the order of the method. Similar results are obtained for ADER type methods.

UNIVERSIDAD DE MÁLAGA

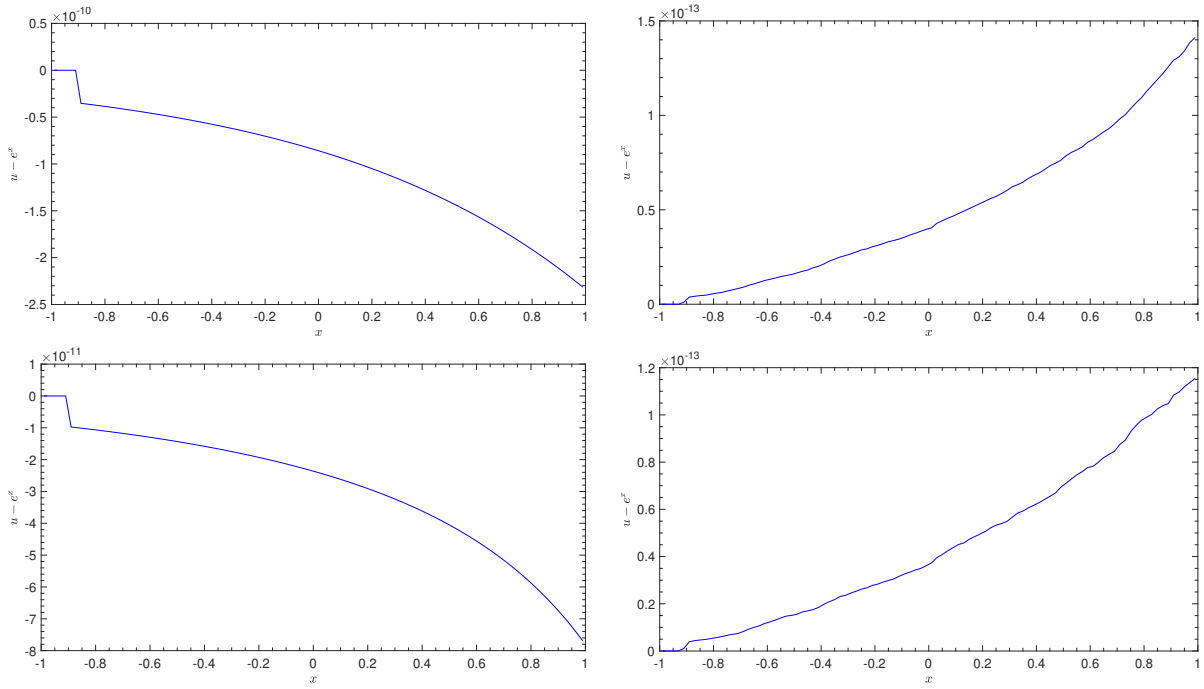


Figure 2.6: Fluctuation with respect to the stationary solution $(u - e^{\sigma})$ (2.3.48) for the Burgers' equation at time $t = 10$ s with the fourth order non well-balanced (left) and well-balanced (right) methods. An uniform Cartesian mesh of $N_s = 100$ elements has been used. Upper and lower panels stands for Runge-Kutta and ADER time marching discretizations respectively.

UNIVERSIDAD DE MÁLAGA

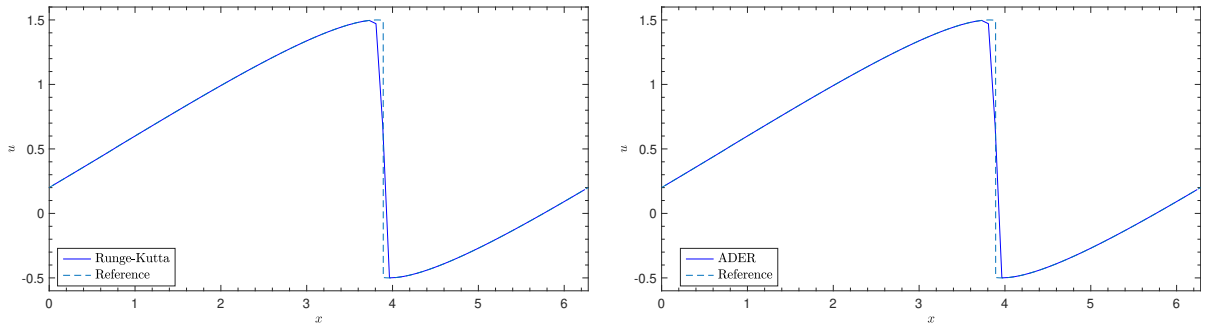


Figure 2.7: Computed variable u for the problem (2.3.49) (Burgers' equation) with the fourth order well-balanced Runge-Kutta (left) and ADER (right) numerical methods at time $t = 1.5$ s. An uniform Cartesian mesh of $N_s = 80$ elements has been used.

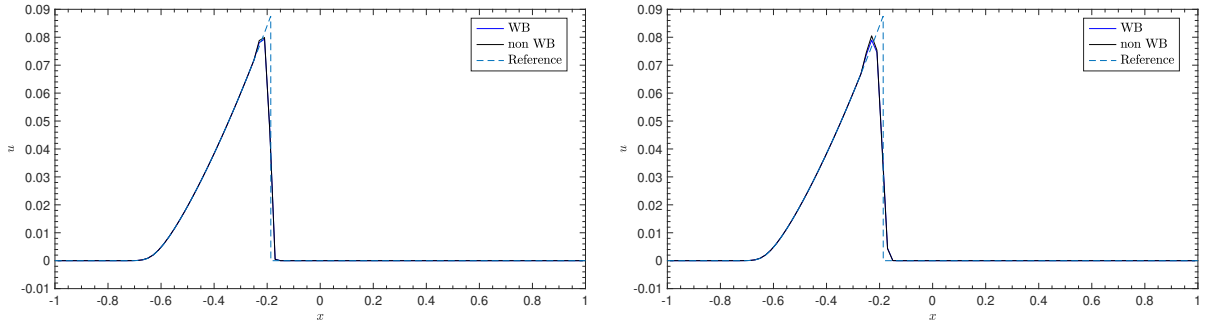


Figure 2.8: Computed variable u for the problem (2.3.50) (Burgers' equation) with the second (left) and third (right) order well-balanced Runge–Kutta numerical methods at time $t = 5$ s. An uniform Cartesian mesh of $N_s = 100$ elements has been used.

UNIVERSIDAD DE MÁLAGA

2.3.6.3 Shallow-water equations

We begin with a brief discussion of the model. The one layer shallow-water equations written in conservative form are formulated as follows:

$$\begin{cases} \partial_t h + \partial_x (hu) = 0, \\ \partial_t (hu) + \partial_x (hu^2 + \frac{1}{2} g h^2) = g h \sigma_x, \end{cases} \tag{2.3.51}$$

where $h = h(x,t)$ is the total water depth, u is the horizontal depth-averaged velocity, and $\sigma(x)$ is the known still water depth that we will suppose to be continuous. Finally, $g = 9.81$ is the gravitational acceleration.

Stationary solutions of the shallow-water equations are described by

$$u = 0, \quad \eta = C, \tag{2.3.52}$$

where $\eta = h - \sigma$ is the free-surface elevation and $C \in \mathbb{R}$ is a constant. Therefore, preserving this kind of solutions and recuperating them after small perturbations will be our goal for this system.

Order of accuracy test The order of accuracy of the numerical scheme for a well-balanced Runge-Kutta and ADER time discretization for a second, third and fourth-order schemes are now checked. The computational domain is $I = [-5, 5]$ and periodic boundary conditions are imposed in all boundaries. As an initial condition, the following function is considered:

$$u = 0, \quad h = \sigma + 0.1 e^{-x^2}, \quad \sigma = 1 - 0.2 e^{-x^2}. \tag{2.3.53}$$

The simulation is stopped at final time $t = 1$ s.

The results for an increasingly refined mesh and different orders can be seen at table 2.2, where the L_1 errors and numerical convergence are posted. As it can be seen, the expected order is reached for both the Runge-Kutta and ADER-DG methods.

Preserving stationary solutions The main objective for this test is to preserve a stationary solution of the form (2.3.52) up to machine precision, while the non well-balanced methods fail to accomplish this. The initial condition is given by,

$$u = 0, \quad h = \sigma, \quad \sigma = 1 - 0.8 e^{-100 x^2}, \tag{2.3.54}$$

in the computational domain $I = [-1, 1]$ with Dirichlet boundary conditions set in all boundaries. Results are depicted in Figure 2.9 and tables 2.3 and 2.4 for the variables h and hu for both Runge-Kutta and ADER time discretization. As it can be seen, the well-balanced scheme is able to preserve the stationary solution at a final time $t = 10$ s, while the non well-balanced one fails regardless of the mesh considered.

UNIVERSIDAD DE MÁLAGA

Table 2.2: Numerical convergence results for the initial condition (2.3.53) (shallow water equations). High-order Runge-Kutta and ADER DG schemes of order $N_p + 1 = 2,\ 3,\ 4$. Uniform Cartesian meshes of N_s elements has been used. The L_1 errors refer to the variables h and hu at a final time $t = 1$ s.

		Runge-Kutta				**ADER**			
		h		***hu***		***h***		***hu***	
N	***N_s***	**Error**	**Order**	**Error**	**Order**	**Error**	**Order**	**Error**	**Order**
1	25	3.93×10^{-03}	-	1.27×10^{-02}		3.60×10^{-03}	-	1.17×10^{-02}	-
	50	5.92×10^{-04}	2.73	1.86×10^{-03}	2.77	8.91×10^{-04}	2.01	2.79×10^{-03}	2.07
	100	1.11×10^{-04}	2.41	3.49×10^{-04}	2.41	2.21×10^{-04}	2.01	6.78×10^{-04}	2.04
	200	2.16×10^{-05}	2.37	6.81×10^{-05}	2.36	5.50×10^{-05}	2.00	1.67×10^{-04}	2.02
2	25	7.37×10^{-05}	-	2.19×10^{-04}	-	2.97×10^{-04}	-	9.46×10^{-04}	-
	50	3.27×10^{-06}	4.49	9.72×10^{-06}	4,50	3.28×10^{-05}	3.18	1.06×10^{-04}	3.15
	100	2.47×10^{-07}	3.73	7.30×10^{-07}	3.74	3.79×10^{-06}	3.11	1.26×10^{-05}	3.08
	200	2.23×10^{-08}	3.47	6.92×10^{-08}	3.40	4.60×10^{-07}	3.04	1.55×10^{-06}	3.02
3	25	3.52×10^{-06}	-	1.02×10^{-05}	-	9.48×10^{-06}	-	3.43×10^{-05}	-
	50	1.41×10^{-07}	4.64	3.52×10^{-07}	4.86	6.60×10^{-07}	3.84	2.24×10^{-06}	3.94
	100	8.95×10^{-09}	3.98	2.21×10^{-08}	3.99	4.41×10^{-08}	3.91	1.44×10^{-07}	3.96
	200	5.63×10^{-10}	3.99	1.39×10^{-09}	3.99	2.75×10^{-09}	4.00	9.12×10^{-09}	3.98

Perturbed stationary solutions We consider now a perturbation of the stationary solution (2.3.54) given by

$$u = 0, \quad h = 0.1e^{-100x^2} + \sigma, \quad \sigma = 1 - 0.8e^{-100x^2}. \tag{2.3.55}$$

We expect to obtain a stationary solution after the perturbation has left the computational domain $I = [-1, 1]$ with $N_s = 100$. Dirichlet and free-outflow boundary conditions are imposed at the right and the left boundaries of the domain, respectively.

As it can be seen in Figure 2.10, the well-balanced scheme can recover the stationary solution at final time $t = 10$ s with the fourth-order DG method for both Runge-Kutta and ADER time discretizations, while the non well-balanced methods fails to do so. Moreover, in order to show the advantages of the well-balanced methods to capture small perturbations, we perform a similar numerical test on a very coarse mesh of just 20 elements and the following initial condition,

$$u = 0, \quad h = 10^{-4}e^{-100x^2} + \sigma, \quad \sigma = 1 - 0.8e^{-100x^2}. \tag{2.3.56}$$

UNIVERSIDAD DE MÁLAGA

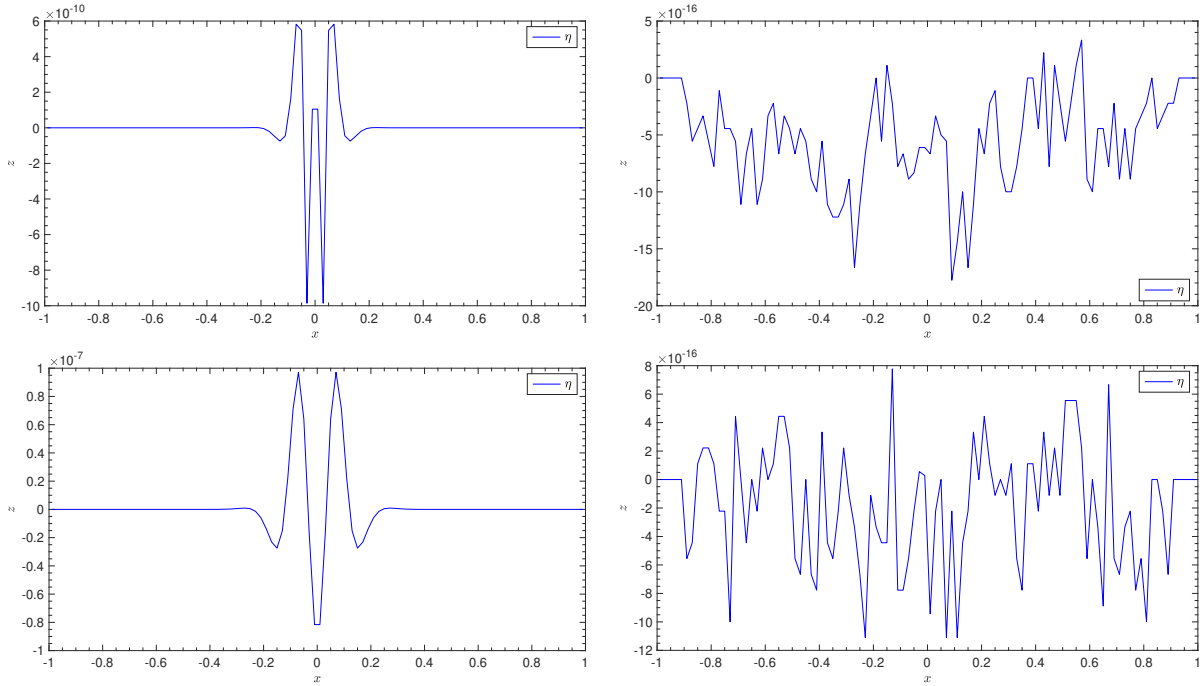


Figure 2.9: Computed free surface for the stationary problem (2.3.54) (sallow-water equations) at time $t = 10$ s with the fourth order non well-balanced (left) and well-balanced (right) methods. An uniform Cartesian mesh of $N_s = 100$ elements has been used. Upper and lower panels stands for Runge-Kutta and ADER time-marching discretizations respectively.

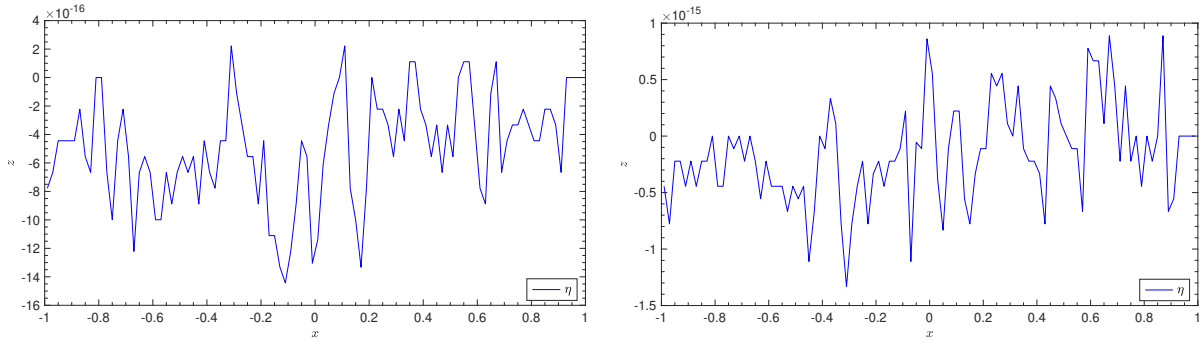


Figure 2.10: Computed free surface of the perturbed stationary solution (2.3.55) (shallow-water equations) with the fourth order well-balanced Runge-Kutta (left) and ADER (right) DG schemes at $t = 10$ s. A uniform Cartesian mesh of $N_s = 100$ elements has been used.

UNIVERSIDAD DE MÁLAGA

Table 2.3: Numerical validation of the well-balanced and non well-balanced methods for the stationary problem (2.3.54) (shallow water equations). Table contains L_1 errors for the water depth variable h at a final time $t = 10$ s for Runge-Kutta and ADER DG schemes of order $N_p + 1 = 1,\ 2,\ 3,\ 4$. Uniform Cartesian meshes of N_s elements has been used.

		Non Well-balanced		**Well-balanced**	
$\boldsymbol{N}$	$\boldsymbol{N_s}$	**Runge-Kutta**	**ADER**	**Runge-Kutta**	**ADER**
0	25	1.92×10^{-02}	1.92×10^{-02}	3.00×10^{-17}	3.00×10^{-17}
	50	6.08×10^{-03}	6.08×10^{-03}	1.27×10^{-16}	3.34×10^{-16}
	100	1.40×10^{-03}	1.40×10^{-03}	4.86×10^{-17}	1.64×10^{-17}
	200	5.37×10^{-04}	5.32×10^{-04}	5.97×10^{-18}	2.30×10^{-17}
1	25	1.16×10^{-03}	2.62×10^{-03}	6.66×10^{-17}	6.88×10^{-16}
	50	2.73×10^{-04}	3.47×10^{-04}	2.06×10^{-16}	8.61×10^{-16}
	100	3.76×10^{-05}	5.48×10^{-05}	2.82×10^{-16}	1.42×10^{-15}
	200	4.81×10^{-06}	8.48×10^{-06}	4.37×10^{-16}	1.61×10^{-15}
2	25	3.22×10^{-04}	4.09×10^{-04}	1.52×10^{-16}	1.36×10^{-16}
	50	3.20×10^{-05}	4.39×10^{-05}	1.85×10^{-16}	1.20×10^{-15}
	100	2.36×10^{-06}	4.47×10^{-06}	2.74×10^{-16}	1.58×10^{-15}
	200	1.51×10^{-07}	4.95×10^{-07}	3.67×10^{-16}	1.96×10^{-15}
3	25	1.84×10^{-05}	4.18×10^{-05}	1.47×10^{-16}	1.63×10^{-15}
	50	1.64×10^{-06}	2.42×10^{-06}	3.11×10^{-16}	2.50×10^{-15}
	100	6.43×10^{-08}	6.25×10^{-08}	5.78×10^{-16}	4.14×10^{-16}
	200	2.08×10^{-09}	3.57×10^{-09}	6.75×10^{-16}	6.98×10^{-16}

Periodic boundary conditions are set and the final computational time is $t = 0.5$ s. Since the perturbation is of the order $(\Delta x)^{N_p+1}$, $N_p + 1 = 4$ being the order of the scheme, we can expect that only the well-balanced method yields satisfactory results. In contrast, the non well-balanced scheme introduces perturbations of that order. Here we only show the numerical results in Figure 2.11 for the Runge-Kutta time discretization Similar results can be observed for ADER time discretizations

Limited simulation The behavior of the well-balanced WENO limiter is now tested. To achieve this, a simulation far from the *equilibria* is considered and therefore a limiting technique becomes mandatory. The computational domain $I = [-5, 5]$ is discretized with a fine mesh of $N_s = 400$ elements. The initial condition is,

$$u = 0, \quad h = 0.1e^{-5x^2} + \sigma, \quad \sigma = 1 - 0.8e^{-x^2}, \tag{2.3.57}$$

UNIVERSIDAD DE MÁLAGA

Table 2.4: Numerical validation of the well-balanced and non well-balanced methods for the stationary problem (2.3.54) (shallow water equations). Table contains L_1 errors for the conserved variable hu at a final time $t = 10$ s for Runge-Kutta and ADER DG schemes of order $N_p + 1 = 1,\ 2,\ 3,\ 4$. Uniform Cartesian meshes of N_s elements has been used.

		Non Well-balanced		**Well-balanced**	
$\boldsymbol{N}$	$\boldsymbol{N_s}$	**Runge-Kutta**	**ADER**	**Runge-Kutta**	**ADER**
0	25	6.12×10^{-02}	6.12×10^{-02}	4.40×10^{-15}	4.39×10^{-15}
	50	3.96×10^{-02}	3.96×10^{-02}	1.06×10^{-15}	1.11×10^{-15}
	100	2.20×10^{-02}	2.21×10^{-02}	4.09×10^{-16}	3.52×10^{-16}
	200	1.20×10^{-02}	1.20×10^{-02}	1.69×10^{-16}	1.79×10^{-16}
1	25	3.56×10^{-03}	6.49×10^{-03}	5.68×10^{-16}	1.07×10^{-14}
	50	3.35×10^{-04}	7.43×10^{-04}	5.33×10^{-16}	2.19×10^{-15}
	100	2.28×10^{-05}	1.38×10^{-04}	8.28×10^{-16}	9.95×10^{-15}
	200	1.44×10^{-06}	3.38×10^{-05}	1.30×10^{-15}	3.89×10^{-15}
2	25	5.27×10^{-04}	7.48×10^{-04}	3.97×10^{-16}	4.50×10^{-16}
	50	1.40×10^{-04}	1.82×10^{-04}	6.83×10^{-16}	4.28×10^{-15}
	100	1.96×10^{-05}	2.69×10^{-05}	7.47×10^{-16}	4.23×10^{-15}
	200	2.52×10^{-06}	3.45×10^{-06}	9.91×10^{-16}	2.79×10^{-15}
3	25	1.23×10^{-04}	3.02×10^{-04}	3.62×10^{-16}	1.51×10^{-14}
	50	2.59×10^{-06}	1.02×10^{-05}	7.01×10^{-16}	4.29×10^{-15}
	100	4.85×10^{-08}	3.25×10^{-07}	9.93×10^{-16}	1.08×10^{-15}
	200	7.75×10^{-10}	2.20×10^{-08}	1.87×10^{-15}	1.55×10^{-15}

and periodic boundary conditions are set. The solution at several times can be seen in Figure 2.12. As it can be seen, the limiter successfully manages to avoid spurious oscillations near discontinuities. Note that the same limiter parameters has been considered for both Runge-Kutta and ADER time discretizations. In practice, an *ad hoc* choice of the parameters taking into account the particularities of each solver can improve the results.

UNIVERSIDAD DE MÁLAGA

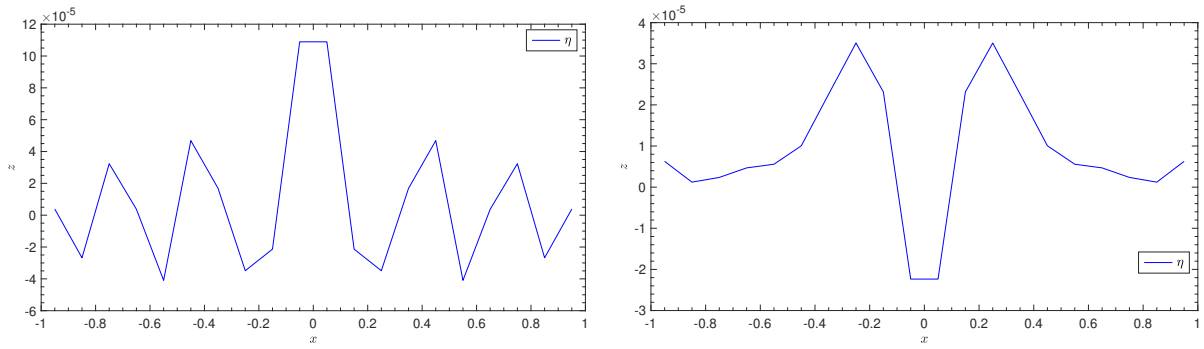


Figure 2.11: Computed free surface of the smaller perturbed stationary solution (2.3.56) (sallow-water equations) with the fourth order non well-balanced (left) and well-balanced (right) Runge-Kutta DG method at $t = 0.5$ s. A coarse uniform Cartesian mesh of $N_s = 20$ elements has been used.

UNIVERSIDAD DE MÁLAGA

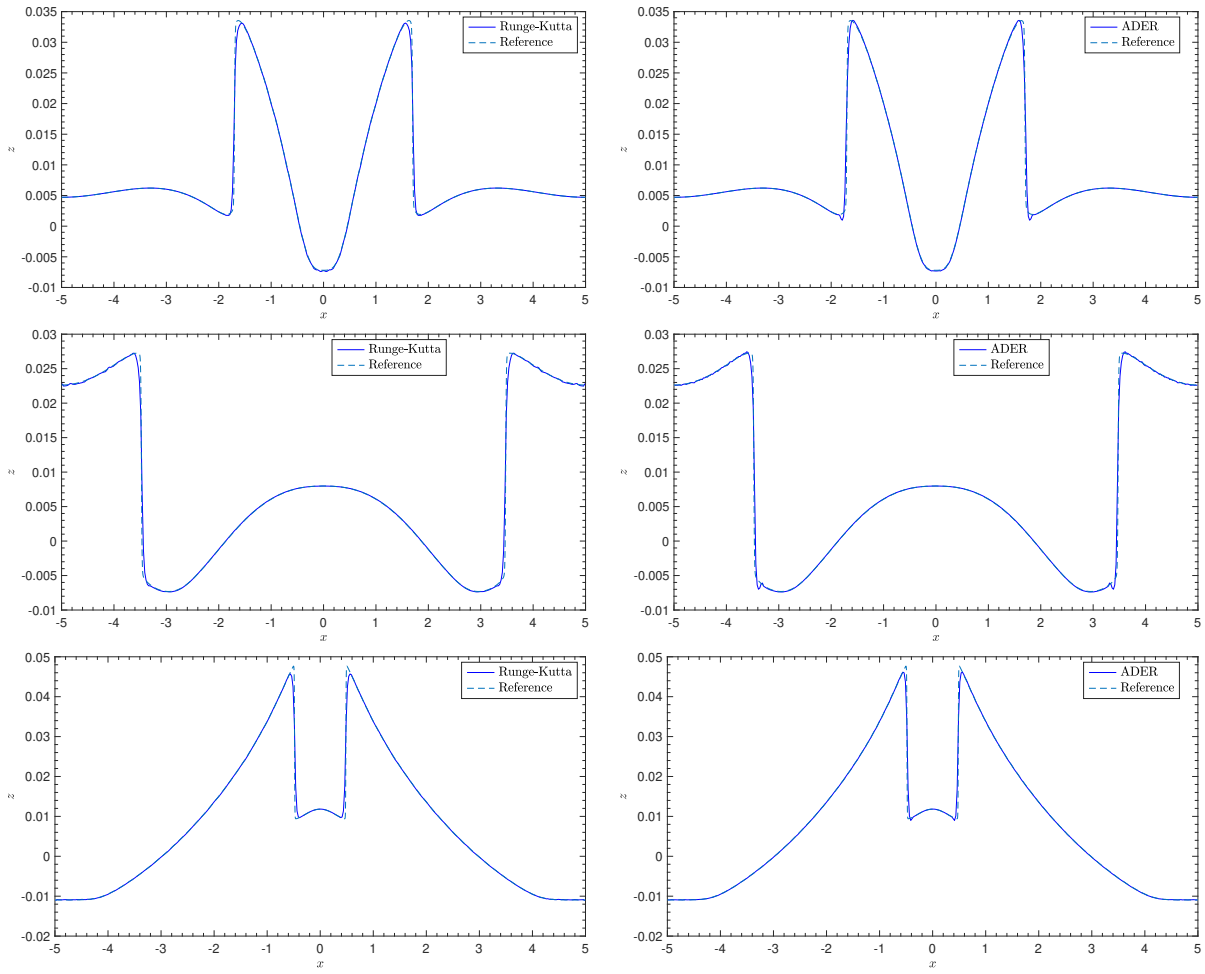


Figure 2.12: Computed free surface for the problem (2.3.57) (shallow water equations) with the fourth order well-balanced Runge-Kutta (left) and ADER (right) numerical methods at times $t = 4,\ 5.5,\ 6.5$ s (from upper to lower panels). An uniform Cartesian mesh of $N_s = 400$ elements has been used.

UNIVERSIDAD DE MÁLAGA

2.3.6.4 Compressible Euler equations with gravitational force

We now consider the gas dynamic Euler system,

$$\begin{cases} \partial_t \rho + \partial_x (\rho u) = 0, \\ \partial_t (\rho u) + \partial_x (\rho u^2 + p) = -\rho \sigma_x, \\ \partial_t E + \partial_x (u(E+p)) = -\rho\, u \sigma_x, \end{cases} \tag{2.3.58}$$

with $\rho \geq 0$ being the density, u the velocity, E the total energy and $\sigma(x)$ a given continuous gravitational potential. Additionally, the internal energy e is given by $\rho e = E - \frac{1}{2}\rho u^2$ while the pressure $p \geq 0$ is given by e through the equation of state (ideal gas),

$$p = (\gamma - 1)\rho e, \tag{2.3.59}$$

where $\gamma > 1$ is the adiabatic constant that will take the value $\frac{5}{3}$ unless stated otherwise.

In this work, we will focus on the hydrostatic steady states solutions given by

$$u = 0, \quad \partial_x p = -\rho \sigma_x. \tag{2.3.60}$$

In [128], the following family of stationary solutions of (2.3.58) with $u = 0$ is given,

$$u = 0, \quad \rho = \alpha(\sigma, C_1) \geq 0, \quad p(x) = \beta(\sigma, C_1, C_2) \geq 0, \quad E = \frac{p(x)}{\gamma - 1}, \tag{2.3.61}$$

where α is a given continuous function and β is given by

$$\partial_\sigma \beta(\sigma, C_1, C_2) = -\alpha(\sigma, C1).$$

A particular solution is chosen to be preserved by the numerical scheme by defining a particular α,

$$\alpha(\sigma, C_1) = C_1 e^{-\sigma}, \tag{2.3.62}$$

that results in the following set of stationary solutions with $u = 0$,

$$u = 0, \quad \rho = C_1 e^{-\sigma(x)} \geq 0, \quad p = C_1 e^{-\sigma(x)} + C_2 \geq 0, \quad E = \frac{p}{\gamma - 1}. \tag{2.3.63}$$

Preserving stationary solutions As with the previous examples, we first seek to preserve a stationary solution of the form (2.3.63). Therefore, the initial condition is given by,

$$u = 0, \quad \rho = e^{-\sigma(x)}, \quad p(x,0) = e^{-\sigma(x)}, \quad E = \frac{p}{\gamma - 1}, \quad \sigma(x) = x. \tag{2.3.64}$$

The computational domain is $I = [-1, 1]$ and Dirichlet boundary conditions are imposed everywhere. The results of the simulations are shown in Figure 2.13 and Table 2.5 for the well-balanced and non well-balanced methods. As expected, only the well-balanced method is able to preserve the stationary solution up to machine precision.

UNIVERSIDAD DE MÁLAGA

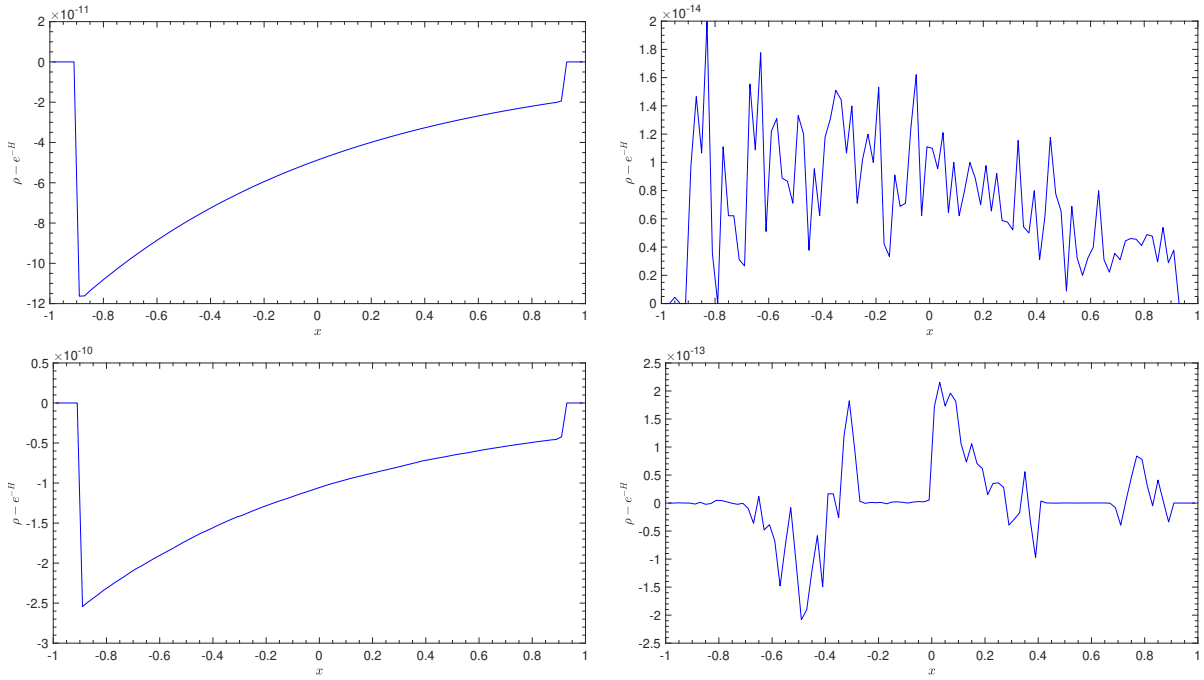


Figure 2.13: Computed density variable $\rho - e^{-\sigma}$ for the stationary problem (2.3.64) (Euler equations) at time $t = 10$ s with the fourth order non well-balanced (left) and well-balanced (right) methods. An uniform Cartesian mesh of $N_s = 100$ elements has been used. Upper and lower panels stands for Runge-Kutta and ADER time-marching discretizations respectively.

Limited simulation Now we consider a standard Riemann problems with $\sigma = 0$ and $\gamma = 1.4$ and the following discontinuous initial condition given by

$$\begin{cases} \left(\rho_L, \quad u_L, \quad p_L\right) & \text{if } x < 0, \\ \\ \left(\rho_R, \quad u_R, \quad p_R\right) & \text{if } x \geq 0. \end{cases} \tag{2.3.65}$$

Then, two well-known numerical tests are contemplated:

- The sod problem [198]:

$$\left(\rho_L, \quad u_L, \quad p_L\right) = \left(1, \quad 0, \quad 1\right), \quad \left(\rho_R, \quad u_R, \quad p_R\right) = \left(0.125 \quad 0 \quad 0.1\right). \tag{2.3.66}$$

- The Lax problem [199]:

$$\left(\rho_L \quad u_L \quad p_L\right) = \left(0.445, \quad 0.698, \quad 3.528\right), \quad \left(\rho_R, \quad u_R, \quad p_R\right) = \left(0.5, \quad 0, \quad 0.571\right). \tag{2.3.67}$$

UNIVERSIDAD DE MÁLAGA

Table 2.5: Numerical validation of the well-balanced and non well-balanced methods for the stationary problem (2.3.64) (Euler equations). Table contains L_1 errors for the density variable ρ at a final time $t = 10$ s for both Runge-Kutta and ADER DG schemes of order $N_p + 1 = 1,\ 2,\ 3,\ 4$. Uniform Cartesian meshes of N_s elements has been used.

		Non Well-balanced		**Well-balanced**	
$\boldsymbol{N}$	$\boldsymbol{N_s}$	**Runge-Kutta**	**ADER**	**Runge-Kutta**	**ADER**
0	25	5.13×10^{-02}	1.03×10^{-02}	7.99×10^{-17}	1.31×10^{-16}
	50	3.18×10^{-02}	5.46×10^{-03}	2.31×10^{-16}	4.42×10^{-16}
	100	1.77×10^{-02}	3.21×10^{-03}	1.94×10^{-16}	6.21×10^{-16}
	200	9.41×10^{-03}	1.85×10^{-03}	3.26×10^{-16}	6.61×10^{-16}
1	25	2.35×10^{-04}	3.08×10^{-04}	7.46×10^{-16}	1.15×10^{-15}
	50	7.10×10^{-05}	8.81×10^{-05}	1.20×10^{-15}	1.47×10^{-15}
	100	1.94×10^{-05}	2.34×10^{-05}	2.15×10^{-15}	2.65×10^{-15}
	200	9.26×10^{-06}	6.03×10^{-06}	3.17×10^{-15}	4.21×10^{-15}
2	25	3.52×10^{-06}	2.98×10^{-06}	1.06×10^{-15}	2.84×10^{-15}
	50	5.40×10^{-07}	4.67×10^{-07}	1.78×10^{-15}	5.01×10^{-15}
	100	7.44×10^{-08}	6.58×10^{-08}	4.40×10^{-15}	9.39×10^{-15}
	200	9.75×10^{-09}	8.73×10^{-09}	4.93×10^{-15}	1.36×10^{-14}
3	25	8.20×10^{-09}	1.81×10^{-08}	1.26×10^{-15}	1.35×10^{-14}
	50	7.14×10^{-10}	1.53×10^{-09}	2.74×10^{-15}	2.83×10^{-14}
	100	5.18×10^{-11}	1.09×10^{-10}	7.71×10^{-15}	4.53×10^{-14}
	200	3.48×10^{-12}	7.00×10^{-12}	9.54×10^{-15}	7.16×10^{-14}

For both problems, the computational domain is $I = [-5, 5]$, discretized with $N_s = 300$ cells and with the limiter parameter M set to 0.01 and the CFL number to 0.5. Finally, open boundary conditions are chosen.

The results are depicted in Figures 2.14 and 2.15 for a fourth order ADER and Runge-Kutta DG method for the Sod and Lax problem respectively. The numerical results are compared against a reference solution that has been computed using a first order method with $N_s = 20000$ elements.

UNIVERSIDAD DE MÁLAGA

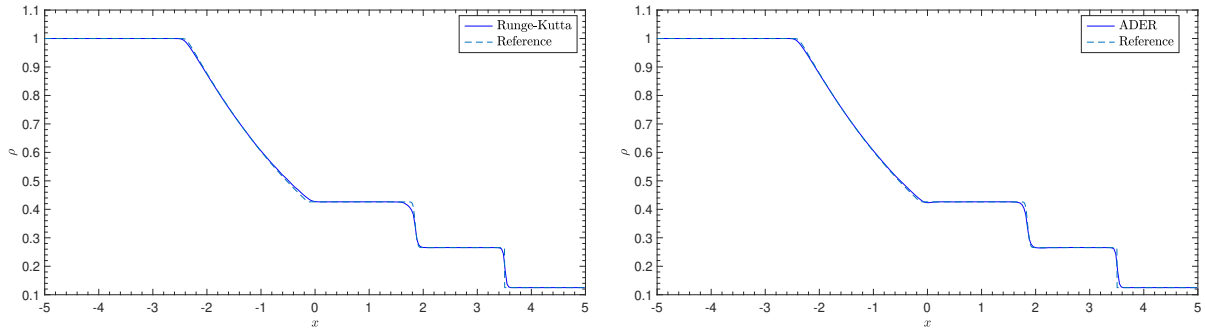


Figure 2.14: Computed density variable ρ for the problem (2.3.66) (Euler equations) with the fourth order well-balanced Runge-Kutta (left) and ADER (right) numerical methods at time $t = 2$ s. An uniform Cartesian mesh of $N_s = 300$ elements has been used.

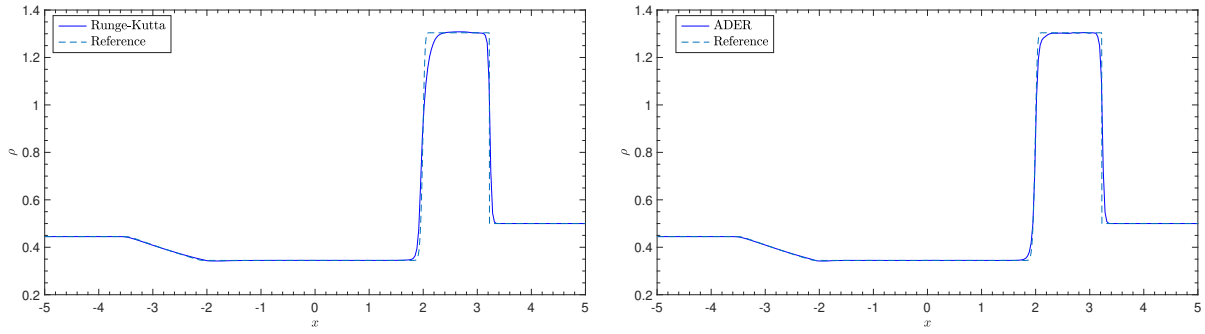


Figure 2.15: Computed density variable ρ for the problem (2.3.67) (Euler equations) with the fourth order well-balanced Runge-Kutta (left) and ADER (right) numerical methods at time $t = 1.3$ s. An uniform Cartesian mesh of $N_s = 300$ elements has been used.

UNIVERSIDAD DE MÁLAGA

Chapter 3

Numerical discretization

3.1 Introduction

It has been previously discussed the general discretization strategies to design high order numerical methods in the finite volume and discontinuous Galerkin frameworks. This chapter aims to describe the particular numerical schemes designed for the multilayer shallow water model with variable pressure introduced in Chapter 1. This model is extremely sensitive to density fluctuations and it is therefore very important for the numerical scheme to be both robust and accurate, especially at high order.

First, we discuss the finite volume method numerical discretization. It is based on a HLL Riemann solver with a hydrostatic reconstruction in order to preserve lake-at-rest stationary solutions, improving the overall stability of the scheme. The upwind discretization of the transfer terms is also discussed. The second order accuracy is achieved by a MUSCL reconstruction: a reconstruction operator is defined based on the combined slopes of the states linking the central cell with its neighbors. Finally, a strategy to preserve a parametric family of stationary solutions corresponding to relative density stratification within the framework of finite volume methods is also discussed.

Additionally, an alternative arbitrary high order in space and time numerical discretization of the multilayer shallow water model with variable pressure is also described. It is performed within the framework of unlimited arbitrary high order accurate (ADER) discontinuous Galerkin (DG) methods. The guidelines for this particular discretization is given by the general framework provided in Chapter 2. The very high order and the enhanced resolution of the DG methods are especially suited for problems where the computational load is an important factor. In this way, it is possible to capture complex behavior with coarse or even very coarse meshes. Examples of this will be given in Chapter 4. Additionally, as in the finite volume method, a strategy to preserve a stratified water at rest solution is also described.

UNIVERSIDAD DE MÁLAGA

3.2 A second order well-balanced finite volume numerical scheme

In this chapter we will detail the numerical discretization of the model presented in Chapter 1 with the techniques presented in Chapter 2. We consider the full system (1.2.7) written in the form of an hyperbolic system with convective fluxes and non conservative products:

$$\partial_t \boldsymbol{w} + \partial_x \boldsymbol{F}_C(\boldsymbol{w}) + \boldsymbol{P}(\boldsymbol{w}, \eta, \partial_x \boldsymbol{w}, \partial_x \eta) - \boldsymbol{T}(\boldsymbol{w}, \partial_x \boldsymbol{w}) = \boldsymbol{0}, \tag{3.2.1}$$

where $\boldsymbol{w}$ is the vector of the state variables,

$$\boldsymbol{w} = (h \,|\, h\theta_\alpha \,|\, h\theta_\alpha u_\alpha)^T \in \mathbb{R}^{2M+1}, \tag{3.2.2}$$

and the Einstein notation has been used. We recall that M is the total number of vertical layers of the model. Additionally, the convective flux is given by,

$$\boldsymbol{F}_C(\boldsymbol{w}) = \left(h \sum_{\beta=1}^{M} l_\beta u_\beta \,\middle|\, h\theta_\alpha u_\alpha \,\middle|\, h\theta_\alpha u_\alpha^2 \right)^T \in \mathbb{R}^{2M+1}, \tag{3.2.3}$$

while the pressure terms, which are dependent on the relative density, the bathymetry function and the water depth, are defined by,

$$\boldsymbol{P}(\boldsymbol{w}, \eta, \partial_x \boldsymbol{w}, \partial_x \eta) = (0 \,|\, \boldsymbol{0} \,|\, P_\alpha) \in \mathbb{R}^{2M+1}, \tag{3.2.4}$$

with

$$P_\alpha = gh\theta_\alpha \partial_x \eta + \frac{g l_\alpha}{2}(h\partial_x(h\theta_\alpha) - h\theta_\alpha \partial_x h) + g \sum_{\beta=\alpha+1}^{M} l_\beta (h\partial_x(h\theta_\beta) - h\theta_\alpha \partial_x h). \tag{3.2.5}$$

Finally, the transfer terms $\boldsymbol{T}(\boldsymbol{w}, \partial_x \boldsymbol{w})$ corresponding to the mass, density and momentum exchange between layers are,

$$\boldsymbol{T}(\boldsymbol{w}, \partial_x \boldsymbol{w}) = \left(0 \,\middle|\, \frac{1}{l_\alpha}(\theta_{\alpha+\frac{1}{2}} G_{\alpha+\frac{1}{2}} - \theta_{\alpha-\frac{1}{2}} G_{\alpha-\frac{1}{2}}) \,\middle|\, \frac{1}{l_\alpha}(u_{\alpha+\frac{1}{2}} \theta_{\alpha+\frac{1}{2}} G_{\alpha+\frac{1}{2}} - u_{\alpha-\frac{1}{2}} \theta_{\alpha-\frac{1}{2}} G_{\alpha-\frac{1}{2}}) \right)^T \in \mathbb{R}^{2M+1}. \tag{3.2.6}$$

We recall that $G_{\alpha+\frac{1}{2}}$ is described by (1.2.8).

As usual in the finite volume framework, the computational domain I is discretized into uniform cells $I_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$, $i = 1, \ldots, N_s$, where N_s is the total number of cells

UNIVERSIDAD DE MÁLAGA

with constant length $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$. In each cell an average approximation of the solution at time $t^n = n\Delta t$ is considered:

$$\boldsymbol{w}_i^n \approx \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \boldsymbol{w}(x, t^n)\, dx.$$

Likewise, the approximation of the bathymetry function at the cell I_i is denoted by $z_{B,i}$ and defined as:

$$z_{B,i} \approx \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} z_B(x)\, dx.$$

Moreover, we introduce a commonly used notation in this section. For a function f, we shall denote its average at the intercell $i+\frac{1}{2}$ as:

$$\overline{f} \equiv \overline{f}_{i+\frac{1}{2}} = \frac{1}{2}\left(f_i + f_{i+1}\right). \tag{3.2.7}$$

In particular, for a function f_α defined within the layer α, we shall write

$$\overline{f_\alpha}_{,i+\frac{1}{2}} = \frac{1}{2}\left(f_{\alpha,i} + f_{\alpha,i+1}\right). \tag{3.2.8}$$

Meanwhile the difference at the intercell $i+\frac{1}{2}$ is written as

$$\Delta f \equiv (\Delta f)_{i+\frac{1}{2}} = f_{i+1} - f_i.$$

Finally, the average of a variable f_α at the layer interface $\alpha+\frac{1}{2}$ is denoted as

$$\langle f \rangle_{\alpha+\frac{1}{2}} = \frac{1}{2}\left(f_{\alpha+1} + f_\alpha\right), \tag{3.2.9}$$

while at the bottom or free surface we assume,

$$\langle f \rangle_{\frac{1}{2}} = f_1, \quad \langle f \rangle_{M+\frac{1}{2}} = f_M.$$

3.2.1 First order HLL-type scheme

As discussed in Chapter 2 and following [48], a HLL-type numerical scheme for general systems of PDE (3.2.1) could be rewritten as follows:

$$\boldsymbol{w}_i^{n+1} = \boldsymbol{w}_i^n - \frac{\Delta t}{\Delta x}\left(\boldsymbol{D}^+_{i-\frac{1}{2}}(\boldsymbol{w}_{i-1}^n, \boldsymbol{w}_i^n, z_{B,i-1}, z_{B,i}) + \boldsymbol{D}^-_{i+\frac{1}{2}}(\boldsymbol{w}_i^n, \boldsymbol{w}_{i+1}^n, z_{B,i}, z_{B,i+1})\right), \tag{3.2.10}$$

where

$$\boldsymbol{D}^-_{i+\frac{1}{2}} = \frac{1}{2}\left((1-\alpha_{1,i+\frac{1}{2}})\boldsymbol{E}_{i+\frac{1}{2}} - \alpha_{0,i+\frac{1}{2}}(\boldsymbol{w}_{i+1} - \boldsymbol{w}_i)\right) + \boldsymbol{F}_C(\boldsymbol{w}_i), \tag{3.2.11}$$

$$\boldsymbol{D}^+_{i+\frac{1}{2}} = \frac{1}{2}\left((1+\alpha_{1,i+\frac{1}{2}})\boldsymbol{E}_{i+\frac{1}{2}} + \alpha_{0,i+\frac{1}{2}}(\boldsymbol{w}_{i+1} - \boldsymbol{w}_i)\right) - \boldsymbol{F}_C(\boldsymbol{w}_{i+1}), \tag{3.2.12}$$

UNIVERSIDAD DE MÁLAGA

and

$$\boldsymbol{E}_{i+\frac{1}{2}} = \boldsymbol{F}_C(\boldsymbol{w}_{i+1}) - \boldsymbol{F}_C(\boldsymbol{w}_i) + \boldsymbol{P}_{i+\frac{1}{2}} + \boldsymbol{T}_{i+\frac{1}{2}},$$

with

$$\boldsymbol{P}_{i+\frac{1}{2}} = \begin{pmatrix} 0 \\ \hline \mathbf{0} \\ \hline g\overline{h\theta_\alpha}\Delta\eta + \dfrac{gl_\alpha}{2}(\overline{h}\Delta(h\theta_\alpha) - \overline{h\theta_\alpha}\Delta h) + g\displaystyle\sum_{\beta=\alpha+1}^{M} l_\beta(\overline{h}\Delta(h\theta_\beta) - \overline{h\theta_\alpha}\Delta h) \end{pmatrix}.$$

Note that the subindex $i+\frac{1}{2}$ is omitted in order to simplify the notation. In this way, $\overline{h}$ stands for $\overline{h}_{i+\frac{1}{2}}$, Δh stands for $(\Delta h)_{i+\frac{1}{2}}$ and so on.

Similarly, the transfer terms between layers at the interfaces $i+\frac{1}{2}$, $\boldsymbol{T}_{i+\frac{1}{2}}$, are defined as,

$$\boldsymbol{T}_{i+\frac{1}{2}} = -\begin{pmatrix} 0 \\ \hline \dfrac{1}{l_\alpha}(\langle\overline{\theta}\rangle_{\alpha+\frac{1}{2}}\widetilde{G_{\alpha+\frac{1}{2}}} - \langle\overline{\theta}\rangle_{\alpha-\frac{1}{2}}\widetilde{G_{\alpha-\frac{1}{2}}}) \\ \hline \dfrac{1}{l_\alpha}(\langle\overline{u}\rangle_{\alpha+\frac{1}{2}}\langle\overline{\theta}\rangle_{\alpha+\frac{1}{2}}\widetilde{G_{\alpha+\frac{1}{2}}} - \langle\overline{u}\rangle_{\alpha-\frac{1}{2}}\langle\overline{\theta}\rangle_{\alpha-\frac{1}{2}}\widetilde{G_{\alpha-\frac{1}{2}}}) \end{pmatrix},$$

where the notation has been simplified again, denoting,

$$\langle\overline{\theta}\rangle_{\alpha+\frac{1}{2}} \equiv \left\langle\overline{\theta}_{i+\frac{1}{2}}\right\rangle_{\alpha+\frac{1}{2}} = \frac{1}{2}\left(\overline{\theta_{\alpha+1,i+\frac{1}{2}}} + \overline{\theta_{\alpha,i+\frac{1}{2}}}\right) = \frac{1}{4}\left(\theta_{\alpha+1,i} + \theta_{\alpha+1,i+1} + \theta_{\alpha,i} + \theta_{\alpha,i+1}\right),$$

and analogously for $\langle\overline{u}\rangle_{\alpha+\frac{1}{2}}$. Finally, the discretization of the transfer terms is,

$$\widetilde{G_{\alpha+\frac{1}{2}}} \equiv (\widetilde{G_{\alpha+\frac{1}{2}}})_{i+\frac{1}{2}} = \sum_{\beta=1}^{\alpha} l_\beta\left(\Delta(hu_\beta) - \Delta\left(\sum_{\gamma=1}^{M} l_\gamma hu_\gamma\right)\right), \quad 1 \le \alpha < M,$$

where we assume $\widetilde{G_{\frac{1}{2}}} = \widetilde{G_{M+\frac{1}{2}}} = 0$.

As discussed in Section 2.2.2, a HLL-type system needs two coefficients, $\alpha_{0,i+\frac{1}{2}}$ and $\alpha_{1,i+\frac{1}{2}}$, related to the numerical viscosity of the scheme and defined in terms of the upper

UNIVERSIDAD DE MÁLAGA

and lower bounds of the maximum and minimum of the waves speed (see [48] for more details):

$$\alpha_{0,i+\frac{1}{2}} = \frac{\lambda^+_{i+\frac{1}{2}}|\lambda^-_{i+\frac{1}{2}}| - \lambda^-_{i+\frac{1}{2}}|\lambda^+_{i+\frac{1}{2}}|}{\lambda^+_{i+\frac{1}{2}} - \lambda^-_{i+\frac{1}{2}}}, \qquad \alpha_{1,i+\frac{1}{2}} = \frac{|\lambda^+_{i+\frac{1}{2}}| - |\lambda^-_{i+\frac{1}{2}}|}{\lambda^+_{i+\frac{1}{2}} - \lambda^-_{i+\frac{1}{2}}},$$

where $\lambda^{\pm}_{i+\frac{1}{2}}$ is an approximation of the maximum and minimum wave speed defined at (1.2.9)-(1.2.11),

$$\lambda^{\pm}_{i+\frac{1}{2}} = \bar{u}_{i+\frac{1}{2}} \pm \Psi_{i+\frac{1}{2}},$$

with

$$\bar{u}_{i+\frac{1}{2}} = \frac{1}{M}\sum_{\alpha=1}^{M} \overline{u_\alpha}_{,i+\frac{1}{2}},$$

and

$$\Psi_{i+\frac{1}{2}} = \sqrt{\frac{2M-1}{2M}\left(2\sum_{\alpha=1}^{M}(\bar{u}_{i+\frac{1}{2}} - \overline{u_\alpha}_{,i+\frac{1}{2}})^2 + g\overline{h_{i+\frac{1}{2}}}\left(1 + \frac{1}{M}\sum_{\beta=1}^{M}(2\beta-1)\overline{\theta_\beta}_{,i+\frac{1}{2}}\right)\right)}.$$

3.2.2 Hydrostatic reconstruction

The numerical scheme (3.2.10) defined in the previous section is not able to preserve lake-at-rest steady states solutions since the term associated with the numerical viscosity $\alpha_{0,i+\frac{1}{2}}(\boldsymbol{w}_{i+1} - \boldsymbol{w}_i)$ does not vanish in such situations. In order to obtain a well-balanced scheme, a modified hydrostatic reconstruction technique [98] is performed. Thus, from $\boldsymbol{w}^n_i$ and $\boldsymbol{w}^n_{i+1}$, and $z_{B,i}$ and $z_{B,i+1}$ we define the reconstructed states at the cell interfaces $\boldsymbol{w}^{HR,\pm}_{i+\frac{1}{2}}$ and $z_{B,i+\frac{1}{2}}$ as follows:

$$z_{B,i+\frac{1}{2}} = \max\left(z_{B,i}, z_{B,i+1}\right), \tag{3.2.13}$$

and the reconstructed states for the water depth,

$$h^{HR,-}_{i+\frac{1}{2}} = \left(h_i + z_{B,i} - z_{B,i+\frac{1}{2}}\right)_+, \qquad h^{HR,+}_{i+\frac{1}{2}} = \left(h_{i+1} + z_{B,i+1} - z_{B,i+\frac{1}{2}}\right)_+, \tag{3.2.14}$$

where $(f)_+$ denotes the positive part of f. In this way, all the unknowns are defined at the intercells as follows:

$$\boldsymbol{w}^{HR,\pm}_{i+\frac{1}{2}} = \left(h^{HR,\pm}_{i+\frac{1}{2}} \,\middle|\, h^{HR,\pm}_{i+\frac{1}{2}}\theta_{\alpha,i} \,\middle|\, h^{HR,\pm}_{i+\frac{1}{2}}\theta_{\alpha,i}u_{\alpha,i}\right)^T \in \mathbb{R}^{2M+1}, \tag{3.2.15}$$

where $\theta_{\alpha,i}$ and $\theta_{\alpha,i}u_{\alpha,i}$ denotes the cell averages on the cell I_i of the corresponding quantities at the layer for $\alpha = 1, \ldots, M$.

UNIVERSIDAD DE MÁLAGA

The notation averages and differences in (3.2.7)–(3.2.9) are now updated in terms of the reconstructed variables as follows,

$$\overline{f} = \frac{1}{2}(f^{HR+}_{i+\frac{1}{2}} + f^{HR-}_{i+\frac{1}{2}}), \quad \Delta f = \frac{1}{2}(f^{HR+}_{i+\frac{1}{2}} - f^{HR-}_{i+\frac{1}{2}}) \tag{3.2.16}$$

and also,

$$\overline{f^+} = \frac{f^{HR,+}_{i+\frac{1}{2}} + f_{i+1}}{2}, \qquad \overline{f^-} = \frac{f_i + f^{HR,-}_{i+\frac{1}{2}}}{2}, \tag{3.2.17}$$

and finally,

$$\Delta f^+ = f^{HR,+}_{i+\frac{1}{2}} - f_{i+1}, \qquad \Delta f^- = f_i - f^{HR,-}_{i+\frac{1}{2}}. \tag{3.2.18}$$

We are now in disposition of redefining the original numerical scheme (3.2.10) as,

$$\begin{aligned} \boldsymbol{w}_i^{n+1} = \boldsymbol{w}_i^n - \frac{\Delta t}{\Delta x}\Big(& \boldsymbol{D}^+_{i-\frac{1}{2}}(\boldsymbol{w}^{HR-}_{i-\frac{1}{2}}, \boldsymbol{w}^{HR+}_{i-\frac{1}{2}}, z_{B,i-\frac{1}{2}}, z_{B,i-\frac{1}{2}}) \\ & + \boldsymbol{D}^-_{i+\frac{1}{2}}(\boldsymbol{w}^{HR-}_{i+\frac{1}{2}}, \boldsymbol{w}^{HR+}_{i+\frac{1}{2}}, z_{B,i+\frac{1}{2}}, z_{B,i+\frac{1}{2}}) + \boldsymbol{S}^+_{i-\frac{1}{2}} + \boldsymbol{S}^-_{i+\frac{1}{2}}\Big), \end{aligned} \tag{3.2.19}$$

where $\boldsymbol{D}^{\pm}_{i+\frac{1}{2}}$ are defined by (3.2.11)-(3.2.12) and the term $\boldsymbol{S}^{\pm}_{i+\frac{1}{2}}$ correspond to the correction associated with the hydrostatic reconstruction and guarantee both the consistency of the scheme and the well-balanced property,

$$\boldsymbol{S}^{\pm}_{i+\frac{1}{2}} = \boldsymbol{P}^{\pm}_{i+\frac{1}{2}} + \boldsymbol{T}^{\pm}_{i+\frac{1}{2}}.$$

Where $\boldsymbol{P}^{\pm}_{i+\frac{1}{2}}$ correspond to the pressure terms,

$$\boldsymbol{P}^{\pm}_{i+\frac{1}{2}} = \left(0 \,\middle|\, \mathbf{0} \,\middle|\, P^{\pm}_{\alpha,i+\frac{1}{2}}\right)^T \in \mathbb{R}^{2M+1},$$

with

$$P^{\pm}_{\alpha,i+\frac{1}{2}} = g \sum_{\beta=\alpha+1}^{M} l_\beta \overline{h^{\pm}} \Delta^{\pm} h(\theta_{\beta,i+(\frac{1}{2}\pm\frac{1}{2})} - \theta_{\alpha,i+(\frac{1}{2}\pm\frac{1}{2})}).$$

Remark 3.2.1. *The pressure term $P^{\pm}_{\alpha,i+\frac{1}{2}}$ comes from the evaluation of the integral of*

$$P = gh\theta_\alpha \partial_x \eta + \frac{g l_\alpha}{2}\left(h\partial_x(h\theta_\alpha) - h\theta_\alpha \partial_x h\right) + g\sum_{\beta=\alpha+1}^{M} l_\beta \left(h\partial_x(h\theta_\beta) - h\theta_\alpha \partial_x h\right),$$

between the center of the cell and the intercell along the path that defines the reconstruction (see [200, 201]). Thanks to the definitions of the state variables (3.2.15), the primitive variables u_α, and θ_α, as well as the free surface η, remain constant along the integral path, justifying the given expression of the pressure terms.

UNIVERSIDAD DE MÁLAGA

The term $\boldsymbol{T}^{\pm}_{i+\frac{1}{2}}$ is defined by

$$\boldsymbol{T}^{\pm}_{i+\frac{1}{2}} = -\begin{pmatrix} 0 \\ \hline \frac{1}{l_\alpha}\left(\left(\langle\theta\rangle_{\alpha+\frac{1}{2}}\right)_{i+(\frac{1}{2}\pm\frac{1}{2})} G^{\pm}_{\alpha+\frac{1}{2}} - \left(\langle\theta\rangle_{\alpha-\frac{1}{2}}\right)_{i+(\frac{1}{2}\pm\frac{1}{2})} G^{\pm}_{\alpha-\frac{1}{2}}\right) \\ \hline \frac{1}{l_\alpha}\left(\left(\langle u\rangle_{\alpha+\frac{1}{2}}\langle\theta\rangle_{\alpha+\frac{1}{2}}\right)_{i+(\frac{1}{2}\pm\frac{1}{2})} G^{\pm}_{\alpha+\frac{1}{2}} - \left(\langle u\rangle_{\alpha-\frac{1}{2}}\langle\theta\rangle_{\alpha-\frac{1}{2}}\right)_{i+(\frac{1}{2}\pm\frac{1}{2})} G^{\pm}_{\alpha-\frac{1}{2}}\right) \end{pmatrix}.$$

We recall that,

$$\left(\langle f\rangle_{\alpha+\frac{1}{2}}\right)_{i+(\frac{1}{2}\pm\frac{1}{2})} = \frac{1}{2}\left(f_{\alpha,i+(\frac{1}{2}\pm\frac{1}{2})} + f_{\alpha+1,i+(\frac{1}{2}\pm\frac{1}{2})}\right),$$

for any variable f. In addition, $G^{\pm}_{\alpha+\frac{1}{2}}$ is defined as

$$G^{\pm}_{\alpha+\frac{1}{2}} = \sum_{\beta=1}^{\alpha} l_\beta \Delta^{\pm} h \left(u_{\beta,i+(\frac{1}{2}\pm\frac{1}{2})} - \sum_{\gamma=1}^{M} l_\gamma u_{\gamma,i+(\frac{1}{2}\pm\frac{1}{2})}\right),$$

with $G^{\pm}_{\frac{1}{2}} = G^{\pm}_{M+\frac{1}{2}} = 0$.

The resulting numerical scheme (3.2.19) with a hydrostatic reconstruction is first order accurate in space and time, and it is able to preserve stationary solutions corresponding to the lake-at-rest with a constant density profile. This is easy to check: indeed, $\boldsymbol{D}^{\pm}_{i+\frac{1}{2}} = \boldsymbol{0}$ as the convective flux and transfer terms are zero and the pressure terms are also zero for a constant relative density and $\Delta h = h^{HR,+}_{i+\frac{1}{2}} - h^{HR,-}_{i+\frac{1}{2}} = 0$ as $h^{HR,+}_{i+\frac{1}{2}} = h^{HR,-}_{i+\frac{1}{2}}$ in a water at rest solution. It is also able to preserve positivity for a Courant number fulfilling $CFL \leq 0.5$, as the solver for the homogeneous system is the HLL scheme (see [48]).

3.2.3 Upwind approximation of the exchange terms between layers

The previous numerical scheme is not able to ensure that $\theta_\alpha \geq 1$ for $1 \leq \alpha \leq M$, which could potentially result in non-physical solutions. To illustrate this, we consider a dam break in relative density problem with $M = 4$ layers,

$$\begin{gathered} u_\alpha = 0, \qquad h(x,0) = 1 - \frac{1}{2}e^{-x^2}, \qquad z_B(x) = \frac{1}{2}e^{-x^2}, \\ \theta_\alpha(x,0) = \begin{cases} 1 & \text{if } x < 0, \\ 1.01 & \text{if } x \geq 0. \end{cases} \end{gathered} \tag{3.2.20}$$

UNIVERSIDAD DE MÁLAGA

The numerical results for this problem at time $t = 30$ seconds can be seen in Figure 3.1. As we can see, the relative density of some layers are clearly below the unity, directly contradicting (1.2.4).

This is explained because the discretization of the transfer terms is incomplete. In Subsection 1.2 it was discussed that the transfer terms must be written with additional jump terms (1.1.29) that are not simplified when the viscous terms are neglected. The upwind discretization proposed now is based on recuperating these terms. Indeed, if the jump terms at the interface are not taken into account, the information relative to the flux direction is lost since they are written as non-conservative products. In order to consider the jump terms, the following upwind discretization is proposed:

$$\boldsymbol{T}_{i+\frac{1}{2}} = \begin{pmatrix} 0 \\ \hline \dfrac{1}{l_\alpha}\left((\theta G)^{UP}_{\alpha+\frac{1}{2},i+\frac{1}{2}} - (\theta G)^{UP}_{\alpha-\frac{1}{2},i+\frac{1}{2}}\right) \\ \hline \dfrac{1}{l_\alpha}\left((u\theta G)^{UP}_{\alpha+\frac{1}{2},i+\frac{1}{2}} - (u\theta G)^{UP}_{\alpha-\frac{1}{2},i+\frac{1}{2}}\right) \end{pmatrix},$$

where

$$(\theta G)^{UP}_{\alpha+\frac{1}{2},i+\frac{1}{2}} = -\langle\theta\rangle_{\alpha+\frac{1}{2},i+\frac{1}{2}}\,\widetilde{G_{\alpha+\frac{1}{2}}} - \frac{1}{2}|\widetilde{G_{\alpha+\frac{1}{2}}}|(\overline{\theta_{\alpha+1}}_{,i+\frac{1}{2}} - \overline{\theta_\alpha}_{,i+\frac{1}{2}}),$$

$$(u\theta G)^{UP}_{\alpha+\frac{1}{2},i+\frac{1}{2}} = -\langle u\theta\rangle_{\alpha+\frac{1}{2},i+\frac{1}{2}}\,\widetilde{G_{\alpha+\frac{1}{2}}} - \frac{1}{2}|\widetilde{G_{\alpha+\frac{1}{2}}}|(\overline{(u\theta)_{\alpha+1}}_{,i+\frac{1}{2}} - \overline{(u\theta)_\alpha}_{,i+\frac{1}{2}}),$$

and

$$\boldsymbol{T}^{\pm}_{i+\frac{1}{2}} = \begin{pmatrix} 0 \\ \hline \dfrac{1}{l_\alpha}\left((\theta G)^{UP^\pm}_{\alpha+\frac{1}{2}} - (\theta G)^{UP^\pm}_{\alpha-\frac{1}{2}}\right) \\ \hline \dfrac{1}{l_\alpha}\left((u\theta G)^{UP^\pm}_{\alpha+\frac{1}{2}} - (u\theta G)^{UP^\pm}_{\alpha-\frac{1}{2}}\right) \end{pmatrix},$$

where

$$(\theta G)^{UP^\pm}_{\alpha+\frac{1}{2}} = -\langle\theta\rangle_{\alpha+\frac{1}{2},i+(\frac{1}{2}\pm\frac{1}{2})}\,G^\pm_{\alpha+\frac{1}{2}} - \frac{1}{2}|G^\pm_{\alpha+\frac{1}{2}}|(\theta_{\alpha+1,i+(\frac{1}{2}\pm\frac{1}{2})} - \theta_{\alpha,i+(\frac{1}{2}\pm\frac{1}{2})}),$$

$$(u\theta G)^{UP^\pm}_{\alpha+\frac{1}{2}} = -\langle u\theta\rangle_{\alpha+\frac{1}{2},i+(\frac{1}{2}\pm\frac{1}{2})}\,G^\pm_{\alpha+\frac{1}{2}} - \frac{1}{2}|G^\pm_{\alpha+\frac{1}{2}}|((u\theta)_{\alpha+1,i+(\frac{1}{2}\pm\frac{1}{2})} - (u\theta)_{\alpha,i+(\frac{1}{2}\pm\frac{1}{2})}),$$

UNIVERSIDAD DE MÁLAGA

with $(\theta G)^{UP}_{\frac{1}{2},i+\frac{1}{2}} = (\theta G)^{UP}_{M+\frac{1}{2},i+\frac{1}{2}} = (u\theta G)^{UP}_{\frac{1}{2},i+\frac{1}{2}} = (u\theta G)^{UP}_{M+\frac{1}{2},i+\frac{1}{2}} = 0$, and $(\theta G)^{UP\pm}_{\frac{1}{2}} = (\theta G)^{UP\pm}_{M+\frac{1}{2}} = (u\theta G)^{UP\pm}_{\frac{1}{2}} = (u\theta G)^{UP\pm}_{M+\frac{1}{2}} = 0$.

Remark 3.2.2. *As it is shown in [7], this upwind strategy can be seen as a centered approximation of the transfer terms plus the following vertical diffusion term for the relative density and momentum equations respectively,*

$$\partial_z \left(h_\alpha |G| \partial_z \theta\right), \qquad \partial_z \left(h_\alpha |G| \partial_z (u\theta)\right).$$

This vertical diffusion is proportional to $(l_\alpha + l_{\alpha+1})/2$ and improves the overall stability of the numerical scheme while also tends to zero when the number of layers tends to infinity.

If we consider again the same problem (3.2.20) with an upwind discretization of the transfer terms, we can see now in Figure 3.2 that the numerical scheme no longer produces non-physical behavior.

The numerical scheme can be reduced to a classical HLL scheme for the variable $\sum_{\alpha=1}^{M} l_\alpha h \theta_\alpha$ by summing up all the equations for $l_\alpha h \theta_\alpha$ and thus canceling the transfer terms $\boldsymbol{T}^{\pm}_{i+\frac{1}{2}}$. Therefore, it is possible to prove that $\sum_{\alpha=1}^{M} l_\alpha \theta_\alpha \geq 1$ following the usual arguments for the HLL scheme. We could not formally prove that $\theta_\alpha \geq 1$. However, this property is verified throughout all the numerical experiments performed.

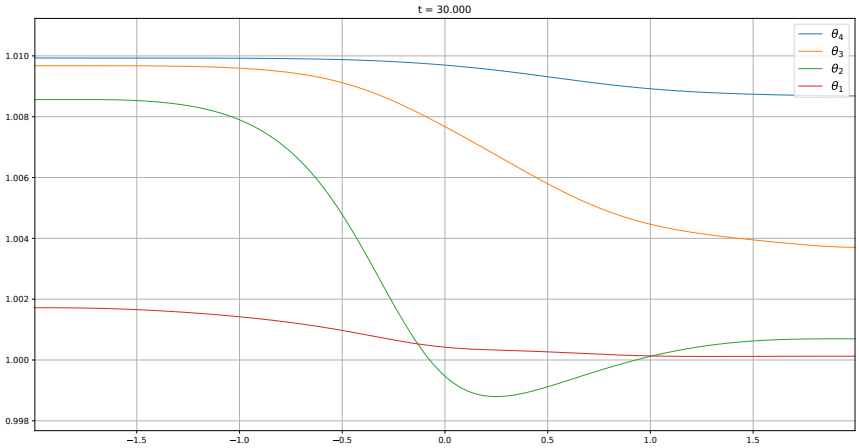


Figure 3.1: Simulation with $M = 4$ number of layers for a version of the code without an upwind approximation. The relative density θ_α is shown.

UNIVERSIDAD DE MÁLAGA

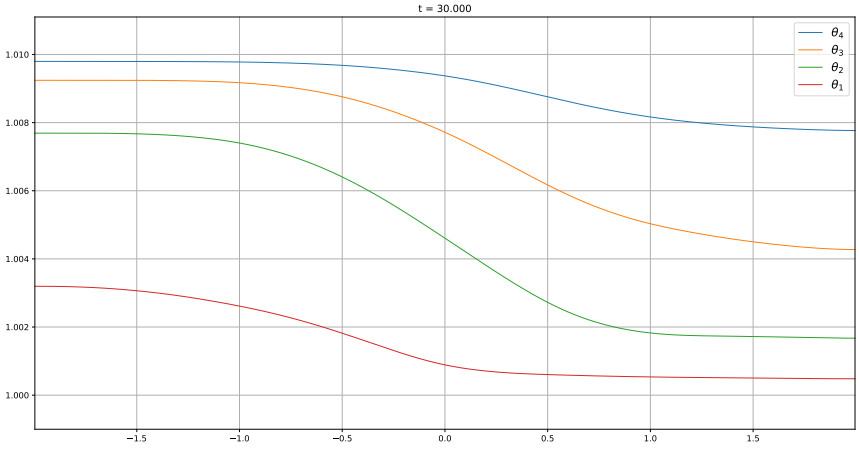


Figure 3.2: Simulation with $M = 4$ number of layers for a version of the code with an upwind approximation. The relative density θ_α is shown.

3.2.4 Second order approximation

To reach second order in space, we combine the first order numerical scheme (3.2.19) with a second order reconstruction operator. In this way, a reconstruction function $\boldsymbol{R}_i^t(x) = \boldsymbol{w}(x,t) + O(\Delta x^2),\ x \in I_i$, is defined inside each cell $I_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ at each time step t^n using the cell averages $\{\boldsymbol{w}_j(t),\ j \in \mathcal{S}_i\}$, where $\mathcal{S}_i$ is the stencil of the reconstruction operator. The following standard notation is also used,

$$\lim_{x \to x^+_{i-\frac{1}{2}}} \boldsymbol{R}_i^t(x) = \boldsymbol{w}^+_{i-\frac{1}{2}}(t), \qquad \lim_{x \to x^-_{i+\frac{1}{2}}} \boldsymbol{R}_i^t(x) = \boldsymbol{w}^-_{i+\frac{1}{2}}(t).$$

As it was discussed in Section 2.2.3 and according to [41], the second order extension of the first order numerical scheme (3.2.19) can be written as,

$$\begin{aligned} \boldsymbol{w}_i'(t) &= -\frac{1}{\Delta x}\left(\boldsymbol{D}^+_{i-\frac{1}{2}}(t) + \boldsymbol{D}^-_{i+\frac{1}{2}}(t)\right) \\ &\quad -\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \left(\boldsymbol{P}(\boldsymbol{R}_i^t, R_i^{\eta,t}, \partial_x \boldsymbol{R}_i^t, \partial_x R_i^{\eta,t}) - \boldsymbol{T}(\boldsymbol{R}_i^t, \partial_x \boldsymbol{R}_i^t)\right)\, dx, \end{aligned} \tag{3.2.21}$$

where

$$\begin{aligned} \boldsymbol{D}^+_{i-\frac{1}{2}}(t) &= \boldsymbol{D}^+_{i-\frac{1}{2}}(\boldsymbol{w}^{HR-}_{i-\frac{1}{2}}(t), \boldsymbol{w}^{HR+}_{i-\frac{1}{2}}(t), z_{B,i-\frac{1}{2}}, z_{B,i-\frac{1}{2}}) + \boldsymbol{S}^+_{i-\frac{1}{2}}, \\ \boldsymbol{D}^-_{i+\frac{1}{2}}(t) &= \boldsymbol{D}^-_{i+\frac{1}{2}}(\boldsymbol{w}^{HR-}_{i+\frac{1}{2}}(t), \boldsymbol{w}^{HR+}_{i+\frac{1}{2}}(t), z_{B,i+\frac{1}{2}}, z_{B,i+\frac{1}{2}}) + \boldsymbol{S}^-_{i+\frac{1}{2}}. \end{aligned}$$

To simplify the notation, the time dependence will be subsequentially dropped. Additionally, the components of $\boldsymbol{R}_i$ will be denoted as $\boldsymbol{R}_i = (R^h | R^{h\theta}_\alpha | R^{h\theta u}_\alpha)$.

UNIVERSIDAD DE MÁLAGA

Remark 3.2.3. *The implementation of the second order reconstruction operator with the hydrostatic reconstruction in Section* 3.2.2 *is performed by applying the hydrostatic reconstruction procedure from the reconstructed states. In this way, the hydrostatic reconstructed states would be,*

$$z_{B,i+\frac{1}{2}} = \max\left(z^-_{B,i+\frac{1}{2}}, z^+_{B,i+\frac{1}{2}}\right) \tag{3.2.22}$$

and

$$h^{HR,-}_{i+\frac{1}{2}} = \left(h^-_{i+\frac{1}{2}} + z^-_{B,i+\frac{1}{2}} - z_{B,i+\frac{1}{2}}\right)_+, \qquad h^{HR,+}_{i+\frac{1}{2}} = \left(h^+_{i+\frac{1}{2}} + z^+_{B,i+\frac{1}{2}} - z_{B,i+\frac{1}{2}}\right)_+, \tag{3.2.23}$$

where again $h^{HR} = (\cdot)_+$ *denotes the positive part and* $z^{\pm}_{B,i+\frac{1}{2}}$ *are the reconstructed values of the bathymetry* z_B *at* $x_{i+\frac{1}{2}}$ *for the cells* I_i *and* I_{i+1}.

As stated in [41], to preserve the well-balanced properties of the first order scheme, the reconstruction operators should also preserve the stationary solution corresponding to the lake-at-rest with constant relative density. Specifically, for a water at rest steady state solution, the reconstruction operator should fulfill,

$$R^h_i(x) + R^{z_B}_i(x) = C,$$

where $C \in \mathbb{R}$ is a constant. Therefore, the reconstruction operator for the bathymetry function has to be properly chosen. Particularly, the reconstruction operator corresponding to the bathymetry function is defined from the free surface reconstruction operator R^η_i, which is likewise defined from the cell averages of the surface $\{\eta_j,\ j \in \mathcal{S}_i\}$. The bathymetry reconstruction operator is then,

$$R^{z_B}_i(x) = R^\eta_i(x) - R^h_i(x). \tag{3.2.24}$$

In this way, the reconstruction operator satisfies that,

$$R^h_i(x) + R^{z_B}_i(x) = R^\eta_i(x),$$

and therefore the well-balanced property is achieved since the reconstructor operator R^η_i is exact if the states η_i are constant in the stencil.

For the finite volume method, a MUSCL reconstruction operator (as in Section 2.2.3.1) is considered. A piece-wise linear operator is defined in each cell I_i with the form,

$$\boldsymbol{R}_i(x) = \boldsymbol{w}_i + \boldsymbol{\delta}_i(x - x_i), \tag{3.2.25}$$

where x_i is the center of the cell I_i and $\boldsymbol{\delta}_i$ provides the slope of the reconstruction for each variable in (3.2.15). As stated before, in the definition of $\boldsymbol{\delta}_i$ a slope limiter must also be defined to avoid spurious oscillations near strong gradients or discontinuities, while

UNIVERSIDAD DE MÁLAGA

preserving the second order accuracy in smooth regions. To achieve this, an average limiter (avg) is used. We recall that,

$$[\boldsymbol{\delta}_i]_k = \text{avg}\left(\frac{[\boldsymbol{w}_{i+1} - \boldsymbol{w}_i]_k}{\Delta x}, \frac{[\boldsymbol{w}_i - \boldsymbol{w}_{i-1}]_k}{\Delta x}\right), \tag{3.2.26}$$

where the subindex k refers the k-th component of the vector and the avg operator is defined in (2.2.70).

The final reconstruction procedure can be summarized as follows,

1. In a first step, the water depth h and the free surface η are reconstructed using their corresponding reconstruction operators $R_i^h(x)$ and $R_i^\eta(x)$. Then, the reconstruction of the bathymetry is recovered using (3.2.24). To guarantee the positivity of the water height during the reconstruction, the technique introduced in [202] is used.

2. In the next step, the primitive variable corresponding to the relative density θ_α are reconstructed using their reconstruction operator $R_i^{\theta_\alpha}(x)$. If we denote by $\sigma_i^{\theta_\alpha}$ the slope of the reconstruction of θ_α and σ_i^h the corresponding slope for the water depth, then the full reconstruction of the conservative variable $h\theta_\alpha$ is given by,

   $$R_i^{h\theta_\alpha}(x) = (h\theta_\alpha)_i + \delta_i^{h\theta_\alpha}(x - x_i)$$

   with

   $$\delta_i^{h\theta_\alpha} = \theta_{\alpha,i}\delta_i^h + h_i\delta_i^{\theta_\alpha}.$$

   Again, we follow [202] to guarantee that $R_i^{\theta_\alpha}(x) \geq 1$, $x \in I_i$.

3. Finally, the velocity u_α is reconstructed at each cell. Similarly to the previous step, we denote by $\delta_i^{u_\alpha}$ the slope of the reconstruction of u_α at the cell I_i, and thus the slope for $h\theta_\alpha u_\alpha$ is

   $$\delta_i^{h\theta_\alpha u_\alpha} = u_{\alpha,i}\delta_i^{h\theta_\alpha} + (h\theta_\alpha)_i\delta_i^{u_\alpha}.$$

   and the final reconstruction operator is defined as,

   $$R_i^{h\theta_\alpha u_\alpha}(x) = (h\theta_\alpha u_\alpha)_i + \delta_i^{h\theta_\alpha u_\alpha}(x - x_i).$$

Note that the definition of $\delta_i^{h\theta_\alpha}$ and $\delta_i^{h\theta_\alpha u_\alpha}$ grants that

$$\delta_i^{h\theta_\alpha} = \partial_x R_i^{h\theta_\alpha} = R_i^h(x_i)\partial_x R_i^{\theta_\alpha} + R_i^{\theta_\alpha}(x_i)\partial_x R_i^h,$$

$$\delta_i^{h\theta_\alpha u_\alpha} = \partial_x R_i^{h\theta_\alpha} = R_i^{u_\alpha}(x_i)\partial_x R_i^{h\theta_\alpha} + R_i^{h\theta_\alpha}(x_i)\partial_x R_i^{u_\alpha}.$$

Finally, the integral term in (3.2.21) is approximated by the middle point quadrature rule, resulting in:

$$\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \left(\boldsymbol{P}(R_i, R_i^\eta, \partial_x R_i, \partial_x R_i^\eta) - \boldsymbol{T}(R_i, \partial_x R_i)\right) dx \approx \boldsymbol{P}_i - \boldsymbol{T}_i,$$

UNIVERSIDAD DE MÁLAGA

with

$$\boldsymbol{P}_i = \begin{pmatrix} 0 \\ \hline \mathbf{0} \\ \hline g(h\theta_\alpha)_i\delta_i^\eta + \dfrac{gl_\alpha}{2}(h_i\delta_i^{h\theta_\alpha} - (h\theta_\alpha)_i\delta_i^h) + g\displaystyle\sum_{\beta=\alpha+1}^{M} l_\beta(h_i\delta_i^{h\theta_\beta} - (h\theta_\alpha)_i\delta_i^h) \end{pmatrix}.$$

$$\boldsymbol{T}_i = \begin{pmatrix} 0 \\ \hline \dfrac{1}{l_\alpha}\left(\overline{\overline{\theta G}}_{\alpha+\frac{1}{2}} - \overline{\overline{\theta G}}_{\alpha-\frac{1}{2}}\right) \\ \hline \dfrac{1}{l_\alpha}\left(\overline{\overline{u\theta G}}_{\alpha+\frac{1}{2}} - \overline{\overline{u\theta G}}_{\alpha-\frac{1}{2}}\right) \end{pmatrix},$$

where

$$\overline{\overline{\theta G}}_{\alpha+\frac{1}{2}} = -\langle\theta\rangle_{\alpha+\frac{1}{2},i}\,\overline{G_{\alpha+\frac{1}{2},i}} - \frac{1}{2}|\overline{G_{\alpha+\frac{1}{2},i}}|(\theta_{\alpha+1,i} - \theta_{\alpha,i}),\ 1 \le \alpha < M,$$

$$\overline{\overline{u\theta G}}_{\alpha+\frac{1}{2}} = -\langle u\theta\rangle_{\alpha+\frac{1}{2},i}\,\overline{G_{\alpha+\frac{1}{2},i}} - \frac{1}{2}|\overline{G_{\alpha+\frac{1}{2},i}}|((u\theta)_{\alpha+1,i} - (u\theta)_{\alpha,i}),\ 1 \le \alpha < M,$$

with

$$\overline{G_{\alpha+\frac{1}{2}\,i}} = \sum_{\beta=1}^{\alpha} l_\beta\left(\delta_i^{hu_\beta} - \sum_{\gamma=1}^{M} l_\gamma\delta_i^{hu_\gamma}\right),$$

and

$$\delta_i^{hu_\alpha} = h_i\delta_i^{u_\alpha} + u_{\alpha,i}\delta_i^h.$$

As usual, the transfer terms associated with the free surface and bottom are considered zero, that is, $\overline{\overline{\theta G}}_{\frac{1}{2}} = \overline{\overline{u\theta G}}_{\frac{1}{2}} = 0$ and $\overline{\overline{\theta G}}_{M+\frac{1}{2}} = \overline{\overline{u\theta G}}_{M+\frac{1}{2}} = 0$.

Finally, the second order in time is achieved via a *total variation diminish* (TVD) Runge–Kutta method (see [178]).

The final numerical scheme (3.2.21) is second order accurate, well-balanced for the stationary solution corresponding to water at rest with constant density and positive preserving for the total water column h and the relative density.

UNIVERSIDAD DE MÁLAGA

3.2.5 Well-balanced for a family of stationary solutions

As discussed in Chapter 2, the well-balanced properties of the previous second order numerical scheme (3.2.21) can be enhanced to preserve a wider range of stationary solutions corresponding to the stationary solutions for non-trivial density profiles reviewed in Section 1.3.

The first stage consists on computing a stationary solution $\boldsymbol{w}_i^*(x, t^n)$ in each cell I_i at each time $t \in [0, T]$. We subsequently drop the time dependency in order to simplify the notation. The stationary solution is fully determined by $h_{e,i}$ and $\theta_{1,e,i}, \dots \theta_{M,e,i}$ since stationary solutions (1.3.6) assume $u_{\alpha,e,i} = 0$, $1 \leq \alpha \leq M$. The family of stationary solutions (1.3.6) is determined by setting some local constants $\bar{\eta}_i$, $\bar{\theta}_{\alpha,i}$, $1 \leq \alpha \leq M$. These constants must be properly chosen in such a way that the conservative properties of the schemes are kept intact. In practice, we may choose $\bar{\eta}_i = \eta_i$ and $\bar{\theta}_\alpha$ determined by (1.3.6). Once these constants are computed, the local stationary solution $\boldsymbol{w}_i^*(x)$ is determined at each cell. Note that, by definition, the local stationary solutions satisfy that the pressure terms (3.2.5) at each cell cancel,

$$\boldsymbol{P}(\boldsymbol{w}_i^*, \partial_x \boldsymbol{w}_i^*, \partial_x \bar{\eta}_i) = 0. \tag{3.2.27}$$

Once the stationary solution $\boldsymbol{w}_i^*(x, t^n)$ is computed in each cell, the reconstruction operator is written in terms of the fluctuation with respect to the stationary solution,

$$\tilde{\boldsymbol{w}}_i^n(t) = \boldsymbol{w}_i^n - \boldsymbol{w}_i^*(x_i),$$

where x_i denotes the center of the cell I_i. Note that $\boldsymbol{w}_i^*(x_i)$ is the evaluation of the stationary solution at the center of the cell and therefore it is a second order approximation of the stationary solution cell average. The next step consists on applying the reconstruction operators defined in the previous section to these fluctuations. The reconstruction operators applied to $\tilde{\boldsymbol{w}}_i^n$ are denoted as $\tilde{\boldsymbol{R}}_i(x)$. Finally, the reconstruction of the state variable $\boldsymbol{w}$ is defined as,

$$\boldsymbol{R}_i(x) = \boldsymbol{w}_i^*(x) + \tilde{\boldsymbol{R}}_i(x).$$

The quadrature formula appearing in the second order numerical scheme (3.2.21) needs to be properly computed. For a second order scheme such as this, the mid point rule is enough to compute the integral up to the desired precision, while ensuring that it is zero up to machine precision when the solution $\boldsymbol{w}_i(x, t^n)$ is a stationary solution.

Now, the second order numerical scheme (3.2.21) is able to preserve a wider range of stationary solutions, as it will be shown in Chapter 4.

UNIVERSIDAD DE MÁLAGA

3.3 An arbitrary high order discontinuous Galerkin numerical scheme

The arbitrary high order explicit one step ADER-DG numerical scheme described in Chapter 2 is now applied to the multilayer shallow-water model discussed in Chapter 1, just as we have done in the framework of finite volumes in the previous section. The system of PDE equations (3.2.1) is approximated applying the family of pure discontinuous Galerkin schemes as described in Section 2.3, providing high order of accuracy in both space and time. Particularly, the DG method is evolved in time using the ADER method, based in the computation of a high order approximation of the solution at the next time step and using this prediction in the DG scheme as a corrector.

As in the finite volume case, we consider the computational domain I to be discretized into a set of conforming elements $I_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ where N_s is the total number of cells with a constant length $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$.

We recall the notation used for the discontinuous Galerkin method. For any variable f defined in I_i, $f^{\pm}_{i+\frac{1}{2}}$ will denote the values at the left and right side of the cell interface $x_{i+\frac{1}{2}}$.

The discrete approximation of the PDE system (3.2.1) at time t^n is again denoted as $\boldsymbol{w}_h(x, t^n)$. However, now $\boldsymbol{w}_h$ is defined in terms of a piecewise polynomial of degree N_p on the spatial direction. As stated in the DG Section 2.3, the solution $\boldsymbol{w}_h$ is continuous within the cell I_i, but possibly discontinuous across cell interfaces. We remind that the space of piecewise polynomials up to degree N_p is denoted by $\mathcal{U}_h$ and that $\boldsymbol{w}_h(\cdot, t^n) \in \mathcal{U}_h$. The same nodal basis as in Section 2.3, defined by the Lagrange interpolation polynomials over the (N_p+1) Gauss-Legendre quadrature nodes on the element I_i, is adopted. In this way, the discrete solution $\boldsymbol{w}_h$ is written in terms of the nodal basis functions $\Phi_l(x)$ and some unknown degrees of freedom $\hat{\boldsymbol{w}}^n_{i,l}$,

$$\boldsymbol{w}_h(x, t^n) = \sum_l \hat{\boldsymbol{w}}^n_{i,l} \Phi_l(x) \coloneqq \hat{\boldsymbol{w}}^n_{i,l} \Phi_l(x), \qquad \text{for } x \in I_i. \tag{3.3.1}$$

In this way, taking into account the structure of $\boldsymbol{w}_h(x, t^n)$ given by (3.2.2), the water depth at time t^n would read,

$$h_h = \sum_l \hat{h}_{i,l} \Phi_l(x) = \hat{h}_{i,l} \Phi_l(x), \qquad \text{for } x \in I_i. \tag{3.3.2}$$

The same is done for the other conserved variables. In these expressions, the Einstein summation convention over two repeated indices has been applied. Regarding the spatial basis functions over a reference interval in space and time, the same considerations as in Section 2.3 are made.

The same DG method (2.3.4) is applied, now particularized to the multilayer shallow-

UNIVERSIDAD DE MÁLAGA

water model with variable pressure (1.2.7),

$$\int_{t^n}^{t^{n+1}} \int_{I_i} \Phi_k \partial_t \boldsymbol{w}\, dxdt + \int_{t^n}^{t^{n+1}} \int_{I_i} \Phi_k \left(\partial_x \boldsymbol{F}_C(\boldsymbol{w})\, dxdt + \boldsymbol{P}(\boldsymbol{w}, \partial_x \boldsymbol{w}, \partial_x \eta) - \boldsymbol{T}(\boldsymbol{w}, \partial_x \boldsymbol{w})\right) dxdt = \boldsymbol{0}. \quad (3.3.3)$$

As before, using (3.3.1), integrating the first term by parts in time and integrating the flux derivative in space and taking into account the local space-time predictor solution $\boldsymbol{q}_h$ instead of $\boldsymbol{w}_h$, the weak formulation (3.3.3) can be rewritten as,

$$\begin{aligned}\left(\int_{I_i} \Phi_k \Phi_l\, dx\right)\left(\hat{\boldsymbol{w}}_{i,l}^{n+1} - \hat{\boldsymbol{w}}_{i,l}^{n}\right) - \int_{t^n}^{t^{n+1}} \int_{I_i^\circ} \Phi_k' \boldsymbol{F}_C(\boldsymbol{q}_h)\, dxdt \\ + \int_{t^n}^{t^{n+1}} \Phi_{k,i+\frac{1}{2}} \boldsymbol{D}_{i+\frac{1}{2}}^{-}\left(\boldsymbol{q}_{h,i+\frac{1}{2}}^{-}, \boldsymbol{q}_{h,i+\frac{1}{2}}^{+}, z_{b}{}_{h,i+\frac{1}{2}}^{-}, z_{b}{}_{h,i+\frac{1}{2}}^{+}\right) dt \\ + \int_{t^n}^{t^{n+1}} \Phi_{k,i-\frac{1}{2}} \boldsymbol{D}_{i-\frac{1}{2}}^{+}\left(\boldsymbol{q}_{h,i-\frac{1}{2}}^{-}, \boldsymbol{q}_{h,i-\frac{1}{2}}^{+}, z_{b}{}_{h,i-\frac{1}{2}}^{-}, z_{b}{}_{h,i-\frac{1}{2}}^{+}\right) dt \\ + \int_{t^n}^{t^{n+1}} \int_{I_i^\circ} \Phi_k \left(\boldsymbol{P}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h, \partial_x \eta_h) - \boldsymbol{T}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h)\right) dxdt = \boldsymbol{0}, \quad (3.3.4)\end{aligned}$$

where I_i° denotes the interior of I_i and η_h is the projection of η onto the space $\mathcal{U}_h$. Moreover, $\boldsymbol{D}_{i\pm\frac{1}{2}}^{\pm}$ stands for the numerical flux approximation at the cell interfaces meanwhile $z_{b}{}_{h,i\pm\frac{1}{2}}^{\pm}$ are the extrapolated values of the bathymetry at the intercells. Additionally, the first integral term in the weak formulation (3.3.4) corresponds to the element mass matrix, which is diagonal since our basis is orthogonal. We would like to stress that, due to the polynomial nature of the representation of the bathymetry function z_b at each cell, $z_{b}{}_{h,i+\frac{1}{2}}^{-} \neq z_{b}{}_{h,i+\frac{1}{2}}^{+}$ in general.

Since the discrete solution is allowed to jump across elements interfaces, it is natural to approach the numerical flux at the cell interfaces by an approximate Riemann solver. Particularly, the hydrostatic path-conservative Riemann solver described in Subsection 3.2.2 is considered. On this occasion, the hydrostatic reconstruction solver is enunciated in terms of the intercell values $\boldsymbol{q}_{h,i+\frac{1}{2}}^{\pm}$. Likewise, for the transfer terms between layers, an upwind discretization is also considered, as described in Section 3.2.3.

3.3.1 ADER-DG space-time predictor

In order to evolve the DG scheme in time, an ADER discretization is chosen. We recall, as stated in Section 2.3, that the predictor solution $\boldsymbol{q}_h(x,t)$ is a high order approximation

UNIVERSIDAD DE MÁLAGA

of the solution at time t^{n+1} based on a weak formulation of the governing PDE system in space and time. This approximation is computed by means of a Cauchy problem *in the small*, i.e, without any interaction with the neighbors states. The first step consists on expanding the predictor solution $\boldsymbol{q}_h$ within an element I_i in terms of a local space-time basis,

$$\boldsymbol{q}_h(x,t) = \sum_l \theta_l(x,t)\hat{\mathbf{q}}_l^i \coloneqq \theta_l(x,t)\hat{\mathbf{q}}_l^i\,, \tag{3.3.5}$$

with the multi-index $l = (l_0, l_1)$ and where the space-time basis function $\theta_l(x,t) = \varphi_{l_0}(\tau)\varphi_{l_1}(\xi)$ is again generated from the same one-dimensional nodal basis functions as before, i.e. the Lagrange interpolation polynomials of degree N passing through $N+1$ Gauss-Legendre quadrature nodes. The same time and spatial mapping for the reference element is used. Finally, the ADER proceeding (2.3.9) and (2.3.10) is applied to the multilayer shallow-water system. Therefore, the PDE system (3.2.1) is multiplied by a space-time test function θ_k and integrated over the space time control volume $I_i \times [t^n, t^{n+1}]$:

$$\begin{aligned}
&\int_{t^n}^{t^{n+1}} \int_{I_i} \theta_k(x,t)\partial_t \boldsymbol{q}_h \, dxdt \\
&\quad + \int_{t^n}^{t^{n+1}} \int_{I_i} \theta_k(x,t) \left(\partial_x \boldsymbol{F}_C(\boldsymbol{q}_h) + \boldsymbol{P}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h, \partial_x \eta_h) - \boldsymbol{T}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h)\right) dxdt = \mathbf{0},
\end{aligned} \tag{3.3.6}$$

where again η_h stands for the projection of the free surface η onto the local space-time polynomials, $\mathcal{U}_h$.

Equation (3.3.6) can be expanded as (2.3.10), yielding the following local expression:

$$\begin{aligned}
&\int_{I_i} \theta_k(x,t^{n+1})\boldsymbol{q}_h(x,t^{n+1})\, dx \\
&\qquad - \int_{I_i} \theta_k(x,t^n)\boldsymbol{q}_h^0(x,t^n)\, dx - \int_{t^n}^{t^{n+1}} \int_{I_i} \partial_t \theta_k(x,t)\boldsymbol{q}_h(x,t)\, dxdt \\
&= -\int_{t^n}^{t^{n+1}} \int_{I_i} \theta_k(x,t) \left(\partial_x \boldsymbol{F}_C(\boldsymbol{q}_h) + \boldsymbol{P}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h, \partial_x \eta_h) - \boldsymbol{T}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h)\right) dxdt.
\end{aligned} \tag{3.3.7}$$

This expression is similar to the general version derived in Subsection 2.3.2. It consists on a local nonlinear system for the unknown degrees of freedom $\hat{\mathbf{q}}_l^i$ of the space-time polynomials $\boldsymbol{q}_h$, the solution of which can be found via a simple and fast converging fixed point iteration algorithm. In the aforementioned subsection, it was discussed how the choice of the initial guess can have a great impact on the convergence speed of the algorithm. In this case, the initial guess is simply $\boldsymbol{q}_h^0(x,t) = \boldsymbol{w}_h(x,t^n)$.

3.3.2 A posteriori subcell finite volume limiter

Thus far, the ADER-DG numerical scheme proposed is unlimited, and consequently there is no mechanism preventing the appearance of spurious oscillations near strong gradients

UNIVERSIDAD DE MÁLAGA

or discontinuities associated with high order. Of course, there is no need to apply any limiter when the solution is smooth, but detecting and limiting cells in the proximity of shocks or steep gradients is of capital importance. In Chapter 2, two different limiting strategies were introduced: the MOOD and the WENO limiter for DG schemes. For the multilayer shallow-water model, the MOOD approach has been chosen. In the following, we proceed to recall its most important features and its particularities for the multilayer shallow-water model with variable pressure.

The limiting technique considers the unlimited approximation of the hyperbolic system (3.2.1) as a candidate solution, denoted as $\boldsymbol{w}_h^c(x, t^{n+1})$. This solution will be evaluated and it will remain unchanged if deemed adequate. However, if the solution is not admissible, then it is overridden by approximating the solution of the PDE system (3.2.1) by a fully discrete second order accurate MUSCL-Hancook finite volume method numerical scheme, as described in Subsection 2.2.3.2.

To achieve this, the solution at time t^n is projected into a subgrid of K_s elements in I_i and denoted by $\mathcal{S}_{i,j}$, verifying that $I_i = \bigcup_j \mathcal{S}_{i,j}$ for $j = 1, \ldots, K_s$. The projection is denoted by $\boldsymbol{v}_h(x, t^n)$, and consists of a set of piecewise constant subcell averages that are an L_2 projection that preserves the mean of $\boldsymbol{w}_h(x, t^n)$ in $\mathcal{S}_{i,j}$,

$$\boldsymbol{v}_h(x, t^n) = \frac{1}{|\mathcal{S}_{i,j}|} \int_{\mathcal{S}_{i,j}} \boldsymbol{w}_h(x, t^n)\, dx, \quad \forall x \in \mathcal{S}_{i,j} \subset I_i. \tag{3.3.8}$$

This subcell averages $\boldsymbol{v}_h(x, t^n)$ are evolved with an explicit second order finite volume solver similar to the one described in Section 3.2. The solution of the finite volume scheme $\boldsymbol{v}_h(x, t^{n+1}) \in I_i$ is a set of piecewise values that has to be reconstructed into a limited N_p degree polynomial. This is achieved through a classical least square reconstruction operator that preserves the average of the projected solutions,

$$\int_{\mathcal{S}_{i,j}} \boldsymbol{w}_h(x, t^{n+1})\, dx = \int_{\mathcal{S}_{i,j}} \boldsymbol{v}_h(x, t^{n+1}) dx, \quad \mathcal{S}_{i,j} \subset I_i. \tag{3.3.9}$$

Since a subcell resolution $K_s > N_p + 1$ is admitted, this problem may be overdetermined. Hence, the following constraint may also be considered,

$$\int_{I_i} \boldsymbol{w}_h(x, t^{n+1})\, dx = \int_{I_i} \boldsymbol{v}_h(x, t^{n+1}) dx. \tag{3.3.10}$$

In this way, in regions where the solution is found unsatisfactory, the reconstructed solution is used instead, while no special action is needed when the solution is smooth. Note that the same cell may be troubled through two consecutive time iterations. In this case, the initial data for the finite volume solver is $\boldsymbol{v}_h(x, t^{n+1})$ and not the projected DG polynomial (3.3.9).

As discussed in Subsection (2.3.2), the MOOD limiter allows to verify a wide range of physical and numerical properties. For the multilayer shallow-water model with variable pressure, one numerical criterion and two physical criteria are considered.

UNIVERSIDAD DE MÁLAGA

Physical admissibility criteria The physical demands for the candidate solution $\boldsymbol{w}_h^c(x, t^{n+1})$ are related with the positivity of both the water column height h and the relative density θ. According to (1.2.4), the relative density θ must always be greater or equal than one.

Numerical admissibility criteria In order to detect discontinuities, a relaxed maximum principle is used (see [91]). This is applied in *a posteriori* manner as follows,

$$\min_{y\in\mathcal{V}_i}(\boldsymbol{v}_h(y,t^n)) - \boldsymbol{\delta} \leq \boldsymbol{v}_h(x,t^{n+1}) \leq \max_{y\in\mathcal{V}_i}(\boldsymbol{v}_h(y,t^n)) + \boldsymbol{\delta}, \quad \forall x \in I_i \tag{3.3.11}$$

where the discrete form of the polynomial $\boldsymbol{w}_h(x, t^{n+1})$ is used, $\boldsymbol{\delta}$ is a small value that relaxes the criterion to allow some very small overshoot or undershoot and avoid roundoff errors that would arise if (3.3.11) is applied strictly. $\mathcal{V}_i$ is a set containing I_i and its neighbors cells.

Finally, the choice of a suitable number of subcells K_s has to be considered. In this work, the subgrid $K_s = 2N_p + 1$ is considered. As discussed in Subsection 2.3.2, this choice is optimal in the sense that it allows to keep the same time step of the DG method while maintaining a CFL number close to the unity for the finite volume solver. Also, note that it is important to update the flux in the non-troubled cells to be consistent with the flux calculated in the troubled cell, so that we keep intact the conservation properties of the numerical scheme.

3.3.3 Preserving stationary solutions in the ADER-DG framework

In this section, we use the technique presented in Section 2.3.4 to preserve a parametric family of stationary solutions of the variable pressure shallow water model given by (1.3.6) in the framework of ADER-DG numerical methods. This family of solutions are a discrete version for the shallow water framework of stratified stationary solutions with $u_\alpha = 0$ for $\alpha = 1, \ldots, M$.

The first step consists on computing a local stationary solution given by (1.3.6). We remark that this stationary solution, denoted by $\boldsymbol{w}_{i,h}^*(x),\ x \in I_i$, is computed in each cell locally for all time step t^n. Nevertheless, the time dependence of the stationary solution that reflects that it is computed in each time step is dropped in order to simplify the notation. The stationary solution (1.3.6) assumes $u_{\alpha,e,i} = 0$, $1 \leq \alpha \leq M$ and is defined in terms of cell constants $\bar{\eta}_i$ and $\bar{\theta}_{1,e,i}, \ldots \bar{\theta}_{M,e,i}$. These constants are computed so that they preserve the conservative properties of the scheme. Particularly, for the free surface this translates as,

$$\bar{\eta}_i = \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} (h_h(x,t^n) + z_{bh}(x))\, dx,$$

UNIVERSIDAD DE MÁLAGA

where we have denoted by f_h the discrete representation of f onto the polynomial space $\mathcal{U}_h$. Similarly, the constants $\bar{\theta}_{1,e,i}, \dots \bar{\theta}_{M,e,i}$ are computed as in (2.2.93) and must fulfill,

$$\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} (h\theta)_{\alpha,e,i}(x, \bar{\eta}_i, \bar{\theta}_{\alpha,i}, \cdots, \bar{\theta}_{1,i})\, dx = \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} (h\theta)_{h,\alpha}(x, t^n)\, dx, \quad 1 \leq \alpha \leq M.$$

With the computation of these constants, it is now possible to determine the stationary solution $\boldsymbol{w}^*_{i,h}(x)$. Note that, by definition, the local stationary solutions satisfies that the pressure terms (1.3.4) vanish at each cell,

$$\boldsymbol{P}(\boldsymbol{w}^*_{i,h}, \partial_x \boldsymbol{w}^*_{i,h}, \partial_x \bar{\eta}_i) = 0. \tag{3.3.12}$$

As in (2.3.37), the final arbitrary high order explicit well-balanced ADER-DG numerical scheme applied to the multilayer shallow-water model with variable pressure is,

$$\begin{aligned}\left(\int_{I_i} \Phi_k \Phi_l\, dx\right)\left(\hat{\boldsymbol{w}}^{n+1}_{i,l} - \hat{\boldsymbol{w}}^{n}_{i,l}\right) - \int_{t^n}^{t^{n+1}}\int_{I_i^\circ} \left(\Phi'_k \boldsymbol{F}_C(\boldsymbol{q}_h) - \Phi_k \boldsymbol{T}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h)\right)\, dxdt \\ + \int_{t^n}^{t^{n+1}} \Phi_{k,i+\frac{1}{2}} \boldsymbol{D}^-_{i+\frac{1}{2}}\left(\boldsymbol{q}^-_{h,i+\frac{1}{2}}, \boldsymbol{q}^+_{h,i+\frac{1}{2}}, z_{b^-_{h,i+\frac{1}{2}}}, z_{b^+_{h,i+\frac{1}{2}}}\right)\, dt \\ + \int_{t^n}^{t^{n+1}} \Phi_{k,i-\frac{1}{2}} \boldsymbol{D}^+_{i-\frac{1}{2}}\left(\boldsymbol{q}^-_{h,i-\frac{1}{2}}, \boldsymbol{q}^+_{h,i-\frac{1}{2}}, z_{b^-_{h,i-\frac{1}{2}}}, z_{b^+_{h,i-\frac{1}{2}}}\right)\, dt \\ + \int_{t^n}^{t^{n+1}}\int_{I_i^\circ} \Phi_k \left(\boldsymbol{P}(\boldsymbol{q}_h, \partial_x \boldsymbol{q}_h, \partial_x \eta_h) - \boldsymbol{P}(\boldsymbol{w}^*_{i,h}\partial_x \boldsymbol{w}^*_{i,h}, \partial_x \bar{\eta}_i)\right)\, dxdt = \boldsymbol{0}.\end{aligned} \tag{3.3.13}$$

The final step to ensure the well-balanced property of the ADER-DG scheme (3.3.13) is to assure that the extrapolated values $\boldsymbol{q}^{\pm}_{h,i\pm\frac{1}{2}}$ are properly computed. This can be achieved by extrapolating the fluctuation with respect to the stationary solution,

$$\tilde{\boldsymbol{q}}^-_{h,i+\frac{1}{2}} = (\boldsymbol{q}_{h,i} - \boldsymbol{w}^*_{i,h})(x_{i+\frac{1}{2}}).$$

The final extrapolated value at the cell interface is recuperated with,

$$\boldsymbol{q}^-_{h,i+\frac{1}{2}} = \boldsymbol{w}^*_i(x_{i+\frac{1}{2}}) + \tilde{\boldsymbol{q}}^-_{h,i+\frac{1}{2}}. \tag{3.3.14}$$

Of course, this procedure has to be performed in the neighbor cell as well,

$$\boldsymbol{q}^+_{h,i+\frac{1}{2}} = \boldsymbol{w}^*_{i+1}(x_{i+\frac{1}{2}}) + \tilde{\boldsymbol{q}}^+_{h,i+\frac{1}{2}}, \tag{3.3.15}$$

where

$$\tilde{\boldsymbol{q}}^+_{h,i+\frac{1}{2}} = (\boldsymbol{q}_{h,i+1} - \boldsymbol{w}^*_{i+1,h})(x_{i+\frac{1}{2}}).$$

UNIVERSIDAD DE MÁLAGA

Finally, for the bathymetry function we have that $z_{b\,h,i+\frac{1}{2}}^{\pm} = z_b(x_{i+\frac{1}{2}})$. Note that if the bathymetry is a smooth function, it remains continuous across the cell interfaces and the hydrostatic reconstruction is no longer needed.

The high order space-time predictor ADER must also be modified to preserve stationary solutions. The algorithm computes now the fluctuation with respect to the local stationary solution $\boldsymbol{w}_{i,h}^*$. Particularly, the following well-balanced space-time predictor is considered,

$$\begin{aligned}\int_{I_i} \theta_k(x,t^{n+1})\tilde{\boldsymbol{q}}_h(x,t^{n+1})\,dx - \int_{I_i} \theta_k(x,t^n)\tilde{\boldsymbol{q}}_h^0(x,t^n)\,dx - \int_{t^n}^{t^{n+1}}\int_{I_i} \partial_t\theta_k(x,t)\tilde{\boldsymbol{q}}_h(x,t)\,dxdt \\ = -\int_{t^n}^{t^{n+1}}\int_{I_i} \theta_k(x,t)\left(\partial_x \boldsymbol{F}_C(\boldsymbol{q}_h) - \boldsymbol{T}(\boldsymbol{q}_h,\partial_x\boldsymbol{q}_h)\right)\,dxdt \\ -\int_{t^n}^{t^{n+1}}\int_{I_i} \theta_k(x,t)\left(\boldsymbol{P}(\boldsymbol{q}_h,\partial_x\boldsymbol{q}_h,\partial_x\eta_h) - \boldsymbol{P}(\boldsymbol{w}_{i,h}^*,\partial_x\boldsymbol{w}_{i,h}^*\partial_x\bar{\eta}_i)\right)\,dxdt, \quad (3.3.16)\end{aligned}$$

where $\tilde{\boldsymbol{q}}_h(x,t)$ is the projection into $\mathcal{U}_h$ of the fluctuation of the stationary solution $\boldsymbol{w}_{i,h}^*(x)$,

$$\tilde{\boldsymbol{q}}_h(x,t) = \boldsymbol{q}_h(x,t) - \boldsymbol{w}_{i,h}^*(x).$$

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD
DE MÁLAGA

Chapter 4

Numerical tests

4.1 Introduction

Throughout this thesis, two different numerical discretizations following the finite volume and discontinuous Galerkin approaches has been derived for a multilayer shallow-water model with variable density. In this chapter, we aim to provide a wide range of experiments to show the numerical properties and general behavior of both schemes.

A number of simulations are considered for the finite volume and DG methods, and they are depicted side by side for a better comparison. The experiments are designed so that they emulate the potential situations that may be found in geophysical flows, especially the most challenging ones. In this way, perturbation of a stationary solution, density driven flows over a bump or the evolution of a smooth distribution of relative density are considered, among others. Additionally, such simulations contribute to show the well-balanced character of the numerical methods. For all test simulation in this chapter, the MOOD limiter strategy reviewed in Section 3.3.2 is used for the DG numerical scheme. For the finite volume case, the simulations are limited using the *avg* limiter (3.2.26).

Of particular interest is a dam-break simulation within a channel were empirical data is available. The data agreement between the numerical approximation and the laboratory solution is excellent, provided the correct number of layers are considered. This experiment, along the others, allow to estimate the conditions for the numerical method to accurately solve geophysical flows.

Finally, note that all units considered in this chapter are in the SI of units.

4.2 Order of accuracy test

We now perform several numerical test to numerically check the order of accuracy of the arbitrary high order ADER-DG method and the finite volume method. Particularly, the

UNIVERSIDAD DE MÁLAGA

second, third and fourth order versions have been chosen for the ADER-DG solver, while the finite volume approach correspond to second order accuracy. A computational domain $I = [-5, 5]$ with $M = 5$ layers is chosen with $25, 50, 100, 200$ and 400 discretization points. The initial condition is described by:

$$u_\alpha(x, 0) = 0, \quad \theta_\alpha(x, 0) = 1 + \frac{1}{100}\, e^{-5x^2}, \quad 1 \leq \alpha \leq 5,$$

and

$$\eta(x, 0) = 1 + \frac{1}{10}\, e^{-10\, x^2},$$

while the bathymetry is given by,

$$z_b(x) = \frac{1}{2}\, e^{-x^2}.$$

The final simulation time is $t = 0.5$ s and periodic boundary conditions are considered. The free surface and the velocity are depicted in Figure 4.1 at the final simulation time $t = 0.5$ s for the fourth order scheme using the 200 cells mesh.

The convergence results are shown in Table 4.1 for the finite volume method and Table 4.2 for the ADER-DG method Note that we have chosen the variables at the bottom because of their greater exposure to pressure effects. Similar results are observed for other variables. The errors and order of accuracy have been obtained by comparing to a reference solution computed with the same scheme with 2400 cells for the ADER-DG method and 3200 cells for the finite volume method. The expected order of accuracy is achieved. Note that the large errors associated with the numerical scheme with very coarse meshes can yield unrealistic results. However, the overall tendency of the numerical scheme to the desired accuracy is not affected.

	h		$h\theta_1$		$h\theta_1 u_1$	
N_s	**Error**	**Order**	**Error**	**Order**	**Error**	**Order**
25	5.97×10^{-02}	-	5.97×10^{-03}	-	1.79×10^{-01}	-
50	4.15×10^{-02}	0.53	3.24×10^{-03}	0.88	1.22×10^{-01}	0.56
100	1.67×10^{-02}	1.31	8.20×10^{-04}	1.99	4.57×10^{-02}	1.41
200	4.50×10^{-03}	1.89	1.32×10^{-04}	2.64	1.18×10^{-02}	1.96
400	1.04×10^{-03}	2.11	1.69×10^{-05}	2.96	2.50×10^{-03}	2.24

Table 4.1: Numerical convergence results of the finite volume scheme of second order.

UNIVERSIDAD DE MÁLAGA

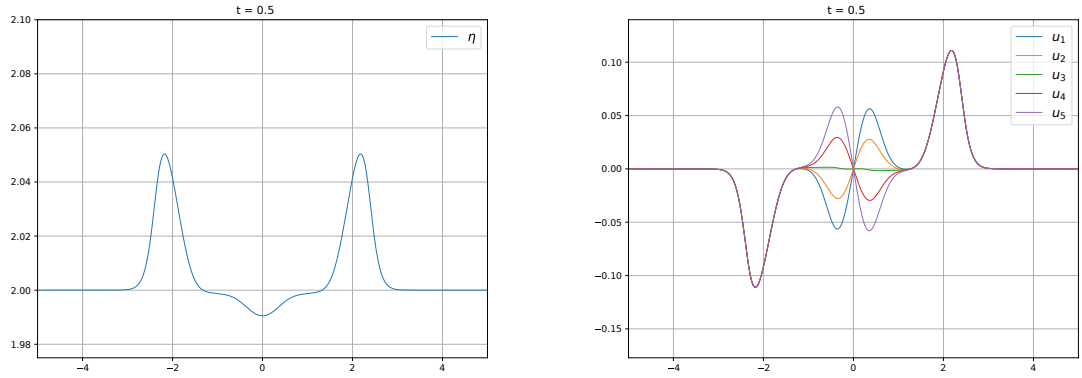


Figure 4.1: Accuracy test for $t = 0.5$ s. Spatial distribution of the free surface (left) and velocity profiles (right) computed with the ADER-DG numerical scheme.

		h		$h\theta_1$		$h\theta_1 u_1$	
N	N_s	Error	Order	Error	Order	Error	Order
1	25	2.24×10^{-02}	-	2.23×10^{-02}	-	9.87×10^{-02}	-
	50	6.45×10^{-03}	1.79	6.50×10^{-03}	1.78	3.00×10^{-02}	1.72
	100	7.19×10^{-04}	3.16	7.30×10^{-04}	3.15	3.30×10^{-03}	3.19
	200	1.03×10^{-04}	2.81	1.05×10^{-04}	2.79	4.69×10^{-04}	2.81
	400	2.02×10^{-05}	2.34	2.09×10^{-05}	2.33	9.23×10^{-05}	2.35
2	25	2.38×10^{-03}	-	2.38×10^{-03}	-	1.08×10^{-02}	-
	50	3.30×10^{-04}	2.85	3.33×10^{-04}	2.84	1.62×10^{-03}	2.74
	100	4.37×10^{-05}	2.92	4.56×10^{-05}	2.87	2.24×10^{-04}	2.86
	200	4.80×10^{-06}	3.19	5.41×10^{-06}	3.07	2.45×10^{-05}	3.19
	400	6.06×10^{-07}	2.98	7.71×10^{-07}	2.81	3.06×10^{-06}	3.00
3	25	1.60×10^{-03}	-	1.60×10^{-03}	-	7.21×10^{-03}	-
	50	9.03×10^{-05}	4.14	9.04×10^{-05}	4.14	4.08×10^{-04}	4.14
	100	4.72×10^{-06}	4.26	4.73×10^{-06}	4.26	2.14×10^{-05}	4.26
	200	2.75×10^{-07}	4.10	2.76×10^{-07}	4.10	1.24×10^{-06}	4.10
	400	1.68×10^{-08}	4.04	1.68×10^{-08}	4.04	7.56×10^{-08}	4.04

Table 4.2: Numerical convergence results of the ADER-DG scheme of order $N+1 = 2, 3, 4$.

UNIVERSIDAD DE MÁLAGA

4.3 Well-balanced tests

The well-balanced property of the finite volume and ADER-DG approaches is tested for the variable density shallow-water model. We begin by a simple lake-at-rest solution with constant density. A computational domain $I = [-5, 5]$ with 100 volumes cells and $M = 3$ layers is used. As initial condition, we fix a constant free surface $\eta = 2$, and bottom topography given by

$$z_b(x) = \frac{1}{2}\, e^{-x^2}.$$

No-slip reflecting boundary conditions are set and the relative density is constant across the domain and equal to one. Figure 4.2 shows the results for the finite volume method. Similar results can be obtained with the ADER-DG solver.

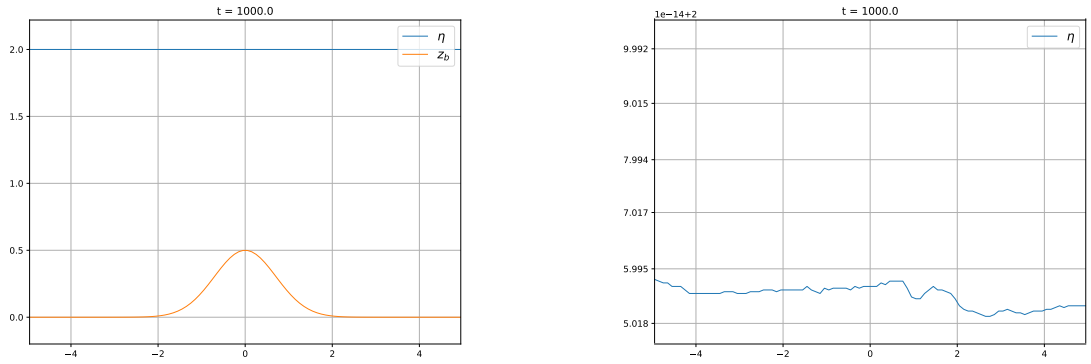


Figure 4.2: Spatial distribution of the free surface and bathymetry (left) and zoom of the free surface (right) at final time $t = 1000$ s for a well-balanced finite volume solver.

A more challenging simulation consist on preserving stratified solutions described by (1.3.6). A particular solution of this family for the three layer case is given by,

$$\begin{aligned} u_\alpha = 0, \quad \eta(x) = 1, \quad z_b(x) = \tfrac{1}{2}\, e^{-x^2}, \;\; h(x) = 1 - z_b(x), \\ \theta_1(x) &= h(x)^4\, \bar{\theta}_3 + 3h(x)^2\, \bar{\theta}_2 + \bar{\theta}_1, \\ \theta_2(x) &= h(x)^2\, \bar{\theta}_2 + \bar{\theta}_1, \\ \theta_3(x) &= \bar{\theta}_1. \end{aligned} \tag{4.3.1}$$

Here, the constant values $\bar{\theta}_i$ are chosen so that a stable stratified profile with $\theta_3(x) \leq \theta_2(x) \leq \theta_1(x)$ is obtained. Particularly, $\bar{\theta}_1 = 1.01$, $\bar{\theta}_2 = 0.02$ and $\bar{\theta}_3 = 0$ are chosen. The rest of the simulation characteristics remain unchanged. Figure 4.3 depicts the initial condition while Figures 4.4-4.5 show how the solver is able to preserve this kind of solution

UNIVERSIDAD DE MÁLAGA

for long simulations time. On this occasion, the fourth order well-balanced ADER-DG scheme has been used, although similar results can be obtained with the finite volume approach.

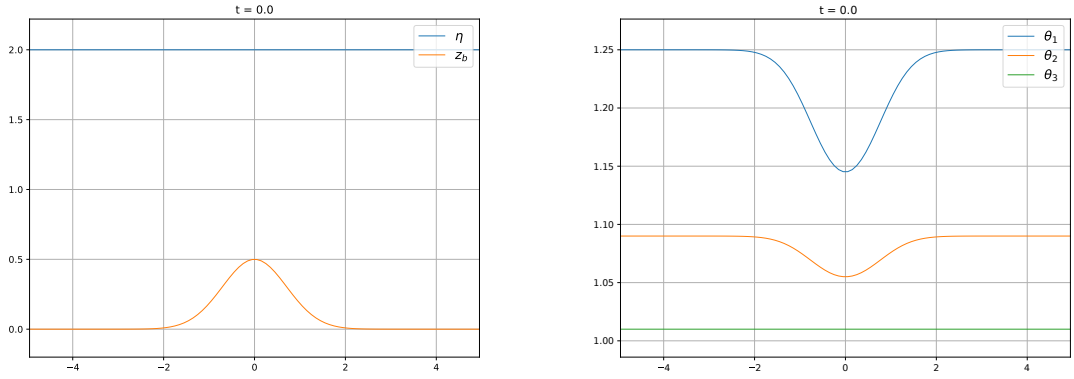


Figure 4.3: Spatial distribution of a lake-at-rest steady state with non-constant density profile. Left: surface and bottom. Right: relative density for each layer.

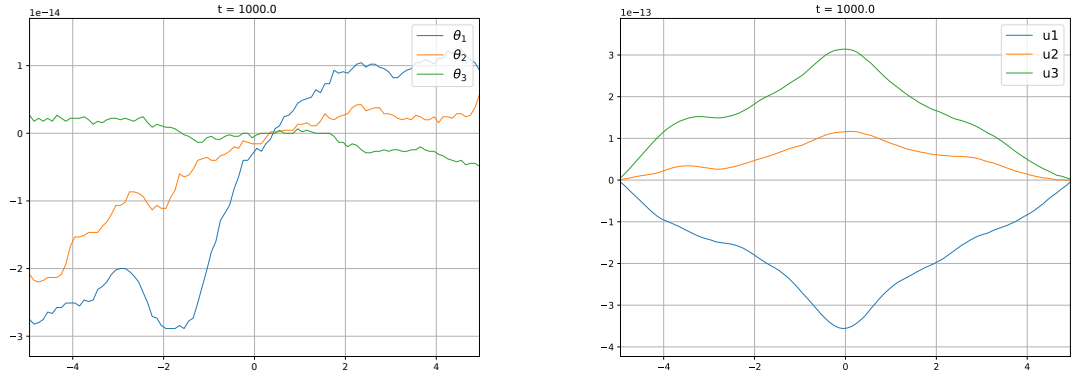


Figure 4.4: Spatial distribution of the difference between computed solution at time $t = 1000$ s and the original steady state for a lake-at-rest steady state with non-constant density profile for the ADER-DG numerical scheme. Left: difference on the relative densities. Right: difference on the velocities.

Next, a new test is considered consisting on a small perturbation of the previous test. The same considerations relative to the initial conditions, boundary conditions and the

UNIVERSIDAD DE MÁLAGA

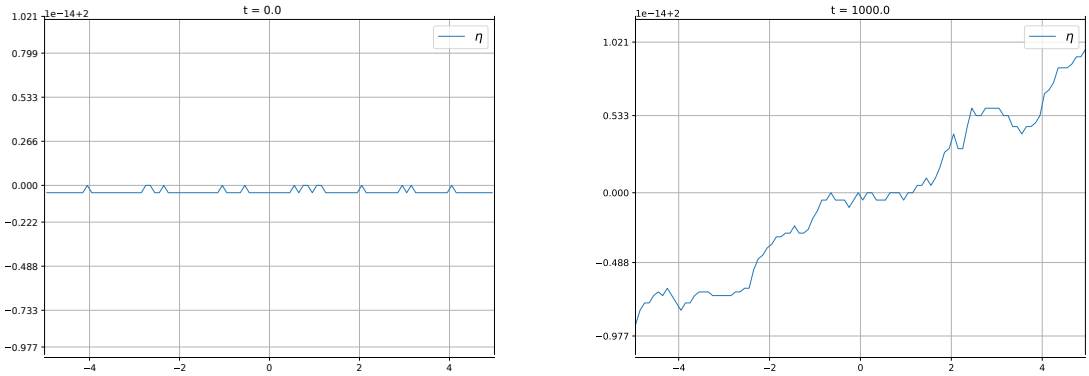


Figure 4.5: Spatial distribution of the free surface at time $t = 0$ s (left) and $t = 1000$ s (right) for a lake-at-rest steady state with non-constant density profile.

numerical scheme are now applied, but the free surface function is now given by

$$\eta(x, 0) = 2 + \frac{1}{10} e^{-5\,x^2}.$$

We use the well-balanced fourth order ADER-DG numerical scheme. Note that in such a high order method, a perturbation can take a very long time to dissipate due to numerical viscosity. For this reason, additional friction terms have been added.

The results are shown in Figures 4.6 to 4.9. As it can be seen, the numerical scheme reaches a stationary solution different from the one initially considered, since this time the perturbation does not leave the domain because of the periodic boundary conditions. As expected, the stationary free surface achieved at final time $t = 1000$ s has increased, due to the initial perturbation (see figure 4.8 left). The final velocity profiles can be seen in Figure 4.8 (right), and they are close to zero, showing that a stationary solution with a new stratification has been reached.

Note that in order to recover a stratified solution, it is necessary to preserve all possible stationary solutions corresponding to a density stratification. Indeed, if only one particular stationary solution is preserved, then the most probable outcome of a perturbation will be to converge to an homogeneous stratification in density. To show this, we perform a simulation with a second order finite volume scheme that only preserve stationary solutions corresponding to a constant free surface and density profile and a simulation with a second order finite volume scheme that only preserve a single stationary solution corresponding to the profile given by (4.3.1). The remaining simulation characteristics are unchanged. The results are depicted in Figure 4.10. We can see that

UNIVERSIDAD DE MÁLAGA

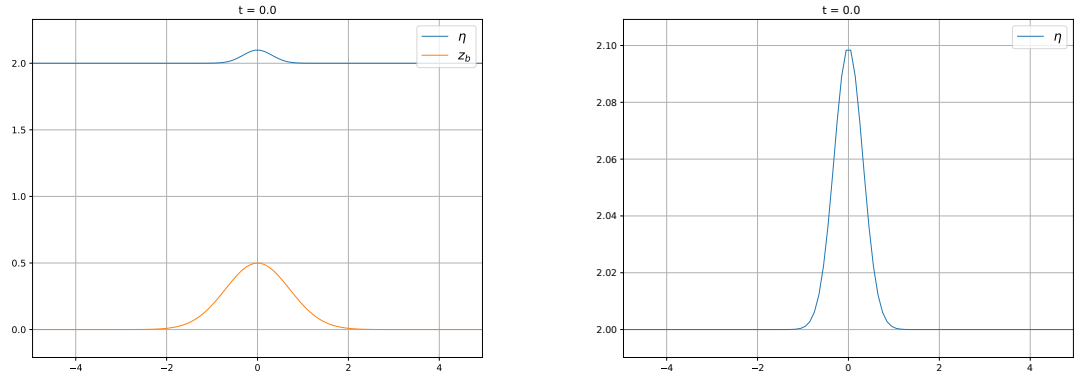


Figure 4.6: Initial condition for a simulation consisting on a perturbation of a steady state with a non-constant density profile. Left: surface and bottom. Right: zoom on the free surface.

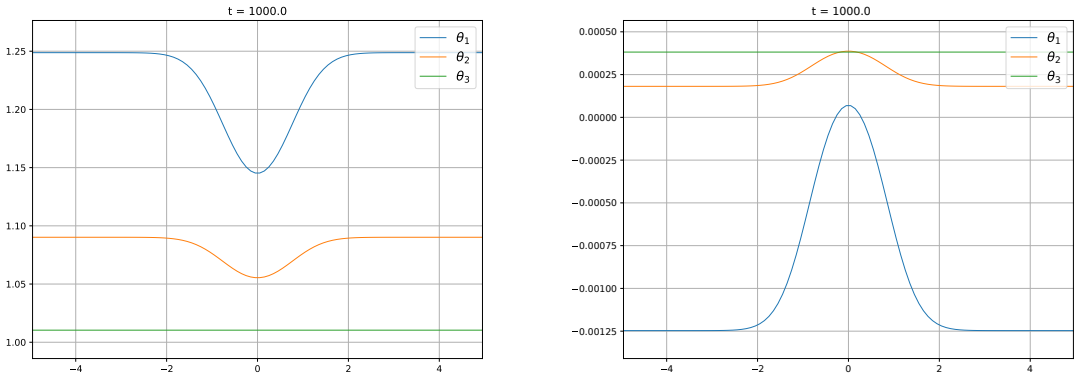


Figure 4.7: Spatial distribution of the relative density profile at final time $t = 1000$ s (left). Difference of relative densities at $t = 0$ and $t = 1000$ s (right). Both figures are computed with the well-balanced fourth order ADER-DG numerical scheme.

the scheme that preserve only the lake-at-rest type stationary solutions actually converges to a homogeneous profile in density faster that the scheme that preserve only one stratified stationary solution.

UNIVERSIDAD DE MÁLAGA

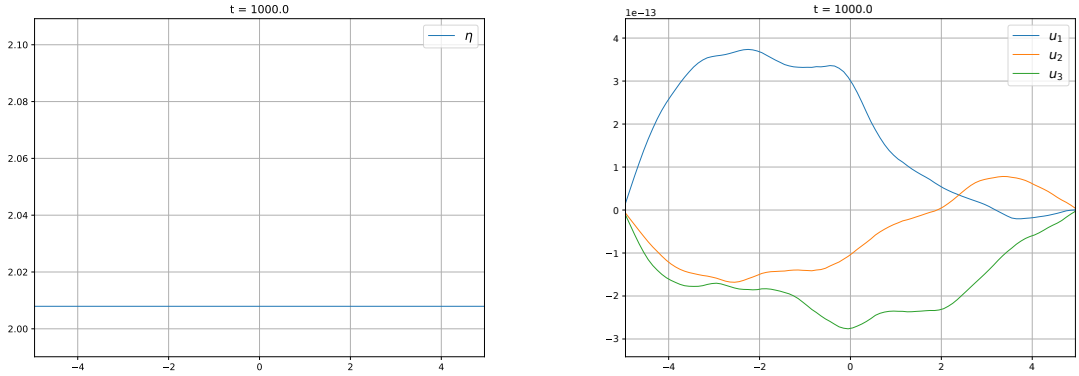


Figure 4.8: Spatial distribution of the free surface and velocities at final time $t = 1000$ s for the fourth order ADER-DG numerical scheme.

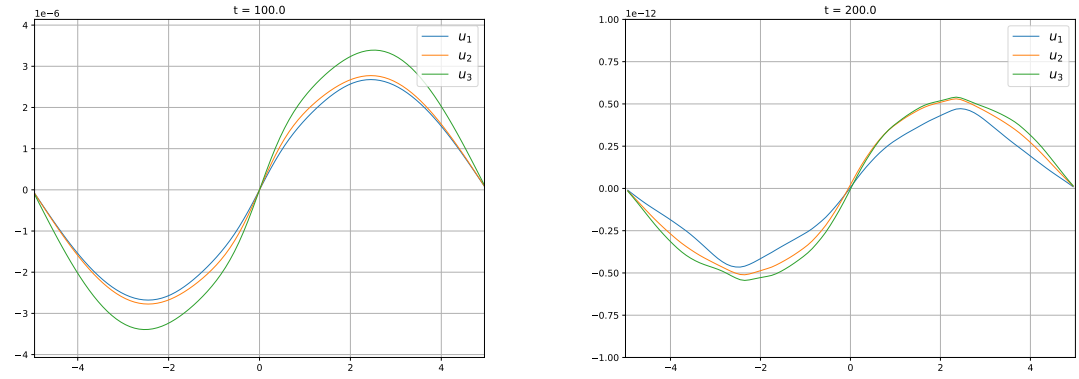


Figure 4.9: Spatial distribution displaying the evolution of the velocity profiles at time $t = 100$ s (left) and $t = 200$ s (right) for the well-balanced fourth order ADER-DG numerical scheme.

UNIVERSIDAD DE MÁLAGA

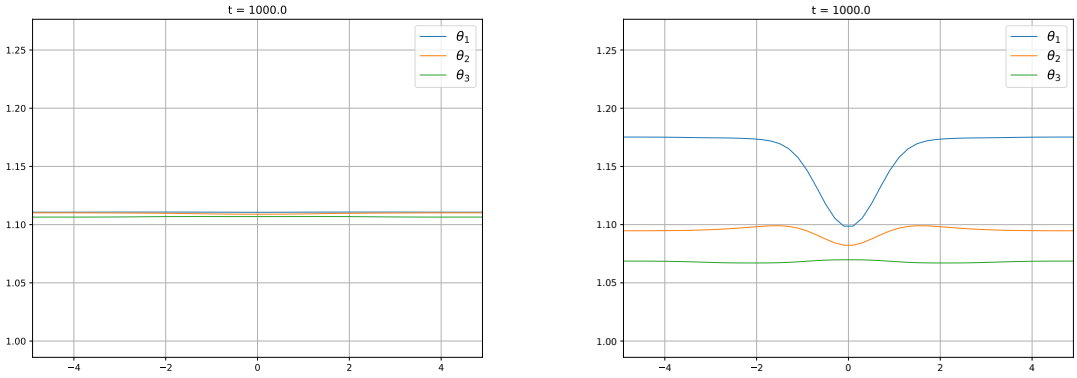


Figure 4.10: Spatial distribution of the density profile for a second order finite volume method well-balanced for the lake-at-rest solution (left) and density profile for a second order finite volume method well-balanced for a particular stationary solution corresponding to (4.3.1) at final time $t = 1000$ s (right).

UNIVERSIDAD DE MÁLAGA

4.4 Simulation for a smooth distribution of relative density

We consider now a smooth profile of relative density in a fixed domain $I = [-4, 4]$ with $M = 10$ layers. The relative density profile for all layers is given by,

$$\theta(x) = 1 + \frac{1}{100}\, e^{-10\, x^2}.$$

The bathymetry and free surface are constant and equal to 0 and 2 respectively. The initial condition can be seen in Figure 4.11. Free flow boundary conditions are set in all boundaries. The results of the simulations are depicted in Figure 4.12 at different simulation times. The relative density tends to expand downwards and to the sides at different speeds due to the difference of pressure between layers, forming a shock. Eventually, the solution tends to a vertical stratification of relative density. Note that the representation is achieved through a heat map of the relative density alongside the density profile for the finite volume and the ADER-DG numerical schemes. Both the finite volume solver and the ADER-DG schemes are depicted together for a better comparison. The finite volume solver is second order accurate with 800 discretization cells while the ADER-DG scheme is a fifth order scheme in space and time with only 80 discretization cells. Note that a parameter β has been included in the Figure to indicate that the cell is considered as troubled when its value is greater that one, and therefore it is limited. The very high order of the ADER-DG scheme accounts for the small oscillations in the solution. Nevertheless, it is able to provide satisfactory results, even for the slow moving shock.

Next, we compare both methods setting second order of accuracy and a coarse mesh of $N_s = 50$ for both of them. As we can see in Figure 4.13, the natural subcell resolution of the DG methods allows to capture significantly more details that the corresponding finite volume method, which is unable to properly capture any significant structure relevant to the problem.

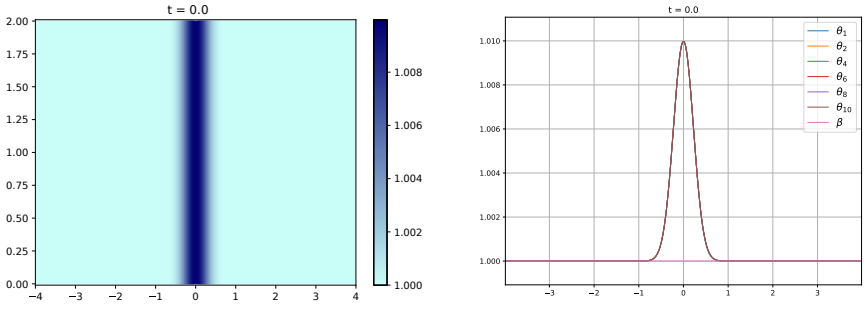


Figure 4.11: Initial condition for a smooth distribution of relative density.

UNIVERSIDAD DE MÁLAGA

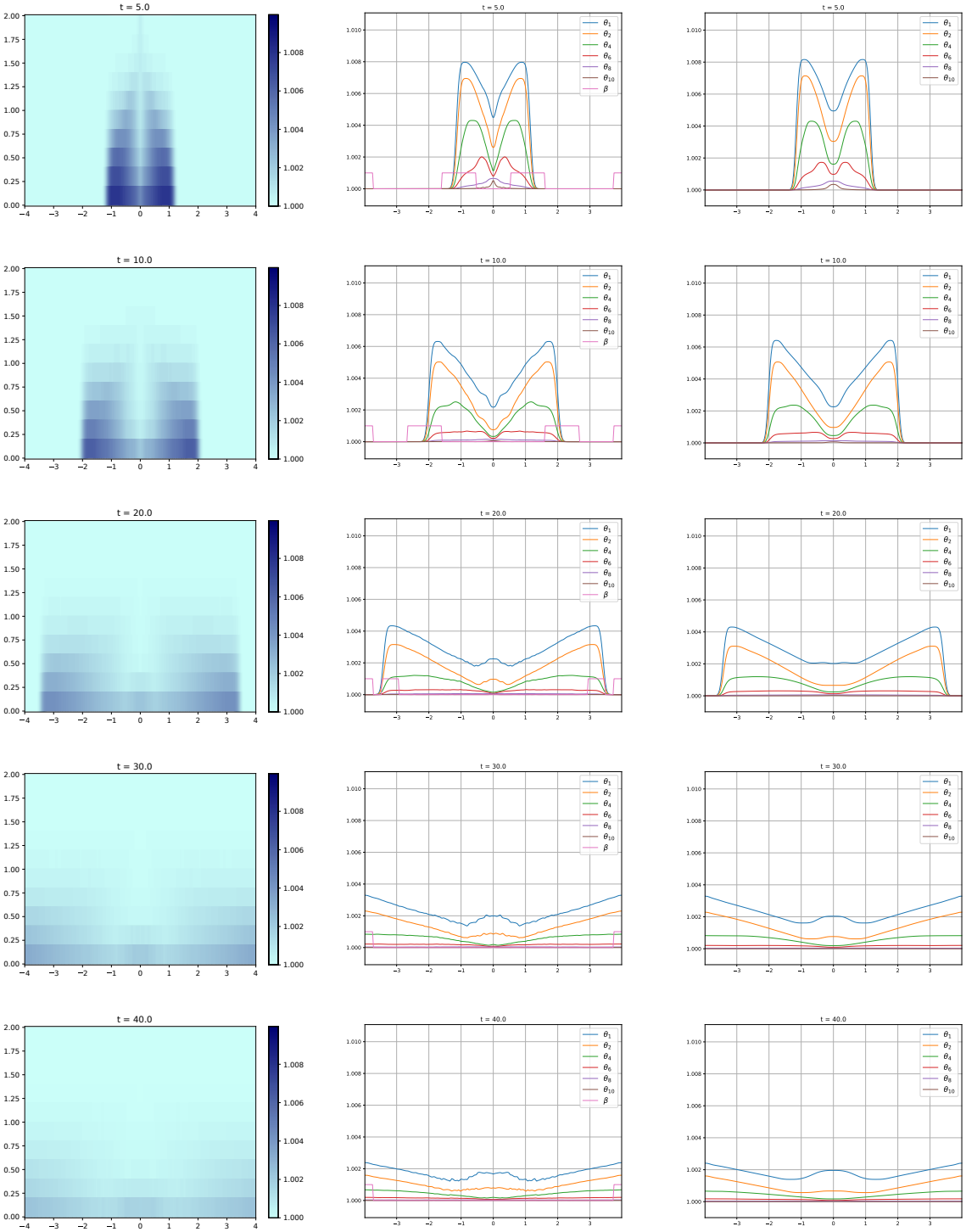


Figure 4.12: Spatial distribution of the evolution of a smooth distribution of relative density at different time steps. The figure depicts the heat map of the relative density computed with the ADER-DG method (left) and the density profile for a selected number of layers for the ADER-DG method (center) with 80 cells and the finite volume method (right) with 800 cells.

UNIVERSIDAD DE MÁLAGA

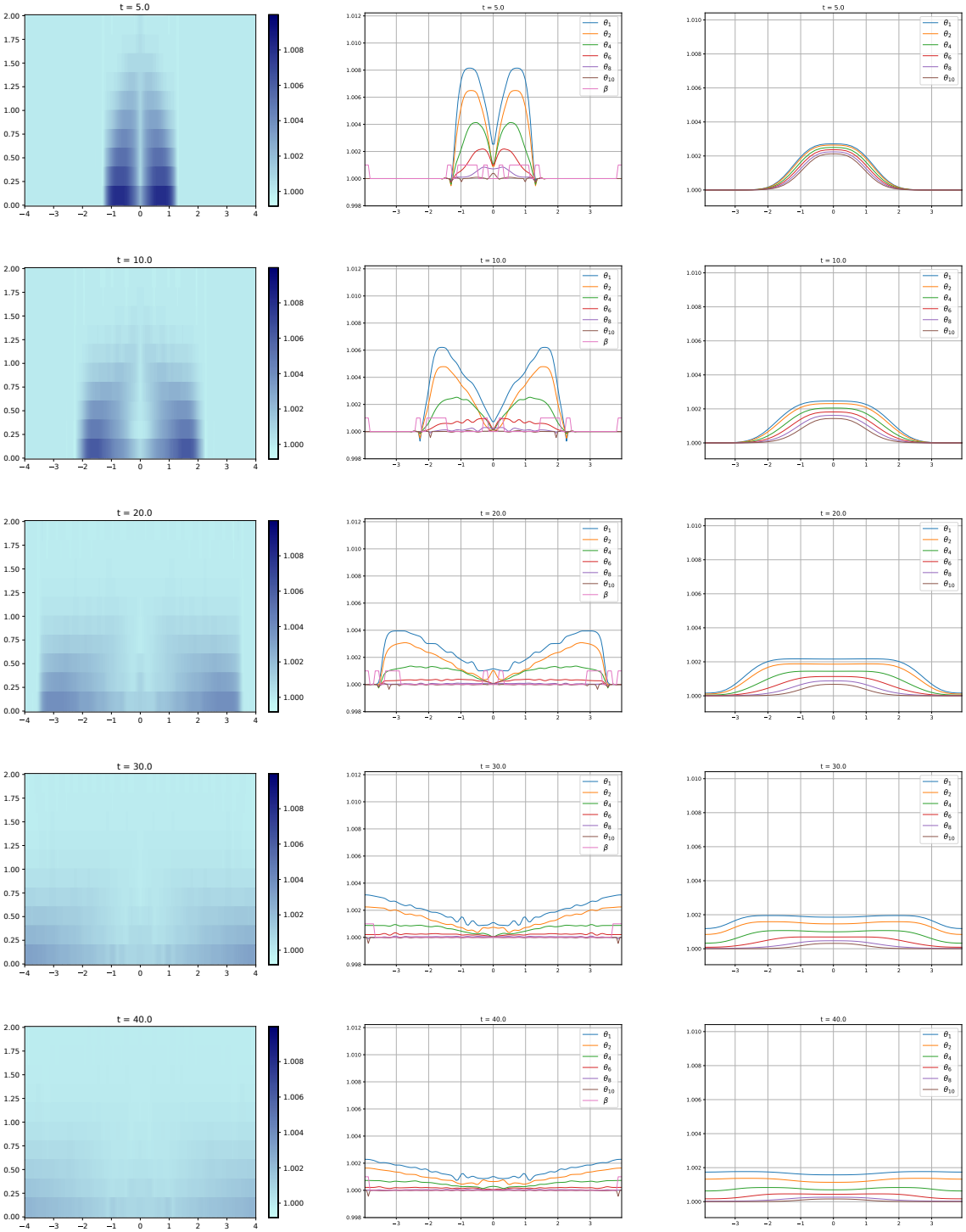


Figure 4.13: Spatial distribution of the evolution of a smooth distribution of relative density at different time steps. The figure depicts the heat map of the relative density computed with the second order ADER-DG method (left) and the density profile for a selected number of layers for the second order ADER-DG method (center) and the second order finite volume method (right), all of them with $N_s = 50$ cells.

UNIVERSIDAD DE MÁLAGA

4.5 Simulation of a lock-exchange in a flat channel

We seek now to validate the proposed model and numerical schemes by reproducing a laboratory experiment where empirical data is available for comparison purposes. The experiment can be consulted in [36]. It consists on a lock-exchange in a three meters long flat channel with a gatebox of length 0.1 meters containing a fluid with density $\rho_1 = 1034$ kg/m^3. This gatebox is placed in the leftmost extreme of the channel and it is opened at initial time into the rest of the channel with density $\rho_0 = 1000$ kg/m^3. Thus, we consider

$$\theta(x) = \begin{cases} 1.034 & \text{if } x \leq 0.1, \\ 1.0 & \text{if } x > 0.1. \end{cases} \tag{4.5.1}$$

The initial free surface and bathymetry are constant and equal to 0.8 and 0.5 meters respectively. The initial condition can be seen in Figure 4.14. The computational domain $I = [0, 3]$ is discretized with 800 volume cells for the second order finite volume method and 80 discretization points for the fourth order ADER-DG numerical method. Reflecting no-slip boundary conditions are set.

Figure 4.15 shows the evolution of the density distribution at different times for the particular case of $M = 40$ layers. As expected, a shock is immediately formed once the gatebox is released, forming the plume front position. Note how the limiter in the ADER-DG simulation keeps track of the discontinuity and prevents most of the spurious oscillations associated with the high order. It is also activated for the traveling wave corresponding to the perturbation of the free surface. Figures 4.16–4.17 and Figures 4.18–4.19 show a comparison of the front position provided in [36] and the numerical results for the finite volume and the ADER-DG schemes respectively. Numerical simulations have been performed for an increasing number of layers. We can observe a better data agreement as we increase the number of layers, up until a convergence around $M = 30$ layers. In any case, both schemes are able to provide excellent data agreement.

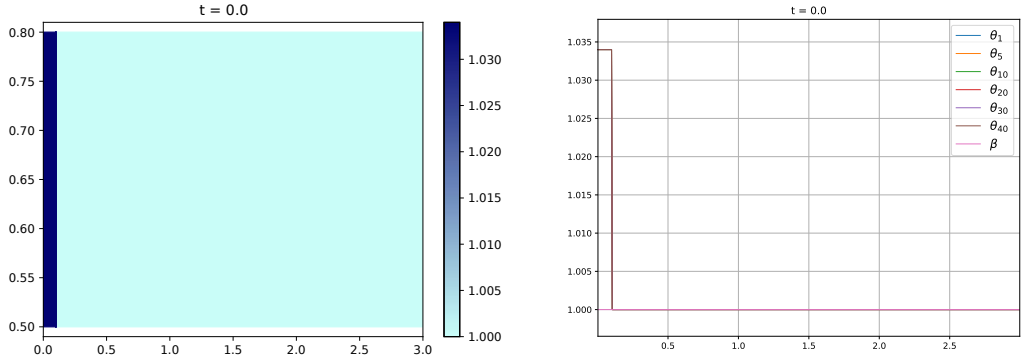


Figure 4.14: Lock-exchange experiment in a flat channel: spatial distribution of the initial condition.

UNIVERSIDAD DE MÁLAGA

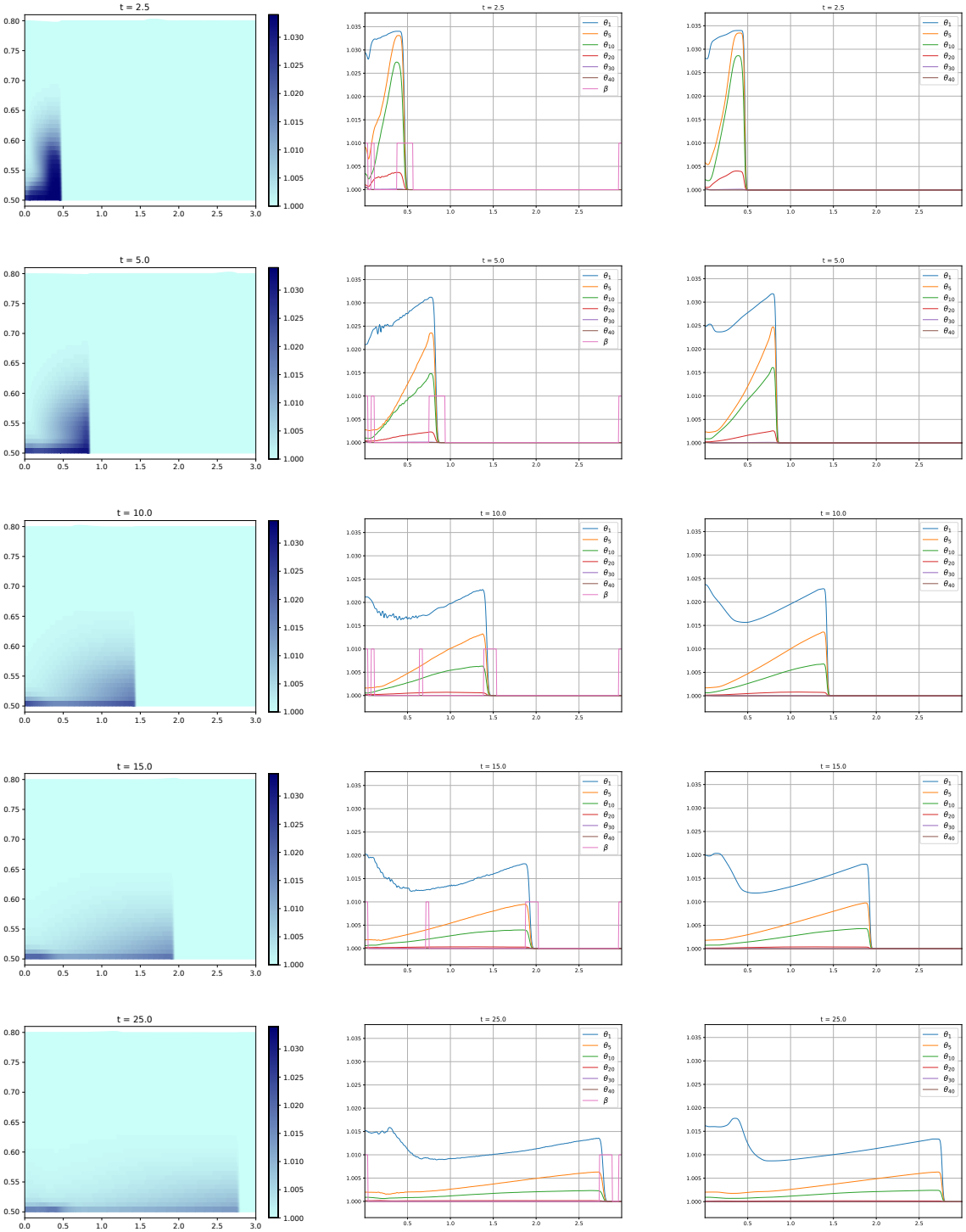


Figure 4.15: Lock-exchange experiment in a flat channel: evolution of the relative density at different time steps. The left Figure shows the relative density computed by the finite volume method through a heat map, while the center and right Figures depicts the results for the fourth order ADER-DG (80 cells) and the second order finite volume (800 cells) methods respectively.

UNIVERSIDAD DE MÁLAGA

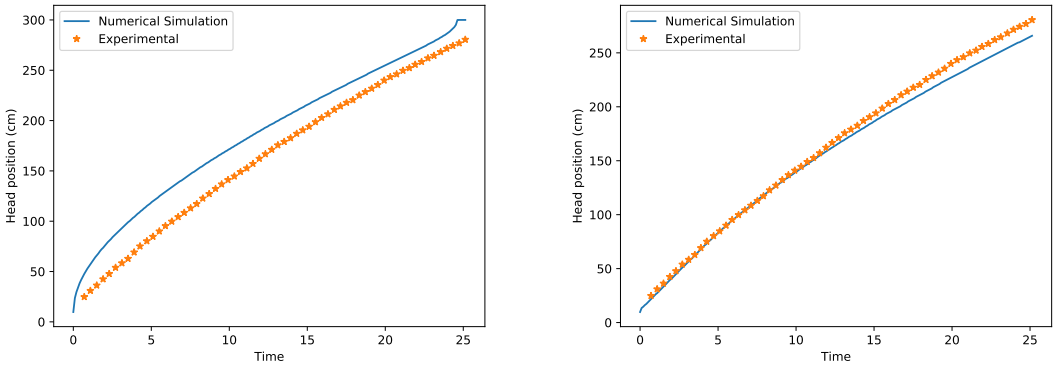


Figure 4.16: Evolution of the front position of the gravity current over time for the second order finite volume solver. The left Figure depicts the problem with $M = 15$ number of layers while the right shows the case with $M = 20$.

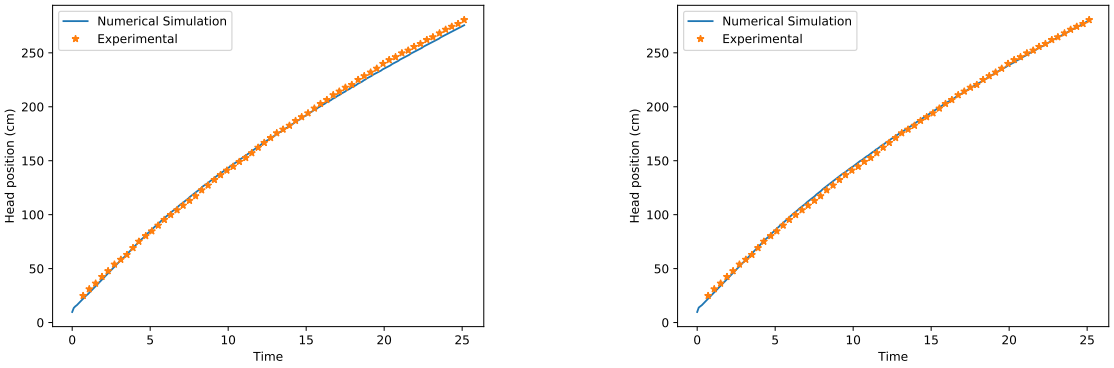


Figure 4.17: Evolution of the front position of the gravity current over time for the second order finite volume solver. The left Figure depicts the problem with $M = 30$ number of layers while the right shows the case with $M = 40$.

UNIVERSIDAD DE MÁLAGA

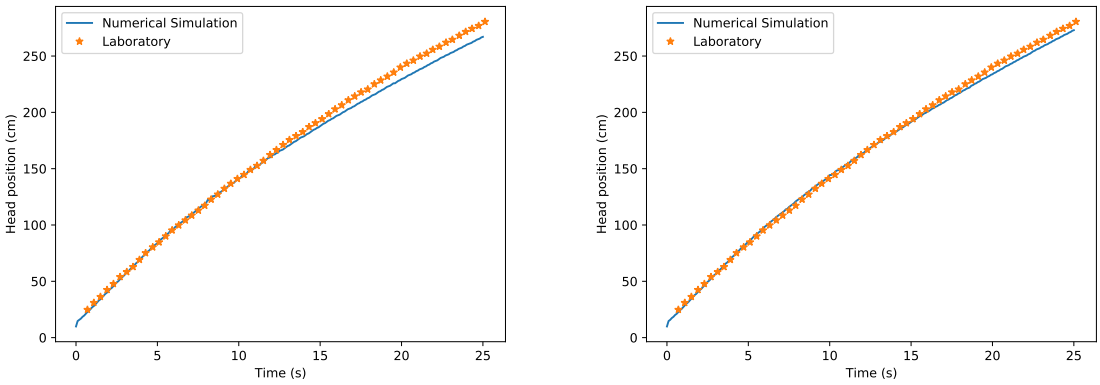


Figure 4.18: Evolution of the front position of the gravity current over time for the fourth order ADER-DG solver. The left Figure depicts the problem with $M = 20$ number of layers while the right shows the case with $M = 25$.

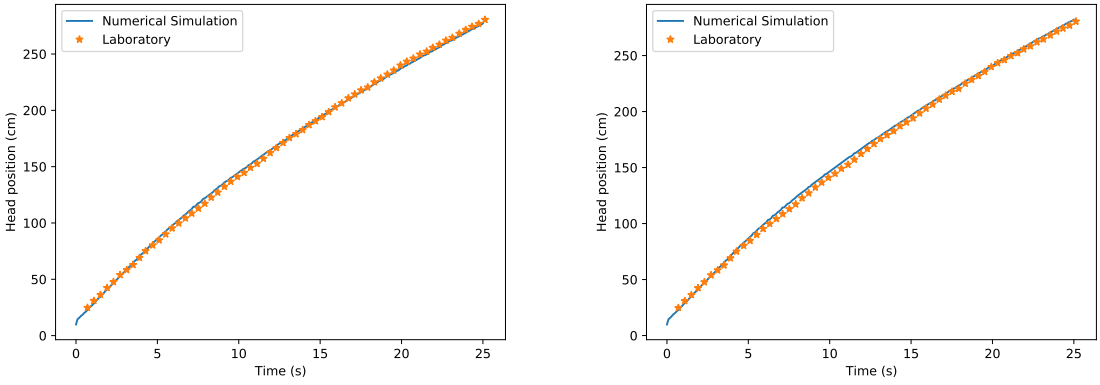


Figure 4.19: Evolution of the front position of the gravity current over time for the fourth order ADER-DG solver. The left Figure depicts the problem with $M = 30$ number of layers while the right shows the case with $M = 40$.

UNIVERSIDAD DE MÁLAGA

4.6 Simulation of a lock exchange problem with a non constant bathymetry function

In order to study a gravity current over an obstacle, a lock exchange in relative density problem with a non-flat bathymetry function is considered,

$$\theta(x) = \begin{cases} 1 & \text{if } x \leq 0, \\ 1.15 & \text{if } x > 0. \end{cases} \tag{4.6.1}$$

The computational domain is $I = [-5, 5]$ with $M = 25$ layers and free-flow boundary conditions are set. Additionally, the bathymetry function is given by

$$z_b(x) = \frac{1}{2} e^{-x^2}, \tag{4.6.2}$$

with the free surface being the constant function $\eta = 2$ meters. The initial condition is shown in Figure 4.20, whereas the evolution of the current is depicted at Figure 4.21 at different simulation times. These results are computed for the first order finite volume and ADER-DG methods with 500 and 50 volume cells respectively. Since both methods are of the same order of accuracy, the higher spatial resolution of the finite volume method yields slightly better results, though they are definitely deficient compared with theirs higher order counterparts. We can see this if we consider a second order finite volume method with 500 cells and a fourth order ADER-DG method with only 50 cells, depicted in Figure 4.22. In this case, both methods deliver similar results, successfully capturing the sharp transition characteristic of these kind of flows.

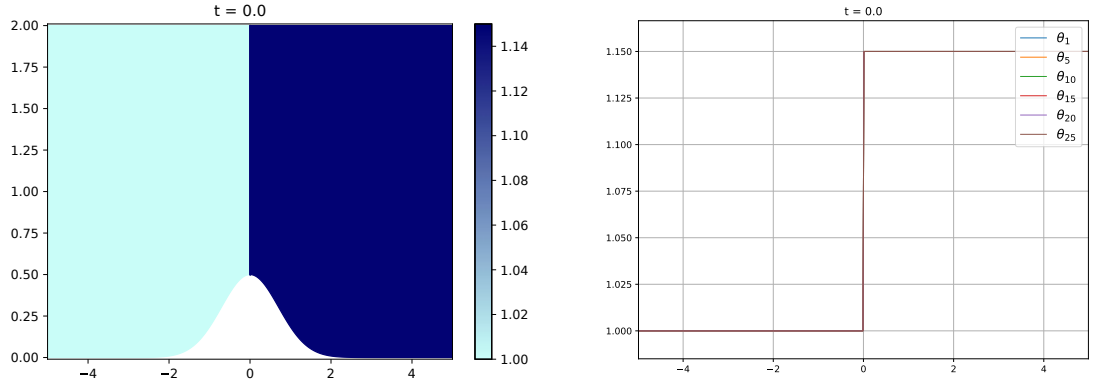


Figure 4.20: Initial condition for the lock exchange problem.

UNIVERSIDAD DE MÁLAGA

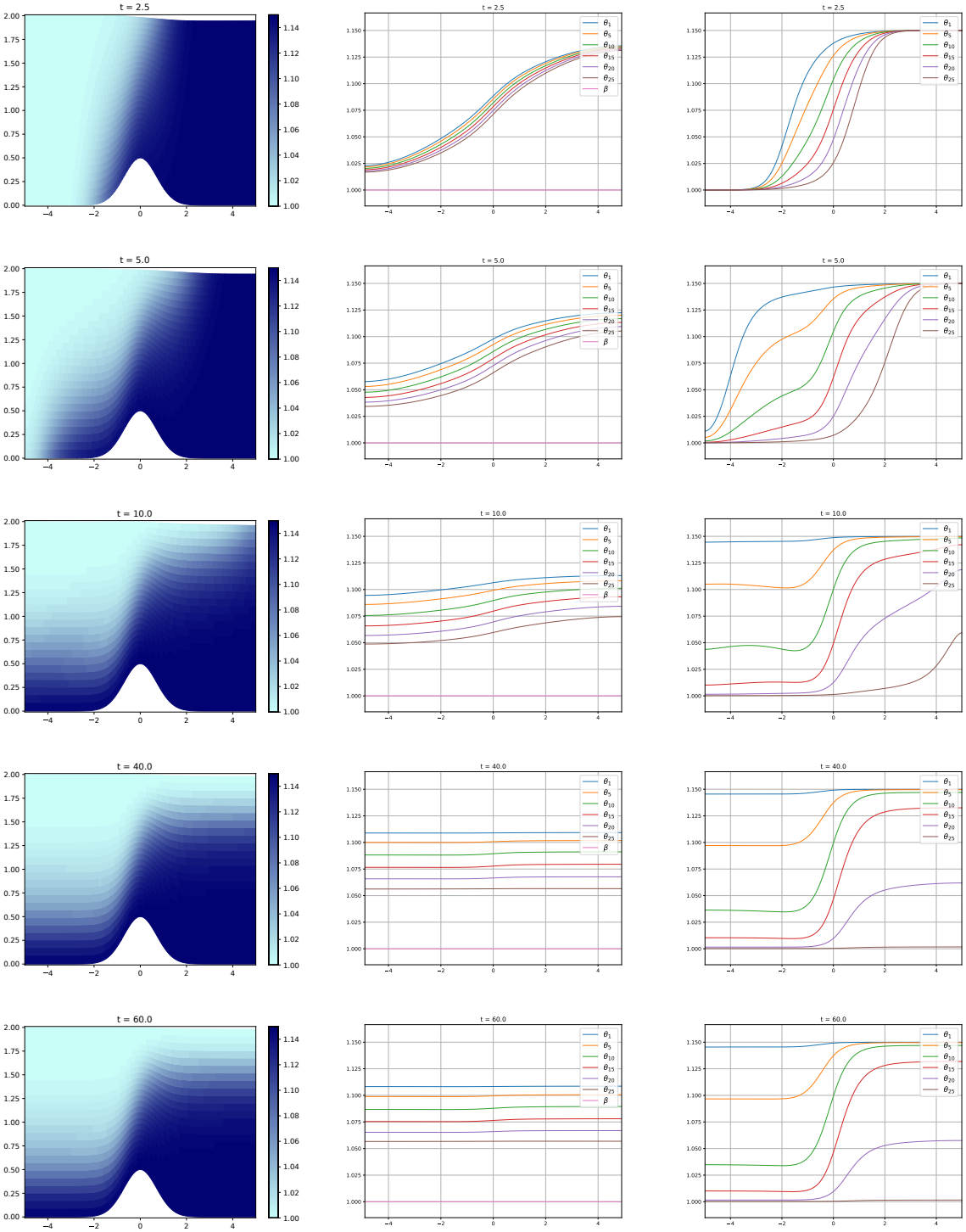


Figure 4.21: Spatial distribution of density profiles for a lock exchange problem at different time steps. The left Figure depicts the relative density through a heat map computed with the first order finite volume numerical scheme with 500 cell, while the center and right Figures correspond to the first order ADER-DG (50 cells) and finite volume (500 cells) solvers respectively.

UNIVERSIDAD DE MÁLAGA

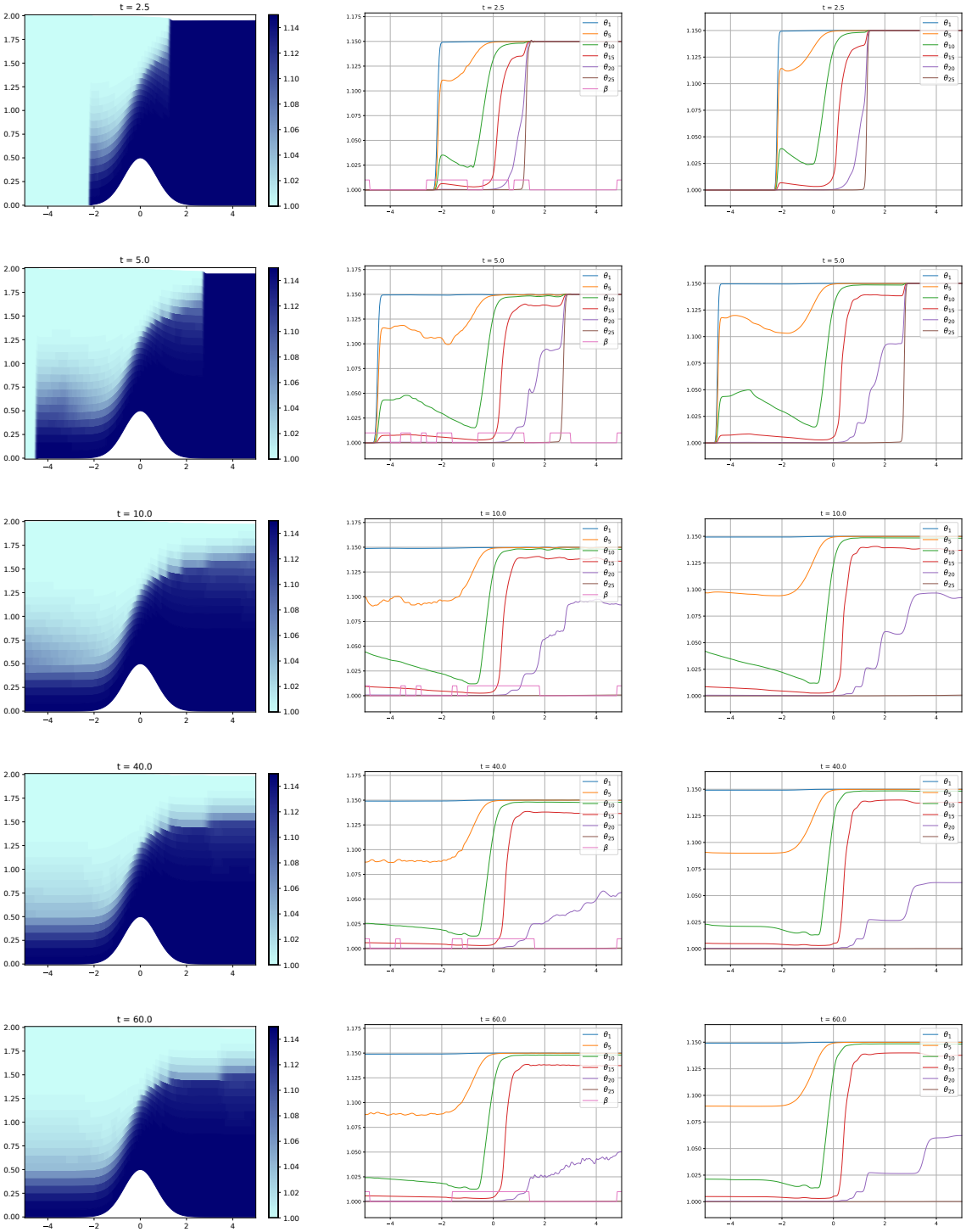


Figure 4.22: Spatial distribution of density profiles for a lock exchange at different time steps. The left Figure depicts the relative density through a heat map computed with the second order finite volume numerical scheme, while the center and right Figures correspond to the fourth order ADER-DG (50 cells) and second order finite volume (500 cells) methods respectively.

UNIVERSIDAD DE MÁLAGA

4.7 Simulation of a lock exchange problem in two dimensions

A similar simulation is now considered in a two-dimension framework, to show how the numerical techniques presented in this thesis can be easily extended to the two dimensional case. The extension is performed by considering the two dimensional shallow-water equations with variable density in a dimension by dimension fashion, following [203]. More information can be found in Appendix A.

Particularly, the domain $I = [-5, 5] \times [-1, 1]$ is considered alongside the following bathymetry function,

$$z_b(x, y) = \frac{1}{2}\, e^{-((x+2)^2+y^2)} + \frac{1}{2}\, e^{-((x-2)^2+y^2)}.$$

The free surface at the initial time is constant and equal to 2 meters, while the relative density is set as

$$\theta(x, y) = \begin{cases} 1.0 & \text{if } x \leq 0, \\ 1.02 & \text{if } x > 0. \end{cases} \tag{4.7.1}$$

The domain is discretized with $36,000$ uniform cells with $\Delta x = 1/60$ and $\Delta y = 1/30$ and the total number of layers is set to 15. Reflecting no-slip boundary conditions are considered for the horizontal boundaries while free-flow boundary conditions are set for the vertical ones. The initial condition can be seen in Figure 4.23 and time evolution of the fluid is shown in Figure 4.24, where vertical cuts along the lines $y = 0$ and $x = -2$ are depicted. Figures 4.25 and 4.26 also show the upper view of the density distribution for the layer 7. The results presented here correspond to the second order finite volume solver that is well-balanced for the stationary solutions corresponding to a constant free surface and a constant density profile, since it is the one which has been implemented for the two dimensional case in this thesis. The DG implementation for the two dimensional case is a work in progress for future applications.

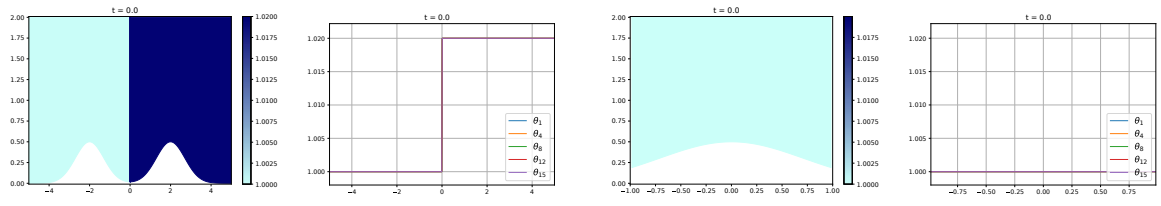


Figure 4.23: Initial condition for the two dimensional lock exchange for a vertical cut in the direction $y = 0$ (first two Figures on the left) and $x = -2$ (last two Figures on the right).

UNIVERSIDAD DE MÁLAGA

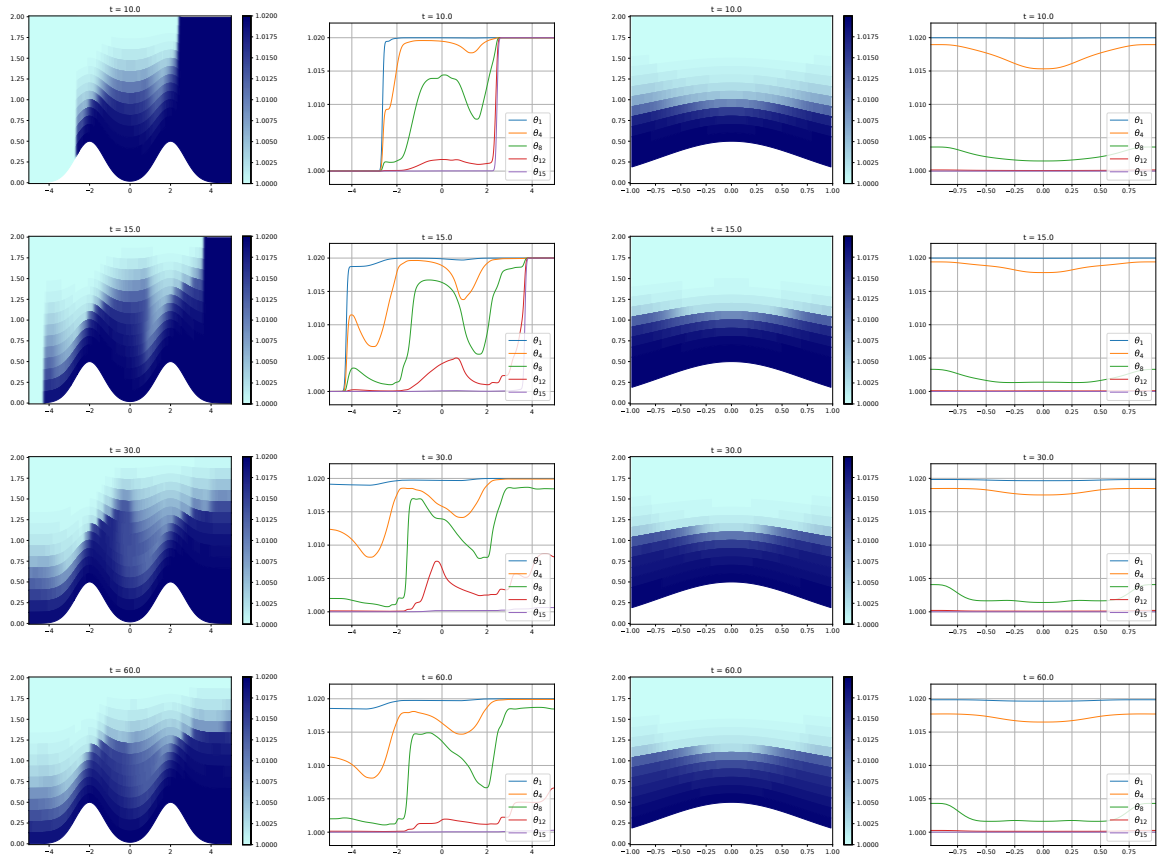


Figure 4.24: lock exchange problem in two dimensions at different time steps. The two Figures on the left depicts the relative density for a vertical cut in the direction $y = 0$ while the two Figures at the right shows the relative density for a vertical cut on the direction $x = -2$.

UNIVERSIDAD DE MALAGA

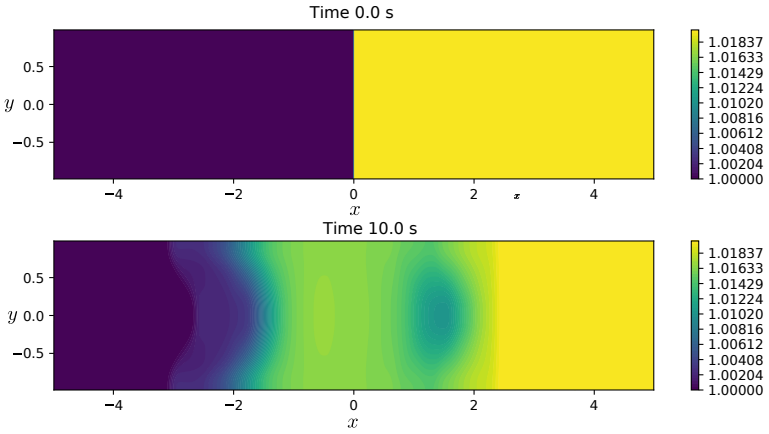


Figure 4.25: Zenithal view of the spatial domain displaying relative density distribution in the layer $M = 7$ at different time steps through a heat map.

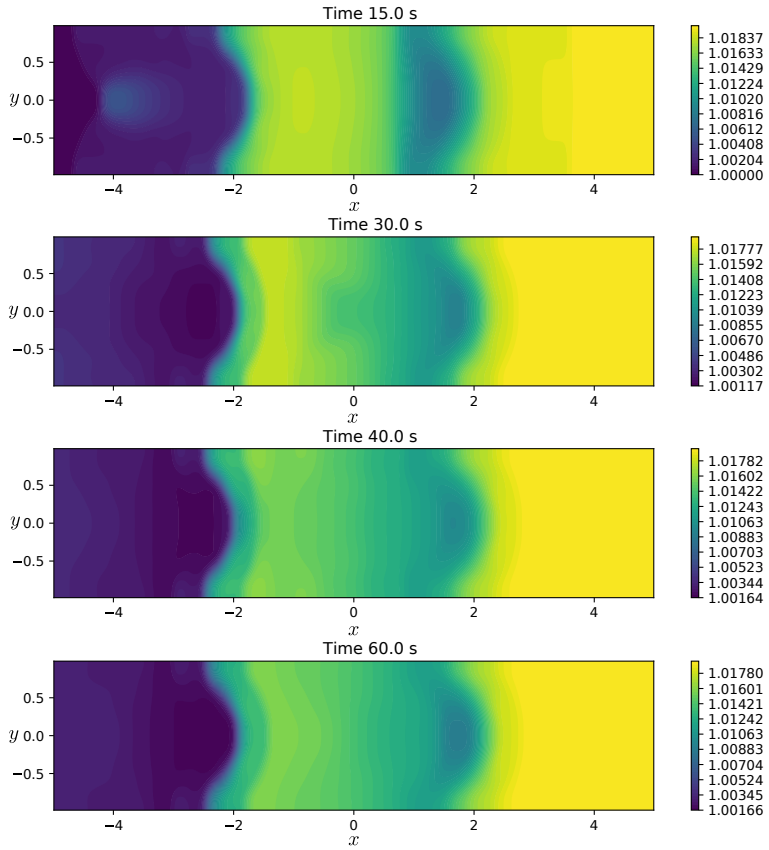


Figure 4.26: Zenithal view of the spatial domain displaying relative density distribution in the layer $M = 7$ at different time steps through a heat map.

UNIVERSIDAD DE MÁLAGA

Chapter 5

Conclusions and future work

To end this thesis, some conclusion and future work are drawn in this chapter, with emphasis on the novel contributions of this work. Additionally, some proposal on possible future work are suggested.

5.1 Conclusions

This thesis incorporates advances on mathematical modeling and numerical analysis that range from general systems of conservation laws to particular discretizations of a state of the art shallow-water type model.

In Chapter 1, a general multilayer shallow water model where density effects are present through a tracer function which is advected with the flow is presented. It aims to address the issue where standard shallow-water models ignore how changes in density may affect the fluid evolution when these effects are relevant. The model is built under the usual hypothesis of multilayer shallow water models, where the velocity and density are constant within a layer. It relies on a hydrostatic expression of the pressure that depends on the relative density. Consequentially, the final model (1.2.7) includes non-conservative terms directly related with those. The model has proven to be extremely sensitive to small changes in relative density, since they can trigger significant fluctuations in the velocity field and free surface. Therefore, a careful numerical treatment is mandatory in order to successfully discretize this model. Another important result obtained in this chapter concerns the study on the stationary solutions of the model. In particular, stationary solutions corresponding with zero velocity are studied and, in particular, we are interested in preserving those corresponding with constant free surface and a stable vertical stratification in density.

In Chapter 2, a general methodology for designing high order well-balanced finite volume and discontinuous Galerkin methods is presented. In particular, a short review on path-conservative finite volume solvers has been performed. Additionally, several strategies to preserve stationary solutions or to improve the overall efficiency of the

UNIVERSIDAD DE MÁLAGA

Riemann solver are given. These results are used to support the review on discontinuous Galerkin methods performed later on. In this framework, a novel contribution is included in the form of a general method for designing well-balanced DG methods. Additionally, a new proposal has been made in the ever present problem of limiting in the framework of DG methods that preserves the well-balanced property of the resulting scheme. The proposed well-balanced scheme and limiter are tested against several systems of conservation laws, resulting in excellent results. These results include order tests, preserving stationary solution, recovering stationary solutions after a perturbation and a simulation far from the *equilibria*. All this simulations yield very satisfactory results, validating the proposed well-balanced strategy. These results gave rise to an article [204].

Chapter 3 is dedicated to an *ad hoc* discretization of the multilayer shallow-water model with variable density introduced in Chapter 1 with the techniques reviewed in Chapter 2. A second order well-balanced finite volume method with an hydrostatic reconstruction is derived, especially tailored for the multilayer shallow-water model with variable density. The numerical scheme is able to preserve non trivial stratified stationary solutions, and recover them after a perturbation. These results were published in the article [100]. An alternative numerical scheme based on the ADER-DG approach is also proposed. This solver presents some evident advantages: allows for arbitrary high order in space and time, defining one-step high order schemes; the natural subcell resolution inherent to DG methods allows for coarse of even very coarse meshes while keeping a great resolution; the proposed well-balanced techniques are applied to preserve non trivial stratified stationary solutions within the framework of DG methods. Moreover, the computational results seem to satisfy a discrete maximum principle for the relative density, thanks to its finite volume based *a posteriori* subcell limiter technique. This limit may check the numerical solution to ensure that all desired physical and numerical properties are met and, in case that this is not the case, it allows to switch to a robust finite volume solver where many powerful tools are available. Again, these results were published in [205].

Finally, Chapter 4 includes several numerical experiments. These experiments have been designed to highlight the advantages and properties of the numerical schemes for the multilayer shallow-water model with variable density designed in this thesis. Both the finite volume and DG solvers share some common tests: an order test to numerically check the order of convergence of the numerical schemes, a number of experiments where the well-balanced property is shown and an experiment where laboratory data is available for comparison purposes, among others. These simulations show the suitability of the model to simulate geophysical flows where density variations play an important role. The experiments also stress the importance of high order solvers as a key feature to gain a greater performance. This is especially true in the case of DG solvers, where it is possible to achieve excellent results even with very coarse meshes. It is of singular interest the ability of both numerical discretizations to correlate with empirical data. The excellent results can be then explained by the number of layers considered and that the effect of the vertical acceleration in the simulation is neglectable for a slow density driven plume.

UNIVERSIDAD DE MÁLAGA

In general, this chapter allows to understand the requirements in terms of discretization and number of vertical layers that are necessary to effectively apply the model in real case scenarios.

5.2 Future work

We present now future research that can be derived from the work presented in this thesis and expand its results. For instance, it would be interesting to rewrite the multilayer shallow-water model with variable density in terms of σ-coordinates. In this way, different numerical approximations of the vertical flow (or exchange term between layers) could be proposed. Additionally, the current model, written in Cartesian coordinates, could be written in terms of spherical coordinates to simulate density driven currents in big spherical domains. Moreover, if Coriolis forces are incorporated to the model, then it would be perfectly feasible to simulate real density driven currents in very large domains.

Other research opportunities reside in improving the numerical discretization. Especially the ADER-DG numerical scheme could be expanded in a dimension by dimension fashion to encompass two dimensional domains. Likewise, the current model can suffer from an aliasing problem when the simulation times are too long, and therefore there is room for improvement in this regard.

Within the field of the shallow-water systems, non hydrostatic models have become very popular for their ability to better capture some physical effects related with the vertical acceleration of the fluid. It would be interesting to extend the present shallow-water model with variable density to a non hydrostatic version to study the influence of these effects in density driven flows. This study could be linked with a more general study relative to the hyperbolicity of the model, an issue that has not been tackled in this thesis.

Finally, this thesis has presented a number of results in the framework of well-balanced numerical schemes that are a resolute step forward in the field, especially in the DG family of solvers. However, a broader effort to preserve stationary solution in a two dimensional framework is also a relevant future work.

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD
DE MÁLAGA

Appendix A

Extension to 2D problems for the finite volume method

We describe now briefly the implementation of the finite volume method for the multilayer shallow-water model with variable density on Chapter 3 in 2D. As in the one dimensional case, the numerical scheme consists essentially on a second order spatial discretization using a MUSCL reconstruction operator and a explicit time discretization using a second order Runge-Kutta method. A HLL type Riemann solver with a hydrostatic reconstruction is also used at the cell interfaces.

First, the two dimensional version of the multilayer shallow-water system with variable density in the horizontal coordinates (x, y) is:

$$\begin{cases} \partial_t h + \partial_x \left(h \sum_{\beta=1}^{M} l_\beta u_{x,\beta} \right) + \partial_y \left(h \sum_{\beta=1}^{M} l_\beta u_{y,\beta} \right) = 0, \\ \partial_t (h\theta_\alpha) + \partial_x (h\theta_\alpha u_{x,\alpha}) + \partial_y (h\theta_\alpha u_{y,\alpha}) = \dfrac{1}{l_\alpha} \left(\theta_{\alpha+\frac{1}{2}} G_{\alpha+\frac{1}{2}} - \theta_{\alpha-\frac{1}{2}} G_{\alpha-\frac{1}{2}} \right), \\ \partial_t (h\theta_\alpha u_{x,\alpha}) + \partial_x (h\theta_\alpha u_{x,\alpha}^2) + \partial_y (h\theta_\alpha u_{x,\alpha} u_{y,\alpha}) + gh\theta_\alpha \partial_x \eta + \dfrac{g l_\alpha}{2} \left(h\partial_x (h\theta_\alpha) - h\theta_\alpha \partial_x h \right) \\ + g \sum_{\beta=\alpha+1}^{M} l_\beta \left(h\partial_x (h\theta_\beta) - h\theta_\alpha \partial_x h \right) = \dfrac{1}{l_\alpha} \left(u_{x,\alpha+\frac{1}{2}} \theta_{\alpha+\frac{1}{2}} G_{\alpha+\frac{1}{2}} - u_{x,\alpha-\frac{1}{2}} \theta_{\alpha-\frac{1}{2}} G_{\alpha-\frac{1}{2}} \right), \\ \partial_t (h\theta_\alpha u_{y,\alpha}) + \partial_y (h\theta_\alpha u_{y,\alpha}^2) + \partial_x (h\theta_\alpha u_{y,\alpha} u_{x,\alpha}) + gh\theta_\alpha \partial_y \eta + \dfrac{g l_\alpha}{2} \left(h\partial_y (h\theta_\alpha) - h\theta_\alpha \partial_y h \right) \\ + g \sum_{\beta=\alpha+1}^{M} l_\beta \left(h\partial_y (h\theta_\beta) - h\theta_\alpha \partial_y h \right) = \dfrac{1}{l_\alpha} \left(u_{y,\alpha+\frac{1}{2}} \theta_{\alpha+\frac{1}{2}} G_{\alpha+\frac{1}{2}} - u_{y,\alpha-\frac{1}{2}} \theta_{\alpha-\frac{1}{2}} G_{\alpha-\frac{1}{2}} \right). \end{cases} \tag{1}$$

We denote by $\boldsymbol{u} = (u_x, u_y)$ the horizontal velocity on the x and y direction and the

UNIVERSIDAD DE MÁLAGA

transference terms are now defined as,

$$G_{\alpha+\frac{1}{2}} = \sum_{\beta=1}^{\alpha} l_\beta \left(\nabla \cdot (h\boldsymbol{u}_\beta) - \nabla \cdot \left(h \sum_{\gamma=1}^{M} l_\gamma \boldsymbol{u}_\gamma \right) \right). \tag{2}$$

The shallow-water system (1) can be written in the more general form,

$$\partial_t \boldsymbol{w} + \partial_x \boldsymbol{F}_C(\boldsymbol{w}) + \partial_y \boldsymbol{G}_C(\boldsymbol{w}) + \boldsymbol{P}_x(\boldsymbol{w}, \eta, \partial_x \boldsymbol{w}, \partial_x \eta) + \boldsymbol{P}_y(\boldsymbol{w}, \eta, \partial_y \boldsymbol{w}, \partial_y \eta) - \boldsymbol{T}(\boldsymbol{w}, \partial_x \boldsymbol{w}, \partial_y \boldsymbol{w}) = \boldsymbol{0}. \tag{3}$$

where the new state variables are,

$$\boldsymbol{w} = (h \,|\, h\theta_\alpha \,|\, h\theta_\alpha u_{x,\alpha} \,|\, h\theta_\alpha u_{y,\alpha})^T \in \mathbb{R}^{3M+1}. \tag{4}$$

Also, the convective fluxes are given by,

$$\boldsymbol{F}_C(\boldsymbol{w}) = \left(h \sum_{\beta=1}^{M} l_\beta u_{x,\beta} \,\middle|\, h\theta_\alpha u_{x,\alpha} \,\middle|\, h\theta_\alpha u_{x,\alpha}^2 \,\middle|\, h\theta_\alpha u_{x,\alpha} u_{y,\alpha} \right)^T \in \mathbb{R}^{3M+1}, \tag{5}$$

$$\boldsymbol{G}_C(\boldsymbol{w}) = \left(h \sum_{\beta=1}^{M} l_\beta u_{y,\beta} \,\middle|\, h\theta_\alpha u_{y,\alpha} \,\middle|\, h\theta_\alpha u_{y,\alpha}^2 \,\middle|\, h\theta_\alpha u_{y,\alpha} u_{x,\alpha} \right)^T \in \mathbb{R}^{3M+1}, \tag{6}$$

while the pressure terms are given by

$$\boldsymbol{P}_x(\boldsymbol{w}, \eta, \partial_x \boldsymbol{w}, \partial_x \eta) = (0 \,|\, \boldsymbol{0} \,|\, P_{x,\alpha} \,|\, \boldsymbol{0}) \in \mathbb{R}^{3M+1}, \tag{7}$$

$$P_{x,\alpha} = gh\theta_\alpha \partial_x \eta + \frac{g l_\alpha}{2} (h\partial_x(h\theta_\alpha) - h\theta_\alpha \partial_x h) + g \sum_{\beta=\alpha+1}^{M} l_\beta (h\partial_x(h\theta_\beta) - h\theta_\alpha \partial_x h), \tag{8}$$

and

$$\boldsymbol{P}_y(\boldsymbol{w}, \eta, \partial_y \boldsymbol{w}, \partial_y \eta) = (0 \,|\, \boldsymbol{0} \,|\, \boldsymbol{0} \,|\, P_{y,\alpha}) \in \mathbb{R}^{3M+1}, \tag{9}$$

$$P_{y,\alpha} = gh\theta_\alpha \partial_y \eta + \frac{g l_\alpha}{2} (h\partial_y(h\theta_\alpha) - h\theta_\alpha \partial_y h) + g \sum_{\beta=\alpha+1}^{M} l_\beta (h\partial_y(h\theta_\beta) - h\theta_\alpha \partial_y h). \tag{10}$$

Finally, the transference terms are given by,

$$\begin{aligned} \boldsymbol{T}(\boldsymbol{w}, \partial_x \boldsymbol{w}, \partial_y \boldsymbol{w}) = \\ \left(0 \,\middle|\, \frac{1}{l_\alpha} (\theta_{\alpha+\frac{1}{2}} G_{\alpha+\frac{1}{2}} - \theta_{\alpha-\frac{1}{2}} G_{\alpha-\frac{1}{2}}) \,\middle|\, \frac{1}{l_\alpha} (u_{x,\alpha+\frac{1}{2}} \theta_{\alpha+\frac{1}{2}} G_{\alpha+\frac{1}{2}} - u_{x,\alpha-\frac{1}{2}} \theta_{\alpha-\frac{1}{2}} G_{\alpha-\frac{1}{2}}) \right. \\ \left. \,\middle|\, \frac{1}{l_\alpha} (u_{y,\alpha+\frac{1}{2}} \theta_{\alpha+\frac{1}{2}} G_{\alpha+\frac{1}{2}} - u_{y,\alpha-\frac{1}{2}} \theta_{\alpha-\frac{1}{2}} G_{\alpha-\frac{1}{2}}) \right)^T \in \mathbb{R}^{3M+1}. \end{aligned} \tag{11}$$

UNIVERSIDAD DE MÁLAGA

We recall that $f_{\alpha+\frac{1}{2}}$ is the arithmetic mean at the interface $\Gamma_{\alpha+\frac{1}{2}}(t)$.

The computational domain is divided along the x and y direction into a series of conforming elements $V_{i,j}$ of length Δx and Δy respectively. The approximation of the solution at a cell $V_{i,j}$ of coordinates (x_i, y_j) is denoted by

$$\boldsymbol{w}_{i,j}^n \approx \frac{1}{\Delta x \Delta y} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \boldsymbol{w}(x, t^n)\, dxdy.$$

System (3) can be solved in a dimension by dimension fashion by extending the second order finite volume numerical scheme (3.2.21) as follows:

$$\begin{aligned} \boldsymbol{w}'_{i,j}(t) \;=\; & -\frac{1}{\Delta x}\left(\boldsymbol{D}^+_{i-\frac{1}{2}}(t) + \boldsymbol{D}^-_{i+\frac{1}{2}}(t)\right) - \frac{1}{\Delta y}\left(\boldsymbol{D}^+_{j-\frac{1}{2}}(t) + \boldsymbol{D}^-_{j+\frac{1}{2}}(t)\right) \\ & -\frac{1}{\Delta x \Delta y}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\int_{y_{i-\frac{1}{2}}}^{y_{i+\frac{1}{2}}} \Big(\boldsymbol{P}_x(\boldsymbol{R}^t_{i,j}, R^{\eta,t}_{i,j}, \partial_x \boldsymbol{R}^t_{i,j}, \partial_x R^{\eta,t}_{i,j}) \\ & +\boldsymbol{P}_y(\boldsymbol{R}^t_{i,j}, R^{\eta,t}_{i,j}, \partial_y \boldsymbol{R}^t_{i,j}, \partial_y R^{\eta,t}_{i,j}) - \boldsymbol{T}(\boldsymbol{R}^t_{i,j}, \partial_x \boldsymbol{R}^t_{i,j}, \partial_y \boldsymbol{R}^t_{i,j}) \Big)\, dx\, dy, \end{aligned} \tag{12}$$

where $\boldsymbol{R}^t_{i,j}(x, y)$ stands for the reconstructor operator (2.2.59) applied in the two dimensional cell $V_{i,j}$,

$$\boldsymbol{R}^t_{i,j}(x, y) = \boldsymbol{R}^t_{i,j}(x, y; \{W_p\}_{p\in\mathcal{S}_p}), \tag{13}$$

with $\mathcal{S}_p$ a set containing the element $V_{i,j}$ together with its neighbor cells that share a common node with $V_{i,j}$. Note that $R^{\eta,t}_{i,j}$ stands for the reconstruction operator applied to the free surface η. Note that

$$\boldsymbol{w}^+_{i-\frac{1}{2}} = \boldsymbol{R}^t_{i,j}(x_{i-\frac{1}{2}}, y_j), \qquad \boldsymbol{w}^-_{i+\frac{1}{2}} = \boldsymbol{R}^t_{i,j}(x_{i+\frac{1}{2}}, y_j), \tag{14}$$

$$\boldsymbol{w}^+_{j-\frac{1}{2}} = \boldsymbol{R}^t_{i,j}(x_i, y_{j-\frac{1}{2}}), \qquad \boldsymbol{w}^-_{j+\frac{1}{2}} = \boldsymbol{R}^t_{i,j}(x_i, y_{j+\frac{1}{2}}). \tag{15}$$

Additionally,

$$\begin{aligned} \boldsymbol{D}^+_{i-\frac{1}{2}}(t) &= \boldsymbol{D}^+_{i-\frac{1}{2}}(\boldsymbol{w}^{HR-}_{i-\frac{1}{2}}(t), \boldsymbol{w}^{HR+}_{i-\frac{1}{2}}(t), z_{B,i-\frac{1}{2}}, z_{B,i-\frac{1}{2}}) + \boldsymbol{S}^+_{i-\frac{1}{2}}, \\ \boldsymbol{D}^-_{i+\frac{1}{2}}(t) &= \boldsymbol{D}^-_{i+\frac{1}{2}}(\boldsymbol{w}^{HR-}_{i+\frac{1}{2}}(t), \boldsymbol{w}^{HR+}_{i+\frac{1}{2}}(t), z_{B,i+\frac{1}{2}}, z_{B,i+\frac{1}{2}}) + \boldsymbol{S}^-_{i+\frac{1}{2}}, \\ \boldsymbol{D}^+_{j-\frac{1}{2}}(t) &= \boldsymbol{D}^+_{j-\frac{1}{2}}(\boldsymbol{w}^{HR-}_{j-\frac{1}{2}}(t), \boldsymbol{w}^{HR+}_{j-\frac{1}{2}}(t), z_{B,j-\frac{1}{2}}, z_{B,j-\frac{1}{2}}) + \boldsymbol{S}^+_{j-\frac{1}{2}}, \\ \boldsymbol{D}^-_{j+\frac{1}{2}}(t) &= \boldsymbol{D}^-_{j+\frac{1}{2}}(\boldsymbol{w}^{HR-}_{j+\frac{1}{2}}(t), \boldsymbol{w}^{HR+}_{j+\frac{1}{2}}(t), z_{B,j+\frac{1}{2}}, z_{B,j+\frac{1}{2}}) + \boldsymbol{S}^-_{j+\frac{1}{2}}. \end{aligned}$$

We recall that a thoroughly description of these terms can be found in Section (3.2).

Finally, as in the one dimensional case, this numerical methods is second order accurate in space and time and inherits the well-balanced property of the 1D solver.

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD
DE MÁLAGA

Appendix B

Parallel implementation

In this appendix, we detail the implementation on GPUs of the finite volume numerical scheme presented in Chapter 3. Parallelization on GPU is able to provide a substantial speed-up with respect to sequential version of the code and it is therefore especially suited when dealing with large computational domains. There are several strategies to develop a parallel code for GPUs. One approximation consists on developing the entire code on CUDA (see [11] and references therein). In fact, there are numerous examples in the literature where different numerical schemes and models have been written directly in CUDA, obtaining extraordinary results from the point of view of computational efficiency even for large computational domains (see for example [3]). Alternatively, one may choose to develop a GPU parallel code using OpenACC directives (see [206]). This approach is a newer paradigm and offers a friendlier introduction to GPU based parallelization, especially when compared to the cumbersome task of parallelizing on GPU using CUDA.

Several authors have studied the results of parallelizing a code under CUDA and OpenACC (see for example [207]). In general, the results of the comparison are slightly favorable to CUDA, thanks to its greater control over data management and load balancing. Nevertheless, OpenACC also presents some advantages over CUDA. In particular, the work demanded for parallelizing code using OpenACC is simpler and less exigent than the one required for CUDA. Indeed, when developing a parallel code in CUDA, the construction of an *accelerator kernel*, the part of the code that will be actually executed by the GPU, must be manually designed by the programmer, taking into account the particular structure of the problem and the target accelerator hardware. As a result, the performance of the resulting CUDA code highly depends on its specific adaptation to the architecture being used. Consequentially, the final acceleration gain relies strongly on the capabilities of the developer in adapting the particular algorithm used, which is not an easy task. OpenACC follows an alternative approximation. In order to build a parallel code on GPU, it relies on the addition of some instruction in the form of *pragmas*. In this way, most of the burden of the parallelization switch from the programmer to the compiler, which is the responsible to translate the information contained in the *pragmas* to build

UNIVERSIDAD DE MÁLAGA

the *accelerator kernel*. In this way, it is not necessary to code again the whole program, but rather instruct the compiler with enough and suitable hints to generate a parallel version of the code using some general heuristic functions. Despite this, it is still possible to achieve similar performance with OpenACC and CUDA for some reference problems (see for instance [208]). However, this requires a heavy personalization of the OpenACC version, potentially losing its accessibility, and getting closer to a CUDA implementation.

Note that, by all means, this does not imply that the usual work required to parallelize a sequential code is not present in OpenACC. However, the tools provided by OpenACC are friendlier than the ones provided by CUDA. Particularly, OpenACC allows easy data transference between the host and the accelerator hardware and an easy parallelization of iterative loops susceptible of being parallelizable. In this sense, it is somehow similar to OpenMP, the programming interface for parallelization on CPUs.

In this thesis, we perform a comparison between an OpenACC and CUDA version of the finite volume numerical scheme of Chapter 3 for the one layer and two dimensional case. For the CUDA code, the procedure described in [141] is followed. The results on performance can be seen in Figure B.1. We can see that the CUDA code is two times faster that the OpenACC code. However, the speed-ups of the OpenACC version are also excellent, up to sixty times faster that the sequential code. Additionally, a comparison of elapsed simulation times using a version of the code parallelized on CPU and several different GPUs architecture has been performed. We can see that the GPU implementation is significantly better that the CPU one and offers great scalability.

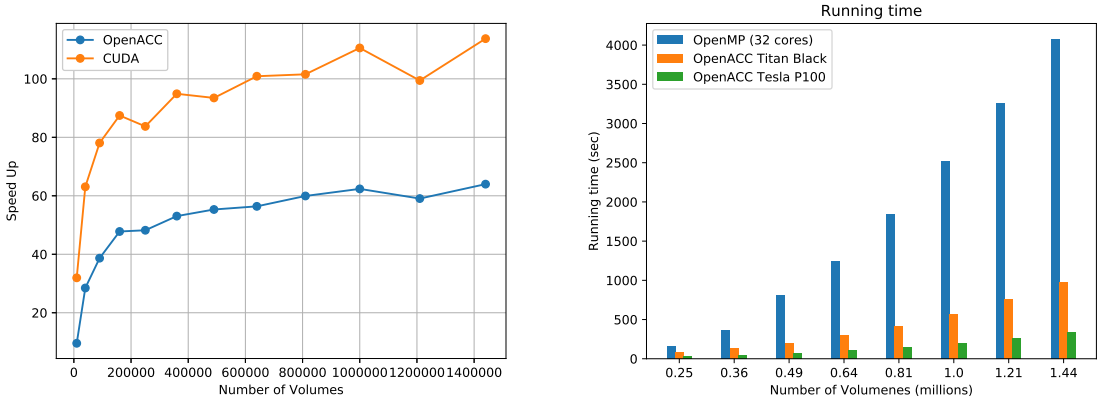


Figure B.1: Speed up of OpenACC vs CUDA (left). Elapsed time of a OpenMP version vs. running time of an OpenACC version in two different graphical processor units.

UNIVERSIDAD DE MÁLAGA

Bibliography

[1] B. De St. Venant. "Théorie du mouvement non permanent des eaux, avec application aux crues des rivières et à l'introduction des marées dans leur lit". In: *Academic de Sci. Comptes Redus* 73.99 (1871), pp. 148–154.

[2] B. Cushman-Roisin and J.-M. Beckers. *Introduction to geophysical fluid dynamics: physical and numerical aspects*. Academic press, 2011.

[3] J. Macias, M. J. Castro, J. M. González-Vida, M. de la Asunción, and S. Ortega. "HySEA: An operational GPU-based model for Tsunami Early Warning Systems". In: *EGU General Assembly Conference Abstracts*. EGU General Assembly Conference Abstracts. 2014, p. 14217.

[4] K. K. Katta, R. D. Nair, and V. Kumar. "High-Order Finite Volume Shallow Water Model on the Cubed-Sphere". In: *Appl. Math. Comput.* 266.C (Sept. 2015), pp. 316–327.

[5] E. Audusse and M.-O. Bristeau. "A well-balanced positivity preserving "second-order" scheme for shallow water flows on unstructured meshes". In: *Journal of Computational Physics* 206.1 (2005), pp. 311–333.

[6] 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". In: *Journal of Computational Physics* 364 (2018), pp. 209–234.

[7] E. Fernández-Nieto, E. H. Koné, and T. Chacón Rebollo. "A Multilayer Method for the Hydrostatic Navier-Stokes Equations: A Particular Weak Solution". In: *Journal of Scientific Computing* 60 (Aug. 2014).

[8] E. Audusse and M.-O. Bristeau. "Finite-Volume Solvers for a Multilayer Saint-Venant System". In: *Applied Mathematics and Computer Science* 17 (Oct. 2007), pp. 311–320.

[9] 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". In: *ESAIM: Mathematical Modelling and Numerical Analysis* 45.1 (2011), pp. 169–200.

UNIVERSIDAD DE MÁLAGA

[10] L. Gosse. "A well-balanced scheme using non-conservative products designed for hyperbolic systems of conservation laws with source terms". In: *Mathematical Models and Methods in Applied Sciences* 11.02 (2001), pp. 339–365.

[11] M. de la Asunción, M. Castro, J. Mantas, and S. Ortega. "Numerical simulation of tsunamis generated by landslides on multiple GPUs". In: *Advances in Engineering Software* 99 (2016), pp. 59–72.

[12] R. J. LeVeque, D. L. George, and M. J. Berger. "Tsunami modelling with adaptively refined finite volume methods". In: *Acta Numerica* 20 (2011), pp. 211–289.

[13] E. Barthélemy. "Nonlinear shallow water theories for coastal waves". In: *Surveys in Geophysics* 25.3-4 (2004), pp. 315–337.

[14] M. Lastra, J. M. Mantas, C. Ureña, M. J. Castro, and J. A. García-Rodríguez. "Simulation of shallow-water systems using graphics processing units". In: *Mathematics and Computers in Simulation* 80.3 (2009), pp. 598–618.

[15] P. Brufau, M. Vázquez-Cendón, and P. García-Navarro. "A numerical model for the flooding and drying of irregular domains". In: *International Journal For Numerical Methods In Fluids* 39.3 (2002), pp. 247–275.

[16] M. Gonzalez-Sanchis, J. Murillo, B. Latorre, F. Comin, and P. Garcia-Navarro. "Transient Two-Dimensional Simulation of Real Flood Events in a Mediterranean Floodplain". In: *Journal Of Hydraulic Engineering-Asce* 138.7 (2012), pp. 629–641.

[17] M. Ersoy, O. Lakkis, and P. Townsend. "A Saint-Venant shallow water model for overland flows with precipitation and recharge". In: *arXiv e-prints*, arXiv:1705.05470 (2017), arXiv:1705.05470.

[18] R. Martínez-Cantó and A. Hidalgo. "A Methodology Based on Numerical Simulation to Study River Floods. Application to Lower River Omaña Basin". In: *Water Resources* 46.6 (2019), pp. 844–852.

[19] M. Ersoy, O. Lakkis, and P. Townsend. "Numerical simulation of flood inundation using a well-balanced kinetic scheme for the shallow water equations with bulk recharge and discharge". In: *EGU General Assembly Conference Abstracts*. Vol. 18. 2016.

[20] C. Dawson et al. "Discontinuous Galerkin methods for modeling hurricane storm surge". In: *Advances in Water Resources* 34.9 (2011), pp. 1165–1176.

[21] K. T. Mandli and C. N. Dawson. "Adaptive mesh refinement for storm surge". In: *Ocean Modelling* 75 (2014), pp. 36–50.

[22] S. Tanaka, S. Bunya, J. J. Westerink, C. Dawson, and R. A. Luettich. "Scalability of an unstructured grid continuous Galerkin based hurricane storm surge model". In: *Journal of Scientific Computing* 46.3 (2011), pp. 329–358.

UNIVERSIDAD DE MÁLAGA

[23] M.-O. Bristeau, A. Mangeney, J. Sainte-Marie, and N. Seguin. "An energy-consistent depth-averaged Euler system: derivation and properties". In: *arXiv preprint arXiv:1406.6565* (2014).

[24] E. D. Fernández-Nieto, M. Parisot, Y. Penel, and J. Sainte-Marie. "A hierarchy of dispersive layer-averaged approximations of Euler equations for free surface flows". EN. In: *Communications in Mathematical Sciences* 16.5 (2018), pp. 1169–1202.

[25] G. Tumolo and L. Bonaventura. "Simulations of Non-hydrostatic Flows by an Efficient and Accurate p-Adaptive DG Method". In: *Numerical Methods for Flows: FEF 2017 Selected Contributions*. Cham: Springer International Publishing, 2020, pp. 41–53.

[26] D. B. Haidvogel, J. L. Wilkin, and R. E. Young. "A semi-spectral primitive equation ocean circulation model using vertical sigma and orthogonal curvilinear horizontal coordinates". In: *Journal of Computational Physics* 94 (1991), pp. 151–185.

[27] V. Casulli. "A semi-implicit finite difference method for non-hydrostatic, free-surface flows". In: *International journal for numerical methods in fluids* 30.4 (1999), pp. 425–440.

[28] V. Casulli and P. Zanolli. "Semi-implicit numerical modeling of nonhydrostatic free-surface flows for environmental problems". In: *Mathematical and computer modelling* 36.9-10 (2002), pp. 1131–1149.

[29] G. Ma, F. Shi, and J. T. Kirby. "Shock-capturing non-hydrostatic model for fully dispersive surface wave processes". In: *Ocean Modelling* 43 (2012), pp. 22–35.

[30] V. Casulli. "Semi-implicit finite difference methods for the two-dimensional shallow water equations". In: *Journal of Computational Physics* 86 (1990), pp. 56–74.

[31] V. Casulli and R. Cheng. "Semi-implicit finite difference methods for three-dimensional shallow water flow". In: *International Journal of Numerical Methods in Fluids* 15 (1992), pp. 629–648.

[32] V. Casulli. "A high-resolution wetting and drying algorithm for free-surface hydrodynamics". In: *International Journal for Numerical Methods in Fluids* 60 (2009), pp. 391–408.

[33] V. Casulli. "A semi-implicit numerical method for the free-surface Navier-Stokes equations". In: *International Journal for Numerical Methods in Fluids* 74 (2014), pp. 605–622.

[34] T. M. de Luna, E. Fernández Nieto, and M. J. Castro Díaz. "Derivation of a Multilayer Approach to Model Suspended Sediment Transport: Application to Hyperpycnal and Hypopycnal Plumes". In: *Communications in Computational Physics* 22.5 (2017), pp. 1439–1485.

UNIVERSIDAD DE MÁLAGA

[35] R. Bürger, E. D. Fernández-Nieto, and V. Andrés Osores. "A dynamic multilayer shallow water model for polydisperse sedimentation." In: *ESAIM: Mathematical Modelling and Numerical Analysis* (Apr. 2019).

[36] C. Adduce, G. Sciortino, and S. Proietti. "Gravity Currents Produced by Lock Exchanges: Experiments and Simulations with a Two-Layer Shallow-Water Model with Entrainment". In: *Journal of Hydraulic Engineering* 138.2 (2012), pp. 111–121.

[37] J. Garres-Díaz and L. Bonaventura. "Flexible and efficient discretizations of multilayer models with variable density". In: *Applied Mathematics and Computation* 402 (2021), p. 126097.

[38] E. Audusse, M.-O. Bristeau, M. Pelanti, and J. Sainte-Marie. "Approximation of the hydrostatic Navier–Stokes system for density stratified flows by a multilayer model: Kinetic interpretation and numerical solution". In: *Journal of Computational Physics* 230.9 (2011), pp. 3453–3478.

[39] G. Dal Maso, P. G. LeFloch, and F. Murat. "Definition and weak stability of nonconservative products". In: *Journal de Mathématiques Pures et Appliquées. Neuvième Série* 74 (Jan. 1995).

[40] C. Parés. "Numerical methods for nonconservative hyperbolic systems: a theoretical framework." In: *SIAM Journal on Numerical Analysis* 44.1 (2006), pp. 300–321.

[41] M. Castro, T. Morales de Luna, and C. Parés. "Chapter 6 - Well-Balanced Schemes and Path-Conservative Numerical Methods". In: *Handbook of Numerical Methods for Hyperbolic Problems*. Ed. by R. Abgrall and C.-W. Shu. Vol. 18. Handbook of Numerical Analysis. Elsevier, 2017, pp. 131–175.

[42] I. Toumi. "A weak formulation of Roe's approximate Riemann solver". In: *J. Comput. Phys.* 102.2 (1992), pp. 360–373.

[43] M. Castro, 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 1D shallow water system". In: *ESAIM: Mathematical Modelling and Numerical Analysis* 35.1 (2001), pp. 107–127.

[44] A. Bermúdez and M. E. Vázquez. "Upwind methods for hyperbolic conservation laws with source terms". In: *Computers & Fluids* 23.8 (1994), pp. 1049–1071.

[45] C. Parés and M. Castro. "On the well-balance property of Roe's method for nonconservative hyperbolic systems. Applications to shallow-water systems". In: *ESAIM: mathematical modelling and numerical analysis* 38.5 (2004), pp. 821–852.

[46] A. Harten and J. M. Hyman. "Self adjusting grid methods for one-dimensional hyperbolic conservation laws". In: *Journal of computational physics* 50.2 (1983), pp. 235–269.

UNIVERSIDAD DE MÁLAGA

[47] B. Einfeldt, C.-D. Munz, P. L. Roe, and B. Sjögreen. "On Godunov-type methods near low densities". In: *Journal of computational physics* 92.2 (1991), pp. 273–295.

[48] M. Castro and E. Fernández-Nieto. "A Class of Computationally Fast First Order Finite Volume Solvers: PVM Methods". In: *SIAM Journal on Scientific Computing* 34 (Jan. 2012).

[49] P. Degond, P.-F. Peyrard, G. Russo, and P. Villedieu. "Polynomial upwind schemes for hyperbolic systems". In: *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics* 328.6 (1999), pp. 479–483.

[50] M. Castro, J. M. Gallardo, and C. Parés. "High Order Finite Volume Schemes Based on Reconstruction of States for Solving Hyperbolic Systems with Nonconservative Products. Applications to Shallow-Water Systems". In: *Mathematics of Computation* 75.255 (2006), pp. 1103–1134.

[51] M. J. Castro, E. D. Fernández-Nieto, A. M. Ferreiro, J. A. García-Rodríguez, and C. Parés. "High order extensions of Roe schemes for two-dimensional nonconservative hyperbolic systems". In: *Journal of Scientific Computing* 39.1 (2009), pp. 67–114.

[52] W. H. Reed and T. Hill. *Triangular mesh methods for the neutron transport equation*. Tech. rep. Los Alamos Scientific Lab., N. Mex.(USA), 1973.

[53] B. Cockburn, S. Hou, and C.-W. Shu. "The Runge-Kutta Local Projection Discontinuous Galerkin Finite Element Method for Conservation Laws. IV: The Multidimensional Case". In: *Mathematics of Computation* 54.190 (1990), pp. 545–581.

[54] B. Cockburn, S.-Y. Lin, and C.-W. Shu. "TVB Runge-Kutta local projection Discontinuous Galerkin Finite Element Method for conservation laws III: One-dimensional systems". In: *Journal of Computational Physics* 84.1 (1989), pp. 90–113.

[55] B. Cockburn and C.-W. Shu. "TVB Runge-Kutta Local Projection Discontinuous Galerkin Finite Element Method for Conservation Laws II: General Framework". In: *Mathematics of Computation* 52.186 (1989), pp. 411–435.

[56] Cockburn, Bernardo and Shu, Chi-Wang. "The Runge-Kutta local projection P^1 Discontinuous Galerkin Finite Element method for scalar conservation laws". In: *ESAIM: M2AN* 25.3 (1991), pp. 337–361.

[57] C. M. Klaij, J. J. van der Vegt, and H. van der Ven. "Space–time discontinuous Galerkin method for the compressible Navier–Stokes equations". In: *Journal of Computational Physics* 217.2 (2006), pp. 589–611.

UNIVERSIDAD DE MÁLAGA

[58] J. van der Vegt and H. van der Ven. "Space–Time Discontinuous Galerkin Finite Element Method with Dynamic Grid Motion for Inviscid Compressible Flows: I. General Formulation". In: *Journal of Computational Physics* 182.2 (2002), pp. 546–585.

[59] H. van der Ven and J. van der Vegt. "Space–time discontinuous Galerkin finite element method with dynamic grid motion for inviscid compressible flows: II. Efficient flux quadrature". In: *Computer Methods in Applied Mechanics and Engineering* 191.41 (2002), pp. 4747–4780.

[60] M. Dumbser and M. Facchini. "A local space-time discontinuous Galerkin method for Boussinesq-type equations". In: *Applied Mathematics and Computation* 272 (2016), pp. 336–346.

[61] A. Taube, M. Dumbser, D. S. Balsara, and C.-D. Munz. "Arbitrary high-order discontinuous Galerkin schemes for the magnetohydrodynamic equations". In: *Journal of Scientific Computing* 30.3 (2007), pp. 441–464.

[62] M. Dumbser and C.-D. Munz. "Building blocks for arbitrary high order discontinuous Galerkin schemes". In: *Journal of Scientific Computing* 27.1-3 (2006), pp. 215–230.

[63] J. Qiu, M. Dumbser, and C.-W. Shu. "The discontinuous Galerkin method with Lax–Wendroff type time discretizations". In: *Computer Methods in Applied Mechanics and Engineering* 194.42 (2005), pp. 4528–4543.

[64] G. Tumolo, L. Bonaventura, and M. Restelli. "A semi-implicit, semi-Lagrangian, p-adaptive discontinuous Galerkin method for the shallow water equations ". In: *Journal of Computational Physics* 232 (2013), pp. 46–67.

[65] M. Dumbser and V. Casulli. "A staggered semi-implicit spectral discontinuous Galerkin scheme for the shallow water equations". In: *Applied Mathematics and Computation* 219 (2013), pp. 8057–8077.

[66] M. Tavelli and M. Dumbser. "A high order semi-implicit discontinuous Galerkin method for the two dimensional shallow water equations on staggered unstructured meshes". In: *Applied Mathematics and Computation* 234 (2014), pp. 623–644.

[67] S. Busto, M. Tavelli, W. Boscheri, and M. Dumbser. "Efficient high order accurate staggered semi-implicit discontinuous Galerkin methods for natural convection problems". In: *Computers & Fluids* 198 (2020), p. 104399.

[68] E. F. Toro, R. Millington, and L. Nejad. "Towards very high order Godunov schemes". In: *Godunov methods*. Springer, 2001, pp. 907–940.

[69] V. Titarev and E. Toro. "ADER: Arbitrary high order Godunov approach". In: *Journal of Scientific Computing* 17.1-4 (2002), pp. 609–618.

UNIVERSIDAD DE MÁLAGA

[70] E. Toro and V. Titarev. "Solution of the generalized Riemann problem for advection-reaction equations". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 458.2018 (2002). cited By 159, pp. 271–281.

[71] V. Titarev and E. Toro. "ADER schemes for three-dimensional non-linear hyperbolic systems". In: *Journal of Computational Physics* 204.2 (2005), pp. 715–736.

[72] E. Toro and V. Titarev. "Derivative Riemann solvers for systems of conservation laws and ADER methods". In: *Journal of Computational Physics* 212.1 (2006), pp. 150–165.

[73] S. Busto, E. F. Toro, and M. E. Vázquez-Cendón. "Design and analysis of ADER-type schemes for model advection–diffusion–reaction equations". In: *Journal of Computational Physics* 327 (2016), pp. 553–575.

[74] M. Dumbser, C. Enaux, and E. F. Toro. "Finite volume schemes of very high order of accuracy for stiff hyperbolic balance laws". In: *Journal of Computational Physics* 227.8 (2008), pp. 3971–4001.

[75] M. Dumbser, D. S. Balsara, E. F. Toro, and C.-D. Munz. "A unified framework for the construction of one-step finite volume and Discontinuous Galerkin schemes on unstructured meshes". In: *Journal of Computational Physics* 227.18 (2008), pp. 8209–8253.

[76] S. Busto, S. Chiocchetti, M. Dumbser, E. Gaburro, and I. Peshkov. "High Order ADER Schemes for Continuum Mechanics". In: *Frontiers in Physics* 8 (2020), p. 32.

[77] N. Wintermeyer, A. R. Winters, G. J. Gassner, and T. Warburton. "An entropy stable discontinuous Galerkin method for the shallow water equations on curvilinear meshes with wet/dry fronts accelerated by GPUs". In: *Journal of Computational Physics* 375 (Dec. 2018), pp. 447–480.

[78] G. Li, L. Song, and J. Gao. "High order well-balanced discontinuous Galerkin methods based on hydrostatic reconstruction for shallow water equations". In: *Journal of Computational and Applied Mathematics* 340 (Oct. 2018), pp. 546–560.

[79] G. Li, J. Li, S. Qian, and J. Gao. "A well-balanced ADER discontinuous Galerkin method based on differential transformation procedure for shallow water equations". In: *Applied Mathematics and Computation* 395 (Apr. 2021), p. 125848.

[80] X. Wu, E. J. Kubatko, and J. Chan. "High-order entropy stable discontinuous Galerkin methods for the shallow water equations: Curved triangular meshes and GPU acceleration". In: *Computers & Mathematics with Applications* 82 (Jan. 2021), pp. 179–199.

[81] M. Dumbser, M. Castro, C. Parés, and E. F. Toro. "ADER schemes on unstructured meshes for nonconservative hyperbolic systems: Applications to geophysical flows". In: *Computers & Fluids* 38.9 (Oct. 2009), pp. 1731–1748.

UNIVERSIDAD DE MÁLAGA

[82] N. Izem, M. Seaid, and M. Wakrim. "A discontinuous Galerkin method for two-layer shallow water equations". In: *Mathematics and Computers in Simulation* 120 (Feb. 2016), pp. 12–23.

[83] Y. Cheng, H. Dong, M. Li, and W. Xian. "A High Order Central DG method of the Two-Layer Shallow Water Equations". In: *Communications in Computational Physics* 28.4 (June 2020), pp. 1437–1463.

[84] R. L. Higdon. "Discontinuous Galerkin methods for multi-layer ocean modeling: Viscosity and thin layers". In: *Journal of Computational Physics* 401 (Jan. 2020), p. 109018.

[85] C. Escalante, T. M. de Luna, and M. Castro. "Non-hydrostatic pressure shallow flows: GPU implementation using finite volume and finite difference scheme". In: *Applied Mathematics and Computation* 338 (2018), pp. 631–659.

[86] C. Escalante, M. Dumbser, and M. Castro. "An efficient hyperbolic relaxation system for dispersive non-hydrostatic water waves and its solution with high order discontinuous Galerkin schemes". In: *J. Comput. Phys.* 394 (2019), pp. 385–416.

[87] C. Bassi, L. Bonaventura, S. Busto, and M. Dumbser. "A hyperbolic reformulation of the Serre-Green-Naghdi model for general bottom topographies". In: *Computers & Fluids* 212 (2020), p. 104716.

[88] S. Busto, M. Dumbser, C. Escalante, S. Gavrilyuk, and N. Favrie. "On high order ADER discontinuous Galerkin schemes for first order hyperbolic reformulations of nonlinear dispersive systems". In: *J. Sci. Comput.* 87 (2021), p. 48.

[89] S. Clain, S. Diot, and R. Loubère. "A high-order finite volume method for systems of conservation laws—Multi-dimensional Optimal Order Detection (MOOD)". In: *Journal of Computational Physics* 230.10 (2011), pp. 4028–4050.

[90] S. Diot, S. Clain, and R. Loubère. "Improved detection criteria for the Multi-dimensional Optimal Order Detection (MOOD) on unstructured meshes with very high-order polynomials". In: *Computers and Fluids* 64 (2012), pp. 43–63.

[91] M. Dumbser, O. Zanotti, R. Loubère, and S. Diot. "A posteriori subcell limiting of the discontinuous Galerkin finite element method for hyperbolic conservation laws". In: *Journal of Computational Physics* 278 (Dec. 2014), pp. 47–75.

[92] B. Cockburn and C.-W. Shu. "The Runge–Kutta Discontinuous Galerkin Method for Conservation Laws V: Multidimensional Systems". In: *Journal of Computational Physics* 141.2 (1998), pp. 199–224.

[93] R. Biswas, K. D. Devine, and J. E. Flaherty. "Parallel, adaptive finite element methods for conservation laws". In: *Applied Numerical Mathematics* 14.1-3 (1994), pp. 255–283.

UNIVERSIDAD DE MÁLAGA

[94] A. Burbeau, P. Sagaut, and C.-H. Bruneau. "A Problem-Independent Limiter for High-Order Runge–Kutta Discontinuous Galerkin Methods". In: *Journal of Computational Physics* 169.1 (2001), pp. 111–150.

[95] X. Zhong and C.-W. Shu. "A simple weighted essentially nonoscillatory limiter for Runge–Kutta discontinuous Galerkin methods". In: *Journal of Computational Physics* 232.1 (2013), pp. 397–415.

[96] J. Qiu and C.-W. Shu. "Runge-Kutta discontinuous Galerkin method using WENO limiters". In: *SIAM Journal on Scientific Computing* 26.3 (2005), pp. 907–929.

[97] J. Shi, C. Hu, and C.-W. Shu. "A technique of treating negative weights in WENO schemes". In: *Journal of Computational Physics* 175.1 (2002), pp. 108–127.

[98] 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". In: *SIAM Journal on Scientific Computing* 25.6 (2004), pp. 2050–2065.

[99] J. P. Berberich, P. Chandrashekar, and C. Klingenberg. "High order well-balanced finite volume methods for multi-dimensional systems of hyperbolic balance laws". In: *Computers & Fluids* 219 (2021), p. 104858.

[100] E. Guerrero Fernández, M. J. Castro-Díaz, and T. M. d. Luna. "A Second-Order Well-Balanced Finite Volume Scheme for the Multilayer Shallow Water Model with Variable Density". In: *Mathematics* 8.5 (2020), p. 848.

[101] A. Bermúdez, X. López, and M. E. Vázquez-Cendón. "Finite volume methods for multi-component Euler equations with source terms". In: *Computers & Fluids* 156 (2017). Ninth International Conference on Computational Fluid Dynamics (ICCFD9), pp. 113–134.

[102] F. Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well–balanced schemes for sources*. Frontiers in Mathematics. Basel: Birkhäuser Verlag, 2004, pp. viii+135.

[103] A. Canestrelli, A. Siviglia, M. Dumbser, and E. F. Toro. "Well-balanced high-order centred schemes for non-conservative hyperbolic systems. Applications to shallow water equations with fixed and mobile bed". In: *Advances in Water Resources* 32.6 (2009), pp. 834–844.

[104] M. J. Castro Díaz, T. Chacón Rebollo, E. D. Fernández-Nieto, and C. Parés. "On Well-Balanced Finite Volume Methods for Nonconservative Nonhomogeneous Hyperbolic Systems". In: *SIAM Journal on Scientific Computing* 29.3 (2007), pp. 1093–1126.

[105] T. Chacón Rebollo, A. Domínguez Delgado, and E. D. Fernández Nieto. "A family of stable numerical solvers for the shallow water equations with source terms". In: *Computer Methods in Applied Mechanics and Engineering* 192.1–2 (Jan. 2003), pp. 203–225.

UNIVERSIDAD DE MÁLAGA

[106] T. Chacón Rebollo, A. Delgado, and E. Fernández-Nieto. "Asymptotically balanced schemes for non-homogeneous hyperbolic systems - Application to the Shallow Water equations". In: *Comptes Rendus Mathematique* 338 (Jan. 2004), pp. 85–90.

[107] P. Chandrashekar and M. Zenk. "Well-balanced nodal discontinuous Galerkin method for Euler equations with gravity". In: *Journal of Scientific Computing* 71.3 (2017), pp. 1062–1093.

[108] 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". en. In: *International Journal for Numerical Methods in Fluids* 81.2 (May 2016), pp. 104–127.

[109] V. Desveaux, M. Zenk, C. Berthon, and C. Klingenberg. "Well–balanced schemes to capture non–explicit steady states: Ripa model". In: *Mathematics of Computation* 85.300 (2016), pp. 1571–1602.

[110] E. Gaburro, M. J. Castro, and M. Dumbser. "Well-balanced Arbitrary-Lagrangian-Eulerian finite volume schemes on moving nonconforming meshes for the Euler equations of gas dynamics with gravity". In: *Monthly Notices of the Royal Astronomical Society* 477.2 (Mar. 2018), pp. 2251–2275.

[111] E. Gaburro, M. J. Castro, and M. Dumbser. "A well balanced diffuse interface method for complex nonhydrostatic free surface flows". In: *Computers & Fluids* 175 (2018), pp. 180–198.

[112] E. Gaburro, M. Dumbser, and M. J. Castro. "Direct Arbitrary-Lagrangian-Eulerian finite volume schemes on moving nonconforming unstructured meshes". In: *Computers & Fluids* 159 (2017), pp. 254–275.

[113] L. Gosse. "A well-balanced flux-vector splitting scheme designed for hyperbolic systems of conservation laws with source terms". In: *Computers & Mathematics with Applications* 39.9 (2000), pp. 135–159.

[114] L. Gosse. "Localization effects and measure source terms in numerical schemes for balance laws". In: *Mathematics of computation* 71.238 (2002), pp. 553–582.

[115] J. Greenberg, A. Leroux, R. Baraille, and A. Noussair. "Analysis and approximation of conservation laws with source terms". In: *SIAM Journal on Numerical Analysis* 34.5 (1997), pp. 1980–2007.

[116] J. M. Greenberg and A.-Y. LeRoux. "A well-balanced scheme for the numerical processing of source terms in hyperbolic equations". In: *SIAM Journal on Numerical Analysis* 33.1 (1996), pp. 1–16.

[117] L. Grosheintz-Laval and R. Käppeli. "High-order well-balanced finite volume schemes for the Euler equations with gravitation". In: *Journal of Computational Physics* 378 (2019), pp. 324–343.

UNIVERSIDAD DE MÁLAGA

[118] R. Käppeli and S. Mishra. "Well-balanced schemes for the Euler equations with gravitation". In: *Journal of Computational Physics* 259 (2014), pp. 199–219.

[119] R. J. LeVeque. "Balancing Source Terms and Flux Gradients in High-Resolution Godunov Methods: The Quasi-Steady Wave-Propagation Algorithm". In: *Journal of Computational Physics* 146.1 (1998), pp. 346–365.

[120] S. Noelle, N. Pankratz, G. Puppo, and J. R. Natvig. "Well-balanced finite volume schemes of arbitrary order of accuracy for shallow water flows". In: *Journal of Computational Physics* 213.2 (2006), pp. 474–499.

[121] S. Noelle, Y. Xing, and C.-W. Shu. "High-order well-balanced finite volume WENO schemes for shallow water equation with moving water". In: *Journal of Computational Physics* 226.1 (2007), pp. 29–58.

[122] B. Perthame and C. Simeoni. "A kinetic scheme for the Saint-Venant system with a source term". In: *Calcolo* 38 (Nov. 2001), pp. 201–231.

[123] B. Perthame and C. Simeoni. "Convergence of the Upwind Interface Source Method for Hyperbolic Conservation Laws". en. In: *Hyperbolic Problems: Theory, Numerics, Applications*. Ed. by T. Y. Hou and E. Tadmor. Springer Berlin Heidelberg, 2003, pp. 61–78.

[124] I. Gómez-Bueno, M. J. Castro, and C. Parés. "High-order well-balanced methods for systems of balance laws: a control-based approach". In: *Applied Mathematics and Computation* 394 (2021), p. 125820.

[125] H. Tang, T. Tang, and K. Xu. "A gas–kinetic scheme for shallow–water equations with source terms". en. In: *Zeitschrift für angewandte Mathematik und Physik ZAMP* 55.3 (May 2004), pp. 365–382.

[126] R. Touma, U. Koley, and C. Klingenberg. "Well-balanced unstaggered central schemes for the Euler equations with gravitation". In: *SIAM Journal on Scientific Computing* 38.5 (2016), B773–B807.

[127] 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". In: *Journal of Computational Physics* 242 (2013), pp. 53–85.

[128] M. J. Castro and C. Parés. "Well-balanced high-order finite volume methods for systems of balance laws". In: *Journal of Scientific Computing* 82.2 (2020), p. 48.

[129] Y. Xing, X. Zhang, and C.-W. Shu. "Positivity-preserving high order well-balanced discontinuous Galerkin methods for the shallow water equations". In: *Advances in Water Resources* 33.12 (2010), pp. 1476–1493.

[130] Y. Xing. "Exactly well-balanced discontinuous Galerkin methods for the shallow water equations with moving water equilibrium". In: *Journal of Computational Physics* 257 (2014), pp. 536–553.

UNIVERSIDAD DE MÁLAGA

[131] A. Ern, S. Piperno, and K. Djadel. "A well-balanced Runge–Kutta discontinuous Galerkin method for the shallow-water equations with flooding and drying". In: *International Journal for Numerical Methods in Fluids* 58.1 (2008), pp. 1–25.

[132] Y. Xing and C.-W. Shu. "High order finite difference WENO schemes with the exact conservation property for the shallow water equations". In: *Journal of Computational Physics* 208.1 (2005), pp. 206–227.

[133] Y. Xing and C.-W. Shu. "High Order Well-Balanced Finite Volume WENO Schemes and Discontinuous Galerkin Methods for a Class of Hyperbolic Systems with Source Terms". In: *J. Comput. Phys.* 214.2 (May 2006), pp. 567–598.

[134] Y. Xing, C.-W. Shu, and S. Noelle. "On the Advantage of Well-Balanced Schemes for Moving-Water Equilibria of the Shallow Water Equations". In: *Journal of Scientific Computing* 48 (2011), pp. 339–349.

[135] A. Bollermann, S. Noelle, and M. Lukacova-Medvidova. "Finite volume evolution Galerkin methods for the shallow water equations with dry beds". In: *Communications in Computational Physics* 10 (Jan. 2010).

[136] Y. Xing and X. Zhang. "Positivity-preserving well-balanced discontinuous Galerkin methods for the shallow water equations on unstructured triangular meshes". In: *Journal of Scientific Computing* 57.1 (2013), pp. 19–41.

[137] G. Li and Y. Xing. "Well-balanced discontinuous Galerkin methods for the Euler equations under gravitational fields". In: *Journal of Scientific Computing* 67.2 (2016), pp. 493–513.

[138] G. Li and Y. Xing. "Well-balanced discontinuous Galerkin methods with hydrostatic reconstruction for the Euler equations with gravitation". In: *Journal of Computational Physics* 352 (2018), pp. 445–462.

[139] S. Vater, N. Beisiegel, and J. Behrens. "A limiter-based well-balanced discontinuous Galerkin method for shallow-water flows with wetting and drying: One-dimensional case". In: *Advances in water resources* 85 (2015), pp. 1–13.

[140] G. J. Gassner, A. R. Winters, and D. A. Kopriva. "A well balanced and entropy conservative discontinuous Galerkin spectral element method for the shallow water equations". In: *Applied Mathematics and Computation* 272 (2016), pp. 291–308.

[141] J. M. Mantas, M. De la Asunción, and M. J. Castro. "An introduction to GPU computing for numerical simulation". In: *Numerical simulation in physics and engineering*. Springer, 2016, pp. 219–251.

[142] E. Audusse, M. Bristeau, and A. Decoene. "Numerical simulations of 3D free surface flows by a multilayer Saint-Venant model". In: *International journal for numerical methods in fluids* 56.3 (2008), pp. 331–350.

UNIVERSIDAD DE MÁLAGA

[143] F. Bouchut and V. Zeitlin. "A robust well-balanced scheme for multi-layer shallow water equations". In: *Discrete and Continuous Dynamical Systems-series B* 13 (June 2010), pp. 739–758.

[144] V. Goloviznin, P. A. Maiorov, P. A. Maiorov, and A. Solovjev. "New Numerical Algorithm for the Multi-Layer Shallow Water Equations Based on the Hyperbolic Decomposition and the CABARET Scheme". In: *Physical Oceanography* 26.6 (2019).

[145] E. Creaco, A. Campisano, A. Khe, C. Modica, and G. Russo. "Head reconstruction method to balance flux and source terms in shallow water equations". In: *Journal of engineering mechanics* 136.4 (2010), pp. 517–523.

[146] G. Russo and A. Khe. "High order well-balanced schemes based on numerical reconstruction of the equilibrium variables". In: *Waves and Stability in Continuous Media*. World Scientific, 2010, pp. 230–241.

[147] R. W. D. Nickalls. "A new bound for polynomials when all the roots are real". In: *The Mathematical Gazette* 95.534 (2011), pp. 520–526.

[148] C. Sánchez–Linares, T. Morales de Luna, and M. J. Castro Díaz. "A HLLC scheme for Ripa model". In: *Applied Mathematics and Computation*. Recent Advances in Numerical Methods for Hyperbolic Partial Differential Equations 272, Part 2 (Jan. 2016), pp. 369–384.

[149] M. Pelanti, F. Bouchut, and A. Mangeney. "A Roe-type scheme for two-phase shallow granular flows over variable topography". In: *ESAIM: Mathematical Modelling and Numerical Analysis* 42.5 (Sept. 2008), pp. 851–885.

[150] 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". In: *J. Comput. Phys.* 242 (June 2013), pp. 53–85.

[151] S. Rhebergen, O. Bokhove, and J. J. W. van der Vegt. "Discontinuous Galerkin finite element methods for hyperbolic nonconservative partial differential equations". In: *Journal of Computational Physics* 227.3 (Jan. 2008), pp. 1887–1922.

[152] M. Dumbser, M. Castro, C. Parés, and E. Toro. "ADER Schemes on Unstructured Meshes for Non-Conservative Hyperbolic Systems: Applications to Geophysical Flows". In: *Computers and Fluids* 38 (2009), pp. 1731–1748.

[153] M. Dumbser, A. Hidalgo, M. Castro, C. Parés, and E. Toro. "FORCE Schemes on Unstructured Meshes II: Non–Conservative Hyperbolic Systems". In: *Computer Methods in Applied Mechanics and Engineering* 199 (2010), pp. 625–647.

[154] C. Berthon and F. Coquel. "Nonlinear Projection Methods for Multi–Entropies Navier–Stokes Systems". In: *Mathematics of Computation* 76.259 (2007), pp. 1163–1194.

UNIVERSIDAD DE MÁLAGA

[155] C. Berthon, F. Coquel, and P.G. LeFloch. "Why many theories of shock waves are necessary: Kinetic functions, equivalent equations, and fourth–order models". In: *J. Comput. Phys.* (2008), pp. 4162–4189.

[156] T. Chacón Rebollo, A. Domínguez Delgado, and E. D. Fernández Nieto. "Asymptotically balanced schemes for non–homogeneous hyperbolic systems. Application to the Shallow Water equations". In: *Comptes Rendus Mathematique* 338.1 (Jan. 2004), pp. 85–90.

[157] F. De Vuyst and Y. Maday. "Schémas nonconservatifs et schémas cinétiques pour la simulation numérique d'écoulements hypersoniques non visqueux en déséquilibre thermochimique". Text. Thése de Doctorat de l'Université Paris VI, 1994.

[158] M. Lukáčová–Medvid'ová, S. Noelle, and M. Kraft. "Well–balanced finite volume evolution Galerkin methods for the shallow water equations". In: *Journal of Computational Physics* 221.1 (Jan. 2007), pp. 122–147.

[159] G. Russo and A. Khe. "High order well balanced schemes for systems of balance laws". In: *Hyperbolic problems: theory, numerics and applications*. Vol. 67. Proc. Sympos. Appl. Math. Amer. Math. Soc., Providence, RI, 2009, pp. 919–928.

[160] R. Käppeli and S. Mishra. "Well–balanced schemes for gravitationally stratified media". In: *ASTRONUM 2014 proceedings* (2014).

[161] P. G. LeFloch. "Shock waves for nonlinear hyperbolic systems in nonconservative form". In: *Institute for Math. and its Appl.* Minneapolis, Preprint 593 (1989).

[162] M. Castro Díaz, E. Fernández–Nieto, and A. Ferreiro. "Sediment transport models in Shallow Water equations and numerical approach by high order finite volume methods". In: *Computers & Fluids* 37.3 (Mar. 2008), pp. 299–316.

[163] 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". In: *Communications in Computational Physics* 6.4 (2009), pp. 848–882.

[164] M. J. Castro, Y. Cheng, A. Chertock, and A. Kurganov. "Solving Two-Mode Shallow Water Equations Using Finite Volume Methods". In: *Commun. Comput. Phys.* 15.5 (2014), pp. 1323–1354.

[165] M. Dumbser, A. Hidalgo, M. J. Castro Díaz, C. Parés, and E. Toro. "FORCE schemes on unstructured meshes II: Nonconservative hyperbolicsystems". In: *Comput. Methods Appl. Mech. Engrg.* 199.9–12 (2010), pp. 625–647.

[166] E. D. Fernández–Nieto, F. Bouchut, D. Bresch, M. J. Castro Díaz, and A. Mangeney. "A new Savage-Hutter type model for submarine avalanches and generated tsunami". In: *Journal of Computational Physics* 227.16 (Aug. 2008), pp. 7720–7754.

UNIVERSIDAD DE MÁLAGA

[167] E. D. Fernández–Nieto, J. M. Gallardo, and P. Vigneaux. "Efficient numerical schemes for viscoplastic avalanches. Part 1: The 1D case". In: *Journal of Computational Physics* 264.1 (May 2014), pp. 55–90.

[168] S. Munkejord, S. Evje, and T. Flatten. "A MUSTA Scheme for a Nonconservative Two-Fluid Model". In: *SIAM Journal on Scientific Computing* 31.4 (Jan. 2009), pp. 2587–2622.

[169] I. Toumi. "A weak formulation of Roe's approximate Riemann solver". In: *J. Comput. Phys.* 102.2 (1992), pp. 360–373.

[170] C. Parés and M. Castro. "On the well-balance property of Roe's method for nonconservative hyperbolic systems. Applications to shallow-water systems". In: *M2AN Math. Model. Numer. Anal.* 38.5 (2004), pp. 821–852.

[171] M. J. Castro, A. Pardo, C. Parés, and E. F. Toro. "On some fast well-balanced first order solvers for nonconservative systems". In: *Math. Comp.* 79.271 (2010), pp. 1427–1472.

[172] M. J. Castro, J. M. Gallardo, and A. Marquina. "A class of incomplete Riemann solvers based on uniform rational approximations to the absolute value function". In: *Journal of Scientific Computing* 60.2 (2014), pp. 363–389.

[173] A. Harten, P. D. Lax, and B. van Leer. "On Upstream Differencing and Godunov-Type Schemes for Hyperbolic Conservation Laws". In: *SIAM Review* 25.1 (Jan. 1983), pp. 35–61.

[174] E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction, 3rd ed.* en. Springer Science & Business Media, Apr. 2013.

[175] P. Goatin and P. G. LeFloch. "The Riemann problem for a class of resonant hyperbolic systems of balance laws". In: *Annales de l'Institut Henri Poincare (C) Non Linear Analysis* 21.6 (Nov. 2004), pp. 881–902.

[176] S. Davis. "Simplified Second-Order Godunov-Type Methods". In: *SIAM Journal on Scientific and Statistical Computing* 9.3 (1988), pp. 445–473.

[177] E. Toro and S. Billett. "Centred TVD schemes for hyperbolic conservation laws". In: *IMA Journal of Numerical Analysis* 20.1 (Jan. 2000), pp. 47–79.

[178] S. Gottlieb and C.-W. Shu. "Total Variation Diminishing Runge-Kutta Schemes." In: *Mathematics of Computation* 67 (Aug. 1996).

[179] C.-W. Shu and S. Osher. "Efficient implementation of essentially non-oscillatory shock-capturing schemes". In: *Journal of Computational Physics* 77.2 (1988), pp. 439–471.

[180] M. J. Castro, C. Parés, G. Puppo, and G. Russo. "Central Schemes for Nonconservative Hyperbolic Systems". In: *SIAM J. Sci. Comput.* 34.5 (2012), pp. 523–558.

UNIVERSIDAD DE MÁLAGA

[181] A. Harten, B. Engquist, S. Osher, and S. R. Chakravarthy. "Uniformly High Order Accurate Essentially Non-oscillatory Schemes, III". In: *Journal of Computational Physics* 131.1 (Feb. 1997), pp. 3–47.

[182] A. Marquina. "Local Piecewise Hyperbolic Reconstruction of Numerical Fluxes for Nonlinear Scalar Conservation Laws". In: *SIAM Journal on Scientific Computing* 15.4 (July 1994), pp. 892–915.

[183] W. Chi–Shu. *Essentially Non-Oscillatory and Weighted Essentially Non-Oscillatory Schemes for Hyperbolic Conservation Laws*. Tech. rep. Institute for Computer Applications in Science and Engineering (ICASE), 1997.

[184] M. Dumbser and M. Käser. "Arbitrary high order non-oscillatory finite volume schemes on unstructured meshes for linear hyperbolic systems". In: *Journal of Computational Physics* 221.2 (Feb. 2007), pp. 693–723.

[185] M. Dumbser, M. Käser, V. A. Titarev, and E. F. Toro. "Quadrature–free non–oscillatory finite volume schemes on unstructured meshes for nonlinear hyperbolic systems". In: *Journal of Computational Physics* 226.1 (Sept. 2007), pp. 204–243.

[186] B. van Leer. "Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method". In: *Journal of Computational Physics* 32.1 (1979), pp. 101–136.

[187] M. Castro, A. Pardo Milanés, and C. Parés. "Well-balanced numerical schemes based on a generalized hydrostatic reconstruction technique". In: *Mathematical Models and Methods in Applied Sciences* 17.12 (2007), pp. 2055–2113.

[188] C. Berthon, A. Duran, F. Foucher, K. Saleh, and J. D. D. Zabsonré. "Improvement of the hydrostatic reconstruction scheme to get fully discrete entropy inequalities". In: *Journal of Scientific Computing* 80.2 (2019), pp. 924–956.

[189] I. Gómez-Bueno, M. J. Castro, and C. Parés. "High-order well-balanced methods for systems of balance laws: a control-based approach". In: *Applied Mathematics and Computation* 394 (2021), p. 125820.

[190] B. Cockburn and C. Shu. "The Runge-Kutta discontinuous Galerkin method for conservation laws V: multidimensional systems ". In: *Journal of Computational Physics* 141.2 (1998), pp. 199–224.

[191] A. Hidalgo and M. Dumbser. "ADER schemes for nonlinear systems of stiff advection–diffusion–reaction equations". In: *Journal of Scientific Computing* 48.1-3 (2011), pp. 173–189.

[192] H. Jackson. "On the eigenvalues of the ADER-WENO Galerkin predictor". In: *Journal of Computational Physics* 333 (2017), pp. 409–413.

[193] O. Zanotti and M. Dumbser. "Efficient conservative ADER schemes based on WENO reconstruction and space-time predictor in primitive variables". In: *Computational Astrophysics and Cosmology* 3.1 (2016), p. 1.

UNIVERSIDAD DE MÁLAGA

[194] F. Fambri, M. Dumbser, S. Köppel, L. Rezzolla, and O. Zanotti. "ADER discontinuous Galerkin schemes for general-relativistic ideal magnetohydrodynamics". In: *Monthly Notices of the Royal Astronomical Society* 477 (2018), pp. 4543–4564.

[195] B. Owren and M. Zennaro. "Derivation of efficient, continuous, explicit Runge-Kutta methods". In: *SIAM J. Sci. and Stat. Comput.* 13 (1992), pp. 1488–1501.

[196] G. Gassner, M. Dumbser, F. Hindenlang, and C. Munz. "Explicit one–step time discretizations for discontinuous Galerkin and finite volume schemes based on local predictors". In: *Journal of Computational Physics* 230.11 (2011), pp. 4232–4247.

[197] D. E. Charrier and T. Weinzierl. "Stop talking to me - a communication-avoiding ADER-DG realisation". In: *ArXiv* abs/1801.08682 (2018).

[198] G. A. Sod. "A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws". In: *Journal of Computational Physics* 27.1 (1978), pp. 1–31.

[199] P. D. Lax. "Weak solutions of nonlinear hyperbolic equations and their numerical computation". In: *Communications on Pure and Applied Mathematics* 7.1 (1954), pp. 159–193.

[200] M. J. Castro, A. Pardo Milanés, and C. Parés. "Well-balanced numerical schemes based on a generalized hydrostatic reconstruction technique". In: *Mathematical Models and Methods in Applied Sciences* 17.12 (2007), pp. 2055–2113.

[201] T. Morales de Luna, M. Castro Díaz, and C. Parés. "Reliability of first order numerical schemes for solving shallow water system over abrupt topography". In: *Applied Mathematics and Computation* 219.17 (2013), pp. 9012–9032.

[202] X. Zhang and C.-W. Shu. "On maximum-principle-satisfying high order schemes for scalar conservation laws". In: *Journal of Computational Physics* 229.9 (2010), pp. 3091–3120.

[203] E. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*. Springer Berlin Heidelberg, 2009.

[204] E. Guerrero Fernández, C. Escalante, and M. J. Castro Díaz. "Well-Balanced High-Order Discontinuous Galerkin Methods for Systems of Balance Laws". In: *Mathematics* 10.1 (2022).

[205] E. G. Fernández, M. J. C. Díaz, M. Dumbser, and T. M. de Luna. "An Arbitrary High Order Well-Balanced ADER-DG Numerical Scheme for the Multilayer Shallow-Water Model with Variable Density". In: *Journal of Scientific Computing* 90.1 (Dec. 2021), p. 52.

[206] S. Chandrasekaran and G. Juckeland. *OpenACC for Programmers: Concepts and Strategies*. Addison-Wesley Professional, 2017.

UNIVERSIDAD DE MÁLAGA

[207] S. Christgau et al. “A comparison of CUDA and OpenACC: Accelerating the Tsunami Simulation EasyWave”. In: *ARCS 2014; 2014 Workshop Proceedings on Architecture of Computing Systems*. Feb. 2014, pp. 1–5.

[208] T. Hoshino, N. Maruyama, S. Matsuoka, and R. Takaki. “CUDA vs OpenACC: Performance Case Studies with Kernel Benchmarks and a Memory-Bound CFD Application”. In: *2013 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing*. 2013, pp. 136–143.

UNIVERSIDAD DE MÁLAGA